270°-Beurteilung von Führungsverhalten: Interperspektivische Übereinstimmung und ihr Zusammenhang mit Erfolg
–
eine Befragung in der Automobilindustrie

Dissertation

zur Erlangung des akademischen Grades
Doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der Philosophischen Fakultät der
Technischen Universität Chemnitz

von Dipl.-Psych. Katrin Rathgeber,
geboren am 01.07.1974 in Nördlingen

Chemnitz, 02.02.2005

Vorbemerkung

In Analogie zum Tanzen, wo Führen und Führen-Lassen gleichermaßen wichtig sind, um die erwünschte Ästhetik und Flüssigkeit zu erreichen, sollten sich auch bei der Personalführung Führende und Geführte gemeinsam in die gleiche Richtung drehen. Damit Führungskräfte und Mitarbeiter aber überhaupt gleiche Ziele anstreben können, müssen sie sich kontinuierlich abstimmen – und diese Abstimmung erfolgt zu einem gewichtigen Teil über das Führungsverhalten. Während der Führende beim Tanz recht unmittelbar bemerkt, wenn sich die Geführte – aus welchen Gründen auch immer – nicht führen lässt und andere Richtungen vorzieht, ist diese Rückkopplung bei der Personalführung verhältnismäßig schwierig. Deshalb kann es sich für Führungskräfte lohnen, neben den „Figuren", die ihre Mitarbeiter beherrschen, auch deren Wahrnehmung des Geführt-Werdens sowie deren Wünsche an die Führung zu kennen. In diesem Sinne muss also den Mitarbeitern das Wort gegeben werden – ein Prozess, der beispielsweise durch standardisierte Befragungen initiiert oder katalysiert werden kann.

Für die Unterstützung bei der Verwirklichung der vorliegenden Arbeit möchte ich Prof. Dr. Klaus Jonas danken – auch dafür, dass er mir durch seine Betreuung immer wieder den Weg geebnet hat. Dr. Günter Keilhofer, Dr. Axel Weber und Dr. Holger Naduschewski gilt mein Dank, weil sie die Durchführung der Studie überhaupt erst ermöglicht sowie später tatkräftig begleitet haben. Auch die kooperative Haltung des Betriebsrats habe ich sehr geschätzt. Allen Befragten danke ich nicht nur für ihre Beteiligung, sondern auch für die zahlreichen Anregungen und Hinweise. Über die Fortführung des Projekts durch Herrn Ullrich Obst und Nadja Senger freue ich mich besonders.
Bei Prof. Dr. Siegfried Gauggel möchte ich mich für die zahlreichen Gespräche und Anregungen bedanken sowie für das Verständnis dafür, dass andere Arbeiten manchmal liegen bleiben mussten. Bei Prof. Dr. Josef Krems und bei Prof. Dr. Lothar Laux bedanke ich mich für die Bereitschaft, die Arbeit zu begutachten. Allen Hilfskräften, die bei der Umsetzung einzelner Projektschritte oder bei der Gestaltung einzelner Abschnitte beteiligt waren, sei ebenfalls herzlich gedankt.
Nicht vergessen will ich meine Eltern sowie alle Freunde und Freundinnen, die mich jeweils auf ihre ganz eigene Art unterstützt haben, insbesondere Sebastian, Iris, Anja, Tina, Nicole, Franz, Johannes, Gerd, Timo, Carsten und Olaf.

1 EINLEITUNG UND AUFBAU DER ARBEIT 10

2 PERSONALFÜHRUNG 15

2.1 Begriffsbestimmung 16
2.2 Führungserfolg 19
2.3 Führungsaufgaben und -instrumente 23
2.3.1 Leistungsbeurteilung 24
2.3.2 Zielsetzung 25
2.3.3 Personalentwicklung 25
2.3.4 Organisieren 26
2.3.5 Entscheiden 26
2.3.6 Führungsgrundsätze 28
2.3.6.1 Definition und Funktion von Führungsgrundsätzen 28
2.3.6.2 Führungsgrundsätze des fokussierten Unternehmens im Abgleich mit klassischen Führungsaufgaben 30
2.4 Traditionelle Führungstheorien 33
2.5 Transformationale Führung 44
2.5.1 Historische Verankerung der transformationalen Führungskonzeption von Bass in der Herrschaftssoziologie von Weber 45
2.5.2 Das neo-charismatische Führungsparadigma 47
2.5.2.1 Zum Begriff „Charisma“ 47
2.5.2.2 Die Wiederentdecker und die „Transformationale Trias“ 51
2.5.2.3 Moderne Charismatische Führungstheorien im engeren Sinne 55
2.5.3 Das transformationale Führungsmodell von Bass 57
2.5.3.1 Die Eckpunkte der Theorie 57
2.5.3.2 Die Führungs-Bausteine im „Full Range of Leadership“: Laissez Faire, transaktionale und transformationale Führung 61
2.5.4 Empirische Befundlage zur Wirkung transformationaler Führung 73
2.5.4.1 Sonnenseiten 73
2.5.4.2 Schattenseiten 83
2.5.5 Kontextfaktoren transformationaler Führung 85
2.5.6 Erlernbarkeit transformationaler Führung 88
2.6 Zusammenfassung und Integration der Führungstheorien und -aufgaben 91
2.7 Messung von Führungsverhalten 103
2.7.1 Methodenüberblick 103
2.7.2 Fragebogen zur Vorgesetzten-Verhaltens-Beschreibung (FVVB) 105
2.7.3 Qualitative Führungsstilanalyse (QFA) und Modifikation 106
2.7.4 Multifactor Leadership Questionnaire (MLQ) 107

3 MULTIPERSPEKTIVISCHE FÜHRUNGSBESCHREIBUNG 112
3.1 Einführung und Begriffsbestimmung 113
3.1.1 Die Überwindung der „Abwärts-Vorherrschaft“ 113
3.1.2 Definition und Begriffsklärung 118
3.1.3 Ablauf und Ziele 120
3.1.4 Verbreitung 123
3.2 Interperspektivische Passung der Beurteilungen von beruflicher Leistung 125
3.2.1 Niveau-Unterschiede 125
3.2.2 Korrelativer Zusammenhang 128
3.2.3 Above-Average-Effekt 132
3.3 Moderatorvariablen für die Passung zwischen Selbst- und Fremdbild 135
3.3.1 Moderatorvariablen im personalen Bereich 137
3.3.2 Moderatorvariablen im tätigkeitsbezogenen und organisationalen Bereich 140
3.3.3 Moderatorvariablen im Bereich der Durchführungsbedingungen 142
3.4 Potentielle Erklärungsmechanismen für die mangelnde Passung zwischen Selbst- und Fremdbild 149
3.4.1 Perspektivenbedingte Diskrepanzen 152
3.4.1.1 Unterschiedliche Verhaltensstichproben 152
3.4.1.2 Mangelnde Beobachtbarkeit und Transparenz von Entscheidungsprozessen 154
3.4.1.3 Unterschiedliche Definitionen von Führung 155
3.4.1.4 Unterschiedliche Gewichtung von Führungsverhaltensweisen 156
3.4.1.5 Unterschiedliche Differenzierungen von Selbst- und Fremdbild 158
3.4.1.6 Akteur-Beobachter-Verzerrung 159
3.4.1.7 Sozialer Vergleich und soziale Identität 161
3.4.2 Idiosynkratisch bedingte Diskrepanzen 163
3.4.2.1 Antworttendenzen und Urteilsfehler 163
3.4.2.2 Mangelndes Feedback 165
3.4.2.3 Selbsttäuschung und Eindrucksmanagement 167
3.4.2.4 Motiv zu Selbstkonsistenz 173
3.4.3 Bewertung der potentiellen Erklärungsmechanismen 174
3.5 Die Rolle einer hohen und geringen Übereinstimmung zwischen Selbst- und Fremdbild 180
3.5.1 Zusammenhang zwischen diskrepanter Führungsbeurteilung und Führungserfolg 181
3.5.2 Potentielle Erklärungsmechanismen für den Zusammenhang zwischen diskrepanter Führungsbeurteilung und Führungsmisserfolg 188
3.6 Feedback-Intervention nach multiperspektivischer Führungsbeurteilung 193

3.6.1 Unter welchen Ausgangsbedingungen wirkt Feedback? 195
3.6.2 Wie lange halten Feedbackveränderungen vor? 199
3.6.3 Wie muss Feedback aussehen, damit es wirkt? 200
3.7 Zusammenfassung und Fazit 205

4 FRAGESTELLUNG 207

5 METHODE 239
5.1 Stichprobe 240
5.2 Untersuchungsverfahren 244
5.2.1 Fragebogen zur Führungsbeschreibung (FzFü) 244
5.2.2 Multifactor Leadership Questionnaire (MLQ) 248
5.2.3 Kurzskalen zur Erfassung der Erfolgskriterien 251
5.2.4 Erfassung biografischer und organisationsbezogener Daten 253
5.2.5 Offene Fragen 253
5.3 Durchführung 256
5.3.1 Grundgerüst 256
5.3.2 Fragebogenverteilung 257
5.3.3 Rücklaufprozess und Rückmeldung 259
5.4 Auswertungsmethoden 261
5.5 Exkurs: Konstruktion und Überprüfung der Fragebogeninstrumente 265
5.5.1 Fragebogen zur Führungsbeschreibung (FzFü) 266
5.5.1.1 Festlegen relevanter Inhalte 266
5.5.1.2 Generieren des Itempools 266
5.5.1.3 Bestimmen des Antwortformats 267
5.5.1.4 Übertragen der Items auf verschiedene Versionen 268
5.5.1.5 Reduzieren des Itempools 269
5.5.1.6 Erstellen einer randomisierten Itemreihenfolge 271
5.5.1.7 Formulieren der Instruktion 271
5.5.1.8 Itemselektion 272
5.5.1.9 Skalenanalysen und inhaltliche Interpretation 272
5.5.1.10 Konstruktvalidität 278
5.5.2 Multifactor Leadership Questionnaire (MLQ) 281
5.5.3 Skalen zur Beurteilung der Erfolgskriterien 283

6 ERGEBNISSE 286
6.1 Deskriptive Statistik und Interrater-Übereinstimmung 287
6.1.1 Deskriptive Statistik 287
6.1.1.1 Fragebogen zur Führungsbeschreibung 287

6.1.1.2 Multifactor Leadership Questionnaire ... 291
6.1.1.3 MLQ-externe Erfolgsvariablen ... 292
6.1.2 Zur Aggregation der Mitarbeiter-Ratings: Ihre Übereinstimmung ... 293
6.2 Beobachtetes versus erwartetes Führungsverhalten ... 296
6.3 Übereinstimmung zwischen verschiedenen Beurteilungsrichtungen ... 304
6.3.1 Korrelative Zusammenhänge zwischen Selbst-, Aufwärts- und Abwärtsbeurteilungen ... 304
6.3.2 Niveau-Unterschiede zwischen Selbst-, Aufwärts- und Abwärtsbeurteilungen ... 314
6.3.3 Moderatoreffekte auf die interperspektivische Übereinstimmung ... 320
6.3.4 Zusammenhang zwischen verschiedenen Übereinstimmungs-Indizes ... 337
6.3.5 Above-Average-Effekt ... 342
6.4 Validität von Führungsbeurteilungen ... 350
6.4.1 Zusammenhang zwischen Führungsverhalten und Erfolg ... 350
6.4.2 Inkrementelle Validität der Führungsbeurteilungen ... 361
6.4.3 Moderatorvariablen für den Zusammenhang zwischen Führungsverhalten und Erfolg ... 375
6.5 Validität der interperspektivischen Übereinstimmung ... 380
6.5.1 Hängt die interperspektivische Übereinstimmung der Führungsbeurteilungen mit Erfolg zusammen? ... 381
6.5.2 Interperspektivische Diskrepanz versus uniperspektivische Beurteilung ... 401
6.5.3 Leistet die interperspektivische Beurteilungsübereinstimmung inkrementelle Beiträge zur Erfolgsaufklärung? ... 407
6.5.4 Bedeutet Erfolg automatisch auch Übereinstimmung? ... 411
6.5.5 Spielt die interperspektivische Übereinstimmung eine moderierende Rolle für den Zusammenhang zwischen Führung und Erfolg? ... 414
6.5.6 Zusammenfassende Diskussion zur Rolle der interperspektivischen Diskrepanz ... 416
6.6 Entstehung und Reduktion von Selbst-Aufwärts-Diskrepanzen ... 421
6.6.1 Hinweise auf die Entstehung von Selbst-Aufwärts-Diskrepanzen ... 421
6.6.2 Ansatzpunkte zur Reduzierung der Beurteilungs-Diskrepanz ... 424

7 DISKUSSION ... 430
7.1 Allgemeine Zusammenfassung ... 431
7.2 Kann man sich auf Führungsbeurteilungen verlassen? Zu ihrer Güte im weitesten Sinne ... 435
7.2.1 Gütekriterien im engeren Sinne ... 435
7.2.2 Interperspektivische Beurteilungsübereinstimmung ... 439

7.2.3 Vergleich der interperspektivischen Korrelationen mit denen anderen Stichproben ... 444
7.2.4 Moderatorvariablen für die interperspektivische Übereinstimmung ... 449
7.2.5 Above-Average-Effekt ... 454
7.3 Wie kann die geringe interperspektivische Übereinstimmung in der Führungsbeurteilung erklärt werden? ... 456
7.4 Wie erfolgsrelevant ist Führung eigentlich? Zur Kriteriumsvalidität der einfachen Führungsbeurteilungen ... 466
7.5 Wieso der Aufwand? Zum Nutzen von multiperspektivischen Führungsbeurteilungen ... 471
7.6 Was bringt die Multidimensionalität? Zum Facettenreichtum von Führung ... 477
7.7 Grenzen der vorliegenden Studie und Ausblick: Eine Baustelle mit Wegweisern für Forschung und Praxis ... 480

8 LITERATURVERZEICHNIS ... 489

9 PERSONENVERZEICHNIS ... 515

10 ABBILDUNGSVERZEICHNIS ... 520

11 TABELLENVERZEICHNIS ... 522

12 ANHANG ... 525

1 Einleitung und Aufbau der Arbeit

Immer häufiger werden berufliche Leistungsbeurteilungen nicht mehr nur dem Vorgesetzten überlassen. Besonders für den Bereich der *Personalführung* werden zunehmend auch Beurteilungen von den geführten Mitarbeitern eingeholt. Da Führungskräfte aber nicht nur ihren Mitarbeitern gegenüber Führungsverantwortung haben, sondern im Zeitalter der „Selbst-Kompetenz“ außerdem das „Sich-Selbst-Führen“ hoch im Kurs steht, werden neben den Fremdbeurteilungen oft auch Selbstbeurteilungen abgegeben. Entsprechend befindet sich die multiperspektivische Führungsbeurteilung unter dem Schlagwort „360°-Feedback“ in einem ungebremsten Aufschwung. Es wird von einem „Zeitgeist“ gesprochen, aus dem die multiperspektivische Leistungsbeurteilung gerade aufgrund eines steigenden Qualitätsbewusstseins und aufgrund zunehmend geforderten Selbstmanagements von Teams und Führungskräften nicht mehr wegzudenken ist (Hedge, Borman & Birkeland, 2001). Entsprechend findet sich kaum ein erfolgreiches Unternehmen (nur 5% der US-amerikanischen Fortune 2000), das sich kein Beurteilungssystem leistet, bei dem nicht mindestens zwei Beurteilungsperspektiven einbezogen werden (Ewen & Edwards, 2001). Für Deutschland ergeben sind vergleichbare Daten (Bungard, Fettel & Jöns, 1997).

Es ist aber nicht die Verbreitung multiperspektivischer Beurteilungen selbst, mit der unweigerlich auch hohe finanzielle Kosten einhergehen,[1] die ein verstärktes Forschungsinteresse fordern lässt, sondern vielmehr die Reichweite der personalpolitischen Entscheidungen, die auf der Basis der Beurteilungsergebnisse getroffen werden (Brutus & Derayeh, 2002). Empirische Studien werden zeigen müssen, ob und unter welchen Bedingungen dies durch eine hinreichend hohe Validität überhaupt gerechtfertigt werden kann.

Besonders dringend sind solche Forschungsprojekte folglich in denjenigen Industriebereichen, in denen sich der Einsatz von multiperspektivischen Führungsbeurteilungen fest etabliert hat, wozu vor allem wirtschaftsstarke Branchen und große Unternehmen, bezogen auf Deutschland unbestreitbar Konzerne der Automobilindustrie, gehören.

Die Relevanz der Automobilindustrie in Deutschland zeigt sich beispielsweise an einem zur wirtschaftlichen Situation gegenläufigen Trend. Trotz der bekanntermaßen ungünstigen generellen Wirtschaftslage nahmen in der Automobilbranche nicht nur die Beschäftigtenzahlen (vor allem bei den Zulieferbetrieben) im Jahre 2003 zu, so dass dort über 770.000 Personen tätig waren, sondern es konnte auch im zehnten Jahr in Folge der Umsatz gesteigert werden (Verband der Automobilindustrie, 2004). Mit der ökonomischen Bedeutung, die diese Branche seit langem trägt (Brunner, 2001), geht auch eine gesellschaftliche Bedeutung und Verantwortung einher. Diese tritt besonders dann zu Tage, wenn das Gütesiegel „Made in Germany“ im Konkurrenzdruck der weltweit produzierenden

[1] Die Kosten werden für die USA auf mehrere Hundert Millionen Dollar jährlich veranschlagt (Conway, Lombardo & Sanders, 2001).

Automobilindustrie zunehmend an Prestige einbüßt (vgl. Kosch, 2004) Dann nämlich werden Konflikte auf medial-öffentlicher Bühne ausgetragen und auf emotionale Weise zu befürchtende Folgen für das ganze Land heraufgeschworen, wie es in kaum einer anderen Branche der Fall ist. Nicht zuletzt vor diesem Hintergrund kommt es auch immer wieder zu politischen Interventionen, was Schröder den Beinamen „Auto-Kanzler" eingebracht hat (z.B. Wendel, 2003). Die Assoziation zwischen Automobilbranche und Wohlstand ist in der Bevölkerung augenscheinlich so stark, dass sich gesunder Journalismus dazu gezwungen sieht, den „Niedergang von Opel & Co" explizit nicht als „Ende für Deutschland" zu interpretieren (Metzger, 2004).

Die Bedeutung der Automobilbranche spiegelt sich aber nicht nur in ihrem positiven Image wider, das beispielsweise durch verhältnismäßig sichere und gut bezahlte Arbeitsplätze oder durch eine hohe Anzahl an Patenten (Verband der Automobilindustrie, 2004) aufrechterhalten wird. Sie zeigt sich auch darin, dass die Automobilindustrie bezüglich verschiedener Organisationspraktiken schon oft eine Vorreiterrolle für andere Unternehmen eingenommen hat (vgl. Wilkens, 1998). Diese betreffen in erster Linie Strukturen und Prozesse der Arbeitsorganisation sowie deren Auswirkungen auf das „Human Resource Management". An dieser Stelle seien lediglich die engen Verbindungen zwischen Ford und der Einführung der Fließbandarbeit, Volvo und den teilautonomen Arbeitsgruppen, Toyota und Lean Management oder – auf der Ebene konkreter Politiken – die Verbindung zwischen Opel und den Rückkehr-/Fehlzeitengesprächen genannt.[2]

Vor diesem Hintergrund wundert man sich fast darüber, dass die psychologische Führungsforschung ihre Fragestellungen verhältnismäßig selten in der Automobilindustrie untersucht. Dies betrifft auch den Bereich der multiperspektivischen Führungsbeurteilung. Da zudem der überwiegende Teil der Studien zur multiperspektivischen Führungsbeurteilung in Organisationen durchgeführt wurde, die ganz offensichtlich andere Führungsstrukturen haben als die klassische Industrie,[3] können deren Ergebnisse nicht blindlings auf die Automobilindustrie oder auf andere produzierende Industrie generalisiert werden.

Die vorliegende Arbeit hat zum Ziel, verschiedene Forschungsfragen der uni- und multiperspektivischen Leistungsbeurteilung im Bereich der Personalführung – und zwar in der Automobilindustrie – empirisch nachzuzeichnen und darüber hinaus weiterzuentwickeln. Schwerpunktmäßig werden zum einen die interperspektivischen Übereinstimmungen[4] und die entsprechenden Moderatorvariablen und zum anderen die Validität der Führungs-

[2] Für eine ausführliche Charakterisierung des Human Resource Managements in der Automobilindustrie wird auf Wilkens (1998) und Jonas, Schaller und Keilhofer (2004) verwiesen. Während erstgenannte Arbeit aus betriebswirtschaftlicher Perspektive verfasst wurde, zeigt die letztgenannte organisationspsychologische Ansatzpunkte.

[3] Dazu gehören der militärische Sektor (z.B. Bass & Yammarino, 1991) genauso wie der öffentliche Sektor, also Dienstleistungsanbieter, Verwaltungen oder Krankenhäuser (z.B. Tartler, Goihl, Kroeger & Felfe, 2003).

[4] Übereinstimmung zwischen den Beurteilungen aus zwei Beurteilungsperspektiven

beurteilungen und der interperspektivischen Übereinstimmung von Führungsbeurteilungen betrachtet, so dass insgesamt eine breit angelegte Fragestellung zur (multiperspektivischen) Führungsbeurteilung bearbeitet wird. Zukünftig müssen die Befunde aus den einzelnen Forschungsteilgebieten, die im Gegensatz zur allgemeinen beruflichen Leistungsbeurteilung erst zögerlich erforscht werden, zu einem übergreifenden Modell integriert werden.

Die gesamte Untersuchung wurde in Form einer Fragebogenerhebung durchgeführt, in der einerseits Führungskräfte und andererseits Mitarbeiter ohne Führungsverantwortung das Führungsverhalten von Managern eines großen Automobilkonzerns beurteilten. Es kann von einer 270°-Beurteilung gesprochen werden, da drei Beurteilungsperspektiven (Selbst-, Aufwärts- und Abwärtsbeurteilungen) vorliegen. In Abbildung 1 wird ein Überblick über den Aufbau der Arbeit gegeben.

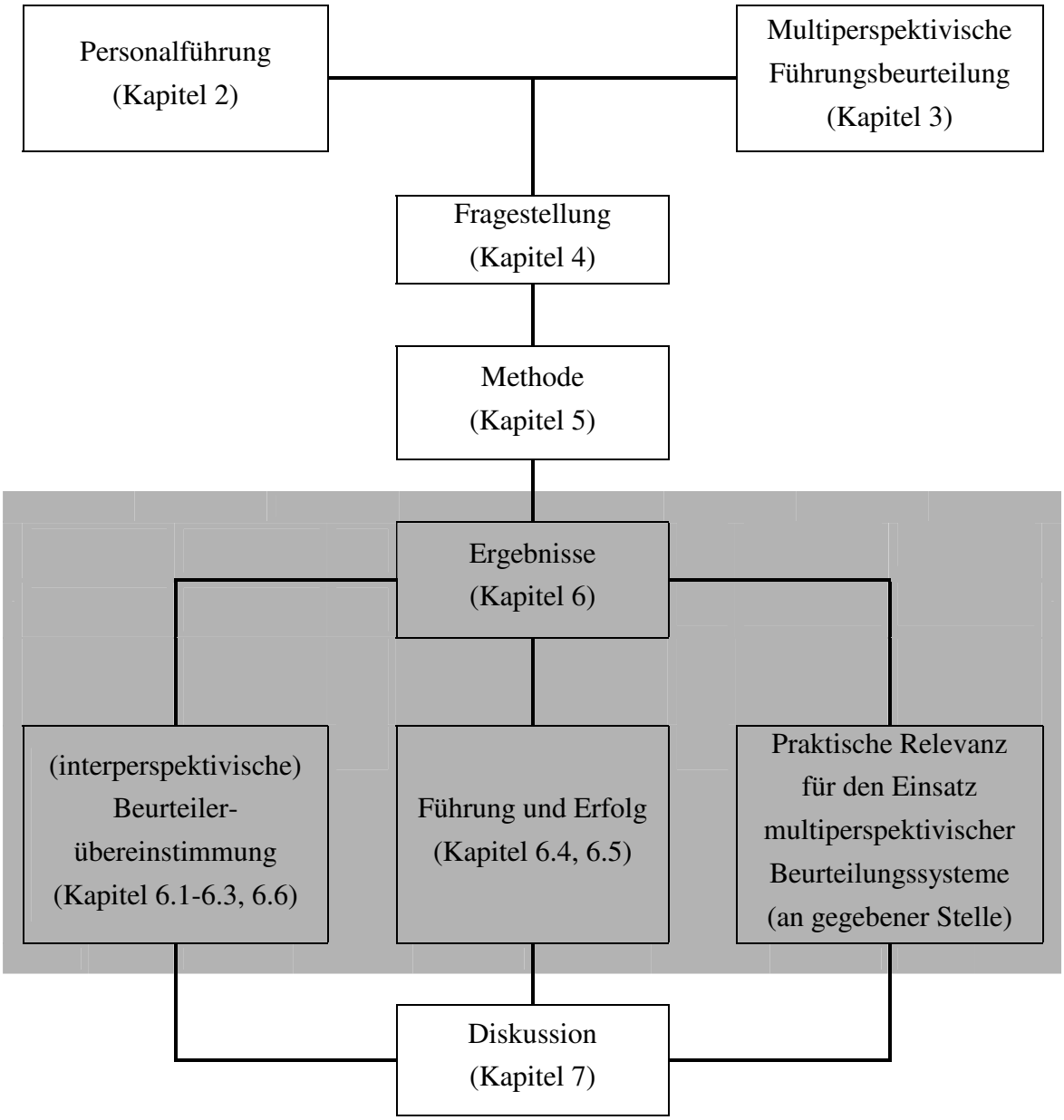


Abbildung 1. Aufbau der Arbeit

In den Kapiteln 2 und 3 werden die Forschungsgebiete der Personalführung und der multiperspektivischen Führungsbeurteilung dargestellt, auf die anschließend in den Ausführungen zur Fragestellung (Kapitel 4) rekurriert wird. Genauer folgen in Kapitel 2 eine Begriffsbestimmung von Personalführung und Führungserfolg sowie ein Abriss über traditionelle, klassische Führungsaufgaben, verhaltenswissenschaftliche Führungstheorien und Führungsmessung. Ausführlich wird das Modell der transformationalen Führung von Bass (1985) behandelt, das die Grundlage für einzelne Hypothesen der vorliegenden Untersuchung darstellt. Kapitel 2 umspannt damit den thematischen Rahmen des allgemeinen Untersuchungsgegenstands – der Personalführung – und begründet damit auch die allgemeine Basis aller bearbeiteten Fragestellungen. Kapitel 3 ist der multiperspektivischen Führungsbeurteilung gewidmet und bereitet dadurch die Mehrzahl der Fragestellungen auf einer konkreteren Ebene vor. Seine Struktur entspricht im Großen und Ganzen der Struktur der Fragestellung und erleichtert dadurch die Einordnung der gewonnenen Ergebnisse. Auffällig ist, dass es an theoretischen Modellen mangelt, so dass dieses Forschungsfeld wie die vorliegende Studie durch eine Aneinanderreihung empirischer Befunde charakterisiert ist. Nach der konkreten Herleitung der Hypothesen im Kapitel zur Fragestellung (Kapitel 4) folgt die Darstellung der Methoden (Kapitel 5), in deren Zusammenhang auch die Skalenkonstruktion beschrieben wird, welche zur Bearbeitung des gesamten Forschungsanliegens notwendig war.

Die Ergebnisse (Kapitel 6) beziehen sich hauptsächlich auf zwei der drei Forschungsgebiete, die Conway und Huffcutt (1997, S. 332) für die multiperspektivische Leistungsbeurteilung ausmachen. Der erste Teil (Kapitel 6.1 bis 6.3) kann der „Grundlagenforschung“ innerhalb der multiperspektivischen Beurteilungen zugeordnet werden. Hier wird das Ausmaß der Übereinstimmung untersucht, das zwischen beobachtetem und erwartetem Führungsverhalten und vor allem zwischen den Führungsbeurteilungen aus verschiedenen Beurteilungsperspektiven besteht. Dabei wird auch die Wirkung potentiell moderierender Variablen überprüft. Der zweite Teil (Kapitel 6.4 und 6.5) untersucht die Kriteriumsvalidität der Führungsbeurteilungen sowie ihrer Übereinstimmung und trägt somit zu einem besseren Verständnis des Zusammenhangs zwischen Führung und Führungserfolg bei. Durchgängig wird versucht, den als „anwendungsbezogen“ charakterisierten dritten Bereich nach Conway und Huffcutt (1997) zu berücksichtigen. Da anwendungsorientierte Forschung auch dem Kriterium der Nützlichkeit verpflichtet ist, werden auf der Basis der gewonnenen Ergebnisse dort, wo es möglich erscheint, immer auch Schlussfolgerungen für den praktischen Einsatz multiperspektivischer Beurteilungssysteme gezogen und der traditionelle Einsatz von multiperspektivischen Beurteilungssystemen insgesamt kritisch betrachtet. Eine Diskussion von potentiellen Erklärungsmechanismen wird zwar am Ende des Ergebnisteils (Kapitel 6.6) bereits vorbereitet, sie stellt dann aber erst in der Diskussion (Kapitel 7) einen zentralen Bestandteil dar. In diesem abschließenden Kapitel werden nicht nur verschiedene

nur verschiedene einzelne Ergebnisse herausgegriffen und integriert, sondern auch Grenzen der vorliegenden Studie mit entsprechenden Verbesserungsideen für die Durchführung zukünftiger Studien aufgezeigt.

2 Personalführung

Führung ist nicht nur ein Begriff, der sich in unserer Alltagssprache häufig wiederfindet. Führung ist auch ein Phänomen, das jeder tagtäglich erlebt und deshalb auch zu kennen und zu verstehen glaubt. Insofern ist Führung ein universelles Phänomen (Bass, 1997, S. 130). Die unübersichtliche Menge an mehr oder weniger wissenschaftlicher Literatur, empirischen Studien und praktischen Ratgebern ist darauf zurückzuführen, dass unterschiedlichste Vorstellungen darüber bestehen, was Führung eigentlich ist und was Führung leisten sollte. Eine Definierung von Führung ist also kein leichtes Unterfangen.

2.1 Begriffsbestimmung

Nähert man sich dem Begriff etymologisch, findet sich als mittelhochdeutsche Wurzel das Wort "vüeren", das sich im Sinne von "fahren" auch ins Althochdeutsche, Altsächsische und Germanische zurück verfolgen lässt. Außerdem scheint nicht ausgeschlossen, dass eine zweite homonyme Wurzel im Sinne von "durchdringen, hinüberbringen" im Sinne von „in Bewegung setzen“ vorliegt (Kluge, 1999). Was sich hier schon andeutet, ist die Komponente der Beeinflussung, die sich durch fast alle neueren Definitionen von Führung zieht (vgl. Neuberger, 2002, S. 12ff.). Der vorliegenden Arbeit wird die Definition von Weibler (2001, S. 29) zugrunde gelegt:

> "Führung heißt andere durch eigenes, sozial akzeptiertes Verhalten so zu beeinflussen, dass dies bei den Beeinflussten mittelbar oder unmittelbar ein intendiertes Verhalten bewirkt."

Diese Definition mag auf den ersten Blick abstrakt und „defensiv“ klingen. Sie setzt aber zwei wichtige Akzente, die die ethische Rechtfertigung von Führung berühren. So wird einerseits die Freiheit der Führenden sowohl in der Wahl der Mittel als auch in der Wahl der Ziele begrenzt, was gleichzeitig andererseits die Stellung der Geführten stärkt. Diese Ausrichtung weist eine hohe Kongruenz mit dem Instrument multiperspektivischer Führungsbeurteilung auf, da auch hier die Stellung der Geführten verbessert wird.

Betrachtet man die Komponenten der Definition im Einzelnen, wird Führung zunächst als eine Art *Beeinflussung* durch eine Person verstanden, die sich auf das Verhalten der Geführten auswirken soll (*Personal*führung). Diese Auswirkung kann entweder darin bestehen, dass gewünschtes Verhalten (häufiger) gezeigt wird oder darin, dass unerwünschtes Verhalten (häufiger) unterlassen wird.

Zusätzlich muss die Einflussnahme von den Geführten *akzeptiert* werden, weil diese erst dann dazu bereit sind, die gewünschten Veränderungen auch durchzuführen (vgl. Reaktanztheorie, Brehm, 1966). Nur unter solchen Bedingungen werden die Vorgesetzten von ihren Mitarbeitern auch als Führungskräfte wahrgenommen (vgl. Social Information Processing Theory; Lord & Maher, 1991). Wird der Führungsanspruch eines Vorgesetzten von seinem Mitarbeiter nicht akzeptiert, liegt nach Weibler (2001, S. 34) lediglich eine Leitungs- und keine Führungsbeziehung vor. Führung ist gemäß dessen also „nie der Beeinflussungsversuch selbst, sondern immer nur der akzeptierte Beeinflussungsversuch, der sich zudem verhaltensmäßig niederschlagen muss“ (Weibler, 2001, S. 34).

Obwohl damit Führung nicht mehr eindeutig an hierarchische Ordnungen geknüpft ist, wird in der vorliegenden Arbeit unter Führung immer die Führung von oben nach unten – entsprechend der formalen Organisationsstrukturen – verstanden. Führungsbeziehungen,

die nicht gleichzeitig Leitungsbeziehungen darstellen, dürfen zur umfassenden Beschreibung eines organisationalen Funktionierens zwar nicht fehlen, bleiben in der vorliegenden Arbeit aber unberücksichtigt.

Durch das Kriterium der *Intentionalität* wird gefordert, dass die Führungskräfte ihre Führung an den Zielen der Organisation und ihrer Mitglieder ausrichten. Dies gewinnt für die vorliegende Arbeit besonders dadurch Bedeutung, dass Führung im Sinne der Führungsgrundsätze operationalisiert wurde und dass letztere die Ziele der Organisation verkörpern.

Indem der Führung nicht nur *unmittelbare*, sondern auch mittelbare, also zeitlich verzögerte Auswirkungen auf die Geführten zugeschrieben werden, wird der Tatsache Rechnung getragen, dass viele Veränderungen, die Führungskräfte anstreben, Zeit und Übung brauchen. Nur sehr direktive Führungsanweisungen schlagen sich sofort in Verhaltensänderungen nieder.

Es ist wichtig festzuhalten, dass es sich bei der Definition lediglich um die Definition von *Personal*führung handelt. Grundsätzlich muss nämlich zwischen personaler und struktureller (entpersonalisierter) Führung unterschieden werden (z.B. Weibler, 2001). Während bei Ersterer die Führung durch Personen (in der Regel Führungskräfte) erfolgt, erfolgt sie bei Letzterer durch bestimmte Organisationsstrukturen. So können die Bürokratie (z.B. Regeln), die Technologie (z.B. Fließband), die hierarchische Differenzierung (z.B. karriereorientierte Ausrichtung des Verhaltens) und die Kultur (z.B. wertorientierte Verhaltensausrichtung an Führungs- und Kooperationsgrundsätzen) viele Arbeitsprozesse auf eine Art und Weise vorgeben, dass die Aufgabenerledigung nur im gewünschten Rahmen möglich ist. Da in der vorliegenden Arbeit jedoch ausschließlich die Personalführung betrachtet wird, soll an dieser Stelle auch die Definition auf diese beschränkt bleiben. Wer an einer ausführlichen Begriffsbetrachtung interessiert ist, sei auf Neuberger (2002) verwiesen, bezüglich struktureller Führung zusätzlich auf Weibler (2001).

Es stellt sich allerdings die Frage, ob Personalführung überhaupt notwendig ist oder ob sie durch die Einführung geeigneter Struktur- und Kontextvariablen kostengünstig und weniger fehlerbehaftet ersetzt werden kann. Und in der Tat: Die „Lücke“, die von den Führungspersonen gefüllt wird, ist nicht groß, wenn man unter Führung auch oben genannte strukturelle Vorgaben fasst, und unter Führungswirkung letztlich jede, auch routinemäßige Arbeitshandlung der Geführten versteht. Doch zum einen werden von den Führungskräften genau die Faktoren gepflegt, durch die strukturelle Führung ausgeübt wird (z.B. Kultur). Zum anderen hat Katz (1964) schon vor vierzig Jahren sehr plausibel dargestellt, dass eine Organisation nicht überleben kann, wenn ihre Mitarbeiter lediglich „Dienst nach Vorschrift“ machen. Das Funktionieren wird erst dadurch gewährleistet, dass Mitarbeiter Verhaltens-

weisen übernehmen, die außerhalb ihrer vorgeschriebenen Rolle liegen. Man könnte mutmaßen, dass gerade diese Zwischenräume durch die immer kürzer werdenden Veränderungszyklen und dadurch steigenden Anpassungsanforderungen zunehmend an Wichtigkeit gewinnen.

In diesem Kontext ist auch die Popularität der transformationalen Führung zu verstehen, die explizit mehr als rigide Aufgabenerfüllung fordert (vgl. Kapitel 2.5). Kaum ein Mitarbeiter wird sich also nur aufgrund der Strukturen dem Unternehmen verbunden fühlen, geschweige denn motiviert zu guter Leistung sein. So sollte die Personalführung nicht durch Begriffe wie "Lückenbüßer" gering geschätzt werden (vgl. Neuberger, 2002, S. 43), denn kein Unternehmen kann sich leisten, sie so zu behandeln. Führungstätigkeit ist nicht vollständig durch Routinen zu gewährleisten. Empirische Ergebnisse charakterisieren die Führungstätigkeit sogar als sehr spontan, reaktiv, konfliktbeladen, politisch und informell (Stewart, 1983) – Anforderungen, die bislang zumindest keine Strukturvariable erfüllen kann. Deshalb kann keinesfalls auf interaktionale Führung, also auf Personalführung verzichtet werden.

Gute Führung gilt sogar „offiziell" als Qualitätsmerkmal von Unternehmen. So geht Führung mit einem Anteil von 10% in die EFQM-Bewertung[5] ein. Hinzu kommen 9% durch Kriterien der Mitarbeiterorientierung (z.B. Zielvereinbarungen), die ebenfalls von Führungskräften geleistet werden müssen. Der Personalführung wird also eine starke Wirkung zugeschrieben. Anders ist dies nicht zu erklären. Darum soll die Begriffsbestimmung beziehungsweise Konzeption von Führungserfolg im folgenden Kapitel in den Mittelpunkt gestellt werden.

[5] Das EFQM-Modell ist das TQM-Modell (Total Quality Management) der European Foundation for Quality Management (EFQM).

2.2 Führungserfolg

"Does leadership make a difference to organizational performance?"
(Titel von Thomas, 1988)

„Leadership: It can make a difference."
(Titel von Smith, Carson & Alexander, 1984).

Die Frage danach, ob sich das ganze Aufsehen um Personalführung überhaupt lohnt, kann neben der negativen Argumentation, dass ihre Wirkung nicht anderweitig erreichbar erscheint (vgl. S. 17), auch positiv und empirisch beantwortet werden. Da Führung in Organisationen darauf ausgerichtet ist, die Ziele der Organisation zu erreichen (vgl. Definition, Kapitel 2.1), ist Führungserfolg für die Organisation von zentraler Bedeutung. Dann muss man sich also fragen, wieviel Varianz Führung aufklären kann.

Auf diese Frage besteht bislang keine befriedigende Antwort, denn zweifellos tragen auch andere Prozesse und strukturelle Führung zum Erfolg bei. Geschätzt wird, dass Führung bis zu 50% des finanziellen Unternehmenserfolgs ausmacht (z.B. Gebert, Steinkamp & Wendler, 1987; Thomas, 1988). Allerdings existieren auch gegenteilige Ansichten, die der Personalführung nur einen mikroskopisch geringen Anteil am Unternehmenserfolg zusprechen (Lieberson & O'Connor, 1972; Salancik & Pfeffer, 1977; Smith, Carson & Alexander, 1984).

Um diesen vermeintlichen Widerspruch aufzulösen, muss zwischen Führungs- und Unternehmenserfolg differenziert werden. Einigkeit besteht nämlich darin, dass das Verhalten einzelner Führungskräfte durchaus weit reichende Wirkungen auf das Verhalten und den Erfolg der Geführten haben kann (Smith et al., 1984, S. 774). Es ist altbekannt, dass verschiedene Personen unter ähnlichen Bedingungen sehr unterschiedliche Leistung erbringen (McCormick & Tiffin, 1974). Genau diese Varianz muss aufgeklärt werden. Konsistente empirische Befunde liegen deshalb nicht vor, weil Führungserfolg im engeren Sinne und Unternehmenserfolg im Sinne der Gesamtorganisation sehr unterschiedlich operationalisiert werden. Man darf nicht den Fehler begehen und lediglich die Wirkung von Vorstandsentscheidungen auf die Leistung des Unternehmens betrachten. So wundert es nicht, dass sich beispielsweise ein Vorstandswechsel (z.B. Iacoccas Wechsel von Ford zu Chrysler) nicht im Profit niederschlägt (z.B. Svyantek & DeShon, 1992), da er kaum direkte Auswirkungen auf die Tätigkeit einzelner Mitarbeiter hat. Eine Änderung in der strategischen Ausrichtung eines Großkonzerns ist einer Änderung in der produzierenden, wertschöpfenden Ebene zeitlich oft lange vorgelagert, so dass sich ein Zusammenhang nur schwer messen lässt (Limpächer, 2003). Viel wichtiger für die Leistung Einzelner (und folglich in ihrer Summe für die Unternehmensleistung) ist der Führungseinfluss auf die direkten Mitarbeiter.

Hinsichtlich der Operationalisierung von Führungserfolg besteht genauso wenig Einigkeit wie über die zugrunde liegenden Wirkmechanismen. Als plausibel erweist sich die Messung von Führungserfolg auf der Seite der *Geführten*.[6] Neuberger (1976) hat dazu verschiedene Betrachtungsebenen vorgeschlagen: die Ebene der Einstellungen (z.B. Zufriedenheit), die Ebene des Verhaltens (z.B. Kooperation im Team) und die Ebene der Leistungsergebnisse (z.B. Qualität). Diese Einteilung geht recht stimmig mit dem Einstellungs-Verhaltens-Modell von Fishbein (1967) einher, nach dem folgende Prozess-Stufen zunehmend stärker mit dem erwarteten Verhalten selbst verbunden sein müssten: Meinungen, Einstellungen, Verhaltensintentionen bis hin zum Verhalten beziehungsweise den Verhaltensergebnissen selbst.

Für eine globale Erfolgsanalyse bietet die Drei-Ebenen-Betrachtung von Kehr (2000) eine pragmatische Lösung. Führung gilt demnach auf der Ebene der Organisation als effektiv, wenn die Leistungsparameter (des Arbeitsteams) positiv ausfallen. Auf der Ebene der Gruppe steht insbesondere die Kohäsion im Mittelpunkt, auf der Ebene des Individuums die Zufriedenheit (vor allem die Arbeitszufriedenheit). Betont werden muss, dass die Ebenen nicht unabhängig voneinander wirken, so dass positive Kennwerte auf allen Ebenen anzustreben sind. Mit beiden Modellen ist die gängige Annahme aus Forschung und Praxis vereinbar, Arbeitszufriedenheits- und Leistungsparameter als Ergebnisse guter Führung zu konzeptualisieren (vgl. z.B. Zusammenschau von Weibler, 2001; vgl. auch Weg-Ziel-Theorie von House, 1971, 1996).

Diese Vielschichtigkeit von Führungszielen findet sich auch in Modellen zum „Human Resource Management“ wieder. So variieren die fokussierten Ergebnisparameter von individuellem Wohlbefinden bis zur Effektivität einer Organisation (Beer, Spector, Lawrence & Mills, 1985). Entsprechend differieren die bislang vorliegenden Studien hinsichtlich der selektiv einbezogenen Kriterien stark (vgl. auch Kapitel 3.1.1). Auch die Erhebungsmethoden sind unterschiedlich, häufig werden die Kriterien von Vorgesetzten eingeschätzt. Nur selten werden Kriterien direkt erhoben (vgl. ausführliche Diskussion bei Lehner, 1995). Stärker effizienzorientierte Erfolgskriterien und -modelle finden insbesondere in der betriebswirtschaftlichen Betrachtung von Personalführung Berücksichtigung. An dieser Stelle wird dazu lediglich auf Witte (1995) verwiesen. Generell gilt, dass aufgrund der unterschiedlichen Operationalisierungen von Führungserfolg – und, wie unten beschrieben wird, auch von Führungsverhalten – der Forschungsstand differenziert betrachtet werden muss, bevor ein umfassendes Modell konstruiert werden kann.

Meindls Arbeitsgruppe griff den Streitpunkt über die tatsächliche Reichweite von Führung auf und prägte ein Label für die vermeintliche Überschätzung des Führungseinflusses: die „Romantisierung der Führung“. Führung werde gewöhnlich als Erklärung für bis dato

[6] Nur selten wird Erfolg über das Gehalt oder die Ranghöhe der Führungskraft selbst operationalisiert.

unerklärte Prozesse oder für das Zustandekommen von Erfolg herangezogen – und dies eben nicht immer gerechtfertigterweise (Meindl & Ehrlich, 1987; Meindl, Ehrlich & Dukerich, 1985). Damit wird jedoch nicht Führung generell als irrelevant verurteilt, es wird „nur“ die relative Wichtigkeit für das Funktionieren von Gruppen und Organisationen angezweifelt, die ihr von den Geführten zugeschrieben wird. Indem Führung als soziale Konstruktion der Geführten konzipiert wird, wird auch die enge Betrachtungsweise konventioneller Führungsmodelle (vgl. Kapitel 2.4) kritisiert (Meindl, 1993, 1998). Damit ist die Kausalitätsfrage insofern aufgeworfen, als in Abhängigkeit vom wahrgenommenen Erfolg unterschiedliche Führungskonstruktionen der Mitarbeiter anzunehmen sind. Die klassischen Führungstheorien gehen umgekehrt davon aus, dass eine bestimmte Führung dazu imstande ist, Erfolg hervorzurufen. Dies ist auch die implizite Grundannahme praktischer und vieler wissenschaftlichen Betrachtungsperspektiven von Führung. Sie liegt ebenso der vorliegenden Studie zugrunde, selbst wenn an gegebenen Stellen die Kausalitätsrichtung diskutiert wird.

Unter Vorausgriff auf die in den Kapiteln 2.4 bis Kapitel 3 dargestellten Befunde muss davon ausgegangen werden, dass das Führungsverhalten mit verschiedenen Variablen auf Geführtenseite zusammenhängt. Auch wenn dieser Zusammenhang möglicherweise nicht vollständig monokausal im Sinne von „Führung erzeugt Erfolg“ erklärt werden kann, ist diese Einflussrichtung stark genug, um Personalführung zu rechtfertigen. Dies bezweifeln selbst kritische Stimmen nicht (Meindl, 1998; Smith et al., 1984).

Auf die Frage, *wie* Führungserfolg erzielt werden kann, gibt es verschiedene Antworten. Aus der Praxis und der wirtschaftswissenschaftlichen Management-Literatur kommen verstärkt normative Kataloge für Führungsaufgaben und -instrumente. Einzelne dieser Aufgaben (z.B. Zielsetzung) sind auch problemlos in psychologische Theorien integrierbar. Dennoch hat die verhaltenswissenschaftliche Führungsforschung weniger Gestaltungsinstrumente entwickelt als vielmehr Theorien aufgestellt, wie Führung (unter bestimmten Bedingungen) ausgeübt werden muss, um maximale Wirkung zeigen zu können. Entsprechend werden in Kapitel 2.3 die gängigsten Führungsaufgaben skizziert und vor dem Hintergrund der Führungsgrundsätze betrachtet, die der Führungsbeurteilung in der vorliegenden Studie zugrunde liegen. Sowohl Kapitel 2.4 als auch Kapitel 2.5 ist den Führungstheorien gewidmet. Da die traditionelle Führung (Kapitel 2.4) nur etwa 16% der Effektivität der Mitarbeiter und Abteilungen erklären kann (Bass, 1985), wurden zunehmend „neue Führungskonzepte“ entwickelt (Kapitel 2.5) – in der Hoffnung, damit mehr Erfolgsvarianz erklären zu können. Führungsinstrumente/-aufgaben und Führungstheorien stehen bislang unverbunden als zwei verschiedene Gestaltungsmittel für erfolgswirksame Führung nebeneinander (z.B. in der Darstellung von Weibler, 2001). Da die Ausgestaltung oder gar der prinzipielle Einsatz bestimmter Führungsaufgaben (z.B. Kritikgespräch, Einsatz von

Symbolen) aber der Umsetzung bestimmter Führungsstile eher dient als der Umsetzung anderer Führungsstile, sind Führungsaufgaben und Führungstheorien in der Praxis indes miteinander verwoben. Insofern sollte auch die Forschung zukünftig einen Versuch unternehmen, verschiedene konkrete Instrumente der Führung in bestehende Führungstheorien zu integrieren. Kapitel 2.6 ist diesem Unterfangen im Sinne einer Zusammenfassung der bis dahin vorgestellten Inhalte gewidmet. Die Unterschiedlichkeit der angenommenen Wirkmechanismen von Führung verweist auf den hohen Forschungsbedarf.

2.3 Führungsaufgaben und -instrumente

Vorab sei angemerkt, dass in der vorliegenden Arbeit ausschließlich die Führungsgrundsätze als klassisches Führungsinstrument beschrieben werden. Die Führungsgrundsätze werden deshalb exklusiv behandelt, weil sie quasi ein fixiertes Abbild erwarteter Führung, also Referenzpunkte für Führungsbeurteilungen darstellen – und Führungsbeurteilungen das thematische Feld der Arbeit bestimmen. Führungsgrundsätze sind somit ein Führungsinstrument, das explizit zur *Verbesserung der Führung* implementiert wird, während andere Führungsinstrumente entweder auf organisationaler Ebene dazu dienen, die strategische Personalpolitik zu regeln (z.B. Vergütungssysteme) oder einzelnen Führungskräften dazu dienen, direkt den Führungsprozess zu unterstützen (z.B. Mitarbeitergespräche, Mitarbeiterentwicklung). Liegt der Fokus auf Letzterem, dann wird in der vorliegenden Arbeit von „Führungsaufgaben" gesprochen, auch wenn eine begriffliche Trennung von Führungsinstrumenten und -aufgaben schwierig ist.[7] Hier sollen Führungsaufgaben also als Spezialfall von Führungsinstrumenten verstanden werden. Es ist einsichtig, dass die Perspektive, aus der man ein Instrument betrachtet, und die Funktion, die man ihm zuschreibt, ausschlaggebend für diese unscharfe Trennung ist. So kann Personalbeurteilung zum einen unmittelbar im Führungsgeschehen wirksam werden (z.B. die Führungskraft beurteilt das Arbeitsverhalten ihrer Mitarbeiter), sie kann aber auch als komplexes Beurteilungssystem verstanden werden, für das die Organisation die entsprechenden Rahmenbedingungen zur Verfügung stellt. In diesem Sinne wäre die Aufwärtsbeurteilung von Führung (vgl. Kapitel 3.1.1) beispielsweise Teil eines komplexen Führungsinstruments.

Für eine umfassendere Darstellung von Führungsinstrumenten und –aufgaben wird auf Malik (2000), Staehle (1999), Weibler (2001), Wunderer (2001) und Rosenstiel, Regnet und Domsch (1999) verwiesen. Im Folgenden werden diejenigen Führungsaufgaben skizziert, deren Relevanz allgemein anerkannt ist (Beurteilen, Ziele setzen, Mitarbeiter entwickeln, Organisieren, Entscheiden). Anschließend wird auf Führungsgrundsätze eingegangen, bevor ein Vergleich zu denjenigen Führungsgrundsätzen vorgenommen wird, die für die hier untersuchte Stichprobe gelten.

[7] Die Unschärfe zwischen den Begriffen spiegelt sich auch in der Literatur wider. So wird beispielsweise die Personalentwicklung bei Wunderer (2001) als Führungsaufgabe, bei Weibler (2001) als Führungsinstrument bezeichnet.

2.3.1 Leistungsbeurteilung

Die Leistungsbeurteilung ist für die organisationsweite Personalpolitik von zentraler Bedeutung. So wird sie beispielsweise im „Human Resource Management"-Ansatz von Tichy, Fombrun und Devanna (1982) als grundlegende Basis für viele personalpolitische Entscheidungen betrachtet (vgl. Abbildung 2).

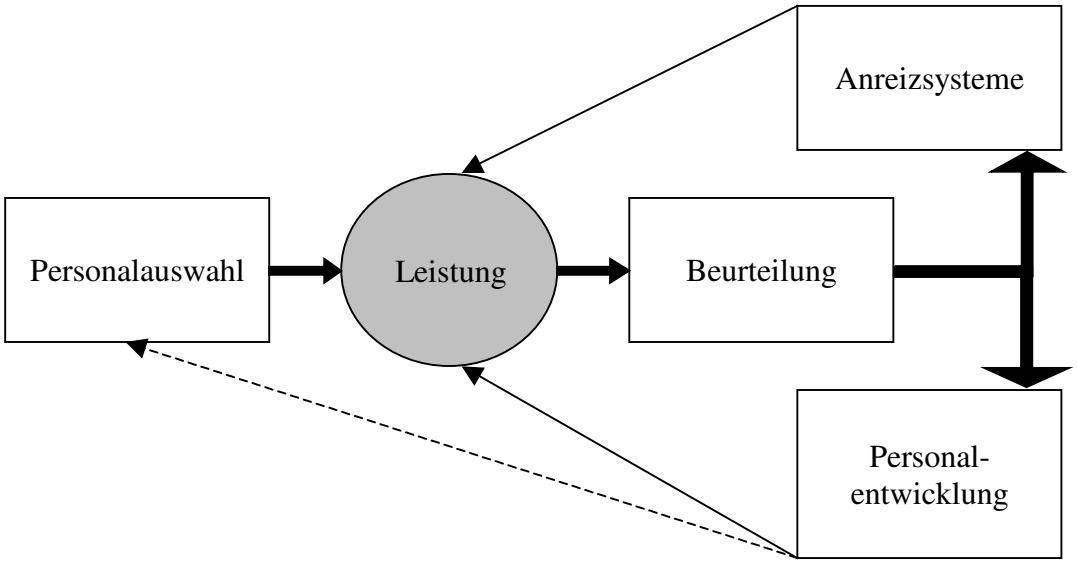


Abbildung 2. Human Resource Management des Michigan-Ansatzes (Tichy et al., 1982)

Nach diesem Modell führt eine positive Leistungsbeurteilung zu belohnenden Konsequenzen, zum Beispiel zu einer Karriereförderung oder zu einer Gehaltserhöhung, während eine negative Beurteilung korrigierende Maßnahmen der Personalentwicklung nach sich zieht. Mit deren Hilfe soll der Mitarbeiter dann in die Lage versetzt werden, die gestellten Leistungsanforderungen erfüllen zu können. Die Leistungsbeurteilung ist also sowohl für die Führungskraft als auch für die Mitarbeiter funktional (Weibler, 2001). Zum einen: Wer seine Mitarbeiter nicht einschätzen kann, kann keine differenzierte Führung leisten. Zum anderen: Die Mitarbeiter werden durch die Beurteilung von oben darin unterstützt, sich selbst zu bewerten, was wiederum Ausgangspunkt für die Ausrichtung ihrer Arbeitsmotivation ist.

Für die Leistungsbeurteilung stehen verschiedene Methoden zur Verfügung. Sie kann mittels freier Eindrucksschilderung oder verschiedener Skalierungsverfahren (Auswahl-, Rangordnung- oder Einstufungsverfahren) vorgenommen werden. Dabei können auf inhaltlicher Ebene prinzipiell Eigenschaften, Fähigkeiten, Verhalten und/oder Ergebnisse betrachtet werden (für einen Überblick: Liebel & Oechsler, 1994; Schuler, 1989).

Aus einer Meta-Perspektive betrachtet, ist die in der vorliegenden Arbeit durchgeführte multiperspektivische Führungsbeurteilung eine spezielle Art von Leistungsbeurteilung (vgl. daher Kapitel 2.7 und Kapitel 3).

2.3.2 Zielsetzung

Die Zielsetzung ist eng mit der Leistungsbeurteilung verbunden. Aus klaren Zielabsprachen ergeben sich einerseits die Beurteilungskriterien, so dass die Zielsetzung als Grundlage glaubwürdiger Beurteilung von Arbeitsverhalten und -ergebnissen betrachtet werden kann. Andererseits sollten auf der Basis jeder Leistungsbeurteilung adäquate Ziele vereinbart werden. Adäquat ist im Sinne der Zielsetzungstheorie von Locke und Latham (1990) so zu verstehen, dass die Ziele spezifisch formuliert und für den Mitarbeiter schwierig (dennoch erreichbar) sind. Im Zielsetzungsgespräch ist außerdem darauf hinzuwirken, dass wichtige Moderatorvariablen bestmöglichst aufzubauen sind (z.B. Zielcommitment, Kompetenzen der Mitarbeiter). Die Zielsetzungstheorie wurde in Form des „Managements by objectives“ (Ferguson, 1973) als Führungs-Tool konkretisiert und populär gemacht. Jedes Zielsetzungsgespräch kann auch vor dem Hintergrund der Erwartungs-Valenz-Modelle oder der VIE-Theorie von Vroom (1964) gestaltet werden. Wenn bei der Zielfindung ein hoher Grad an Partizipation möglich ist, ist der Übergang zur Weg-Ziel-Theorie der Führung (House, 1971, 1996) fließend.

2.3.3 Personalentwicklung

Abbildung 2 zeigt eine enge Verknüpfung zwischen Leistungsbeurteilung und Personalentwicklung. Eine so verstandene Personalentwicklung zielt auf die Behebung aktueller Defizite ab. Versteht man Leistungsbeurteilung auch als Potentialbeurteilung, ist dadurch zudem das Ziel abgedeckt, die Mitarbeiter auf zukünftige Aufgaben vorzubereiten. Zu Letzterem passt, dass Malik (2000) vorschlägt, die Entwicklung von Fertigkeiten durch neue Herausforderungen und Aufgabenbereiche zu erreichen. Dadurch wird die Verbindung zu soeben besprochenen Zielsetzungen deutlich. Letztlich dient die Personalentwicklung in strategischer Hinsicht der langfristigen Deckung des Personalbedarfs (vgl. Liebel & Oechsler, 1994). Nicht zu vergessen ist die individuumsbezogene Perspektive, die eine persönliche Entwicklung jedes Einzelnen anstrebt. In diesem Sinne spricht beispielsweise Malik (2000) davon, nicht Mitarbeiter, sondern „Menschen“ zu entwickeln. Eine ausführliche Darstellung der Personalentwicklung findet sich bei Staufenbiel (1999) oder Sonntag (1999).

Generell können alle Entwicklungsmaßnahmen auf die Basis der Lerntheorien gestellt werden. Zunehmend diskutiert werden auf individueller Ebene Konzepte des Selbstmanagements (vgl. Kanfer, Reinecker & Schmelzer, 2000) sowie auf Organisationsebene Konzepte der lernenden Organisation (vgl. Argyris & Schön, 1999; Senge, 1999).

2.3.4 Organisieren

Organisieren kann auf einer abstrakten Ebene bedeuten, die Gesamtorganisation so zu gestalten, dass deren Ziele erreicht werden. Auf der hier betrachteten konkreten Ebene umfasst Organisieren all die Tätigkeiten einer Führungskraft, die unmittelbar oder mittelbar dazu beitragen, dass die Mitarbeiter mehr Klarheit über Ziele, Prozesse und Strukturen haben. Es geht dabei nicht nur um die Bereitstellung klarer Ziele, sondern auch um die Beseitigung möglicher Hindernisse auf dem Weg zum Ziel. Dazu gehört auch das Implementieren von vernetzten Kommunikationsstrukturen. In Orientierung an Erwartungs*Wert-Motivationsmodellen (Vroom, 1964) muss so organisiert werden, dass das jeweilige Ziel mit hoher Bedeutung versehen wird und dass die Mitarbeiter den Weg so klar vor Augen haben, dass ihre Erwartung, das Ziel zu erreichen, maximal ist. Organisieren bedeutet ausdrücklich nicht ein Maximieren an Besprechungen und abteilungsübergreifender Kommunikation. Eine erfolgreiche Organisation zeichnet sich vielmehr dadurch aus, dass diese Art von Austausch zwar problemlos möglich ist, aber gering gehalten werden kann (Malik, 2000).

2.3.5 Entscheiden

Dass Entscheiden eine zentrale Aufgabe von Führungskräften ist, haben Vroom und Yetton (1973) ausführlich hergeleitet und in einem normativen Entscheidungsmodell der Führung umgesetzt. Im Wesentlichen ergeben sich zwei zentrale Aussagen. Zum einen konzipieren sie Personalführung als kontinuierliche Wahl (Entscheidung!) eines situationsadäquaten Führungsstils (vgl. Kapitel 2.4). Zum anderen ist der Führungsstil insbesondere durch das Ausmaß der Partizipation charakterisiert. Die Wahl eines Führungsstils geht also mit einer Entscheidung für das Ausmaß einher, zu dem Mitarbeiter in Entscheidungen einbezogen werden.

Unter Partizipation ist

- das Zusammenwirken
- bestimmter Personen und/oder Personengruppen (Partizipationssubjekt)

- aus bestimmten Hierarchieebenen (Partizipationsebene)
- bei der Wahrnehmung bestimmter Aufgaben im Unternehmen in allen Phasen des Entscheidungsprozesses (Partizipationsobjekt)
- in einer bestimmten Ausprägung (Partizipationsgrad)
- durch den Einsatz bestimmter Verfahren (Partizipationsinstrumente)
- auf der Basis bestimmter Legitimationen (Partizipationsgrundlage)

zu verstehen (Domsch & Reinecke, 1982, S. 68).

Es zeigte sich, dass durch Mitarbeiterpartizipation die Produktivität (Miller & Monge, 1986) gesteigert werden kann, während ein Mangel an Partizipation mit geringerer Arbeitszufriedenheit und höheren Absentismus- und Fluktuationsraten einhergeht (Brose, 1983). Da derart positive Zusammenhänge zwischen Partizipation und Erfolg nicht durchgängig gefunden werden, wurden bereits Bedingungen erarbeitet, unter denen Partizipation die erwartete Wirkung zeigt (Heller, Pusic, Strauss & Wilpert, 1998). Dazu gehört unter anderem, dass von allen beteiligten Personen Partizipation gewünscht wird und dass sie mit anderen Politiken des „Human Resource Managements“, beispielsweise dem Anreizsystem und Trainingsmodellen, abgestimmt ist. Insbesondere bei der Einführung von Innovationen zeigen Mitarbeiter, die nicht einbezogen wurden, Widerstände (Ram & Jung, 1991). Dies sollte auch bei der Einführung von neuen, beispielsweise multiperspektivischen Beurteilungssystemen bedacht werden.

Betrachtet man das Ausmaß der Partizipation dimensional, dann steht am Pol der stärksten Ausprägung die *Delegation*. Sie umfasst die Übertragung von komplexen Aufgabenstellungen mit den dazu erforderlichen Rechten, Entscheidungsbefugnissen und Ressourcen sowie die daraus resultierende Handlungsverantwortung (Gebert, 2002). Zur Delegation kommt es gewöhnlich erst dann, wenn die bis dahin gezeigte Leistung als positiv bewertet wurde. Insofern steht auch die Delegation mit der Leistungsbeurteilung (Kapitel 2.3.1) in einem abhängigen Verhältnis (vgl. Bauer & Green, 1996). Umgekehrt wird erwartet, dass sich durch die herausfordernden, neuen Aufgaben und Ziele (siehe Zielsetzung) die Leistung verbessert.

In Abhängigkeit vom Partizipationsgrad kann der Mitarbeiter Autonomie erleben, was gemäß des „Job Characteristics Models“ (Hackman & Oldham, 1976, 1980) zu einer Steigerung des Motivationspotentials beitragen kann. Ähnlich kann die durch die neuen Aufgabengebiete gesteigerte Variabilität wirken. Unter gar nicht unwahrscheinlichen Umständen werden durch die delegierten Aufgaben zudem zwei weitere Variablen dieses Modells positiv beeinflusst, die Sinnhaftigkeit und die Ganzheitlichkeit. Partizipation kann also durchaus als Motivationsfaktor bezeichnet werden. Wenn außerdem die letzte kritische Variable des Modells, die Rückmeldung, angesprochen werden soll, ist insbesondere auf die Funktion der Leistungsbeurteilung zu verweisen.

Führungsaufgaben, wie sie auf den vergangenen Seiten angerissen wurden, sind als normative Anforderungen an Führungskräfte zu verstehen, für die zum Teil Kataloge mit entsprechenden Führungswerkzeugen erstellt wurden (Malik, 2000).[8] Aufgrund ihrer starken Verwurzelung im praktischen Management haben sie für Führungskräfteschulungen oder als Grundlage für Selektionsinstrumente seit jeher eine hohe Popularität und entsprechende Bedeutung für die organisationspsychologische Praxis (Sarges, 1995).

2.3.6 Führungsgrundsätze

Im vorliegenden Kapitel wird zunächst allgemein das Führungsinstrument der Führungsgrundsätze beschrieben (Kapitel 2.3.6.1). Anschließend werden die Inhaltsbereiche derjenigen Führungsgrundsätze, die für das Unternehmen gelten, in dem die vorliegende Studie durchgeführt wurde, mit oben vorgestellten Führungsaufgaben abgeglichen (Kapitel 2.3.6.2).

2.3.6.1 Definition und Funktion von Führungsgrundsätzen

Führungsaufgaben, wie sie auf den vergangenen Seiten angerissen wurden, können in Form von Führungsgrundsätzen zusammengefasst werden. Dies hat den Vorteil, dass ein negatives Charakteristikum von Führungsaufgaben, ihre Unstrukturiertheit (Neuberger, 2002, S. 43), im Sinne der Unternehmensanforderungen abgeschwächt wird, indem ein mehr oder weniger differenzierter Regel- und Zielkatalog für das Führungsverhalten entworfen wird. Wunderer (2001, S. 385) definiert Führungsgrundsätze folgendermaßen:

> „Führungsgrundsätze beschreiben und/oder normieren die Führungsbeziehungen zwischen Vorgesetzten und Mitarbeitern im Rahmen einer ziel- und wertorientierten Führungskonzeption zur Förderung eines erwünschten organisations- und mitgliedergerechten Sozial- und Leistungsverhaltens. Sie können in schriftlicher Form (explizit) verbindlich fixiert werden oder als ungeschriebene Normen zur Verhaltensorientierung in den Führungsbeziehungen dienen.“

Führungsgrundsätze stellen also eine Orientierungshilfe für Führungskräfte dar, indem sie mehr oder weniger detailliert vorgeben, wie Führung ausgestaltet werden soll. Insofern verdeutlichen sie auch eine Wertebasis und sind auf das Erreichen von Organisationszielen

[8] Die Anfänge dieser Strömung sind bei Fayol (1949 [1916]) zu suchen.

ausgerichtet. Damit zählen Führungsgrundsätze, die in nahezu allen großen deutschen Unternehmen vorliegen (Gabele, Liebel & Oechsler, 1992), zu den wichtigsten Führungsinstrumenten. Dass deshalb aber nicht unbedingt auf einen entsprechenden betrieblichen Alltag zu schließen ist, bestätigen Aussagen von Führungskräften, nach denen solche Systeme bei höchstens einem Drittel der befragten Betriebe auch umgesetzt werden (Zimolong & Elke, 2001). Im Hinblick auf den Organisationserfolg wäre dies allerdings erstrebenswert, da in den überdurchschnittlich erfolgreichen sowie in den sich verbessernden Unternehmen mehr Personalsysteme umgesetzt werden als in den sich verschlechternden beziehungsweise in den unterdurchschnittlich erfolgreichen Unternehmen (Beckmann, Zimolong, Stapp & Elke, 2001).[9]

Entsprechend ihrer zugeschriebenen Relevanz finden die Führungsgrundsätze auch Platz im „7S-Modell“ von Pascale und Athos (1981), in dem die wichtigen Elemente exzellenter Unternehmen zusammengefasst sind. Die Führungsgrundsätze sind darin dem „harten S“ *Systems* zuzuordnen, das sehr eng mit dem „weichen S“ *Style* (im vorliegenden Unternehmen dem kooperativen Führen selbständiger Mitarbeiter) in Verbindung steht (vgl. Anhang A). Für Mitarbeiter stellen sie eine Grundlage dafür dar, welche Führung sie erwarten und gegebenenfalls von ihren Führungskräften fordern können.

Um wirksam zu werden, genügt es nicht, Führungsgrundsätze aufzustellen. Den Führungskräften kann es nur dann gelingen, sich entsprechend zu verhalten, wenn sie erstens die Inhalte kennen, zweitens die entsprechenden Kompetenzen ihrer Umsetzung besitzen und drittens dazu motiviert sind, die geforderte Führung auch umzusetzen. So ist zu erklären, weshalb Führungsgrundsätze mit Bezug zur Zukunft formuliert sein sollten, unter Einbeziehung von Führungskräften und Mitarbeitern entwickelt werden und mit anderen Führungsinstrumenten abgestimmt sein sollten (Wunderer, 2001). Wenn die Ziele der geforderten Führung akzeptiert werden, kann auch mit einer entsprechenden Handlungsmotivation gerechnet werden. In diesem Zusammenhang kritisiert Weibler (2001, S. 373) die mangelnde Sanktionsfähigkeit von Führungsgrundsätzen in der Praxis.

Für den interessierten Leser bietet Wunderer (1983) eine ausführliche Betrachtung von Führungsgrundsätzen aus unterschiedlichen Perspektiven (Wissenschaft, Organisationsentwicklung, Verbände, Wirtschaft, öffentliche Verwaltung). Stärker an den Funktionen orientiert bleibt die Darstellung bei Gabele et al. (1992), wo Führungs- und Unternehmensgrundsätze allerdings nicht getrennt werden und daher nicht nur Führungskräfte und Mitarbeiter als Adressaten aufgenommen sind.

[9] Erfolg wurde in diesem Forschungsprojekt als mittlere Größe verschiedener Erfolgsindikatoren im Arbeits- und Gesundheitsschutz operationalisiert.

2.3.6.2 Führungsgrundsätze des fokussierten Unternehmens im Abgleich mit klassischen Führungsaufgaben

Die acht schriftlich fixierten Führungsgrundsätze des fokussierten Unternehmens umfassen folgende Führungsaufgaben, deren Original-Beschreibungen in Anhang B nachzulesen sind:

- Qualifizieren der Mitarbeiter für ihre Aufgaben (Fähigkeiten entwickeln)
- Qualifizieren für höherwertige Aufgaben (Mitarbeiter fördern)
- Aufgaben und Befugnisse übertragen (Delegieren)
- Stellvertretung
- Zusammenarbeiten
- Ziele erarbeiten und setzen
- Arbeitsergebnisse, Arbeits- und Führungsverhalten beurteilen
- Arbeitsergebnisse, Arbeits- und Führungsverhalten steuern

Die anvisierten Führungsaufgaben decken sich sehr gut mit den klassischen Führungsaufgaben, die im vorangehenden Kapitel 2.3 beschrieben wurden. So betreffen die beiden ersten Führungsgrundsätze die Personalentwicklung und die beiden letzten die Leistungsbeurteilung. Leistungsbeurteilung wird bei Malik (2000) weiter gefasst und als „Kontrollieren“ konzipiert, was noch deutlicher macht, dass das Steuern von Arbeitsergebnissen, Arbeits- und Führungsverhalten zu diesem Bereich gehört. Das Delegieren wurde im Zuge des Entscheidens als Endpol von Partizipation dargestellt. Die Stellvertretung kann sowohl als Personalentwicklung betrachtet werden als auch im Sinne von Organisieren verstanden werden. Das Zusammenarbeiten spricht insbesondere Punkte der Partizipation und des Organisierens von Arbeitsabläufen an, geht aber vor allem dort über die oben beschriebenen Aufgaben hinaus, wo es um die Forderung von unterstützend-kooperativem Arbeitsverhalten geht. Die Zielsetzung wurde explizit als Führungsaufgabe beschrieben. Bei der Schilderung der einzelnen Führungsaufgaben wurde auf psychologische Theorien Bezug genommen, die ihre Wirkung erklären können, so dass prinzipiell von der Effektivität der aufgestellten Führungsgrundsätze auszugehen ist.

Auffällig ist, dass keine unternehmens- oder branchenspezifischen Führungsaufgaben fixiert sind, so dass von einer generellen Funktionalität der vorliegenden Führungsgrundsätze und demzufolge auch des konstruierten Führungsfragebogens (vgl. Kapitel 5.5.1) auszugehen ist. Mit Ausnahme der Stellvertretung sind auch alle Führungsgrundsätze in die Kategorien einzuordnen, die eine Dokumentenanalyse von Führungsgrundsätzen ergab (Schilling, 2004).

Betrachtet man die Definition von Führungsgrundsätzen sowie die Anforderungen an ihre Ausrichtung (vgl. Kapitel 2.3.6), dann ist eine kenntliche Wertebasis gefordert. Diese wird in den konkret vorliegenden Führungsgrundsätzen in Form eines kooperativen Führungsverhaltens sehr deutlich propagiert. Dass dies in vielen Führungsgrundsätzen der Fall ist, stellt die Dokumentenanalyse von Schilling (2004) heraus. Nicht herausgearbeitet sind die Ziele, die durch das geforderte Führungsverhalten erreicht werden sollen. Generell ist keine Ausrichtung auf die Zukunft gegeben, so dass die Führungskräfte daraus keine Motivation zur Umsetzung der Verhaltensrichtlinien schöpfen können. Wo die Ziele den Führungskräften aber nicht bekannt sind, können sie auch den Mitarbeitern nicht vermittelt werden, so dass sich ein Motivationsdefizit über die Ebenen hinweg übertragen kann.

Indem sich die Fragebogenkonstruktion auf die Führungsgrundsätze stützte (vgl. Kapitel 5.5.1.1) und der Fragebogen zur Führungsbeurteilung eingesetzt wurde, rücken die Führungsgrundsätze in einen bewertbaren Raum. Dies erhöht zum einen die Konsistenz personalpolitischer Instrumente, zum anderen wird die Umsetzung der Führungsgrundsätze sanktionierbar.

Normative Modelle, wie sie in diesem Kapitel vorgestellt wurden, provozieren ein Bild von Führung, wie es nur teilweise durch die so genannte „work activity"-Forschung bestätigt wird. Dieser Forschung geht es um eine möglichst detailgetreue Beschreibung dessen, was Führungskräfte tatsächlich tun. Es zeigt sich beispielsweise, dass das Arbeitsverhalten von Führungskräften viel fragmentierter ist, stärker auf informell-politischen Wegen stattfindet, stärker konfliktbeladen sowie weniger planbar ist als in funktionalen Konzepten vorgesehen. Dass in der Unternehmenspraxis diese Lücke zwischen Norm und wahrgenommener Realität bekannt ist, zeigt sich in Fragestellungen derart, wie sie auch dem vorliegenden Projekt voraus gingen: „Wir haben Führungsgrundsätze eingeführt. Aber werden sie denn auch gelebt?" (vgl. S. 29).

Grundsätzlich ist aufgrund dieser Diskrepanz jedoch zu fragen, ob sich die normativen Führungsaufgaben überhaupt dazu eignen, ein vernünftiges Soll darzustellen. Diese Frage reicht allerdings über die vorliegende Arbeit hinaus. Zumindest muss davon Abschied genommen werden, Führungsaufgaben als allgemeingültig zu betrachten. Stattdessen unterliegen sie einem zunehmenden Veränderungsdruck und müssen beispielsweise dann verändert werden, wenn die Arbeitsorganisation umgestaltet wird. Exemplarisch muss Leistungsbeurteilung in einer traditionellen Linienorganisation anders ausgestaltet werden als in einer Projektorganisation.

Relativ unbeeinflusst von den normativen Führungskonzeptionen blieb die verhaltenswissenschaftliche oder sozialpsychologische Führungsforschung. Sie entwickelte verschiedene Führungstheorien, mit denen sich die beiden folgenden Kapitel 2.4 und 2.5 beschäftigen.

2.4 Traditionelle Führungstheorien

Die Führungstheorien werden im Folgenden mit unterschiedlichem Detaillierungsgrad beschrieben. Dort, wo es für die Einordnung der vorliegenden Arbeit gewinnbringend erscheint, werden sie ausführlicher dargestellt. Dies betrifft insbesondere die transformationale Führung, der deshalb auch ein gesondertes Kapitel gewidmet wird (Kapitel 2.5), zumal die Befunde bislang nicht entsprechend zusammengetragen wurden. Abschließend wird in Kapitel 2.6 eine kritische Würdigung verschiedener Theorien vorgenommen. Für deren umfassendere Darstellung kann auf Neuberger (2002), Weibler (2001) und Wunderer (2001) verwiesen werden.

Generell können die Führungstheorien nach ihrem Fokus klassifiziert werden. So können personen-, interaktions-, positions- und situationszentrierte (kontingenztheoretische) Ansätze unterschieden werden (Abbildung 3).

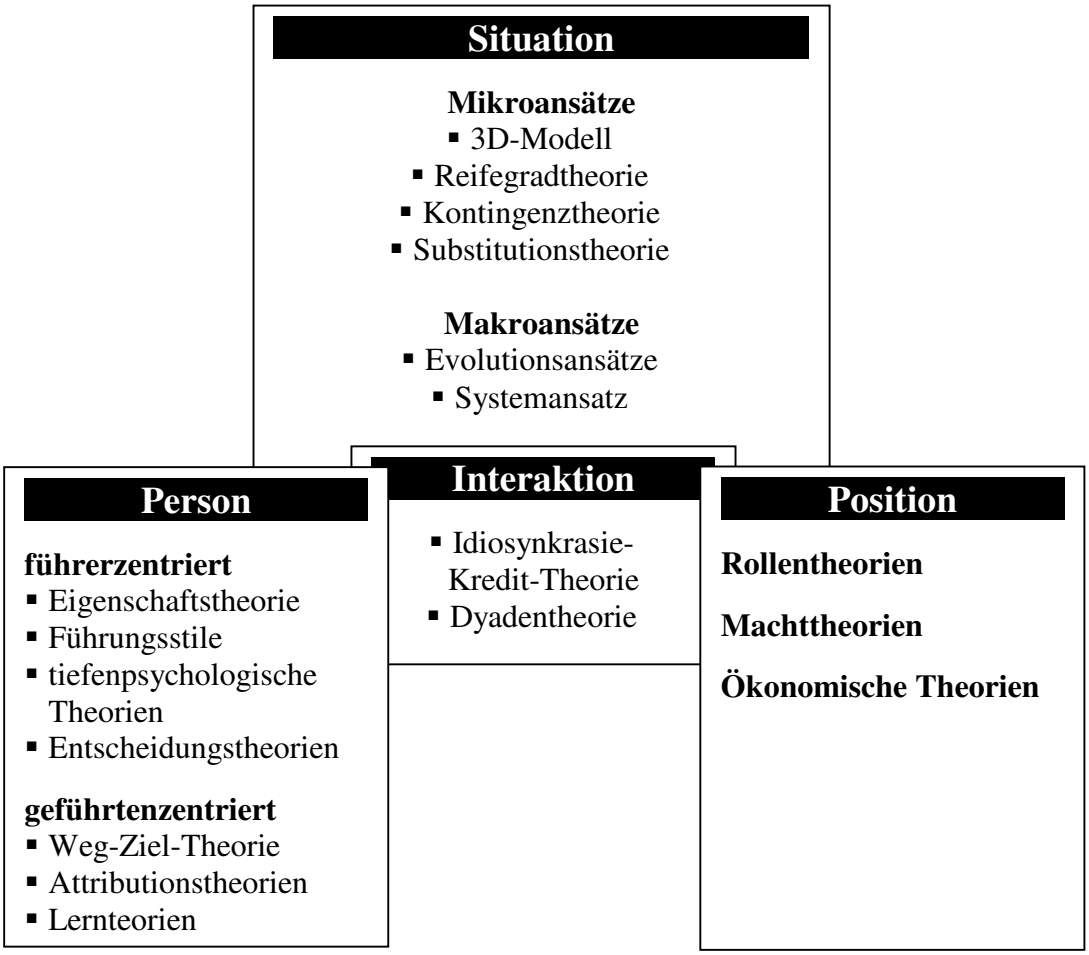


Abbildung 3. Überblick über führungstheoretische Ansätze (modifiziert nach Wunderer, 2001, S. 274)

In der vorliegenden Arbeit werden weder positionsorientierte Ansätze noch situationsbezogene Makroansätze vorgestellt. Auch aus den anderen Gruppen an Führungstheorien werden jeweils nur einzelne Modelle exemplarisch angerissen, um das jeweils zugrunde liegende Prinzip zu illustrieren (Abbildung 4, Abbildung 5, Abbildung 7 und Abbildung 8). Ein anderes Ziel wird im Rahmen dieser Arbeit nicht verfolgt. Die Einordnung der transformationalen Führung wird im Anschluss an ihre Darstellung in Kapitel 2.6 diskutiert.

Unter dem **eigenschaftstheoretischen Ansatz** der Führung werden alle Ansätze verstanden, die der Persönlichkeit des Führenden die zentrale Bedeutung für die Generierung von Führungserfolg zuschreiben (vgl. Abbildung 4). Es geht also um die Frage, welche Eigenschaften erfolgreiche von nicht erfolgreichen Führungskräften unterscheiden. Eine solche Vorstellung ist monokausal und ignoriert beispielsweise die Situation oder die Interaktion zwischen Führendem und Geführten. Nicht zu vergessen ist, dass Management zu einem Massenberuf geworden ist (Malik, 2000) und eine entsprechend hohe Anzahl an Menschen mit ähnlich herausragenden, spezifischen Eigenschaften kaum zu erwarten ist. Insofern verliert die Suche nach einem spezifischen Eigenschaftsprofil ihren praktischen Nutzen. Dieser ist bei den Modellen höher, die eine Erlernbarkeit der erfolgskritischen Variablen postulieren.

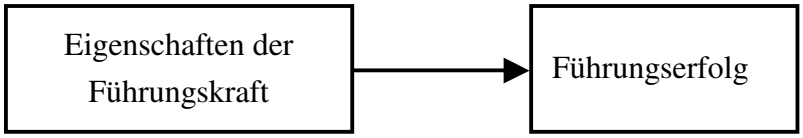


Abbildung 4. Prinzip der Eigenschaftstheorie der Führung (nach Neuberger, 2002)

Inzwischen liegen Überblicksartikel und Meta-Analysen vor (Kirkpatrick & Locke, 1996; Lord, DeVader & Alliger, 1986; Lord & Hall, 1992; Stogdill, 1948). Die zentralen Erkenntnisse sind folgende:

- Eine Vielzahl an Eigenschaften korreliert mit Führungserfolg; es konnte kein spezifisches Persönlichkeitsmuster extrahiert werden.
- Die Zusammenhänge sind in der Regel gering. Auch der Zusammenhang zwischen Intelligenz und Führung, der lange Zeit als stark bedeutsam betrachtet wurde (27% Varianzaufklärung in der Meta-Analyse von Lord et al., 1986), muss relativiert werden. Nach den Ergebnissen der differenzierten Meta-Analyse von Judge, Colbert und Ilies (2004) korreliert Intelligenz (Papier-Bleistift-Verfahren) mit wahrgenommener Emergenz von Führung „nur“ zu $\rho = .25$, mit wahrgenommener Effektivität zu $\rho = .17$ und mit objektiver Effektivität zu $\rho = .33$ (korrigierte Indizes

um Unreliabilität und Rangerestriktion). Da Intelligenz generell mit beruflicher Leistung korreliert (Schmidt & Hunter, 1998), verliert sie dann auch ihren spezifischen Wert.

- Die untersuchten Eigenschaften der Führungskräfte treten mit einem hohen Überlappungsbereich auch bei den Geführten auf.
- Die Erklärungsleistung vieler Eigenschaftskonstrukte variiert zwischen den Studien stark.
- Die Führungskräfte variieren hinsichtlich der Ausprägung bestimmter Eigenschaften deutlich.

Die Ergebnisse sind einerseits ernüchternd. Andererseits bedeuten sie nicht, dass Persönlichkeitseigenschaften als irrelevant zurückgewiesen werden sollen. Es ist weder erstaunlich, dass einzelne Eigenschaften keine hohe Varianzaufklärung erzielen, noch verwundert es, dass die Eigenschaften von Führenden und Geführten nicht stark differieren, da fast alle Führenden auch Geführte sind. Außerdem ist zu diskutieren, wie das Kriterium für den Führungserfolg operationalisiert wird. Häufig wurde in diesen Studien der Karriereerfolg (z.B. Beförderungsgeschwindigkeit, Gehalt) herangezogen. Die Brauchbarkeit dieses Kriteriums ist anzuzweifeln, da es von vielen Faktoren abhängt und keineswegs stabil ist. Zukünftig geht es um eine Integration von Persönlichkeitsvariablen (z.B. Big Five, Dominanz, Leistungsmotivation, Intelligenz) in komplexe Führungsmodelle. Die Big Five erklären gemeinsam 28% der Varianz an Führungsemergenz und 15% der Varianz an Führungseffektivität (korrigierte Indizes der Meta-Analyse von Judge, Bono, Ilies & Gerhardt, 2002).[10]

Zu Beginn der 50er Jahre verlagerte sich der Schwerpunkt der Führungsforschung auf die Untersuchung des **Führungsstils** (operationalisiert als Führungsverhalten). Dieses wurde zunächst in Form von relativ situationsunabhängigen und zeitstabilen Führungsstilen konzipiert, die mehr oder weniger starken Erfolg versprechen (vgl. Abbildung 5). Die Tradition der empirischen Führungsstilforschung reicht zurück bis zur sehr bekannt gewordenen Arbeit von Lewin, Lippitt und White (1939), den sogenannten Iowa-Studien. Dort wurde der Führungsstil experimentell manipuliert (Laissez Faire, demokratischer und autoritärer Stil) und in seiner Auswirkung auf die geführten Gruppen (amerikanische Jungen, Bastelgruppe) untersucht. Laissez Faire und autoritäre Führung können dabei als Endpunkte einer Dimension hinsichtlich der Ausprägung von Mitbestimmung bzw. Kontrolle betrachtet werden, der kooperative Führungsstil liegt in der Mitte. Es kristallisierte sich heraus, dass der demokratische/kooperative Führungsstil die Einstellungsmaße (vor allem Zufriedenheit) am positivsten beeinflusst. Was Leistungsmaße angeht, sind der autoritäre

[10] Die größten Anteile gehen auf Extraversion, Offenheit und Gewissenhaftigkeit zurück.

und demokratische Führungsstil ähnlich wirkungsvoll (vgl. Überblick Neuberger, 1972, 2002). Die *Anwesenheit* der Führungskraft scheint moderierende Wirkung zu haben.

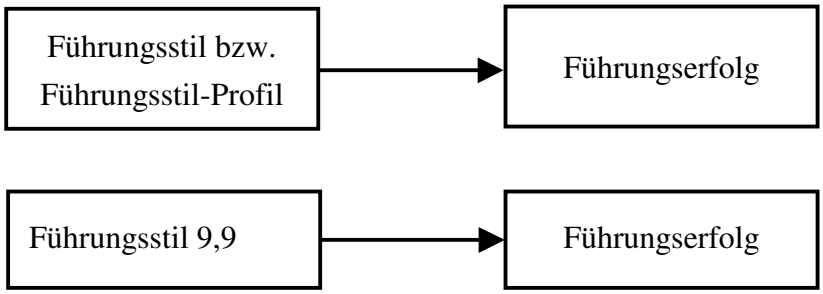


Abbildung 5. Prinzip der Führungsstil-Modelle und beispielhafte Übertragung auf das Managerial Grid von Blake und Mouton (1964)

Mit der Führungsstilforschung kam es zu einem enormen Anstieg an empirischen Untersuchungen. Während bei Lewin et al. (1939) die Führungsstile vorab, quasi am Schreibtisch konzipiert wurden, extrahierten sowohl die Ohio-Arbeitsgruppe um Fleishman als auch die Michigan-Arbeitsgruppe um Likert aus Befragungsergebnissen einander sehr ähnliche Zwei-Faktoren-Modelle. Die Führungsstile, die sich in der Folge in vielen verschiedenen Führungstheorien wiederfanden (Abbildung 6), waren geboren. Bei Fleishman (1953, 1973) wurden sie Initiating Structure und Consideration getauft, bei Likert (1961) Aufgaben/Produktions- und Mitarbeiterorientierung. Im einen Fall geht es jeweils um die Organisation von verschiedenen Abläufen zur Unterstützung der Aufgabenerfüllung, wozu durchaus sehr aktiv-direktives Führungsverhalten genutzt werden kann. Im anderen Fall geht es um eine vertrauensvolle Beziehung zwischen Führungskraft und Mitarbeiter und insgesamt eher um Klima-Variablen. Von der Ohio-Gruppe (nicht von der Michigan-Gruppe) werden die beiden Dimensionen als unabhängig voneinander konzipiert, so dass auch beide Führungsstile von einer Führungskraft gezeigt werden können. So kann das Zwei-Faktoren-Modell der Ohio-Gruppe als unmittelbarer Vorläufer des „Managerial Grid“ betrachtet werden, denn dieses Gitter an potentiellen Führungsstilen ergibt sich aus einem Koordinatensystem mit zwei Achsen – der jeweils neunfach abgestuften produktions- und menschbetonten Führung (Blake & Mouton, 1964).

Es zeichnet sich die Tendenz ab, dass hochproduktive Gruppen eher eine mitarbeiterorientierte Führungskraft haben, während gering produktive Gruppen eher eine leistungsorientierte Führungskraft haben (Likert, 1961). Verschiedene Studien weisen aber auch darauf hin, dass eine hohe Ausprägung von Consideration und zugleich von Initiating Structure am effektivsten ist, während eine niedrige Ausprägung beider Faktoren am wenigsten effektiv ist (Überblicksarbeit von Fleishman, 1973). Im Managerial Grid wird eine Maximalausprägung auf beiden Dimensionen als ideale Führung betrachtet (Führungsstil

9,9, vgl. Abbildung 5). Im Zweifelsfall soll eher in Richtung Mitarbeiterorientierung abgewichen werden (vgl. Zusammenfassung bei Gabele et al., 1992).[11]

Für die Mitarbeiterorientierung zeigt sich eine verhältnismäßig konsistente Befundlage dahingehend, dass deutlich positive Zusammenhänge mit Erfolgsvariablen, insbesondere Zufriedenheit, gefunden werden. Die demgegenüber eher inkonsistenten Befunde zum Zusammenhang zwischen Initiating Structure und Leistung/Effektivität führten dazu, dass die kontingenztheoretische Sichtweise angestoßen wurde (vgl. Fiedler, 1973; vgl. Abbildung 6). So wurde die Ursache für die nicht immer zufriedenstellende Erfolgsaufklärung durch eigenschafts- oder reine führungsstiltheoretische Annahmen darin gesehen, dass zentrale Randbedingungen unberücksichtigt blieben.

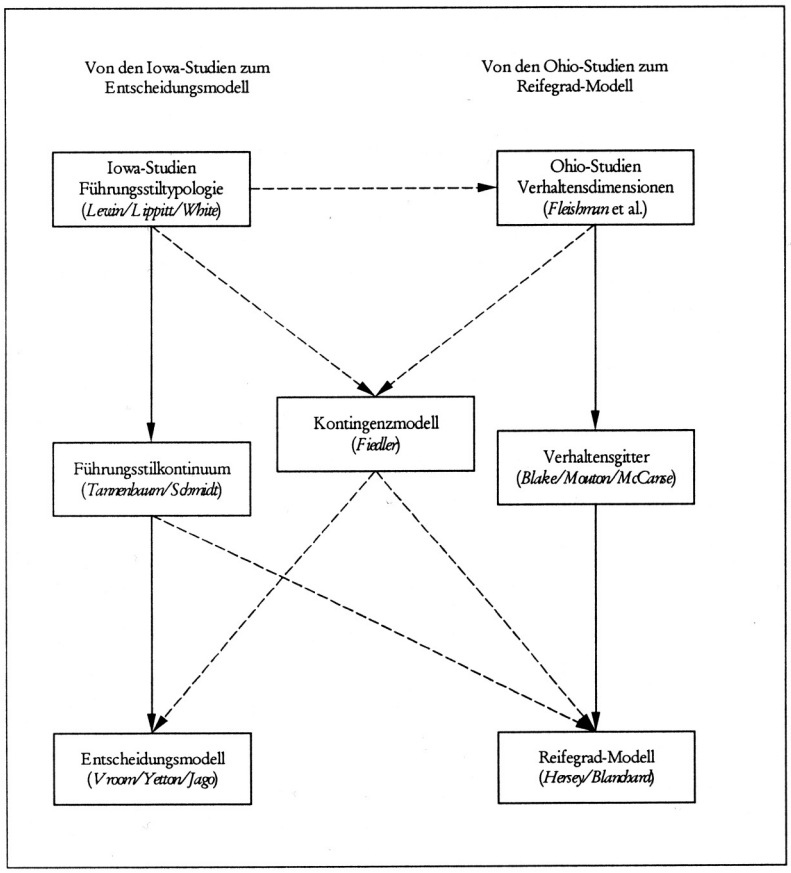


Abbildung 6. Entwicklungslinien der Führungsstilforschung (aus Weibler, 2001, S. 293)

[11] Dadurch, dass für das Managerial Grid ein Trainingsprogramm entwickelt wurde (Blake & McCanse, 1995), wird explizit auf die Erlernbarkeit von erfolgreichem Führungsverhalten verwiesen. Dies steht im deutlichen Unterschied zu den Eigenschaftsansätzen.

Die Mitbestimmungsdimension von Lewin et al. (1939) wurde mit vier weiteren Zwischenstufen zum Führungsstilkontinuum erweitert (Tannenbaum & Schmidt, 1958; leicht modifiziert durch Wunderer, 2001, S. 209). Das Kontinuum reicht in diesem Modell von autoritär über patriarchalisch, informierend, beratend, kooperativ und delegativ bis autonom. Der größte Unterschied zu Vorläufermodellen bestand darin, dass alle Führungsstile im potentiellen Verhaltensraum jeder Führungskraft liegen sollten. Sie sollten in Abhängigkeit von den Bedürfnissen der Mitarbeiter, von den eigenen und den Organisations-Zielen sowie insgesamt von situativen Anforderungen gewechselt werden. Der bestmöglichsten Passung, die allerdings nicht weiter konkretisiert ist, wird eine optimale Erfolgsausschöpfung zugesprochen. Hier kann der Beginn der Kontingenzforschung ausgemacht werden, auch wenn keine systematische Forschung betrieben wurde.

Sogenannte **Kontingenzansätze** spezifizieren das erfolgreiche Zusammenspiel von Führungsverhalten und Situationsvariablen genauer (vgl. Abbildung 7). Beispielhafte Vertreter dieser Gruppe sind das Kontingenzmodell von Fiedler (1967, 1971), das 3-D-Modell von Reddin (1977), das Reifegradmodell von Hersey und Blanchard (1982) oder das normative Entscheidungsmodell von Vroom und Yetton (1973).[12]

Sowohl das Reifegradmodell (Hersey & Blanchard, 1982) als auch das 3-D-Modell (Reddin, 1977) und das Kontingenzmodell (Fiedler, 1967) ziehen die Mitarbeiter- und Aufgabenorientierung als Führungsvariablen heran, deren Erfolgseinfluss vom Reifegrad der Mitarbeiter[13] respektive der situativen Bedingungen/situativen Günstigkeit[14] moderiert wird. Dabei geht das Reifegradmodell über die pure Situationsadäquanz der Führung hinaus, indem es einen idealen Entwicklungsverlauf hin zu delegativer Führung von Mitarbeitern mit hoher Reife vorskizziert (ursprünglicher Begriff der Theorie: „life cycle theory of leadership“, Hersey & Blanchard, 1969).

Trotz vielfacher Kritik am Kontingenzmodell von Fiedler (z.B. Ambrosch, Lippert & Schneider, 1978; Bungard,1984; Weibler, 2001) findet es in seinen Grundzügen auch meta-analytisch Unterstützung (Schriesheim, Tepper & Tetrault, 1994; Strube & Garcia, 1981). Auch das Reifegradmodell von Hersey und Blanchard fand zumindest partielle Bestätigung (z.B. Pascarella & Lunenburg, 1988; Silverthorne & Wang, 2001; Vecchio, 1987). Es wurde in letzter Zeit auf verschiedene Ebenen erweitert. So wurde ein analoges

[12] Auch Fleishman als einer der ersten Führungsstil-Forscher nimmt später Bezug auf Fiedlers Kontingenztheorie und erforscht selbst moderierende Situationsvariablen insbesondere für den Zusammenhang zwischen Initiating Structure und Effektivität (Fleishman, 1973).

[13] Der Reifegrad wird durch die Fähigkeit und Bereitschaft, eine Aufgabe eigenverantwortlich auszuführen, bestimmt.

[14] Als situative Bedingungen (Reddin, 1977) werden die Arbeitsanforderungen, der Führungsstil des nächsthöheren Vorgesetzten, die Kollegen und Unterstellten sowie die Normen der Organisation betrachtet. Spezifischer ist die situative Günstigkeit im Kontingenzmodell von Fiedler (1967) beschrieben. Diese ergibt sich aus der Kombination folgender, binär konzipierten Variablen: Beziehung zwischen Führer und Geführten, Aufgabenstruktur, Positionsmacht (Reihenfolge in abnehmender Wichtigkeit).

Modell zur Teamführung und zur organisationalen Transformation entwickelt (Hersey, Blanchard & Johnson, 2001).

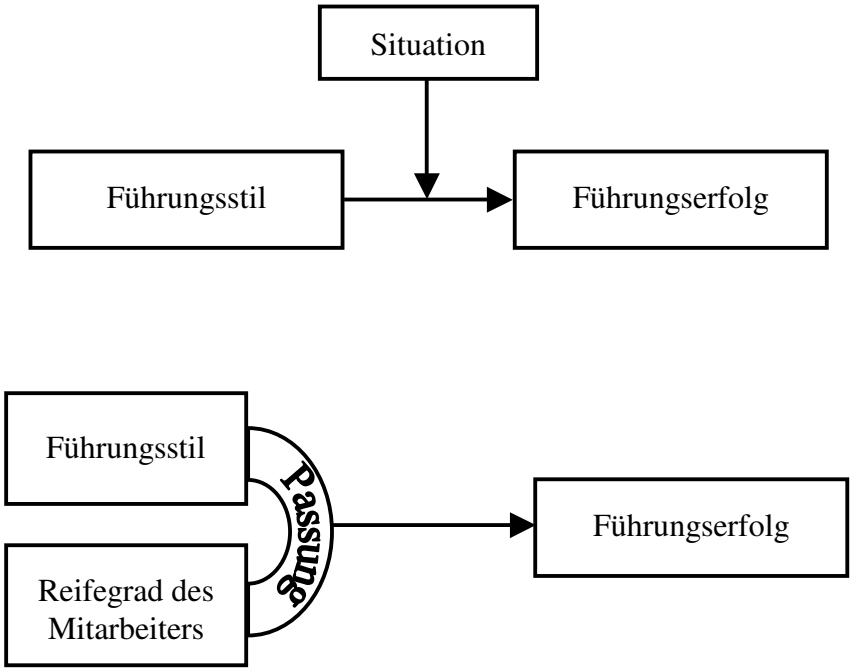


Abbildung 7. Prinzip der klassischen Kontingenzansätze (nach Neuberger, 2002) und beispielhafte Übertragung auf das Reifegradmodell nach Hersey und Blanchard (1969)

Eine genauere Situationsanalyse nimmt das bereits oben angerissene normative Entscheidungsmodell von Vroom und Yetton (1973) vor. Abhängig von der Situationskonstellation werden unterschiedliche Führungsstile empfohlen, die nach dem Grad der Partizipation differieren.[15] Die Situation wird mit Hilfe von sieben Leitfragen beurteilt und führt anhand eines Entscheidungsbaumes zu einem oder mehreren brauchbaren Führungsstilen. Als Ziele werden die Qualität, die Zeiteffizienz und die Akzeptanz der Entscheidung formuliert. Ziele der Führungskraft oder der Mitarbeiter werden nicht als Erfolgskriterien postuliert. Trotz verschiedener Vorteile steht eine Unterstützung des Modells, welche über retrospektive Angaben von Führungskräften hinausgeht, weitgehend aus (zur Validierung vgl. Jago, 1995). Weiterentwicklungen (Jago, 1995; Vroom & Jago, 1991) – auch im Hinblick auf die Trainierbarkeit (Böhnisch, 1991) – des Modells existieren. Aus praktischer Perspektive ist von einer Auseinandersetzung mit dem Modell mindestens zu erwarten, dass die Sensitivität der Führungskräfte für situative Einflüsse steigt.

[15] Es liegt also ein eindimensionales Führungsstil-Konzept zugrunde.

Abschließend soll ein instrumentalistischer Kontingenzansatz angesprochen werden, die Weg-Ziel-Theorie von House (1971, 1996). Die grundlegende Annahme besteht darin, dass sich eine Führungskraft erstens in ihre Mitarbeiter versetzen und deren Bedürfnisse und Ziele erkennen kann und zweitens ihr Führungsverhalten entsprechend dieser „Situationsdiagnose“ wählt und individuell ausrichtet. Dabei wird die postulierte Wirksamkeit der Verhaltensweisen unter Rückgriff auf Erwartungs*Wert-Überlegungen auf eine allgemeinpsychologische Basis gestellt. Während in der ursprünglichen Version (1971) das Führungsverhalten zweidimensional konzipiert war (Consideration, Initiating Structure), wurden in verschiedenen Reformulierungen bis zu neun Führungsverhaltensweisen konkretisiert (House, 1996). Insgesamt finden dort interaktive Verhaltensweisen in der Arbeitseinheit stärkere Bedeutung. Besonders hervorgehoben wird die wertebasierte Führung, die im Zusammenhang der neuen Führungsmodelle zu betrachten ist, die im folgenden Kapitel 2.5 ausführlich beschrieben. Die empirische Befundlage ist inkonsistent.

Spätestens seit der Berücksichtigung situativer Bedingungen wird der Begriff "Führungsstil" zunehmend auch nicht mehr als situationsunabhängig und zeitlich stabil definiert (Gabele et al., 1992). Das Konsistenzkriterium wird auf ähnliche situative Rahmenbedingungen begrenzt. So soll der Begriff auch in der vorliegenden Arbeit verstanden werden. Nur Fiedler (1967) geht von einer situationsunabhängigen, persönlichkeitsspezifischen Leitlinie aus. Dies hat selbstverständlich Auswirkungen auf die Gestaltungsfreiheit der Führungskräfte sowie auf die Ziele von Führungstrainings.

Indem jede situative Führungstheorie aber immer nur einen sehr begrenzten Ausschnitt potentiell denkbarer Situationen betrachtet und zudem potentiell einsetzbares Führungsverhalten häufig auf die klassische Zwei-Dimensionalität der Führungsstilforschung reduziert wird, wird Führung ungerechtfertigterweise vereinfacht und Führungskräfte insgeheim wieder zu „Great Men“, die in der richtigen Situation das Richtige tun (vgl. Neuberger, 2002). Diese implizite Annahme ist den **interaktionsorientierten Theorien** fern, da sie auch den Mitarbeitern Einflussmöglichkeiten zusprechen. So konzentriert sich die dyadische Führungstheorie beispielsweise im „Leader-Member-Exchange“-Ansatz (LMX; Graen & Scandura, 1987a, 1987b) auf die Zweierbeziehung zwischen Führungskraft und Mitarbeiter, da sie grundlegend davon ausgeht, dass die Mitarbeiter nicht als Mittelwert geführt werden, sondern als individuelle Personen. Das Idiosynkrasie-Kredit-Modell (Hollander, 1995) integriert die austauschtheoretische Interaktion insofern, dass Führungskräfte durch besonders gute Leistungserfüllung oder konformes Gruppenverhalten Kredits sammeln können, die es ihnen erlauben, von den Erwartungen der Gruppe abzuweichen, ohne Sanktionen befürchten zu müssen. Diese Abweichungen können für Transformationen genutzt werden. Eine andere Art transformationaler Führung wird im folgenden Kapitel dargestellt.

Keine der bislang dargestellten Theorien hinterfragt die prinzipielle Einflussnahme von Führung (Weibler, 2001). Die Attributionstheorie von Führung (Calder, 1977) sowie die Substitutionstheorie der Führung (Kerr, 1977; Kerr & Jermier, 1978) gehen aber einen Schritt weiter. Calders (1977) Führungstheorie basiert auf der Grundannahme, dass Führung nicht per se existiert (vgl. Abbildung 8 oben). Führung wird vielmehr als Zuschreibung durch die Mitarbeiter konzeptualisiert. Genau dieser Zuschreibungsprozess wird im Modell mittels verschiedener Fragen spezifiziert, die an klassische Attributionstheorien anschließen. Dass bestehende implizite Führungstheorien nicht geleugnet werden dürfen, zeigen auch die Arbeiten von der Arbeitsgruppe von Meindl (vgl. S. 21). Die umgekehrte Perspektive wird im Attributionsmodell von Green und Mitchell (1979) betrachtet. Dort wird das Verhalten der Führungskräfte davon abhängig gemacht, wie sie das Verhalten der Mitarbeiter wahrnehmen und worauf sie deren Leistung attribuieren (vgl. Abbildung 8 unten). Calders (1977) Arbeit ist für die vorliegende Studie insofern von Bedeutung, weil bei der Interpretation der Ergebnisse immer auch attributionstheoretische Überlegungen berücksichtigt werden müssen.

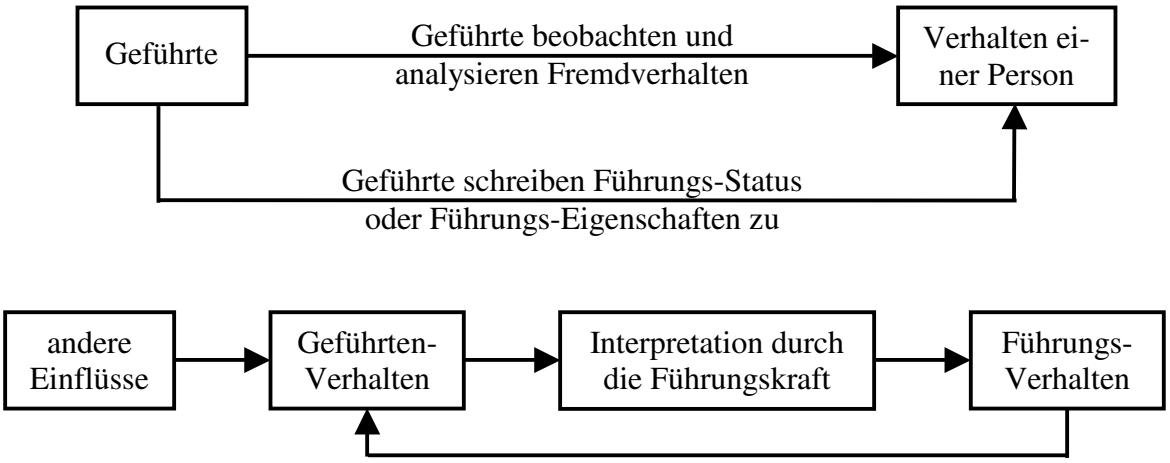


Abbildung 8. Prinzip der attributionstheoretischen Betrachtung von Geführten- beziehungsweise Führungsverhalten (nach Neuberger, 2002)

Die Substitutionstheorie der Führung (Kerr & Jermier, 1978) geht auf andere Weise über die grundlegende Wirksamkeitsannahme von Führung hinaus. Anderen Personen oder Strukturen wird die Qualität zugesprochen, den Führungseinfluss der Führungskraft zu ersetzen (Substitute, z.B. autonome Arbeitsgruppen) oder zunichte zu machen beziehungsweise ihm sogar entgegen zu wirken (Neutralisierer, z.B. räumliche Distanz). Im Prinzip sind diese Variablen als extrem starke Moderatorvariablen für den Zusammenhang zwischen Führung und Erfolg zu betrachten, so dass die Substitutionstheorie als besonders komplexe Kontingenztheorie verstanden werden kann.

Mit Ausnahme dieser beiden zuletzt genannten Theorien können alle vorgestellten Führungskonzepte auch austauschtheoretisch interpretiert werden. Sowohl Mitarbeiter als auch Führungskräfte haben bestimmte Ziele, wobei die Mitarbeiter die Ziele ihrer Führungskräfte deshalb verfolgen, weil sie dann mit der Erfüllung ihrer Bedürfnisse rechnen können und umgekehrt. Consideration kann beispielsweise folgendermaßen zu einem Tauschgut werden: „Durch Freundlichkeit vergilt es die Führungskraft ihren Unterstellten, [...] wenn sie die kaum einklagbare und nicht zu befehlende loyale und konstruktive Mit-Arbeit praktizieren.“ (Neuberger 2002, S. 405). Derartige Transaktionsideen wurden in den 70er Jahren allmählich durch einen Paradigmenwechsel versucht zu komplementieren. Ein neues Führungs-Genre wurde entwickelt, das sich insbesondere durch charismatische Elemente auszeichnet und das sich in den folgenden Jahrzehnten zunehmend in Praxis und Wissenschaft Aufmerksamkeit verschaffen konnte. Im folgenden Kapitel wird die transformationale Führung – auch in ihrer Gegenüberstellung zur transaktionalen Führung – ausführlich beschrieben. Verschiedene Hypothesen der vorliegenden Studie sind auf das Modell der transaktionalen und transformationalen Führung von Bass (1985) ausgerichtet.

2.5 Transformationale Führung

„Der erste Schritt zum Erfolg ist, sich in die Arbeit zu verlieben."
(Kalenderspruch)

Die Popularität charismatischer Führungskonzepte im weitesten Sinne spiegelt sich in der wachsenden Anzahl an Studien (vgl. Kapitel 2.5.4), in der zunehmenden praktischen Bedeutung verwandter Konzepte (vgl. Kapitel 2.6) und in ihrer verstärkten Berücksichtigung durch die Führungsliteratur wider. So widmet Neuberger (2002) der charismatischen, transformationalen und visionären Führung in der sechsten Auflage seines Standardwerks zur Führung ein eigenständiges Kapitel mit einem Textanteil von nahezu 10% des Buches. Ein anderes Beispiel: „Leadership Quarterly" reserviert das Sommerheft 1999 ausschließlich für Artikel zur charismatischen und transformationalen Führung. Insgesamt wird die im vorliegenden Kapitel dargestellte Führungsströmung als „new leadership" bezeichnet (Bryman, 1992).

Insbesondere in der US-amerikanischen Management-Literatur zeichnet sich dieser Trend seit der Revitalisierung des Charisma-Konzepts vor circa 30 Jahren deutlich ab (vgl. z.B. Kapitel 2.5.2; Bass, 1985; Bass, 1990; Burns, 1978; House, 1977; Shamir, House & Arthur, 1993; Yukl, 1999a). Auch in der vorliegenden Arbeit fällt dieser Führungsströmung – genauer der transformationalen Führung – besonderes Gewicht zu, indem ihr Einfluss auf Erfolg mit dem verglichen wird, der durch klassische Führungsaufgaben in einem traditionell organisierten Industrieunternehmen erzielt wird. Im vorliegenden Kapitel wird deshalb der Fokus auf die transformationale Führung gelegt, die jedoch ohne eine Betrachtung der visionären und insbesondere der charismatischen Führung nicht umfassend begreifbar ist. In einer skizzenhaften Darstellung wird Letztere deshalb der Beschreibung der transformationalen Führung vorangestellt (Kapitel 2.5.1). Einen umfassenden Überblick findet man beispielsweise bei Neuberger (2002, S. 141-221). Welche Resultate die transformationale Führung erbringt, welche Umweltbedingungen für sie förderlich sind und wie sie erlernbar ist, wird in den Kapiteln 2.5.4 bis 2.5.6 beschrieben. Abschließend wird sie mit anderen und verwandten Konzepten verglichen (Kapitel 2.6).

2.5.1 Historische Verankerung der transformationalen Führungskonzeption von Bass in der Herrschaftssoziologie von Weber

Die transformationale Führung integriert in hohem Ausmaß charismatisches Führungsverhalten (vgl. Kapitel 2.5.3) und wird deshalb in einem weiten Sinne auch zu den charismatischen Führungstheorien gezählt. Die historischen Wurzeln der transformationalen Führung liegen daher dort, wo auch die Wurzeln der charismatischen Führung zu suchen

sind: in der Herrschaftssoziologie von Max Weber. Auch wenn der Typus der charismatischen Herrschaft zuerst vom Kirchenrechtler Rudolf Sohm (1892) für die altchristliche Gemeinde (unter Rückgriff auf das Neue Testament) entwickelt wurde, war es Max Weber (1924), der das Konstrukt „Charisma“ erweiterte, es auf politische Führung im Allgemeinen transferierte und letztlich das Konzept der charismatischen *Führung* verbreitete und popularisierte (Winckelmann, 1952).

In seiner Herrschaftssoziologie differenziert Weber (1864-1920) drei reine Herrschaftstypen nach einem subjektiven Kriterium, nämlich nach den Motiven, die jeweilige politische Ordnung als rechtmäßig zu akzeptieren und dem Führer Gehorsam entgegen zu bringen.[16] Die Grundlage für Herrschaft ist also ihre Legitimität und Legitimität entsteht aus dem *Glauben* an Legitimität. Entsprechend der Weberschen Typologie kann Herrschaft ihr Geltungsrecht aus rational-legalen, traditionalen oder charismatischen Prinzipien ziehen. Charismatischer Charakter wird der Herrschaft zugeschrieben, wenn sie "auf der außeralltäglichen Hingabe an die Heiligkeit oder die Heldenkraft oder die Vorbildlichkeit einer Person und der durch sie offenbarten oder geschaffenen Ordnungen" (Weber, 1980 [1921], S. 124) beruht. Charisma selbst beinhaltet eine "außeralltäglich [...] geltende Qualität einer Persönlichkeit [...], um derentwillen sie als mit übernatürlichen oder übermenschlichen oder mindestens spezifisch außeralltäglichen, nicht jedem andern zugänglichen Kräften oder Eigenschaften [begabt] oder als gottgesandt oder als vorbildlich und deshalb als 'Führer' gewertet wird. Wie die betreffende Qualität von irgendeinem ethischen, ästhetischen oder sonstigen Standpunkt aus 'objektiv' richtig zu bewerten sein würde, ist natürlich dabei begrifflich völlig gleichgültig: darauf allein, wie sie tatsächlich von den charismatisch Beherrschten, den 'Anhängern', bewertet wird, kommt es an." (Weber, 1980 [1921], S. 140) Es ist also die Attribution auf Charisma, die die Anerkennung und die Legitimation des Führers bestimmt. Weber betrachtet dies jedoch weniger kausal als vielmehr final, indem er aus der Legitimation den Zweck – ja sogar die "Pflicht" – ableitet, zum Wohle der Geführten zu führen. Im Zusammenhang mit der "rationalen Lebensführung" ist es für ihn unabdingbar, dass sich jede charismatische Führung spätestens dann *veralltäglichen* muss, wenn sie für die Geführten kein "Wohlergehen" mehr mit sich bringt. Dann muss sie durch Ordnungsprinzipien der legalen Herrschaft ersetzt werden.

Die Mystik und das schwer Fassbare am Charismakonzept bleiben bei Weber bestehen, indem er charismatischen Führern messianische Fähigkeiten zuschreibt und ihnen einen Machtgebrauch verbietet, wenn "der charismatisch Begnadete von seinem Gott oder seiner magischen oder Heldenkraft verlassen" ist (Weber, 1980 [1921], S. 140). So verwendet Weber "Charismatische Herrschaft" streckenweise genuin als "Gottesgnadentum", immer jedoch reserviert er sie für begnadete Menschen und für Übergangszeiten besonderer

[16] Herrschaft ist definiert als die "Chance [...], für spezifische (oder: für alle) Befehle bei einer angebbaren Gruppe von Menschen Gehorsam zu finden." (Weber, 1980 [1921], S. 122)

"Begeisterung oder Not" (Weber, 1980 [1921], S. 142). Die begriffliche Leistung Webers zeigt sich darin, dass er die göttliche Bedeutung von Charisma im Sinne besonderer Gnadengaben allmählich aufbricht und das Konzept zusätzlich als Zuschreibung außeralltäglicher Fähigkeiten auch für den politischen Kontext brauchbar macht.

Bis in den Charisma-Faktor der transformationalen Führung hinein (Kapitel 2.5.3.2) wird transportiert, dass der charismatische Führungseinfluss nicht auf fachlichen Qualitäten beruht, sondern ganz allein "kraft persönlichen Vertrauens, Offenbarung, Heldentum oder Vorbildlichkeit" (Weber, 1980 [1921], S. 124) sowie des "ewig Neue[n], Außerwerktägliche[n], Niedagewesene[n] und [der] emotionale[n] Hingenommenheit" besteht (Weber, 1966, S. 104).

Bei der Frage, wie man Charisma erlangen kann, unterscheidet Weber zwischen Charisma, das aufgrund der gezeigten Verhaltensweisen des Führers erreicht wird (reines Charisma) und Charisma, das durch formale oder erbliche Nachfolgedesignation zugeschrieben wird. Die im folgenden Kapitel 2.5.2 skizzierten Neocharismatiker, zu denen auch der charismatische Faktor der transformationalen Führung gezählt werden kann, betrachten ausschließlich das reine Charisma.

Dass Weber immer wieder die Vaterrolle der Neocharismatiker zuteil wird, ist zudem deshalb fast verwunderlich, weil er erstens die charismatische Führung lediglich als eine Art „Lückenfüller“ für die Überbrückung einer schweren Zeit betrachtet, während das eigentliche Führungsideal in der rational-legalen (bürokratischen) Führung besteht. Zweitens steht bei ihm Charisma unvereinbar neben, beziehungsweise konträr zu ökonomischem Denken, da er Beruf im eigentlichen Wortsinn als "Sendung" oder "innere 'Aufgabe'" begreift (vgl. Neuberger, 2002, S. 148): Reines Charisma „verschmäht und verwirft [...] die ökonomische Verwertung der Gnadengaben als Einkommensquelle" (Weber, 1980 [1921], S. 142). Für einen Rückgriff auf Weber spricht aber selbstverständlich die erstmalige konzeptionelle Erarbeitung charismatischer Führung im Rahmen der Sozialwissenschaften sowie vielleicht sogar gerade die erwähnte Lückenfüller-Rolle, wenn dadurch die situationale Abhängigkeit der charismatischen Führung betont wird. Hier greift Weber in gewisser Weise der Interaktionismus-Strömung voraus, die erst in den 60er Jahren in die psychologische Führungsforschung Eingang fand (vgl. Kontingenzansätze, S. 38f.) und im Rahmen der modernen charismatisch-transformationalen Führungsgeschichte bis heute nur selten fokussiert wird (z.B. Pawar & Eastman, 1997; vgl. auch Kapitel 2.5.5).[17]

[17] Erstaunlich ist, dass sich die Neo-Charismatiker nicht auch auf Downtons „rebel leadership“ (1973) beziehen. Dort wird nämlich – in enger Anlehnung an Weber – die Führungslegitimation auf die Basis von Loyalität gestellt. Im Unterschied zu Weber arbeitet er heraus, dass *verschiedenste* Führungsverhaltensweisen Vertrauen erwecken und Commitment der Geführten hervorrufen. Dazu zählt er die transaktionale, die inspirierende sowie die charismatische Führung. Auch wenn verschiedene Führer diese verschieden stark gewichten, beschreibt er sie als sich gegenseitig ergänzend und nimmt dadurch einen wichtigen Gedanken des

2.5.2 Das neo-charismatische Führungsparadigma

„Großartige Tänzer sind nicht aufgrund ihrer Technik großartig,
sondern aufgrund ihrer Leidenschaft.“ (Martha Graham)

Nachdem die behavioral und kognitiv ausgerichteten Strömungen innerhalb der Psychologie Emotionskonzepte eher verkannten und die traditionellen Führungstheorien hervorbrachten (Kapitel 2.4), begann die Renaissance der charismatischen Führung. Die entsprechenden Konzepte rekurrierten auf die soeben dargestellte Konzeption von Max Weber. Der Paradigmenwechsel in der verhaltenswissenschaftlichen Führungsforschung wurde in der Mitte der 70er Jahre allmählich eingeleitet und durch das im vorliegenden Kapitel vorgestellte Genre an Führungstheorien bestimmt.

2.5.2.1 Zum Begriff „Charisma“

Handelt es sich bei Charisma um eine göttliche Gabe, ein mystisches Phänomen, eine Kombination an Eigenschaften, um ein Verhaltensprofil der charismatischen Person oder um eine besondere Wirklichkeitskonstruktion der anderen?

Etymologisch lässt sich das Wort selbst auf „charis“ (altgriechisch für Gnade) und das Suffix –„ma“ (Gabe oder Geschenk) zurückführen. Die Definition als Gnadengabe findet sich zwar alternativ zu einer eigenschaftsorientierten Begriffsbeschreibung bei Weber, doch die neueren Ansätze versuchen, Charisma mit anderen Elementen zu greifen. Nahezu alle neo-charismatischen Autoren behelfen sich mit einer Aufzählung verschiedener Führungsfertigkeiten. Die Aufzählung in Tabelle 1 soll hier paradigmatisch für alle stehen. Unter charismatischer Führung ist also zu verstehen, dass Führungskräfte ihre Mitarbeiter über Verhaltensweisen beeinflussen, die dort aufgelistet sind. Auf der Seite der Mitarbeiter ist von einer Zuschreibung von Charisma, Akzeptanz der Führung und insgesamt wohlwollenden Attributionen auszugehen.

Führungsmodells von Bass vorweg, der bei den im Folgenden dargestellten Neo-Charismatikern unberücksichtigt blieb: Echtes Charisma kann und soll nicht die *einzige* Führungslegitimation sein.

Tabelle 1. Charakterisierung von charismatischen Führungskräften (nach House & Shamir, 1995, S. 881ff.)

Charismatische Führer ...
▪ haben eine Vision einer besseren Zukunft, ▪ sind ihrer Vision ergeben, ▪ verfügen über Selbstvertrauen, Entschlossenheit und Ausdauer, ▪ wecken wichtige Motive bei den Geführten, ▪ haben eine außergewöhnliche Bereitschaft zum Risiko, scheuen keine persönlichen Wagnisse, ▪ haben hohe Erwartungen an die Geführten und zugleich hohes Vertrauen in die Geführten, ▪ bewerten die Geführten grundsätzlich positiv, ▪ bemühen sich um die Entwicklung der Geführten, ▪ zeigen symbolische Verhaltensweisen, ▪ verstehen sich in der Selbstdarstellung und in der Schaffung eines positiven Images, ▪ leben ihre Vision vor, ▪ zeichnen sich durch moralische Integrität aus, ▪ fungieren als Sprachrohr der Gemeinschaft, ▪ zeigen oft ein außergewöhnliches Verhalten und ▪ sind anregende Kommunikatoren, die Botschaften einfallsreich und emotional ansprechend transportieren.

Nachdem House und Aditya (1997) rückblickend vier gemeinsame Charakteristika für eine Reihe an Theorien, die um das Charisma-Konzept rangieren, heraus gearbeitet hatten,[18] wurden sie unter dem Label „neo-charismatisch“ (S. 439) zusammengefasst. Die gemeinsamen Kennzeichen sind folgende:

Die Theorien versuchen zu erklären, wie Führungskräfte ihre Organisationen zu *außergewöhnlichen* Leistungen und Veränderungen führen können.

Weiter soll erklärt werden, wie Führungskräfte bei ihren Mitarbeitern *außergewöhnliches* Vertrauen und Commitment (Bass, 1985; House, 1977; Shamir et al., 1993) sowie *außergewöhnliche* Loyalität (House, 1977), Begeisterung, Motivation und Leistung hervorrufen können (Bass, 1985). Führungskräfte dieser Strömung verändern (transformieren) die Motive, Werte und Präferenzen ihrer Mitarbeiter – weg von Selbstinteresse, hin zu Gruppeninteresse (Bass, 1985; House, 1977; Shamir et al., 1993). „Further, they cause followers [...] to perform above and beyond the call of duty.” (House & Shamir, 1993, p. 82)

[18] Sie werden bereits bei House und Shamir (1993, S. 82f.) herausgestellt. Erst vier Jahre später – sozusagen mit einem besseren Über- und Rückblick über das neue Führungsparadigma – werden sie jedoch explizit als Charakteristika dargestellt, die eine Abgrenzung zur traditionellen Führungsforschung und einen Rahmen für die neue Führungsforschung markieren.

Dabei werden sowohl diejenigen Führungsverhaltensweisen betrachtet, die symbolische, visionäre oder unterstützende Kraft haben, als auch diejenigen, die auf Flexibilität, Sensitivität für die Umwelt oder Perspektivenwechsel zielen. Emotional als auch kognitiv orientiertes Verhalten wird also gleichermaßen von der neuen Führung erwartet. Obgleich die Großzahl der Neocharismatiker die typischen Verhaltensweisen und die besondere Bindung zwischen Führer und Geführten ins Zentrum stellen, erkennen sie dennoch den Beitrag bestimmter Persönlichkeitsfaktoren an (vgl. S. 56).

Die Gruppe der Variablen, die durch Führung beeinflusst werden sollen, wurde in den neocharismatischen Theorien deutlich erweitert. Zu den traditionellen Erfolgsmaßen der Arbeitszufriedenheit und Leistung von Mitarbeitern traten insbesondere auch Konzepte mit affektiven Anteilen hinzu. So werden beispielsweise (affektives) Commitment, das Selbstwertgefühl der Mitarbeiter, ihre emotionale Bindung an die Führungskraft oder ihre Identifikation mit einer Vision als Kriteriumsvariablen für erfolgreiche Führung herangezogen. An dieser Stelle soll klar gestellt sein, dass neben den affektiven selbstverständlich auch kognitive und behaviorale Aspekte eine wichtige Rolle spielen.

Charisma wird also selten losgelöst von seinen Konsequenzen definiert. Tritt nämlich keine Leistung über die Erwartungen hinaus[19] auf, dann bewilligen wir der entsprechenden Führungskraft auch kein Charisma. In einem gewissen Sinne retrospektiv erscheint der Definitionsversuch, der eben diese starke, die Erwartungen übertreffende Wirkung als Indikator für charismatische Führung heranzieht. Charisma existiert also nicht in der Führungsperson als solcher. Wilson (1975, S. 7) macht dies in überspitzter Weise deutlich: „If man runs naked down the street proclaiming that he alone can save others from impending doom, and if he immediately wins a following, then he is a charismatic leader. A social relationship has come into being. If he does not win a following, he is simply a lunatic."

Gerade vor dem Hintergrund dieser sehr hohen Erwartungen an Führung ist die Frage aufzuwerfen, ob eine solche Führung als Sinnvermittlung durch exemplarisches Handeln überhaupt umsetzbar ist. Möglicherweise spiegelt sich in diesen Erwartungen eine romantisierende Sichtweise von Führung (Meindl et al., 1985, vgl. S. 21), die uns glauben lässt, dass die Führungskräfte Führungsverantwortung übernehmen und starke Effekte hervorbringen. Zuschreibungen werden von Tatsachen also nicht getrennt. Eine solche „euphemistisch-romantisierende" Sichtweise der Relevanz von Führung und ihrer motivierenden Wirkung auf Mitarbeiter dürfte aber in erster Linie den Anforderungen eines neuen Bestseller-Genres standhalten, nicht jedoch empirischer Führungsforschung. Von den neuen Führungstheorien ist das Modell der transformationalen Führung von Bass (1985; Beschreibung in Kapitel 2.5.3) am besten überprüft (Befunde in Kapitel 2.5.4).

[19] „performance beyond expectations" (Bass, 1985)

2.5.2.2 Die Wiederentdecker und die „Transformationale Trias“

> „Amid these pronunciations of calamity, there appeared two bright spots of especial interest here. Interestingly, the impact of these bright spots was felt only later. The first of these seminal publications was House's (1977) original charismatic theory, while the second was Burn's (1978) now classic work on transformational leadership.”
> (Hunt, 1999, p. 133)

Aber nun zurück zu den Anfängen: Mit seinem Aufsatz von 1977 kann Robert House als Vorreiter der charismatischen Moderne betrachtet werden. Im gleichen Atemzug muss aber auch James MacGregor Burns genannt werden, der 1978 erstmals transaktionale und transformationale Führungskonzepte präsentierte und einander gegenüberstellte. Beide Wegbereiter gehen zwar auf ähnliche Vorläufer ein, z.B. Max Weber, nehmen aber in ihren ersten Arbeiten noch keinen Bezug aufeinander.

House (1977) gründet seine „1976 theory of charismatic leadership“ auf der Sichtung der politischen, soziologischen und sozialpsychologischen Literatur zu charismatischer Führung (empirische Ergebnisse lagen bis dato kaum vor), um dann erstmals zu einer eher spekulativen Theorie zu kommen (House, 1977, p. 190).[20] Er fokussiert in seiner Theorie darauf, wie eine Führungskraft von ihren Mitarbeitern als nachahmenswertes Vorbild und insgesamt vorteilhaft wahrgenommen werden kann.[21] Erst wenn dies gegeben ist, können sich bei den Mitarbeitern die erwünschten Konsequenzen entwickeln: Vertrauen gegenüber der Führungskraft, die Übernahme derer Werte, die Akzeptanz herausfordernder Ziele, erhöhte Selbstwirksamkeits-Erwartung hinsichtlich dieser Ziele und letztlich hohe Leistung.

Burns (1978) war der erste, der zwischen transaktionaler und transformationaler Führung differenzierte. Seine Arbeit beruht auf einer qualitativen Analyse von Politikerbiografien, die in verschiedensten Ländern (China, Deutschland, Frankreich, Russland, USA und andere) weltpolitisch bedeutende Transformationen herbeiführten. Dieser inhaltsanalytischen Arbeit zufolge basiert die transaktionale Führung auf dem sozialen Austausch (vgl. Homans, 1972a, 1972b), vereinfachend also auf dem Prinzip von Geben und Nehmen. Zur Erfüllung dieses Prinzips müssen beide Parteien ihre Motive und Tauschgüter gegenseitig erkennen. Folglich spielt die Informationspolitik in Austauschbeziehungen eine zentrale Rolle. Die ursprüngliche Beschreibung der Beziehung zwischen einer transaktionalen Führungskraft und ihren Mitarbeitern sieht nach Burns (1978, p. 19f.) folgendermaßen aus:

---

[20] Trotz prinzipieller Würdigung dieser Pionierleistung folgt unmittelbare Kritik von Sashkin (1977). Weiterentwicklungen verändern die Theorie vor dem Hintergrund der Weg-Ziel-Theorie der Führung hin zur wertbasierten Führungstheorie (House, 1996).

[21] Diesen Prozess beschreibt House (1977) als Interaktion von Führermerkmalen (vgl. S. 56 zu Führermerkmalen) und folgenden Führungsverhaltensweisen: Zielsetzung, persönliche Darstellung, Modellverhalten bezüglich eines Wertesystems, Motivierung, Kommunikation hoher Leistungserwartung, Vertrauen in das Leistungsvermögen der Mitarbeiter.

„Their purposes are related, at least to the extent that the purposes stand within the bargaining process and can be advanced by maintaining that process. […] A leadership act took place, but it was not one that binds leader and follower together in a mutual and continuing pursuit of a higher purpose." Die Beziehung geht nicht über einen Austausch hinaus. Konträr dazu konzipiert Burns (1978, S. 20) die transformationale Führung: „Such leadership occurs when one or more persons *engage* with others in such a way that leaders and followers raise one another to higher levels of motivation and morality. […] it raises the level of human conduct and ethical aspiration of both leader and led, and thus it has a transforming effect on both. […] leaders throw themselves into a relationship with followers who will feel "elevated" by it and often become more active themselves". Bereits hier wird deutlich: Führer und Geführte folgen einer gemeinsamen, hoch gesetzten Vision und sind bereit, dafür ihr Anspruchsniveau und damit auch ihr Engagement wesentlich zu erhöhen. Dass Führung auch sinnstiftend wirken und über reine Transaktionen hinausgehen soll, war das Anliegen der wachsenden, neuen Führungsgeneration.

Bennis und Nanus (1990, S. 26) beziehen sich ausdrücklich auf die Konzeption transformationaler Führung von Burns, die sie als – man könnte fast sagen weltanschauliche – Grundlage dafür heranziehen, Führung von Management abzugrenzen. Die Autoren halten beides für unerlässlich, fordern aber von einer Führungskraft nicht nur Leitung und Organisation der bestehenden Aufgaben (Management), sondern die Rolle eines Leit-Sterns. Eine Führungskraft soll im Gegensatz zu einem Manager die Richtung vorgeben, eine Vision weisen (Bennis & Nanus, 1985; Bennis, 1989)[22]. So wird das polare Spannungsgefüge zwischen „harten" und „weichen" Faktoren geteilt und im Zuge der Emotionalisierung der Führung insbesondere zu letzteren angemahnt. Als Fazit ziehen die Autoren immer wieder den Schluss: „Manager machen die Dinge richtig, Führende tun die richtigen Dinge." (Bennis & Nanus, 1990, S. 28f.)

In dieser Konzeption von Führung ist die visionäre Führung bereits angelegt, denn Führungskräfte werden als „soziale Architekten" konstruiert, die mit Hilfe einer Vision für die Gegenwart und die Zukunft eine Vielzahl an bis dato wahrgenommenen Zwängen überwinden sollen (Bennis, 1976).[23] In letzter Konsequenz begründet die visionäre Führung ein Führungsethos, das Mitarbeitern einen „Stern am Himmel" weist, wobei die Mitarbeiter diesen als eigenes Ziel akzeptieren, dann aber keinen verlängerten Arm ihrer Führungskräfte darstellen, sondern ihre eigenen Höhenflüge entfalten können. Bennis und Biederman (1998, S. 32) bringen es metaphorisch auf den Punkt:

---

[22] Vorab gab es nur wenige Versuche, Management und Führung als unterschiedliche Konzepte zu betrachten (z.B. Selznick, 1957). Den umgekehrten Weg eines Plausibilitätschecks ging Rost (1998, S. 106), indem er eine Liste an Management-Strategien aufstellte (z.B. "Have the buses, trains, airplanes run late."), die Führung nur wenig effizient ersetzen könnte.

[23] In der Folge tritt die Führung durch Visionen einen Siegeszug in die und quer durch die Management-Literatur an (chronologisch: Roberts, 1985; Kouzes & Posner, 1987; Sashkin, 1988; Nanus, 1992; Bennis & Biederman, 1998) und findet auch in Deutschland wichtige Vertreter (Kienbaum, 1992).

„Große Führungspersönlichkeiten scheinen ihren Traum zu verkörpern und mit ihm zu verschmelzen.“ Die Brücke zu den Mitarbeitern wird gespannt, indem sie „ihren Traum so leidenschaftlich verfolgen, daß die andere einfach mitreißen.“ (Bennis & Biederman, 1998, S. 38). Geradezu wohltuend nüchtern klingen vor dem Hintergrund dieser fast romanhaften Sprache neue Definitionsversuche. So soll auch in der vorliegenden Arbeit unter Vision „ein klares und konsistentes Bild einer positiv belegten Zukunft, aus dem heraus sich Handlungsziele ableiten lassen“, verstanden werden (Scholz, 2000, S. 957). Eine wichtige Funktion von Visionen ist es, die eigene Tätigkeit in einen größeren Sinnzusammenhang zu stellen und ihr damit mehr Bedeutung zu verleihen.[24] Verschiedene empirische Untersuchungen bringen die Klarheit von Visionen mit Unternehmenserfolg in Verbindung (Baum, Locke & Kirkpatrick, 1998; Hoch, Roeding, Purkert & Lindner, 1999). In einer Übersichtsarbeit wird auf den positiven Zusammenhang zwischen Visionen und Produktinnovationen hingewiesen (Brown & Eisenhardt, 1995).

Bass (1985) integriert sowohl die charismatische als auch die visionäre Führung in sein Konzept der transformationalen Führung. Dabei bezieht er sich unmittelbar auf Burns (1978) – er widmet diesem sogar sein Ausgangswerk (1985) – und transferiert dessen Führungsdimensionen aus dem historisch-politischen Kontext auch in andere Führungsbereiche. Er will damit endgültig das Zeitalter von „overmanaged and underled“ (Bennis, siehe oben) besiegeln. Gemeinsam mit Burns und Bennis bildet Bass die Trias der Basisforscher (vgl. Scholz, 2000) zur Transformationalen Führung. Bevor diese in Kapitel 2.5.3 dargestellt wird, folgt ein Abriss weiterer neo-charismatischer Theorien, auf die sich Bass in seiner kontinuierlichen Arbeit erstaunlich selten bezieht. Doch ohne diese Vielzahl an Forschungslinien – so ist anzunehmen – wäre weder die Hausse so langlebig noch die aktuelle Befundlage so erdrückend positiv. Um das Modell von Bass besser einordnen zu können, werden im folgenden Kapitel ausgewählte neo-charismatische Führungstheorien skizziert.

[24] Das Ziel, den beruflichen Tätigkeiten Sinn und Bedeutung zuzusprechen, ist eine zentrale Forderung an die „neuen Führungskräfte“, die immer wieder sogar als „manager of meaning“ bezeichnet werde (Bryman, 1996).

2.5.2.3 Moderne Charismatische Führungstheorien im engeren Sinne

„To be a leader means, especially, having the opportunity to make a meaningful difference in the lives of those who permit leaders to lead.“ (DePree, 1998, p. 132).

Nach dem Phasenmodell von Conger und Kanungo (1987) können neo-charismatische Verhaltensweisen erst dann gezeigt werden, wenn der Status quo als veränderungsbedürftig evaluiert wurde (Phase 1). Dann engagieren sich charismatische Führungskräfte persönlich in Innovationsprojekten und verfolgen damit in aller Regel Visionen, die mit den bisherigen Normen brechen. Während sie unkonventionelle Wege beschreiten, verfolgen sie enthusiastisch die artikulierte Vision und riskieren persönliche Verluste (Phase 2). Da Charisma als zugeschriebenes Merkmal betrachtet wird, werden für die Führungskraft verschiedene Faktoren herausgearbeitet, die eine Attribution von Charisma auslösen. Dazu gehören Selbstbewusstsein, soziale Sensitivität, Management-Fertigkeiten und Fachkenntnis sowie Auftreten als Reformer/Revolutionär.

Shamir et al. (1993) liefern einen der ersten Versuche, charismatische Ansätze miteinander zu integrieren. Ihre Theorie ist zudem eine ausgesprochen psychologische. Im Zentrum steht dabei der Zusammenhang zwischen Motivationsmechanismen und Selbstkonzept. Die Führungskraft soll durch die Stärkung des Selbst (der Mitarbeiter) und nicht durch externe Belohnungen ihr Team erfolgreich führen. Dem zugrunde gelegt sind die Annahmen, dass Menschen erstens das Bedürfnis zur Selbstverwirklichung haben und zweitens dieser näher kommen, je besser es ihnen gelingt, ihr Selbstwertgefühl und ihre Selbstachtung zu steigern. Um dies zu erreichen, sollten Führungskräfte verschiedene Motivationsprozesse anstoßen: Zunächst sollten sie gemeinsam hoch geschätzte Werthaltungen kontinuierlich unterstützen, um ihnen visionäre Kraft zu verleihen. Dann sollten sie wichtige Motive eruieren und wecken – vor allem solche, die der Verwirklichung der Vision dienen. Zu diesem Zweck verhält sich die Führungskraft auf verschiedensten Dimensionen modellhaft (fachlich, emotional, Lebensstil und Präferenzen etc.). Das bedeutet auch, dass eine charismatische Führungskraft ihre Werte und Visionen glaubwürdig und als erstrebenswert kommunizieren muss. Letztlich erleichtern es so geschaffene Rahmenbedingungen, dass die Mitarbeiter ein gestärktes Selbstwertgefühl haben, ihr Selbstkonzept dem der Führungskraft ähnlich wahrnehmen, eine hohe Konsistenz zwischen ihrem Selbstkonzept und ihrem Handeln wahrnehmen und eine positive soziale Identität entwickeln. Unklar bleibt, ob tatsächlich nur solche Selbstkonzeptmodifikationen für motivationale Wirkungen verantwortlich sind. Jedenfalls sind sie ein zentraler Bestandteil der Theorie. Als konkretes Lehrbuch für Führungskräfte sind die theoretischen

Als konkretes Lehrbuch für Führungskräfte sind die theoretischen Ansätze aufgrund ihres hohen Abstraktheitsgrad bislang wenig tauglich.

Unter dem Blickwinkel einer strategischen Unternehmensführung erarbeiten Tichy und Devanna (1986a) mittels Experteninterviews (mit „Leaders“) verschiedene Kennzeichen, die einen transformationalen Führungsprozess beschreiben. Transformational ist hier im Sinne einer Veränderung der Organisation zu verstehen: Eine Führungskraft erkennt und erzeugt Veränderungsbedarf, entwickelt eine neue Vision, integriert den Wandel in den Alltag und institutionalisiert diesen als kontinuierliches Change Management. Da diese strategischen Führungsaufgaben in erster Linie von Führungskräften auf hohen und mittleren Hierarchieebenen übernommen werden, wird transformationale Führung auch lediglich für Führungskräfte dieser Ebenen forciert (Tichy & Devanna, 1986b). Dennoch extrahieren die Autoren ähnliche Unterschiede zwischen transformationalen und transaktionalen Führern wie Bass (siehe Kapitel 2.5.3): Transformationale Führer vermitteln beispielsweise Visionen, bringen dem Einzelnen Vertrauen in seine Fähigkeiten entgegen oder initiieren eine Werthaltung, die auch risikobereites Handeln im Einsatz für die Vision und die Gruppe rechtfertigt (Tichy & Devanna, 1986b).

Nur selten wird charismatische Führung komplett losgelöst von den Personenmerkmalen der Führungskräfte konzipiert. Folgende vier Personenmerkmale werden zur Charakterisierung neo-charismatischer Führungskräfte immer wieder angebracht (seit House, 1977): hohe Selbstsicherheit, ein hohes Dominanzstreben, eine starke Überzeugung von den eigenen Ideen und ihrer moralischen Richtigkeit sowie ein hohes Bedürfnis, andere davon zu überzeugen. In der GLOBE[25]-Studie (House et al., 1998) wird charismatische, wertebasierte Führung mit den Attributen visionär, inspirierend, selbstaufopfernd, integer, entschlossen und leistungsorientiert beschrieben. Viele der bisher zitierten Autoren gehen davon aus, dass verschiedene Persönlichkeitsvariablen eine begünstigende Wirkung auf die Genese einer charismatischen Führung haben. Doch diese lassen sich nicht von den Persönlichkeitsfaktoren erfolgreicher Führungskräfte im Allgemeinen unterscheiden. Mit Neuberger (2002, S. 158) muss eher von einem „Set von Begriffen, die gehäuft verwendet werden,“ gesprochen werden, als von einer „trennscharfe[n] Identifikation charismatischer Eigenschaften“. Durch die Big Five scheint die transformationale Führung jedenfalls deutlich schlechter prognostizierbar zu sein als anderes Führungsverhalten (9% erklärte Varianz in der Meta-Analyse von Bono & Judge, 2004). Wie bereits oben erwähnt, ist charismatische *Führung* nicht (allein) durch Eigenschaften der Führungskraft gekennzeichnet, sondern durch die Art der Beziehung zwischen Führungskraft und Geführten, die selbstverständlich nicht durch Personenmerkmale allein konstituiert ist.

[25] Akronym für “Global Leadership and Organizational Behaviour Effectiveness”

2.5.3 Das transformationale Führungsmodell von Bass

„Alles Große in der Welt geschieht nur, weil jemand mehr tut, als er muss.“
(Kalenderspruch/Volksweisheit)

Das Konzept der transformationalen Führung von Bass ebnete den Einzug der „neuen Führung“ in die Organisationspsychologie, da es auf psychologische Theorien rekurrierte. Die andauernde Arbeit an einer Operationalisierung der Führung in Fragebogenitems (MLQ; Kapitel 2.7.4) löste zudem einen Forschungsboom aus, so dass erstmals eine Vielzahl empirischer Ergebnisse erarbeitet werden konnte. Dennoch musste die Theorie seit 1985 nicht bedeutend verändert werden (Bass, 1998; Bass, 1999; Bass & Avolio, 1994). Die Terminologie der beiden großen berücksichtigten Führungsdimensionen – die transaktionale und die transformationale Führung – geht unmittelbar auf Burns (1978) zurück. Bei beiden Autoren werden diese Dimensionen sowohl durch charakteristische Verhaltensweisen (Kapitel 2.5.3.2) als auch durch die Art der erzielten Erfolge (Kapitel 2.5.4) differenziert.

2.5.3.1 Die Eckpunkte der Theorie

Bass (1985) charakterisiert einen transaktionalen Führer als einen, der erkennt, welche Gegenleistungen seine Mitarbeiter von ihrer Arbeit erwarten, und der bemüht ist, diese Gegenleistungen bei angemessener Arbeitsleistung auch zu erfüllen. Es geht also in erster Linie um eine Austauschbeziehung zwischen Führungskraft und Mitarbeiter. Als Ergebnis ist mit der vereinbarten, also erwarteten Leistung zu rechnen. Quasi als Gegenspieler der transaktionalen Führung wird die transformationale Führung gehandelt. In dieser Führungsbeziehung gehen beide Beteiligten eine Art Pflichtgemeinschaft ein, um sich gegenseitig auf höhere moralische Ebenen weiter zu entwickeln und um eine Transformation hin zu einer visionären Zukunft zu vollziehen. Dadurch, dass eine Führungskraft die unten beschriebenen Bausteine der transformationalen Führung einsetzt, verändern ihre Mitarbeiter ihr Anspruchsniveau und setzen sich für höhere Ziele ein, so dass letztlich Leistungen über die Erwartungen hinaus erzielt werden. Dominieren im einen Fall die Informationsbeziehung und das Erforschen und Instrumentalisieren der Mitarbeitermotive, sind es im anderen Fall die Beziehungsstruktur zwischen Mitarbeiter und Führungskraft sowie die Veränderung kognitiver Strukturen des Mitarbeiters.

Full Range of Leadership

„The Best of Leaders Are Both Transformational and Transactional; the Worst Are Neither; the Worst Avoid Displaying Leadership“ (Bass & Avolio, 1993, p. 72)

Während Burns (1978) transaktionale und transformationale Führung an entgegengesetzten Enden eines Führungskontinuums positioniert und daher nicht zulässt, dass eine Führungskraft beide Führungsweisen ausüben kann, hebt Bass (1985) diese Bipolarität auf. Er konzipiert transaktionale und transformationale Führung als voneinander unabhängige Dimensionen, die sich jeweils in empirisch gewonnene Subfaktoren untergliedern (genauer Kapitel 2.5.3.2). Weiter geht er davon aus, dass jede Führungskraft über ein breites Repertoire an Führungsverhalten verfügen sollte, um sich entsprechend der situativen Anforderungen verhalten zu können. Laissez Faire, transaktionale und transformationale Führung werden zu diesem Zweck auf einem Kontinuum, das den Aktivitätsgrad des Führungsverhaltens beschreibt, arrangiert. In diesem Bogen von Passivität bis Aktivität sieht Bass den „Full Range of Leadership“[26] verankert und postuliert damit, durch sein Führungsmodell ein sehr breites Führungsspektrum abzudecken (siehe Abbildung 9).

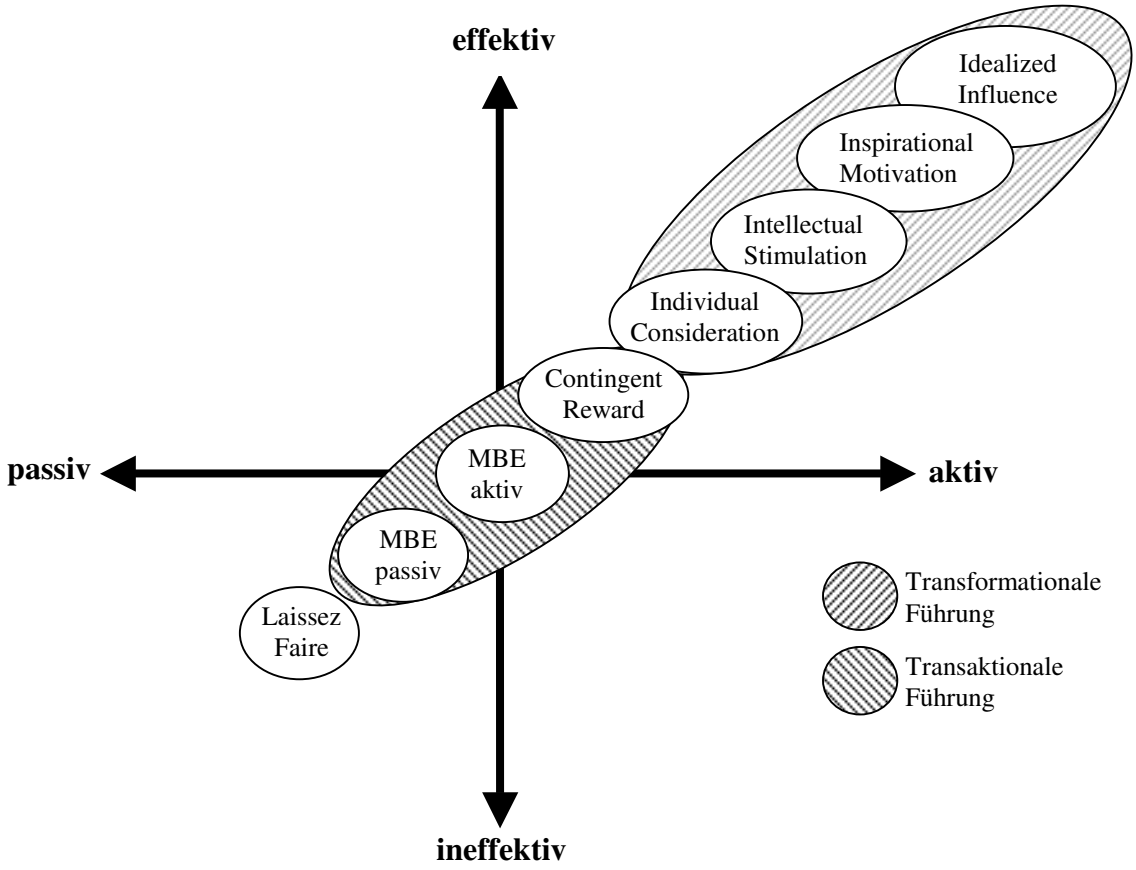


Abbildung 9. „Full Range of Leadership“ (nach Bass & Avolio, 1994)

[26] Eine Diskussion darüber, ob der Terminus tatsächlich angemessen ist, ist überflüssig. Denn Bass und Avolio (2000) selbst erklären, dass sie damit „nur“ dazu anregen wollten, ein breites Führungsrepertoire zu untersuchen sowie dazu herausfordern, das Denken über Führung zu verbreitern.

Letztlich schlägt er ein dreidimensionales Modell vor, dem zufolge erstens die bereits erwähnte Aktivität des Führungsverhaltens, zweitens die Effektivität und drittens – dies folgt dann für den Idealfall – die Einsatzhäufigkeit in folgender Reihenfolge steigen (Bass, 1994, S. 4f.): Laissez Faire, passives „Management by exception“, aktives „Management by exception“, „Contingent Reward“ (Kontingente Belohnung) und letztlich die vier Bausteine der transformationalen Führung. Effektiv bedeutet: „motivating others to achieve higher levels of development and performance“ (Bass & Avolio, 1998, p. 137).
Genauer bestehen folgende Effektivitätskriterien: „Extra Effort“ (Bereitschaft der Mitarbeiter zu zusätzlicher Anstrengung), „Effectiveness“ (wahrgenommene Effektivität des Führungsverhaltens für Einzelnen, Gruppe und Organisation) und „Satisfaction“ (Zufriedenheit mit der Führungskraft). Insgesamt bleibt dabei jeder Führungskraft der Raum, die einzelnen Führungsdimensionen eher partizipativ oder direktiv umzusetzen (Beispiele bei Bass & Avolio, 1993). Das ermöglicht auch, dass eine Führungskraft, die gewöhnlich transformational führt, auf ihren Einfluss in bestimmten Situationen gänzlich verzichtet und sich passiv verhält.

Augmentationshypothese

Erst durch den ergänzenden, additiven Effekt verschiedener Führungsverhaltensweisen sind also außergewöhnlich hohe Leistungen zu erwarten. Transformationaler Führung wird ein sogenannter Augmentationseffekt[27] zugeschrieben, der die Effektivität transaktionaler Führung – erwartete Anstrengung und folglich erwartete Leistung – erhöht und letztlich aufgrund zusätzlicher Anstrengungsbereitschaft zu Erfolg über die Erwartungen hinaus führt (Bass, 1985; vgl. Abbildung 10).

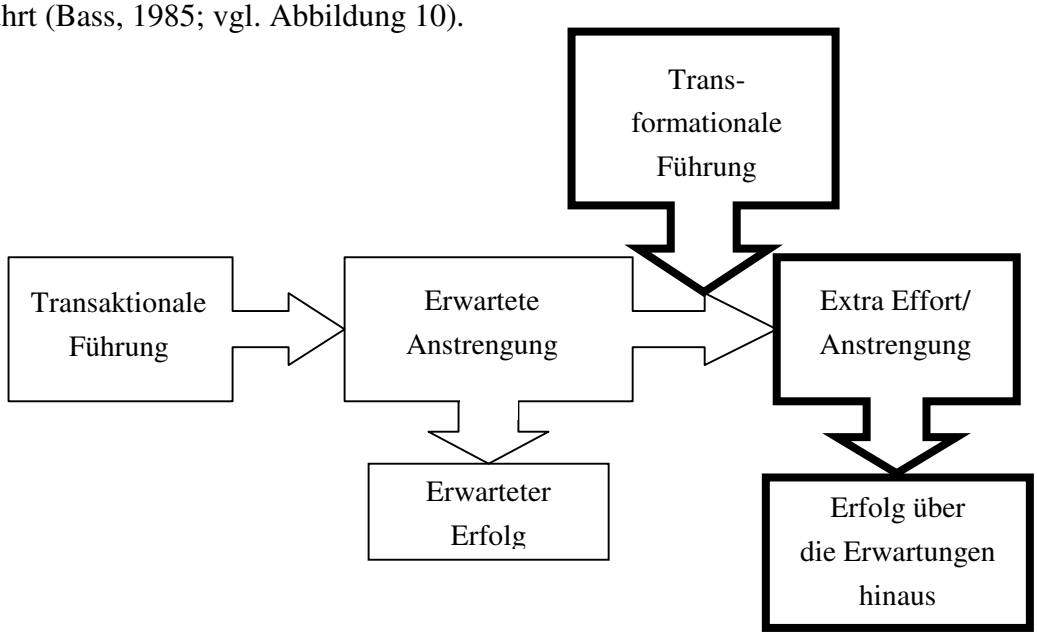


Abbildung 10. Augmentationshypothese transformationaler Führung

[27] In frühen Arbeiten (z.B. Avolio & Bass, 1988) wird diese These als „add-on effect“ bezeichnet.

Wenn die transformationale Führung aber einen eigenen, zusätzlichen Beitrag (zusätzlich zu dem der transaktionalen Führung) zur Varianzaufklärung von Erfolgsvariablen leisten soll, wandelt sich die Frage von einem Entweder-Oder[28] zu einem Wieviel-Wovon, was bereits im "Full Range of Leadership" angesprochen ist. Empirische Unterstützung[29] erhält die Augmentationshypothese durch verschiedene Studien, die durch hierarchische Regressionsanalysen zeigen, dass die transformationale Führung – über kontingente Belohnung beziehungsweise über transaktionale Führung im Allgemeinen hinaus – einen zusätzlichen Beitrag zur Vorhersage folgender Erfolgsmaße leistet (aber nicht umgekehrt): Extra Effort, Effectiveness, Satisfaction, subjektiv geschätzte Arbeitsleistung der Mitarbeiter, Arbeitszufriedenheit, Gesundheit, affektives Commitment, Intention zum Arbeitsplatzwechsel und Kunden- und Marktausschöpfung im Bankgewerbe (Avolio & Bass, 1988; Bass, 1985; Bycio, Hackett & Allen, 1995; Geyer & Steyrer, 1998; Hater & Bass, 1988; Martin & Epitropaki, 2001[30]; Vandenberghe et al., 2002; Waldman, Bass & Yammarino, 1990).

Seltzer und Bass (1990) gehen einen Schritt weiter und finden einen Augmentationseffekt transformationaler Führung über Initiating Structure und Consideration (LBDQ) hinaus. Nicht transformationale Führung allein, sondern transformationale Führung auf der Basis transaktionaler Führung ist modellgemäß also erfolgreich. Was könnte man schon von hoch motivierten Mitarbeitern erwarten, deren Tagesablauf aber aufgrund fehlender Klarheit im Chaos versinken würde? Bass schlägt daher vor, bei Routinetätigkeiten eher transaktional, bei neuen Prozessen jedoch auch transformational zu führen. Effektive transaktionale Führung sollte also die Bedingungen schaffen, die dann tiefere Ebenen des Vertrauens und der Zusammenarbeit ermöglichen (Avolio, 1999).

Einzelne Studien, deren Resultate gegen ein hierarchisches Führungsmodell sprechen und statt dessen darauf hindeuten, dass transaktionale Führung einen ebenso hohen Beitrag zu Erfolgsmaßen leistet wie transformationale Führung und dass letztere keinen Augmentationseffekt aufweist, bleiben in der Minderheit (z.B. Bass et al., 2003). Häufig sind solche inkonsistenten Ergebnisse entweder durch Besonderheiten der Stichprobe oder der Operationalisierung der abhängigen Variablen erklärbar. So wurden bei Bass et al. (2003)

---

[28] transaktionale *oder* transformationale Führung bei Burns

[29] Immer wieder werden neben die empirischen Ergebnisse Fallbeispiele großer Führungspersönlichkeiten (z.B. Franklin Roosevelt oder Charles de Gaulle) gestellt, um zu unterstreichen, dass sich die Komplementarität der Führungsdimensionen mit unterschiedlicher Schwerpunktsetzung erstens überhaupt und zweitens mit erfolgreichen Konsequenzen realisieren lässt (z.B. Waldman et al., 1990). Es sei daran erinnert, dass sich die neo-charismatischen Führungstheorien insgesamt aus qualitativen – historischen, biografischen oder führungsliteratur-gestützten – Arbeiten entwickelt haben.

[30] In der Studie von Martin und Epitropaki (2001) zeigt sich, dass die transformationale Führung zusätzlich mehr Varianz für diejenigen aufklären kann, die sich stark mit der Organisation identifizieren als für diejenigen, die sich schwach identifizieren. Die Autoren erklären dies folgendermaßen (S. 259): „People with a low organizational identification tend to resist being 'pushed' beyond the performance standards set by their work contract and if they are motivated to work hard, it is usually to satisfy self-interest needs rather than the collective good of their work-group or organization."

Leistungen in Gefechtssituationen als Kriterien herangezogen, wo man sich fragen muss, ob es in solchen Situationen überhaupt um mehr als Anweisungen gehen kann.

2.5.3.2 Die Führungs-Bausteine im „Full Range of Leadership“: Laissez Faire, transaktionale und transformationale Führung

Die Bausteine, die den "Full Range of Leadership" abdecken, werden im Folgenden charakterisiert (Überblick Tabelle 2).

Tabelle 2. Charakteristika der verschiedenen Führungsbausteine

Laissez Faire: vermeidet Entscheidungen, übernimmt keine Verantwortung
Transaktionale Führung
Kontingente Belohnung: erkennt Fähigkeiten und Leistung, belohnt gute Leistung
Management by exception (aktiv): sucht Fehler und Abweichungen, korrigiert diese
Management by exception (passiv): korrigiert dann, wenn Fehler eingetreten ist
Transformationale Führung („4i“)
Idealisierte Beeinflussung/Charisma: verbreitet Stolz, gewinnt Vertrauen
Inspirierende Motivierung: gibt Visionen, kommuniziert hohe Erwartungen, verwendet Symbole
Intellektuelle Stimulierung: unterstützt und fordert divergentes Denken und Problemlösen
Individualisierte Berücksichtigung: schenkt jedem Mitarbeiter Beachtung, entwickelt individuell

Laissez Faire

Dieser untere Endpol des "Full Range of Leadership" Modells ist durch extreme Passivität gekennzeichnet und kann daher auch als Nicht-Führung beziehungsweise Verzicht oder Verweigerung von Führung und Einfluss bezeichnet werden. Im Gegensatz zur transaktionalen Führung ist Laissez Faire durch fehlende Transaktion gekennzeichnet (Bass, 1994, S. 4). Es ist auf der zweiten Modelldimension durch maximale Ineffektivität beschrieben, da anstehende Aufgaben nicht bewältigt, gleichzeitig aber sogar Probleme erzeugt werden, indem beispielsweise nicht interveniert oder nicht entschieden wird. Es sei an dieser Stelle nur vermerkt, dass Laissez Faire bereits bei Lewin, Lippitt und White (1939) den am wenigsten erfolgreichsten Führungsstil darstellte.

Transaktionale Führung

„Let's make a deal."

Als „Basis" guter Führung gilt das, womit sich die traditionelle Führungsforschung ausführlich beschäftigt hat: der Austauschprozess zwischen Führungskraft und Mitarbeiter (Avolio & Bass, 1988; Bass, 1985). Die Geführten verfügen über Ressourcen, die der Führende nachfragt, und umgekehrt verfügt der Führende (im Auftrag der Organisation) über Ressourcen, die die Geführten nachfragen. Empirisch ergeben sich im Rahmen der Modellentwicklung zwei relevante Faktoren zur Beschreibung dieses Austauschprozesses: Kontingente Belohnung und "Management by Exception". Ersteres wird auch als „konstruktive Transaktion", Letzteres als „korrektive Transaktion" bezeichnet (Bass & Avolio, 1997).

Wenn die Führungskraft für ihre Geführten Klarheit und Transparenz schafft, dann erleichtert dies die Erreichung der angestrebten Ziele. Akzeptieren die Geführten die Art der Belohnungen, dann verhalten sie sich gemäß der Erwartungen, um diese zu erreichen oder um Bestrafung zu vermeiden. „Simply put, transactional leadership is contingent reinforcement. […] Contingent avoidance reinforcement may be seen in management-by-exception." (Avolio & Bass, 1988, S. 30)

So stellt Bass die transaktionale Führung also auf die Basis der Lern- und Austauschtheorien und versucht dann unter Rückgriff auf die Weg-Ziel-Theorie beziehungsweise Erwartungs-Valenz-Modelle die Motivation der Geführten zu erklären, das gesetzte Ziel (erwartete Anstrengung und erwarteten Erfolg) erreichen zu wollen (vgl. Abbildung 11).

Genauer: Vermittelt eine Führungskraft klar, welchen Weg ein Mitarbeiter zur Zielerreichung gehen muss, stärkt dies seine subjektive Erfolgserwartung, ans Ziel zu gelangen. Erkennt eine Führungskraft weiter, welche Bedürfnisse ein Mitarbeiter hat und macht sie zusätzlich transparent, wie diese im Austausch für adäquate Leistung (also Zielerreichung) befriedigt werden können, dann steigt für den Mitarbeiter der Wert des gesetzten Ziels. Letztlich sind also beide Komponenten der Erwartungs-Valenz-Theorie der Motivation von Vroom (1964) erhöht – und damit die Motivation, das Ziel zu erreichen. Außerdem wird in der Theorie dort auf kognitive Emotionstheorien zurück gegriffen, wo es um die Erklärung des Prozesses zwischen „charismatic qualities" und „follower excitement" geht (Avolio & Bass, 1988).

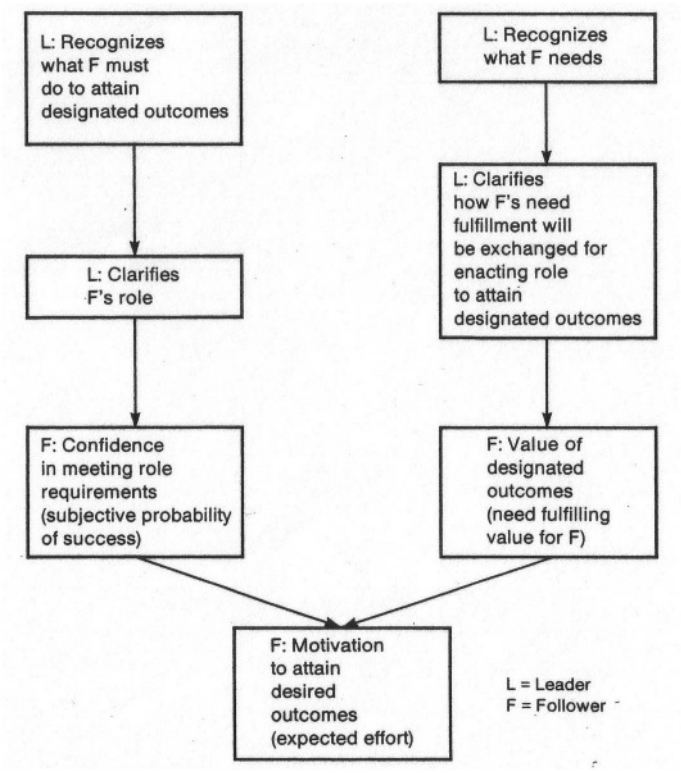


Abbildung 11. Transaktionale Führung und Anstrengung der Geführten (aus Bass, 1985, p. 12)

Im Einzelnen gehören folgende Bausteine zur transaktionalen Führung:

Kontingente Belohnung ("Contingent Reward")

Aufbauend auf dem Prinzip des operanten Lernens ist davon auszugehen, dass kontingente Belohnung auch in Führungsbeziehungen wirkt (z.B. Luthans & Kreitner, 1975; Oldham, 1976; Podsakoff et al., 1982; Reitz, 1971; Sims, 1977). Kontingent lässt sich in diesem Zusammenhang am besten mit leistungsorientiert beschreiben. Diese Führungsdimension zielt insbesondere darauf, Rollen- und Leistungsanforderungen zu klären. Wenn der Mitarbeiter in der Folge diese Erwartungen erfüllt, dann folgt eine adäquate Belohnung, die wiederum die Erwartungen (Bedürfnisse) der Mitarbeiter trifft. Einzuwenden ist an dieser Stelle, dass nicht-kontingente Belohnung ähnlich wirksam ist wie kontingente Belohnung (Podsakoff, Todor & Skov, 1982).

Management by exception

Den zweiten transaktionalen Faktor, der im Laufe der Zeit in einen aktiven und passiven Faktor getrennt wurde, repräsentiert "Management by exception".[31] Im aktiven Part sucht die Führungskraft nach Normabweichungen der Mitarbeiter, während sie im passiven Part wartet, bis diese eingetreten sind und dann erst entsprechend interveniert.

Aktives „Management by exception": Führungskräfte überwachen die Arbeitsprozesse insofern, dass sie sich auf Fehler, Risikosituationen, Probleme oder allgemein von der Norm abweichende Ausnahmen konzentrieren. In Situationen, die bereits sehr routiniert verlaufen, kann eine derartige aktive Kontrolle angemessen sein. Geht es jedoch darum, besondere Ziele zu erreichen, erweist sie sich als ineffektiv.

Passives „Management by exception" (Führung durch Eingreifen nur im Ausnahmefall): Im Regelfall zeichnet sich diese Führungsrolle durch Nicht-Führung aus; die Führungskraft interveniert nämlich erst dann, wenn ein Ausnahmefall, Fehler oder Problem aufgetreten ist. Dies kann folglich nur korrektiv sein. Im Gegensatz zu Laissez Faire wird hier allerdings der Status Quo nicht ignoriert, sondern beobachtet, reglementiert und korrigiert.

Das generelle Prinzip lässt sich folgendermaßen zusammen fassen. Verlässt ein Mitarbeiter auf seinem Weg zum Ziel seine definierte Rolle der Austauschbeziehung, tritt also ein Ausnahmezustand ein, dann unterbleibt die in Aussicht gestellte Belohnung bzw. erfolgt eine Bestrafung. Dies führt nach operanten Prinzipien dazu, dass die Auftretenswahrscheinlichkeit derartiger Ausnahmen sinkt. Da die Führung im Ausnahmefall sich an ungewünschten Verhaltensweisen der Mitarbeiter orientiert, kann nicht belohnend, sondern nur bestrafend reagiert werden. Uneinheitlich ist die Ergebnislage zu kontingenter Bestrafung (Übersicht siehe Bass & Avolio, 1997), wohingegen *nicht*-kontingente Bestrafung negativ mit Leistung korreliert (Podsakoff & Schriesheim, 1985). Die klinische Psychologie liefert in diesem Zusammenhang das Konzept der erlernten Hilflosigkeit (Seligman, 1975).

Die Wirkungsbeschreibung transaktionaler Führung erinnert nicht zu Unrecht an die Weg-Ziel-Theorie der Führung (Evans, 1974; House, 1971; House & Mitchell, 1974; vgl. Kapitel 2.4). Bass (1985) greift aber nicht nur explizit auf deren Erklärungsmechanismen zurück, sondern auch auf die einschränkenden Gedanken von House und Mitchell (1974), dass diese Art von Führung nicht in jedem Bedingungsgefüge höchste Nützlichkeit verspricht – dann nämlich nicht, wenn Ziele, Belohnungen oder „guidance“ bereits deutlich

[31] Management-by-Konzepte hatten in den 80er Jahren einen florierenden Markt. Insofern ist die Namensgebung sicherlich aus diesem Boom der Führungstechniken heraus zu verstehen.

geklärt sind. Es ist auch immer dann nicht zu empfehlen – so lässt sich hinzufügen – wenn das Menschenbild des homo oeconomicus nicht die alleinige Führungsbasis darstellen soll.

Selbst wenn durch eine bessere Nutzung der transaktionalen Führung Effektivitätssteigerungen zu erwarten wären (Bass & Avolio, 1997), nennen Avolio und Bass (1988) einige Schwachstellen, die bestehen bleiben: Erstens können die Verstärker nicht immer individuell abgestimmt werden, so dass sie streckenweise ohne Konsequenzen bleiben. Zweitens haben externe Leistungsverstärker konkurrierende Kontingenzen im Feedback von anderen Personen oder in selbstbelohnenden Konsequenzen, die direkt aus der Arbeit bezogen werden, und verlieren dadurch ebenso ihre Wirksamkeit. Drittens forcieren Führungskräfte, die die kontingente Verstärkung als wichtigstes Führungsmittel favorisieren, insbesondere kurzfristige Zielverfolgung – zulasten von langfristiger. Dieser Effekt tritt vor allem dann ein, wenn Mitarbeiter den Eindruck gewinnen, das Belohnungsverhalten ihres Vorgesetzten manipulieren zu können. Avolio und Bass (1988) schließen, dass kontingente Verstärkung sinnvoll ist, um Rollenverhalten zu erlernen, weniger um Extra Effort zu generieren. Es gibt mehr als transaktionale Führung – das war der Ruf der neuen Führungsära. Hin zu außergewöhnlichen Zielen muss darüber hinaus auch transformational geführt werden. Zu starke transaktionale Führung kann dafür sogar hinderlich sein (Avolio & Bass, 1988, S. 30).

Transformationale Führung

Transformationale Führungskräfte sollen in erster Linie zu Anstrengungen über die ursprünglichen Erwartungen hinaus motivieren (Bass, 1985; Bass & Avolio, 1998). Um dies zu erreichen, sollen die Mitarbeiter so beeinflusst werden, dass sie die angestrebten Ziele (Visionen; siehe unten) als sehr erstrebenswert und insgesamt als wertvoll sowie als erreichbar erachten (vgl. Abbildung 12).[32] Im Idealfall ordnen die Mitarbeiter ihre Selbstinteressen dann unter die Teaminteressen und verändern ihre Bedürfnisse in Richtung Selbst-Aktualisierung[33]. Da dies durch eine Austauschbeziehung mit “normaler” Compliance, “normalem” Einsatz und “normalen” Erfolgen nicht erreicht werden kann, muss Führung auch den Selbstwert der Geführten steigern sowie eine echte Vertrauensbeziehung zwischen Führenden und Geführten im Sinne hohen "Involvements" und gegenseitiger Verpflichtung hoher Ziele darstellen. Insofern geht die transformationale Führung also bedeutend über die transaktionale hinaus (Bass, 1998).[34]

[32] Auch hier wird also auf Erwartungs-Valenz-Modelle zurück gegriffen.
[33] Hier bezieht sich Bass (1985) ausdrücklich auf die Bedürfnispyramide von Maslow, ohne Kritik an dieser aufzunehmen.
[34] Auf die Nähe zur charismatischen Theorie von Shamir et al. (1993) sei verwiesen.

Genauer: Alle vier transformationalen Führungsbausteine – die vier I's (Bass & Avolio, 1994) - sind im "Full Range of Leadership"-Modell durch hohe Aktivität und hohe Effektivität gekennzeichnet. Sie sind stichpunktartig durch charismatische, emotional-inspirierende, fordernde und fördernde Führung zu charakterisieren und werden im Folgenden genauer vorgestellt.

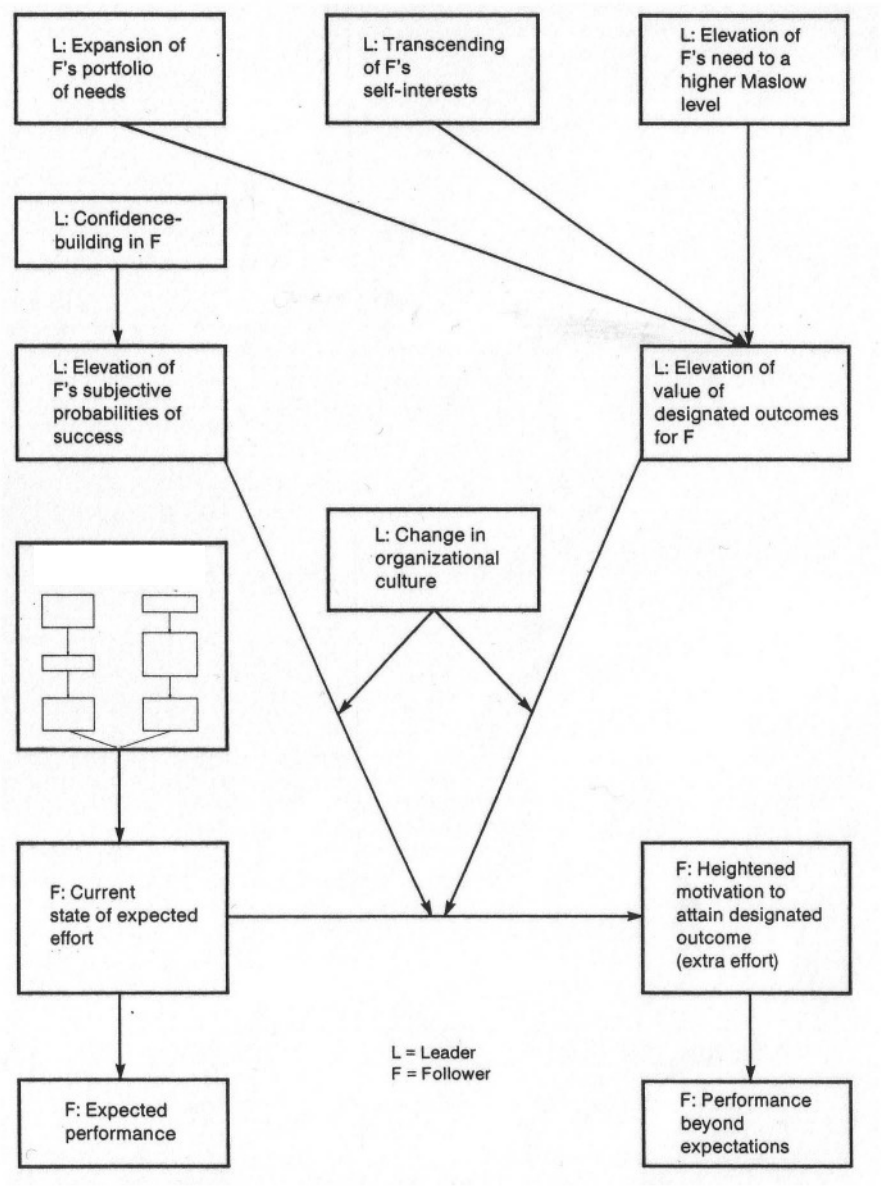


Abbildung 12. Transformationale Führung und Extra Effort der Geführten (aus Bass, 1985, p. 23)

Idealized Influence

"Idealized Influence" kann verstanden werden als „ideally, influential leadership“ (Bass & Avolio, 1994, S. 1), als Beeinflussung hin zu einem Ideal[35] oder als Einfluss von einer idealisierten Quelle. Im engeren Sinne geht es hier um charismatische Führung, die in einen attributionalen und behavioralen Teil untergliedert wird. Diese Trennung erscheint auch aufgrund der Definition von Charisma sinnvoll, da dies nicht allein durch charakteristische Verhaltensweisen, sondern immer auch durch sein Ergebnis definiert ist, also letztlich dadurch, ob das Ergebnis ausreichend außergewöhnlich ist, so dass dem vermeintlichen Verursacher auch das außergewöhnliche Attribut Charisma zugesprochen wird. Es geht also um eine spezifische Ausstrahlungskraft einer Führungsperson, die unabhängig von deren fachlichen Kompetenzen wirkt. Die Führungskraft wird bewundert, ihr wird Respekt und Vertrauenswürdigkeit zugeschrieben (attributional) – sie wird zu einer Identifikationsfigur und vermittelt Enthusiasmus. Auf der Verhaltensebene ist zu beobachten, dass die Führungskräfte von ihren Geführten hohe Leistungs- und ethische Standards erwarten, die sie aber auch an ihr eigenes – modellhaftes – Verhalten anlegen. Dafür nehmen sie nicht nur Risiken in Kauf, sondern ordnen auch die eigenen Bedürfnisse unter die anderer. Sie strahlen aus, genau das Richtige zu tun und bewirken letztlich, dass sie von den Geführten gerne nachgeahmt werden (behavioral). Es ist also der Einfluss angesprochen, der mit Vorbildwirkung und Glaubwürdigkeit zu begründen ist.

Inspirational Motivation

> „Ich machte unermüdlich Propaganda für alles, was wir taten [...] Wenn ich eine Idee oder Botschaft hatte, die ich in der Organisation etablieren wollte, konnte ich nicht oft genug darauf hinweisen. Ich wiederholte sie immer und immer wieder [...], um hunderttausende Menschen restlos davon zu überzeugen.“
> (J. Welch, 2003, S. 401)

Das Konzept der Inspirierenden Motivierung geht auf Yukls (1981) „inspirational leadership“ zurück und integriert die Ansätze zur visionären Führung. Diese Führung wirkt dann transformierend, wenn erstens die Zielkomponente und zweitens das zugrunde liegende Bedürfnis so beeinflusst werden, dass sie Begeisterung und Motivation für dasselbe Ziel fördern. Es handelt sich generell um einen „ansteckenden“ Teil der Führung, durch den Enthusiasmus für ein Ziel verbreitet wird. Inspirierende Führungskräfte haben Visionen, kommunizieren sie, symbolisieren und leben sie. So wird die Überzeugung vermittelt, dass die Zukunft deutlich attraktiver ist als die Gegenwart. Wichtig ist, dass die Führungskräfte durchwegs optimistisch und zuversichtlich sind, diese Vision gemeinsam mit ihrem

[35] Diese Bedeutung ist in der aktuellen Konzeption der transformationalen Führung schwerpunktmäßig im zweiten Faktor, der inspirierenden Motivierung, integriert, der im ursprünglichen Modell noch gänzlich unter Idealized Influence subsumiert war.

Team auch erreichen zu können. Indem sie solche Herausforderungen an die Arbeit der Geführten richten, werden dieser Bedeutsamkeit, Sinn und Ganzheitlichkeit zugesprochen. Auf diese Weise wächst die Motivation, sich für die Vision zu engagieren.

Nach Brodbeck et al. (2000; GLOBE) finden sich sowohl inspirierende als auch visionäre Führung unter den Führungsattributen, die in fast allen 22 untersuchten europäischen Ländern als Attribut hervorragender Führung gelten. Ebenso werden neo-charismatische Verhaltensweisen in den 60 weltweit betrachteten Ländern als prototypisch (auf einer 7er Skala > 6) für hocheffektive Führungskräfte beschrieben (House et al., 1998).

Intellectual Stimulation

Intellektuelle Stimulierung umfasst fordernde Führungsverhaltensweisen und sollte von einer Führungskraft insbesondere dann verfolgt werden (Bass, 1985), wenn ihre Mitarbeiter schlecht definierte Probleme zu lösen haben. Dann regen Führungskräfte ihre Mitarbeiter zu kreativem und innovativem Denken an, indem sie gewohnte Annahmen in Frage stellen, Probleme in einen neuen Bezugsrahmen stellen und von verschiedenen, auch unkonventionellen Perspektiven aus betrachten. Genau diese Art zu denken fordern sie auch von ihren Mitarbeitern (Bass & Avolio, 1994).[36] Insgesamt nimmt eine intellektuell stimulierende Führungskraft eine Rolle ein, die Entscheidungen hinterfragt und die andere lehrt, Entscheidungen in Frage zu stellen (Keller, 1992). Kreativitätsförderliche Rahmenbedingungen werden dadurch geschaffen, dass explizit zu kreativen Ideen ermuntert wird, ohne Fehler oder Kritik öffentlich zu äußern. Es geht dabei darum, die Anzahl sowie die Vielfalt der Ideen zu steigern. Hohe Anzahl beziehungsweise unterschiedliche Denkrichtungen können mit Guilfords (1950) divergenten Denkoperationen der Flüssigkeit bzw. Flexibilität umrissen werden. Mit der Anzahl vorgeschlagener Wege steigt die Wahrscheinlichkeit für einen optimalen Lösungsweg. Dass auch neuen, dem Gros widersprechenden Lösungsvorschlägen Raum gewährt wird, kann mit sozialpsychologischen Befunden untermauert werden. Durch Minoritätseinfluss steigt nämlich sowohl die generierte Ideenanzahl als auch die Wahrscheinlichkeit richtiger und origineller Problemlösung (Nemeth & Kwan, 1985, 1987; Nemeth & Wachtler, 1983). Zieht man noch einmal die konzeptuelle Trennung zwischen Management und Führung heran (vgl. S. 52), sind Manager als Problemlöser, Führer als Problemsucher zu beschreiben, die somit Quellen für visionäre Trends überhaupt erst wahrnehmen. Guilford (1950) würde diese Kompetenz als Problemsensitivität bezeichnen.

[36] Auch wenn Intellektuelle Stimulierung der Faktor mit der engsten Beziehung zu kreativem Outcome ist, können auch andere transformationale Faktoren zu erhöhter Kreativität beitragen. So bestätigen Sosik, Avolio und Kahai (1998), dass Inspirierende Motivierung einen signifikanten Beitrag zu kreativen Leistungen leistet. Mit Shamir et al. (1993) kann erklärt werden, dass transformationale/inspirierende Führer die intrinsische Motivation ihrer Geführten erhöhen, indem sie deren Selbstkonzept mit dem Gruppeninteresse und der gemeinsamen Mission verknüpfen. Dass intrinsische Motivation aber auch die Kreativität fördert, zeigt wiederum Amabile (1983).

Individualized Consideration

> „Ich merkte, dass dies nicht einfach dahingesagt war, er meinte es wirklich so. Und obwohl ich nur zufällig Zeugin dieser Szene geworden war, konnte ich erkennen, wie das Gesicht des Hausmeisters vor Stolz aufleuchtete. Echte Führungskräfte, das erkannte ich an diesem Morgen, müssen in der Lage sein, Grenzen zu überschreiten. Mein Vater sagte immer: >Die Leute arbeiten nicht für eine Firma. Sie arbeiten für andere Menschen.<“ (Morgan, 1998, S. 99)

Zentral für die Individualisierte Berücksichtigung ist ein Führungsverständnis, das für individuelle Unterschiede insofern aufgeschlossen ist, dass jeder Mitarbeiter individuell gefördert wird. Indem transformationale Führungskräfte effektiv und aktiv zuhören, erkennen sie die persönlichen Bedürfnisse ihrer Mitarbeiter. Dabei geht es vor allem auch um Bedürfnisse nach Weiterbildung und Wachstum, so dass Führungskräfte das Potential der einzelnen Mitarbeiter gezielt fördern sollten. Zu diesem Zweck wird ein Arbeitsklima geschaffen, in dem Delegation als Lernchance begriffen wird (vgl. auch Kuhnert, 1994). Insgesamt verbirgt sich hinter diesem Baustein die Führungsrolle als Coach für alle Mitarbeiter.

Zum Überblick sind in Tabelle 3 Charakteristika der transaktionalen Führung denen der transformationalen Führung gegenüber gestellt.[37]

[37] Zur Konzeption von Burns lassen sich drei Unterschiede festhalten (Bass, 1985): In der Konzeption von Bass sollen erstens die Bedürfnisse der Geführten stärker erweitert werden, zweitens muss transformationale Führung nicht notwendigerweise wohlwohlende Ziele verfolgen (Ausdifferenzierung bei Bass folgt später, vgl. Kapitel 2.5.4.2) und drittens können Führungskräfte ein ganzes Spektrum an Führungsverhaltensweisen zeigen und sind nicht – wie bei Burns – auf ein Entweder-Oder eingeschränkt.

Tabelle 3. Vergleich transaktionaler und transformationaler Führung (modifiziert nach Tichy & Devanna, 1986a sowie Bass & Avolio, 1994)

Vergleichsdimension	**Transaktionale Führung**	**Transformationale Führung**
Zeitorientierung	kurz, unmittelbar	lang, zukünftig
Koordinierung über	Regeln	Ziele und Wertkongruenz
Kommunikation	vertikal	multidirektional
Fokus	finanzielle Ziele	externe und interne Kunden
Belohnungssystem	organisational, extrinsisch	individuell, intrinsisch
Machtlegitimation	Belohnung, Position	Anerkennung durch Mitarbeiter Identifikation
Entscheidungsfindung	zentralisiert, von oben nach unten	dezentralisiert, von unten nach oben
Mitarbeiter	ersetzbar	wertvolle Ressource
Compliance-Mechanismus	Direktiv	Verständnis, Übernahme der Ziele
Veränderungsbereitschaft	nicht notwendig	unabdingbar
Führen durch	Anreiz, Profit	Vision, Werte
Kontrolle	Rigidität, Konformität	Selbstkontrolle
Aufgaben	ausschnitthaft, individuell zugeordnet	Job enrichment, Gruppenaufgaben

Trotz der Unterschiede zwischen prototypischer transformationaler und transaktionaler Führung kann die transformationale Führung insofern als Spezialfall der transaktionalen Führung verstanden werden, als der Mitarbeiter die Transaktion quasi selbstregulierend durchführt (vgl. Avolio & Bass, 1988). Die Prozesse, durch die sich Mitarbeiter für die Verfolgung einer Vision motivieren, unterliegen nicht externalen, sondern selbstverstärkenden Mechanismen. So hat die Annäherung an die Vision selbst, aber auch die Kompetenzerweiterung auf dem Weg dorthin Belohungscharakter. In gewisser Hinsicht könnte die Dimension der "Tauschgüter" über den Austausch von Vertrauen und Commitment hinaus erweitert werden.

Zusammenfassend stellt Tabelle 4 dar, dass die transformationale Führung erstens deutlich mehr umfasst als die Ausübung klarer Führungsaufgaben und dass für sie zweitens ein breiteres Wirkungsspektrum formuliert ist als für die transaktionale Führung (siehe oben), traditionelle Führungsaufgaben und -theorien (Kapitel 2.3 und 2.4).

Tabelle 4. Transformationale Führung und dadurch angestrebte Wirkung auf Seiten der Mitarbeiter (nach Tosi et al., 2000, S. 474)

Transformationale Führungskraft	Mitarbeiter
Aufgabe	
starke Visionen und hohe Ziele	erhöhtes Bewusstsein für neue Möglichkeiten
Aufmerksamkeitslenkung auf Veränderungsbedarf	
hohe Erwartungen und gewisses Maß an Risiko	Bewusstsein für neue Leistungsstufen
Vertrauen in andere (im Bereich des Möglichen)	Selbstwirksamkeitserwartung
uneigennützige Förderung	identifiziert sich mit einer Sache
Kommunikation von Problemlösungen/ Strategien, innovative Strategien	Anregung aufgabenrelevanter Motive „harte Arbeit“
Sozioemotional	
enthusiastisch	Gefühl, auf der richtigen Seite zu stehen
gibt sich für die proklamierte „Sache“ her	Identifikation mit FK und Aufgabe
ansprechbar, sympathisch	bringt FK Zuneigung, Bewunderung, Vertrauen entgegen
spricht tiefe Werte und Hoffnungen an	eifert der FK nach
verletzt entgegen gebrachtes Vertrauen nicht, individualisierte Berücksichtigung	fühlt sich persönlich akzeptiert, unterstützt und verstanden
fördert Gruppenkohäsion	Wunsch, in der Gruppe zu bleiben
Macht	
Experte, aufgabenbezogene Macht	breite Basis für Compliance
Identifikationsmacht	
bestimmtes Maß an Belohnungsmacht	
Bereitschaft, mit Hilfe von sozialisierter Macht etwas zu bewirken (vgl. S. 83)	von FK als Person und von Aufgabe (Mission) beeindruckt und diesen verpflichtet
verschiedene emotionale Angebote, inspirierende Anregungen	Vertrauen Selbstwirksamkeit
persönliche Präsenz (Hier und Jetzt), Bedürfnisse und Werte als Grundlage	persönliche Involviertheit, Mitgestaltung (intellektuell, emotional, sozial, Umwelt)

Anmerkung. FK = Führungskraft (transformational).

2.5.4 Empirische Befundlage zur Wirkung transformationaler Führung

Seitdem erste quantitativ-empirische Ergebnisse von Bass (1985) vorgelegt wurden, erfuhr das Modell in einer Vielzahl an Studien Bestätigung. Es zeigte sich, dass in den verschiedensten Settings, Hierarchieebenen und Nationen transformational geführt wird (vgl. Kapitel 2.5.4.1). Im GLOBE-Projekt erreichten die Attribute transformationaler Führung im Vergleich zu anderen Führungsmustern den höchsten Mittelwert (Brodbeck, 2000; Den Hartog et al., 1999).

Aber was wird überhaupt als Outcome transformationaler Führung betrachtet? „The outcomes of transformistic organizations are not solely exceptional products, services, or profits (although these should indeed result); they are qualitative changes in the well-being of society. Transformation of human capabilities within organizations that change the larger society could be tantamount to a new social movement for the 21st century." (Hickman, 1998). Auch wenn bei Hickman sehr weit reichende Veränderungen auf die transformationale Führung attribuiert werden, soll an dieser Stelle betont werden, dass es auch auf personeller Ebene vor allem die subjektiven Faktoren sind, in denen sich das außergewöhnliche Engagement widerspiegelt (Avolio & Bass, 1988). So galten zu Beginn dieser Forschungslinie insbesondere subjektive Einschätzungen der Mitarbeiter zu Extra Effort, Effectiveness und Satisfaction als abhängige Variablen.

Inzwischen sind jedoch unterschiedlichste Erfolgsindikatoren untersucht, auch solche, die unabhängig von der Mitarbeiterperspektive gemessen werden. Dazu zählen subjektive Erfolgseinschätzungen anderer Personen, beispielsweise der Vorgesetzten der jeweiligen Führungskräfte (z.B. Atwater, Ostroff, Yammarino & Fleenor, 1998), genauso wie harte Kennziffern (z.B. Marktanteil, Aktienkurs, Gewinn und Eigenkapitalüberdeckung in der Management-Simulation von Avolio, Waldman & Einstein, 1988 oder Kunden- und Marktausschöpfung in der Felduntersuchung im Bankgewerbe von Geyer & Steyrer, 1998).

2.5.4.1 Sonnenseiten

Empirische Daten lassen mit fast überraschender Konsistenz erkennen, dass transformationale Führungskräfte in sehr verschiedenen organisatorischen Settings als effektiv gelten (vgl. Tabelle 5), so zum Beispiel im Militär (z.B. Bass, 1985; Bass, Avolio, Jung & Berson, 2003; Dvir, Eden, Avolio & Shamir, 2002; Masi & Cooke, 2000), in der Industrie (z.B. Hater & Bass, 1988; Martin & Epitropaki, 2001; Waldman et al., 1987), im Einzelhandel (z.B. Koene, Pennings & Schreuder, 1992), in der Politik (z.B. Pillai, Williams, Lowe & Jung, 2003) und politischen Verwaltung (z.B. House, Spangler & Woycke, 1991),

im Bank- und Dienstleistungsgewerbe (z.B. Geyer & Steyrer, 1998; Rai & Sinha, 2000), im Bildungsbereich (z.B. Koh, Steers & Terborg, 1995; Roberts, 1985), im Sport (z.B. Charbonneau, Barling & Kelloway, 2001) im klassischen Gesundheitswesen (Bycio et al., 1995; Vandenberghe, Stordeur & D'hoore, 2002) und deren Randbereichen wie beispielsweise der Alcoholic Rehabilitation Organization (Trice & Beyer, 1986). Diese Aufzählung zeigt, dass Daten weit über den militärischen Sektor hinaus vorliegen, in dem insbesondere von der Arbeitsgruppe von Bass die Mehrzahl der Studien durchgeführt wurde.

Trennt man die betrachteten Organisationen nach ihrer Zugehörigkeit zum öffentlichen (z.B. Militär) oder privaten Sektor (z.B. Industrie), dann zeigen meta-analytische Ergebnisse (Lowe, Kroeck & Sivasubramaniam, 1996), dass der Zusammenhang zwischen allen transformationalen Faktoren (mit Ausnahme von Individualisierter Berücksichtigung) und Erfolgsmaßen für Organisationen im privaten Sektor deutlich geringer ausfällt als für solche im öffentlichen Sektor (vgl. Fußnote 41). Diese Ergebnisse stimmen damit überein, dass sich für militärische Führer, die bei Lowe et al. (1996) einen Großteil des öffentlichen Sektors ausmachen, ein stärkerer Zusammenhang zwischen charismatischer Führung und Erfolgsvariablen ergibt als für zivile Führungskräfte (Fuller et al., 1996). Möglicherweise erfordert die militärische Aufgabenstruktur, die oft von langfristigen und zugleich unsicheren Zielen und Krisensituationen bestimmt ist, eher als nicht-militärische Strukturen ein Vertrauen in den richtigen Weg und eine Führungsperson, die eine erstrebenswerte Vision weist, so dass letztlich transformationale Führer häufiger selektiert werden und dann auch erfolgreicher sind. Der häufige Kontakt zwischen Führer und Geführten in militärischen Einheiten bietet den Führern zudem mehr Möglichkeiten der Einflussnahme.

Eine andere Erklärung für den engeren Erfolgszusammenhang in militärischen Organisationen liegt in der Organisationskultur: Während es die Militärkultur mit sich bringt, dass die angestellten Personen häufig während ihres gesamten Berufslebens starke personenbezogene Macht erfahren, bedeutet Macht im zivilen, industriellen Sektor häufig nur Belohnungsmacht, die durch schnelle Führungskräftenachfolgen kurzfristig und änderbar erscheint. Außerdem haben die Angehörigen militärischer Organisationen häufig ohnehin ähnliche Werte (Patriotismus, Pflichtbewusstsein), die Ansatzpunkte für transformationale wertebasierte Visionen bieten.

Bezüglich der meisten Settings scheint sich die Wirksamkeit transformationaler Führung jedoch nicht zu unterscheiden (Meta-Analyse von Judge & Piccolo, 2004). Auch die Höhe der Führungsebene spielt kaum eine Rolle. (Fuller et al., 1996; Lowe et al., 1996).[38]

[38] Lediglich die Dimension der Intellektuellen Stimulierung weist für Führungskräfte auf höheren Hierarchieebenen einen stärkeren Erfolgszusammenhang auf als für Führungskräfte auf niedrigeren Hierarchieebenen (Lowe et al., 1996). Erklärt werden kann dies beispielsweise damit, dass kreative Entscheidungen auf höheren Ebenen in der Regel eine größere Reichweite aufweisen als auf niedrigeren Ebenen, da hier der begrenzte Handlungsspielraum die Wirkungstiefe stark begrenzt.

Es wurde bestätigt, dass transformationale Führung hinsichtlich der verschiedensten Outcome-Variablen effektiver ist als transaktionale Führung, was die hierarchische Struktur des Führungsmodells bestätigt. Dieser Befund konnte inzwischen durch Meta-Analysen (Dumdum, Lowe & Avolio, 2002; Fuller et al., 1996; Judge & Piccolo, 2004; Lowe et al., 1996) gesichert werden. Betrachtet man die einzelnen Dimensionen des Führungsmodells von Bass, lassen sich in der Regel folgende Ergebnisse finden: Charismatische und Inspirierende Führung (die häufig gemeinsam behandelt werden) weisen die höchsten positiven Zusammenhänge mit den Erfolgsmaßen auf. Es folgen auf dem zweiten und dritten Rang Individuelle Berücksichtigung beziehungsweise Intellektuelle Stimulierung. Etwas geringer, dennoch deutlich positiv, korreliert auch kontingente Belohnung als erster Faktor der transaktionalen Führung mit den Erfolgsindikatoren. Für "Management by exception" muss der aktive vom passiven Teil getrennt werden, denn während ersterer positive Zusammenhänge mit Erfolg aufweist, sind es bei letzterem negative. Genauso regelmäßig findet sich eine deutlich negative Korrelation zwischen Laissez Faire und den Erfolgsmaßen. Auch wenn aktive Bausteine der transaktionalen Führung mit den genannten Outcome-Variablen positiv korrelieren, so ist es doch die transformationale Führung, die einen stärkeren Beitrag liefert.

Abbildung 13 verdeutlicht diesen Trend anhand der Befunde der Meta-Analyse von Lowe et al. (1996) getrennt für „subjektive“ und „objektive“ Erfolgsmaße.

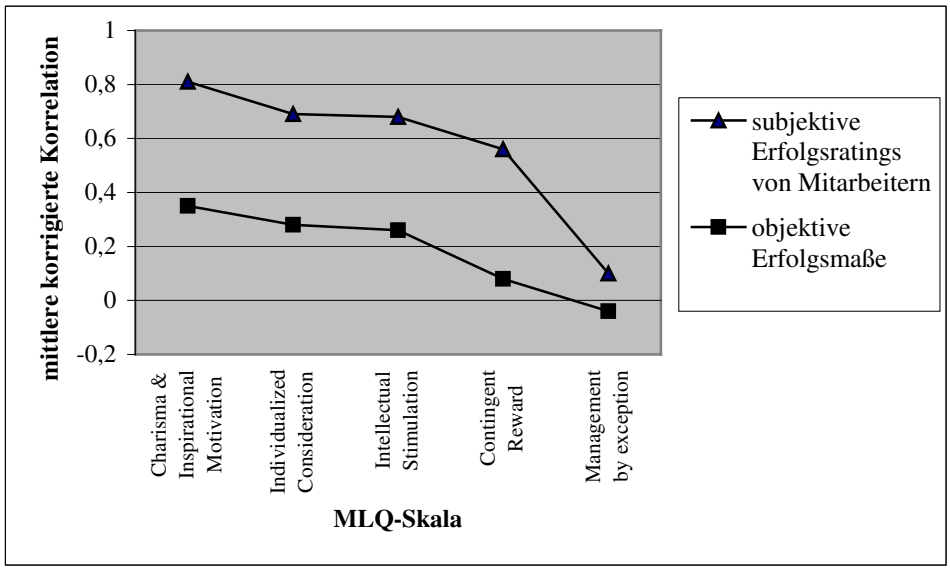


Abbildung 13. Korrelation der transaktionalen und transformationalen Führungsdimensionen mit subjektiven und objektiven Erfolgsmaßen (nach Lowe, Kroeck & Sivasubramaniam, 1996, S. 410)

Anmerkungen. Charisma und Inspirierende Motivierung werden genauso wie der aktive und passive Teil von „Management by exception“ in Anlehnung an die ersten Versionen des MLQ zu jeweils einem Faktor zusammengefasst. Laissez Faire wurde in einer Vielzahl an Studien, die der Berechnung zugrunde liegen, nicht erfasst und deshalb gänzlich ausgeschlossen.

Allerdings lassen sich für subjektive Mitarbeitereinschätzungen deutlich höhere Zusammenhänge mit den verschiedenen Führungsdimensionen messen als für objektive Erfolgsmaße. Zieht man beispielhaft den ersten Faktor – Charisma und Inspirierende Motivierung – heran, dann ergeben sich mittlere korrigierte Korrelationen von .35 mit objektiven beziehungsweise von .81 mit subjektiven Erfolgsmaßen. Vergleichbare Ergebnisse erbringt die Meta-Analyse von Dumdum et al. (2002), die den Erfolgszusammenhang mit allen MLQ-Dimensionen untersuchte. In der Meta-Analyse von Fuller et al. (1996) zeigt sich ein äquivalentes Bild: Die mittleren Korrelationen zwischen Charisma und subjektiven Erfolgsratings der Mitarbeiter liegen bei $r = .78$ (subjektiv bewertete Effektivität der Führung) beziehungsweise $r = .80$ (subjektive Zufriedenheit mit der Führungskraft). Objektiv erfasste Leistung (harte Kennwerte und Fremdratings der Erfolgsvariablen) korreliert mit Charisma zu $r = .34$. Der Effektivitätsunterschied zwischen subjektiven und objektiven Maßen kann zum einen darauf beruhen, dass der Zusammenhang mit subjektiven Maßen durch einen "Single Source Bias" künstlich erhöht (aber nicht vollständig erzeugt) wird (Avolio, Yammarino & Bass, 1991), da die subjektiven Erfolgsratings in aller Regel von den Personen stammen, die auch die Führung beurteilen. Zum anderen kann der Zusammenhang zu den harten Erfolgsfaktoren deshalb geringer ausfallen, weil diese von mehr Einflussgrößen, die nicht im Führungsverhalten selbst liegen, abhängen. So werden beispielsweise Verkaufs- oder Finanzzahlen immer auch zusätzlich durch verschiedene konjunkturelle und gesellschaftliche Variablen bedingt.

Weiter ist es in diesem Zusammenhang unerlässlich zu erwähnen, dass die meisten Einzelstudien, die in die Meta-Analyse eingingen, Führung und Führungserfolg zum gleichen Zeitpunkt erfragen. Erste Befunde zeigen jedoch, dass die transformationale Führung auch Erfolgsmaße vorhersagen kann, die einige Wochen respektive Monate nach der Führungsmessung erhoben werden (z.B. Bass et al., 2003; Howell & Avolio, 1993). Dies spricht für die erwartete Kausalrichtung.

Tabelle 5 gibt einen Überblick darüber, welche Erfolgsmaße im Einzelnen mit der transformationalen Führung einhergehen.

Tabelle 5. Überblick über die Effektivität transformationaler Führung (modifiziert nach Rathgeber & Jonas, 2003)

Effektivität	Überblick	Literatur (Auszug)
Settings, in denen die Effektivität transformationaler Führung nachgewiesen wurde	• Militär • Industrie • Politik & politische Verwaltung • Bank- & Dienstleistungsgewerbe • Sport • Bildungsbereich • Gesundheitswesen • verschiedene Randbereiche	• Bass, 1985, 1998 • Hater & Bass, 1988; Martin & Epitropaki, 2001; Waldman et al., 1987 • Pillai et al., 2003; • Geyer & Steyrer, 1998; Rai & Sinha, 2000 • Charbonneau et al., 2001 • Koh et al., 1995 • Bycio et al., 1995; Vandenberghe et al., 2002 • Trice & Beyer, 1986
Erfolgsmaße, auf die sich transformationale Führung auswirkt	• Extra Effort • wahrgenommene Effektivität • Zufriedenheit mit Führung • Arbeitszufriedenheit • Commitment • Leistung, z.B. Verkaufszahlen im Kreditgeschäft • Kreativität und Innovation • Qualität von Forschungsprojekten • Wohlbefinden, Burnout	• Bass, 1985 • Jung & Avolio, 2000; Pillai, Schriesheim & Williams, 1999 • Felfe & Goihl, 2002 • Bycio et al., 1995 • Geyer & Steyrer, 1998; Rai & Sinha, 2000 • Janssen, 2002 • Keller, 1992 • Martin & Epitropaki, 2001; Seltzer, Numerof & Bass, 1989
Betrachtungsebenen, auf denen Effektivität messbar ist	• Dyade • Informelle Kleingruppe • Arbeitsgruppe • Organisationsseinheiten • Gesamtorganisation • verschiedene Hierarchieebenen	• Howell & Frost, 1989 • Howell & Higgins, 1990 • Hater & Bass, 1988 • Howell & Avolio, 1993 • Roberts, 1985 • Lowe et al., 1996
Methoden, mit denen die Effektivität transformationaler Führung ermittelt wurde	• Fallstudien • Fragebogenuntersuchung • Beobachtungsstudien • Interviewstudien • Inhaltsanalysen von historischem Datenmaterial • Simulationsstudien im Labor • andere Laborstudien	• Roberts, 1985 • Hater & Bass, 1988; Rai & Sinha, 2000 • Trice & Beyer, 1986 • Bennis & Nanus, 1985; Howell & Higgins, 1990 • House et al., 1991 • Avolio et al., 1988 • Howell & Frost, 1989
Länder, in denen die Effektivität transformationaler Führung nachgewiesen wurde	• USA • Kanada • Deutschsprachiger Raum • Belgien • Niederlande • Israel • Indien • Russland • Japan, Taiwan, Neuseeland	• Bass, 1985 • Howell & Avolio, 1993 • Felfe & Goihl, 2002; Geyer & Steyrer, 1998 • Vandenberghe et al., 2002 • Koene et al., 1992 • Dvir et al., 2002 • Rai & Sinha, 2000 • Elenkov, 2002 • Übersicht: Bass, 1997

Vielfach belegt ist der Zusammenhang zu Extra Effort, wahrgenommener Effektivität und Zufriedenheit mit der Führung, da diese Kriterien Subskalen des MLQ darstellen (z.B. Avolio, 1999; Bass, 1985; Bass, 1998; Judge & Piccolo, 2004). Die Zufriedenheit mit der Führung scheint dabei etwas besser vorhersagbar zu sein als die wahrgenommene Effektivität (Meta-Analyse von Dumdum et al., 2002). Dies verwundert nicht, da transformationale Führungsverhaltensweisen hoch erwünscht sind. Der Zusammenhang zwischen transformationaler Führung und Arbeitszufriedenheit ist etwas geringer als der zwischen transformationaler Führung und Zufriedenheit mit der Führung (Meta-Analysen von Dumdum et al., 2002 bzw. von Judge & Piccolo, 2004; Einzelstudien im deutschen Sprachraum von Felfe & Goihl, 2002 bzw. von Tartler, Goihl et al., 2003). Auch dieser Unterschied ist erwartungsgemäß, da das Kriterium der generellen Arbeitszufriedenheit weniger nah am Prädiktor (transformationale Führung) liegt.

Weiter löst transformationale Führung bei den Mitarbeitern hohes affektives Commitment aus (Bycio et al., 1995; Rai & Sinha, 2000: Commitment insgesamt) sowie die Bereitschaft, für das gemeinsame Ziel persönliche Opfer zu bringen (Avolio, 1999). Der Einfluss auf das affektive Commitment scheint etwa doppelt so stark zu sein wie der von „Consideration“ (Meta-Analyse von Mathieu & Zajac, 1990).

Positive Wirkungen sind auch auf Leistungsparameter festzuhalten. Die Effektgrößen hierfür variieren zwischen d = .35 und d = .50 (Fiol, Harris & House, 1999). Sie steigen über d = .80 (z.B. Waldman, Bass & Yammarino, 1990), wenn man subjektive Erfolgsratings der Mitarbeiter heranzieht. Was Leistungsvariablen allgemein betrifft, so zeigen sowohl Feldstudien (Geyer & Steyrer, 1998; Hater & Bass, 1988; Rai & Sinha, 2000) als auch Laborstudien (Howell & Frost, 1989)[39], dass transformationale Führung Leistungsparametern (Verkaufszahlen etc.) steigert. Letztlich dürfen Laborstudien mit Studierenden-Stichproben aufgrund der geringen externen Validität aber nicht blindlings auf reale Führungssituationen generalisiert werden.[40]

Gezielt sollen zwei Variablen betrachtet werden, für die auch unerwartete Ergebnisse berichtet wurden. So findet man verschiedene Befunde dafür, dass transformationale Führung stärker mit Kreativität und Innovation zusammenhängt als transaktionale Führung (Janssen, 2002; Jung, 2001) oder dass die Qualität in denjenigen Forschungsprojekten

[39] In der Studie von Howell und Frost (1989) erwies sich transformationale Führung als effektiver als direktive oder mitarbeiterorientierte Führung. Zu kritisieren ist die Gegenüberstellung dieser Führungsstile, denn die Theorie transformationaler Führung legt nicht fest, wie direktiv bzw. mitarbeiterorientiert sie umgesetzt wird (vgl. Kapitel 2.5.3). Um die verschiedenen Führungsstile mit klarer Rollenbeschreibung glaubwürdig im Labor zu simulieren, wurden professionelle Schauspieler engagiert. Charisma wurde folgendermaßen operationalisiert: Artikulation eines hohen Ziels, Formulierung hoher Leistungserwartungen, Vermittlung von Zuversicht in die Fähigkeiten der Teilnehmer und Verständnis für die Bedürfnisse der Teilnehmer.

[40] Der Zusammenhang zwischen charismatischer Führung und Erfolgsmaßen ist im Labor so hoch wie im militärischen Kontext (Fuller et al., 1996).

durch transformationale Führung erhöht werden kann, in denen mehr kreative Leistung erfordert wird (Keller, 1992). Informelle Innovationsführer führen zudem selbst stärker transformational als weniger innovative Führungskräfte (Howell & Higgins, 1990). Dahingegen finden Sosik, Avolio und Kahai (1998) einen negativen Zusammenhang zwischen Intellektueller Stimulierung und Kreativität. Letztgenannte Autorengruppe bescheinigt der transaktionalen Führung unter bestimmten Bedingungen sogar mehr Auswirkungen auf Flowerlebnisse als der transformationalen Führung (Sosik, Kahai und Avolio, 1999).

Diese erklärungsbedürftigen Diskrepanzen könnte man auf die unterschiedlichen Operationalisierungen der abhängigen Variable zurückführen. Plausibel erscheint jedoch auch der Rückgriff auf andere unerwartete Befunde zur Intellektuellen Stimulierung. Danach scheint der transformationale Baustein, der im eigentlichen Sinne unmittelbar auf kreatives Denken abzielt, zwei Komponenten zu beinhalten. Die eine ist eng verbunden mit charismatischer Führung, die andere aber enthält einen Stressor, indem sie schwierige Denkprozesse fordert und neben der Ideengenerierung auch deren argumentative Untermauerung (Seltzer, Numerof & Bass, 1989). Stress wiederum steht in negativer Verbindung mit Innovationsverhalten (Harter, Schmidt & Keyes, 2003).[41]

Damit ist bereits eine weitere abhängige Variable mit inkonsistenter Ergebnislage angesprochen: das Wohlbefinden (für eine Zusammenfassung der Befunde: Rathgeber & Jonas, 2003). Seltzer et al. (1989) finden zwar erwartungsgemäß negative Zusammenhänge zwischen transformationaler Führung im Allgemeinen und Stress beziehungsweise Burnout. Jedoch steht Intellektuelle Stimulierung positiv mit beiden erwähnten Kriterien in Verbindung. Dies untermauert den stark fordernden, möglicherweise überfordernden Anteil der intellektuellen Stimulierung. Fordernde Führung wurde auch in anderem Kontext bereits als begünstigender Faktor für Burnout extrahiert (Seltzer & Numerof, 1988).[42] Transformationale Führung im Allgemeinen hängt jedoch positiv mit Gesundheit zusammen (Martin & Epitropaki, 2001).

Da unterschiedliche Studien auf unterschiedliche Untersuchungs*einheiten* zurückgreifen, kann der Haupteffekt von transformationaler Führung auf eine Vielzahl an Erfolgsva-

[41] Lowe et al. (1996; S. 415, 416) bezeichnen die Intellektuelle Stimulierung als „third child“ oder „critical leadership construct“ der transformationalen Führung, nachdem ihre Meta-Analyse unerwartet einen signifikant engeren Erfolgszusammenhang für öffentliche als für private Organisationen erbrachte. Als Erklärung dafür ziehen die Autoren die Vermutung heran, dass im Privatsektor intellektuell stimulierendes Verhalten notgedrungen weiter verbreitet ist und deshalb für Mitarbeiter weniger schnell salient wird als für Mitarbeiter einer öffentlichen Organisation, die in der Regel sehr bürokratisch und regelorientiert organisiert ist. Dies muss selbstverständlich nicht für jede einzelne Organisation so zutreffen. Was andere transformationale Skalen betrifft, findet sich in der Meta-Analyse lediglich für Charisma ein höherer Zusammenhang zu Mitarbeiter-Ratings in öffentlichen als privaten Institutionen. Dies ist deshalb weniger überraschend, weil die Charismaskala unter den transformationalen Bausteinen am ehesten einen Gesamteindruck der transformationalen Führung erfasst und dieser sich in insgesamt besseren subjektiven Wahrnehmungen widerspiegelt.

[42] Je weniger Initiating Structure (fordernde Führung) und je mehr Consideration (fördernde Führung) einer Führungskraft zugeschrieben wird, desto weniger Burnout berichten deren Mitarbeiter.

riablen und für eine Bandbreite an Analyseebenen als gesichert gelten (vgl. Tabelle 5). Erfolgreich betrachtet wurden Dyaden (Laborstudie von Howell & Frost, 1989), kleine informelle Gruppen (Howell & Higgins, 1990), formale Arbeitseinheiten (Hater & Bass, 1988), größere Organisationseinheiten von Großunternehmen (Howell & Avolio, 1993) oder auch die Gesamtleistung von Organisationen (Roberts, 1985; Trice & Beyer, 1986). Darüber hinaus werden die Ergebnisse durch eine Methodenvielfalt gestützt (vgl. Tabelle 5), denn unter den Untersuchungen finden sich Fallstudien (z.B. Roberts, 1985), Längsschnitt-Beobachtungsstudien (z.B. Trice & Beyer, 1986), Fragebogenstudien im Feld (z.B. Atwater et al., 1998; Hater & Bass, 1988; Rai & Sinha, 2000), Simulationsstudien im Labor (z.B. Avolio et al., 1988), andere Laborexperimente (z.B. Howell & Frost, 1989), Interviewstudien (z.B. Bennis & Nanus, 1985; Howell & Higgins, 1990) und Inhaltsanalysen von historischem Material (für charismatische Führung insgesamt: House et al., 1991).

Zusammenfassend lässt sich festhalten, dass insbesondere in den USA, wo die Mehrzahl der empirischen Ergebnisse gewonnen wurde, aufgrund der Stichproben-, Analyseebenen-, Methoden- sowie Kriteriumsheterogenität von einem weit verbreiteten Effekt transformationaler Führungsverhaltensweisen ausgegangen werden kann. Verschiedene (neuere) Studien mit nicht-amerikanischen Stichproben weisen aber darauf hin, dass für den Hauptbefund interkulturell eine konsistente Befunde erbracht werden konnten (vgl. Tabelle 5). Zudem gibt es Hinweise darauf, dass der korrelative Zusammenhang zwischen charismatischer Führung und Erfolgskriterien im Gegensatz zu Consideration und Initiating Structure nicht von den Kulturvariablen nach Hofstede moderiert wird (Koene et al., 1992).

Eine transformationale Führung verspricht also zur Erreichung der verschiedensten Erfolge auf personaler, Gruppen- und Organisationsebene hohen Nutzen. Die Effektivität von transformationaler Führung wird durch den Domino-Effekt noch verstärkt (Bass, Waldman, Avolio & Bebb, 1987). Dieser beinhaltet nämlich, dass Führungskräfte eher transformational führen, wenn sie selbst auch transformational geführt werden.

Nicht zu vergessen ist, dass die Wirkung der transformationalen Führung auf der der transaktionalen Führung aufbaut (vgl. Augmentationseffekt in Kapitel 2.5.3.1). Fragebogenuntersuchungen bestätigen regelmäßig die theoretische Annahme, dass die gleiche Führungskraft sowohl transaktionales als auch transformationales Führungsverhalten zeigen kann (Lowe et al., 1996). Erst in jüngster Zeit werden Prozessvariablen untersucht, die zwischen charismatischem Führungsverhalten und Outcome vermitteln.[43] Eine äußerst wichtige Rolle scheinen dabei sprachliche Konstruktionen sowie Kommunikationsstrategien zu spielen, die charismatische Führungskräfte einsetzen (Fiol et al., 1999; Kegan &

[43] Zur transformationalen Führung im Sinne von Bass wurde bisher keine derartige Studie durchgeführt.

Lahey, 2001). Da die Relevanz für die vorliegende Arbeit gering ist, wird auf diesen Forschungsbereich lediglich verwiesen.

2.5.4.2 Schattenseiten

„We cannot say, for example, that in seeking to harvest a charisma like Iacocca's we will not end up with a yield of destruction and violence like Jonesville[44]."
(Roberts & Bradley, 1988, S. 273)

Ein gemeinsamer Tenor der neuen Führungsära ist, dass die neue Führung prinzipiell gewinnbringend ist – für Führende und Geführte. Deshalb blieb die Kehrseite der Medaille lange Zeit unbeachtet, doch inzwischen sind die negativen Konsequenzen charismatischer/transformationaler Führung ausführlich beschrieben. Dabei geht es nicht nur um die bereits von Weber angerissenen Probleme, die sich für die Nachfolger von charismatischen Führungskräften ergeben, da sie in der Regel eine große, eben außeralltägliche Lücke zu füllen haben (Bryman, 1992). Bezogen auf die negativen Seiten charismatischer Führung selbst können in der Geschichte verschiedene negative „Lehrbuchbeispiele" studiert werden, beispielsweise Hitler.

Charismatische Führung kann hinsichtlich der zugrunde liegenden Machtmotivation (nach McClelland, 1985) in „zwei Gesichter" getrennt werden (Howell, 1988).[45] Demnach äußert sich ein personalisiertes Machtmotiv unter anderem in der Artikulation von Zielen, die den persönlichen Motiven des Führers entsprechen, während ein sozialisiertes Machtmotiv in der Artikulation von Zielen kulminiert, die den Geführtenbedürfnissen genügen. Im ersten Fall entsteht eine Abhängigkeit von der Führungsperson, Gehorsam oder Unterwerfung; nur im zweiten Fall sind die Konsequenzen mit der Konzeption der transformationalen Führung von Bass vereinbar: "Empowerment", autonomes "Followership", Unabhängigkeit von der Führungsperson. Hier zeigt sich ein Unterschied zwischen charismatischer und transformationaler Denkweise: Während auch nach der Linie von Bass der „pure charismatic" personalisiertes Charisma im Sinne von Howell (1988) aufweisen kann, ist die charismatisch-transformationale Führung im Sinne des sozialisierten Charisma zu betrachten (Avolio & Gibbons, 1988, S. 277). Bass (1998) differenziert sein ursprüngliches Konzept, indem er der echten transformationalen Führung („sozialisiert") eine pseudotransformationale („personalisiert") gegenüberstellt (Bass, 1998, S. 24f.).

---

[44] In Jonesville (Guyana) kamen 1978 bei einem Massensuizid in Guyana über 900 Menschen ums Leben, nachdem sie ihrem als charismatisch bezeichneten Führer Jones (The People's Church) Folge leisten wollten. Selbstverständlich ist auch hier die Wirkung charismatischer Führung nicht allein auf Persönlichkeitscharakteristika zurückzuführen, sondern ebenso auf situationale Bedingungen sowie einen Nachahmungseffekt (vgl. Cialdini, 1997).

[45] Dies geht Hand in Hand mit der Beschreibung von zwei Seiten von Charisma in einigen Ländern der GLOBE-Studie (z.B. Mexiko; Den Hartog et al., 1999).

Als typisches Beispielverhalten einer pseudotransformationalen Führungskraft nennt er in einem Interview Krokodilstränen im Falle von Entlassungen und gleichzeitige Erhöhung des eigenen Gehalts (Hooijberg & Choi, 2000). Beispiele für ethische und unethische (pseudotransformationale) Ausprägungen der einzelnen Führungsbausteine sind bei Bass und Steidlmeier (1999) beschrieben. Empirische Unterstützung gewinnen diese Diskussionspapiere zum Beispiel durch den Zusammenhang zwischen moralischem Denken und transformationaler Führung, die für transaktionale Führung ausbleiben (Turner, Barling, Epitropaki, Butcher & Milner, 2002).

Letztlich muss aber die alleinige Last von den Schultern der Führungskräfte genommen und auch auf Mitarbeiter verteilt werden. So können auch Mitarbeiter Handlungen unternehmen, die auf negative Konsequenzen für die Führungskraft oder das Team abzielen. Insbesondere, wenn sich die „dark side of leadership“ und die „dark side of followership“ summierten, könnten individuelle und organisationale Katastrophen entstehen (Clements & Washbush, 1999). Beide Seiten sind auch dann beteiligt, wenn charismatische Führung Gruppendenken im Sinne von Janis begünstigt (Conger, 1990).

Festzuhalten bleibt, dass die Ethikdiskussion auf modernen westlichen Moralvorstellungen basiert und diejenigen Führer als „pseudo“ eingestuft werden, die diesen Prinzipien nicht genügen (Bass & Steidlmeier, 1999). Bei Vorliegen einer echten transformationalen Führung ist der Sprung zum alten benediktischen Führungsmodell nicht weit und das Unternehmen kann fast als „Heiligtum“ betrachtet werden: „ein Ort der Heiterkeit, Inspiration und Liebe, ein Ort, an dem sich jeder persönlich entfalten kann, ein Ort, an dem unsere Seele angesprochen und beflügelt wird (Grün, 2003, S. 12).

2.5.5 Kontextfaktoren transformationaler Führung

> „Im Grunde kann also nur eine allgemeine Trivialität formuliert werden: Wenn es den Leuten schlecht geht, suchen sie einen Ausweg; alles, was Hilfe verspricht, ist willkommen – auch eine Führungsperson, die eine goldene Zukunft verheißt.“ (Neuberger, 2002, S. 193)

Sichtet man die Literatur zu transformationaler Führung, könnte man glauben, dass sie universell effektiv ist (z.B. Bass, 1997). In einem Interview (Hooijberg & Choi, 2000) äußert sich Bass eindeutig für eine universelle Perspektive zu Lasten einer kontingenztheoretischen Sichtweise. Zur Unterstützung zieht er Studien heran, die auf die Vererbbarkeit von Führungsqualifikationen hindeuten (Rose, 1995, zit. nach Hooijberg & Choi, 2000).[46]

[46] In dieser Zwillingsstudie wird beispielsweise auf 40% Vererbung geschlossen (Rose, 1995).

Die alte Idee vom „one best way“ drängt sich auf. Dieser Eindruck wird zudem von den Ergebnissen der GLOBE-Studie unterstützt, die implizieren, dass sich die Wirksamkeit charismatischer Führung universell bestätigt (House, Wright & Aditya, 1997). Verschiedene Autoren weisen aber explizit darauf hin, dass – wie jede andere Art und Weise der Führungsausübung – auch die transformationale Führung in organisationale Bedingungen eingebettet ist, die ihrerseits deren Auftretenswahrscheinlichkeit (beziehungsweise Attributionswahrscheinlichkeit) und Effektivität mitbedingen (Pawar & Eastman, 1997; Waldman, Ramírez, House & Puranam, 2001; Yukl & Howell, 1999).

Seit Weber wird immer wieder auf die Krise als Nährboden der charismatischen Führung verwiesen, weil Menschen in von Unsicherheit gekennzeichneten Situationen eine klare, sich glaubhaft verbessernde Richtung suchen. Obgleich es verschiedene Hinweise aus Labor- (Pillai, 1996) und Feldstudien (House, Spangler & Woycke, 1991; Roberts & Bradley, 1988) dafür gibt, dass eine Krise das Auftreten charismatischer Führung wahrscheinlicher macht, bestehen keine Belege dafür, dass sie als eine notwendige geschweige denn hinreichende Bedingung für die Entstehung charismatischer Führung fungiert (z.B. Conger, 1989; Yukl & Howell, 1999). Vereinzelt finden sich sogar negative Korrelationen zwischen Krise und transformationaler Führung (Pillai & Meindl, 1998).

Burns (1978) beschreibt in seinen politischen Fallstudien zwar durchgängig Umbruchsituationen, die entstanden, bevor eine charismatische Führungspersönlichkeit mächtig wurde. Das Besondere sei aber, dass die Entwicklungsphase von diesen Personen nicht passiv hingenommen, sondern aktiv beeinflusst wurde. Insofern kann eine Krise als förderlich gesehen werden, um sich „in Szene zu setzen“. Die Krisensituation wird dann also zugunsten der eigenen Macht am Leben gehalten. In diesem Sinne trägt ein aktives Krisenmanagement zur Charisma-Attribution bei.

Andere Autoren schlagen daher vor, anstelle (oder zusätzlich zu) der Krise auf andere Kontextfaktoren der Emergenz charismatischer Führung zu fokussieren, die teilweise auch unter einem sehr weit gefassten Krisenbegriff zu subsumieren sind (zum Beispiel bei Bass, 1985).[47] Beispiele sind: schnelle Umweltveränderungen durch Einführung neuer Strategien, Märkte, Produkte oder Technologien, frühe Stadien des Organisationszyklus, herausfordernde und komplexe Aufgaben für Einzelne und das Team, die Höhe der Hierarchieebene (Yukl & Howell, 1999), organische Organisationsstrukturen (Howell, 1997) oder die Fokussierung auf Effizienz anstatt Adaption, die Dominanz grenzübergreifender Einheiten über klar umgrenzte technische Zentraleinheiten, einfache und flexible Organisationsformen sowie Netzwerk- statt Bürokratiestruktur (Pawar & Eastman, 1997). Trotz unsicherer Umweltbedingungen ist eine funktionstüchtige, grundlegende (Infra-)Struktur günstig, um transformationale Ideen verbreiten zu können (Neuberger, 2002).

[47] Bei der Bearbeitung der Thematik fällt auf, dass in der entsprechenden Literatur Krise sehr unterschiedlich definiert wird.

House (1996) stellt neben die Erfahrung von Krise, Bedrohung oder Stress vier Bedingungen, unter denen neo-charismatische, wertebasierte Führung besser initiiert und schließlich erfolgreicher ausgeführt werden kann: erstens die Gelegenheit für die Führungskraft, eine ideologische Vision zu kommunizieren, zweitens die Gelegenheit für Führungskraft und Geführte, sich vollständig zu involvieren, drittens außergewöhnliche Anstrengung und Opfer von Führungskraft und Geführte und viertens die Kompatibilität der Vision mit den Werthaltungen der Beteiligten.[48] Letztgenannter Punkt führte dazu, dass in letzter Zeit auch Passungen zwischen Führungskraft und Geführten auf ihre Moderatorwirkung hin untersucht wurden (Ostroff, Atwater & Feinberg, 2004).

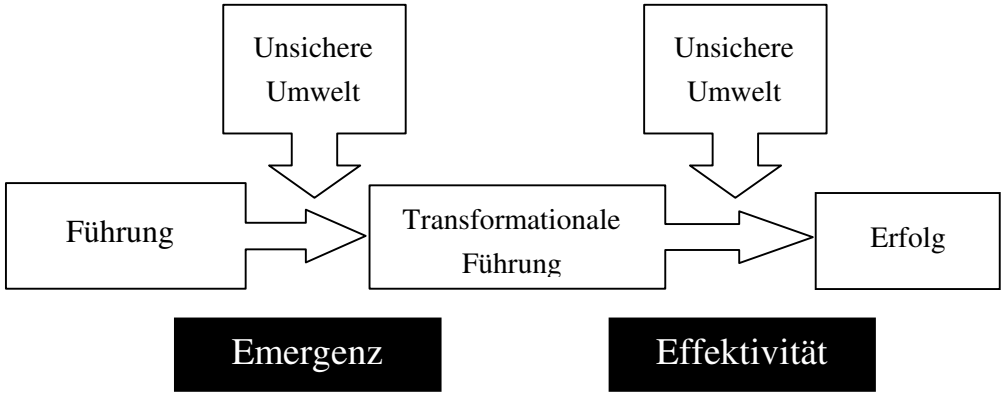


Abbildung 14. Skizzenhafte Darstellung: Die Situation moderiert das Auftreten und den Erfolg transformationaler Führung

Dass neben der Auftretenswahrscheinlichkeit auch die Effektivität transformationaler Führung durch die Situation beeinflusst wird (vgl. Abbildung 14), zeigen Waldman, Ramírez, House und Puranam (2001) in den größten amerikanischen Unternehmen. Die Ergebnisse bestätigen die Annahme, dass bei hoher Umweltunsicherheit der Unternehmensgewinn höher ausfällt, wenn die Führung als überdurchschnittlich charismatisch-visionär beschrieben wird. Im Unterschied dazu scheint Charisma/Vision in sehr sicheren Zeiten sogar einen gegenläufigen Effekt zu generieren und den ökonomischen Erfolg zu schmälern. Festzuhalten bleibt, dass der positive Zusammenhang zwischen charismatischer Führung und Unternehmenserfolg insbesondere gilt, wenn unsichere Zeiten herrschen. Auch unter solchen krisenhaften Kontextbedingungen konnte die Augmentationshypothese bestätigt werden. Neben dem Wirtschaftsbereich erwies sich ebenso im politischen Bereich der Zusammenhang zwischen transformationaler Führung und Erfolg in unsicheren Zeiten als besonders hoch (House, Spangler & Woycke, 1991).

[48] Insbesondere der dritte Punkt ist in einer solchen Konzeptualisierung zu kritisieren, da besondere Anstrengung als Ziel transformationaler Führung gilt.

Neben der Krise im weitesten Sinne sind auch die Geführten selbst als Rahmenbedingung für transformationale Führung nicht zu vernachlässigen (Conger, 1999).[49] Nach den Theorien von Bass (1985) und von Shamir et al. (1993) lassen sich insbesondere diejenigen Personen von einer transformationalen/charismatischen Führungskraft engagieren, die in der Führungsbeziehung die Gelegenheit sehen, ihre hohen Bedürfnisse zu befriedigen bzw. ihr Selbstkonzept stützen zu können.

Betrachtet man die Führungsnachfolge, dann gilt, dass neuen Führungskräften mit höherer Wahrscheinlichkeit charismatische Führung zugesprochen wird als etablierten Führungskräften und dies vor allem dann, wenn sie nicht-charismatischen Führungskräften folgen (Yukl & Howell, 1999).

Auch wenn eine Umbruchsituation die Emergenz von transformationaler Führung begünstigt, heißt das nicht, dass sie in stetigen Zeiten keine Chance hat. Für transaktionale Führung ist eine stabile Zeit geradezu ideal, denn sie bewegt sich innerhalb der bestehenden Strukturen und Organisationskultur. Da transformationale Führungskräfte genau diese aber hinterfragen und teilweise auch verändern, müssen sie in einer ruhigen Phase dafür härter arbeiten als in einer reformwilligen. Für die Emergenz von Individualisierter Berücksichtigung oder Intellektueller Stimulierung allein scheinen eher Unternehmensziele oder die Sozialisation und Entwicklung der Führungskräfte – und weniger eine Phase, die von Unsicherheit gekennzeichnet ist – vorteilhaft zu sein (Bass, 1985).

2.5.6 Erlernbarkeit transformationaler Führung

Für die Entwicklung transformationaler Führungsfertigkeiten sind zum Beispiel hohe moralische Standards der Eltern sowie positive Erfahrungen an der ersten Arbeitsstelle hilfreich. Da aber genetische Dispositionen und frühe Lernerfahrungen zu dem Zeitpunkt nicht mehr beeinflussbar sind, zu dem eine Person im Berufsleben ihre Führungsleistung verbessern will oder muss, werden im vorliegenden Kapitel lediglich solche Befunde skizziert, die die Trainierbarkeit transformationaler Führungsqualitäten betreffen. Begünstigende Erziehungs- und Sozialisationsfaktoren für transformationale Führung werden nicht weiter betrachtet (vgl. dazu Avolio & Gibbons, 1988; Zacharatos, Barling & Kelloway, 2000).

Es ist naheliegend, dass der Ruf nach Trainings für transformationale Führung lauter wird, seitdem von ihrer Effektivität auszugehen ist. Wichtig zu wissen ist, dass das Konzept von Anfang an verhaltensbezogen ausgelegt war und von Beginn an die Entwicklung

[49] Analog zu Reifegradmodell, vgl. Kapitel 2.4

des Führungsverhaltens als wichtiger Bestandteil verstanden wurde.[50] Das bedeutet jedoch nicht, dass ein „one best way“ der Führung anvisiert wird. Das Gegenteil ist der Fall: Wie im "Full Range of Leadership" Modell angelegt ist, bleibt jeder Führungskraft die Art und Weise, wie sie verschiedene Führungsmuster umsetzt, individuell überlassen. Doch kann man wirklich lernen, transformational zu führen? In einem Interview sagt Bass, dass zunächst transaktionales Führungsverhalten (insbesondere CR) gelernt werden muss (Hooijberg & Choi, 2000). Er schlägt zwei prinzipielle und kombinierbare Entwicklungsmaßnahmen vor: individuelle Beratungsgespräche und Workshops (Bass, 1990).

Dem entsprechend konnten auch in empirischen Trainingsstudien keine Unterschiede in der Wirksamkeit von Einzelberatung, Workshop oder einer Kombination aus beiden festgestellt werden (Kelloway, Barling & Helleur, 2000; allerdings bei sehr geringen Fallzahlen). Alle drei Trainingsvarianten zeigten deutliche Verbesserungen im Führungsverhalten im Vergleich zur Kontrollgruppe ohne Trainingsmaßnahme.

Verschiedene Faktoren der transformationalen Führung konnten gezielt durch Trainings verbessert werden. So verwenden Führungskräfte (Industrie-Stichprobe) nach einem Training zur Inspirierenden Motivation mehr entsprechende kommunikative Elemente als vor dem Training (Frese, Beimel & Schoenborn, 2003). Durch Gruppentraining mit besonderer Betonung der Dimension der Intellektuellen Stimulierung und anschließendem Einzelcoaching (vier Sitzungen) konnte das Führungsverhalten von Bankmanagern so beeinflusst werden, dass deren Mitarbeiter 5 Monate nach Trainingsende ein deutlich stärker transformationales Führungsverhalten (insbesondere Intellektuelle Stimulierung) wahrnahmen und deutlich höheres Commitment zeigten als die Mitarbeiter von nicht trainierten Führungskräften. Die Finanzkennwerte der Abteilungen veränderten sich in die erwartete Richtung, wenn auch nur mit geringen Effekten (Barling, Weber & Kelloway, 1996). Auch im Militär mündet ein Training transformationaler Führung 6 Monate nach Trainingsende in bedeutenden, aber geringen Verbesserungen (Wissen und Einsatz) auf Seiten der Geführten (Dvir et al., 2002). In einer Laborstudie konnte ein stärkerer Effekt auf die Leistung (Qualität eines Briefs) derjenigen Studierenden festgestellt werden, die das Video eines vorher hinsichtlich charismatischen Führungsverhaltens trainierten Studierenden gesehen haben (Towler, 2003). Die ersten Trainingsstudien zeigten also bereits ein verhältnismäßig breites Wirkungsspektrum der Führungstrainings. Im Übrigen ist als vorläufiges Fazit nicht nur zu ziehen, dass transformationale Führung prinzipiell trainierbar ist, sondern auch, dass die interpretierte Kausalrichtung der in Kapitel 2.5.4.1 berichteten Zusammenhänge weiter unterstützt wird.

[50] Es sind inzwischen sogar aus der Forschungsgruppe von Bass Bücher publiziert, die explizit zum Selbst-Lernen anregen sollen (Avolio, 1999).

Betrachtet man die Trainingsinhalte genauer, fallen die individuumszentrierte Vorgehensweise und die Einbettung in den alltäglichen Führungskontext auf, was auch bei workshopbasiertem Training nicht vernachlässigt wird (z.B. Bass & Avolio, 1990):

- Information über transformationale Führung, den "Full Range of Leadership" und Zielsetzung
- Reflexion des eigenen Führungsverhaltens und -anspruchs
- Feedback über eigenes Führungsverhalten (nur bei Einzelberatung)
- Erarbeiten individueller, herausfordernder Ziele bezüglich des eigenen Führungsverhaltens (eventuell videogestützt)

So verfolgen die zur transformationalen Führung durchgeführten Trainings zwei klassische Funktionen (vgl. Conger & Kanungo, 1988). Zum einen informieren sie über und sensibilisieren für Führungsprozesse, so dass ein Rahmen für eine diesbezügliche Selbstbewertung gegeben wird. Zum anderen werden entsprechende Fertigkeiten anschließend trainiert. Solche Trainingsgrundlagen sprechen insofern gegen ein starres Trainingsprogramm, als sie nicht den Kreuzzug für einen bestimmten Führungsstil antreten. Außerdem ist es von bedeutendem Vorteil, dass sich Führungskräfte ihre individuellen Ziele stecken, denn lediglich diese fokussierten Führungsdimensionen konnten beispielsweise in der Studie von Avolio und Bass (1998) erfolgreich verbessert werden (vgl. Zielsetzungstheorie von Locke und Latham, 1990). Genauso wie die Individualität der Führungskraft beim Training ist jedoch auch die Individualität der Mitarbeiter in der Führungsbeziehung nicht zu vernachlässigen. Illustriert wird dies durch einen Titel von Avolio und Bass (1998): „You can drag a horse to water but you can’t make it drink unless it is thirsty“.

2.6 Zusammenfassung und Integration der Führungstheorien und -aufgaben

Es wurde eine Bandbreite an Führungsaufgaben und -theorien skizziert, die einen groben Abriss der psychologischen Führungsforschung gaben. Diese sollen im vorliegenden Kapitel zusammenfassend betrachtet werden, um übergeordnete Elemente herausarbeiten zu können. Grundlegend gilt für alle Theorien, dass sie die Kriterien der Definition von Weibler (S. 16) erfüllen. So soll jeweils die *Beeinflussung* von Mitarbeitern hin zu einem *intendierten* Verhalten/Ergebnis erklärt werden. Ob es um *mittelbare oder unmittelbare* Veränderung geht, wird in der Regel nicht festgelegt. Insgeheim sind durch unterschiedliche Führungsstile aber unterschiedlich vermittelte Effekte zu erwarten. Keine der Theorien geht explizit davon aus, dass das jeweils anvisierte Verhalten der Führungskraft erst dann als Führung bezeichnet werden sollte, wenn es von den Mitarbeitern *akzeptiert* wird. Die geführtenzentrierten Ansätze (z.B. Weg-Ziel-Theorien, Attributionstheorien) und die transformationale Führung kommen dem allerdings sehr nahe. Tabelle 6 gibt zunächst einen Überblick über die Ähnlichkeit der vorgestellten Führungstheorien.

Es fällt auf, dass Mitarbeiter- und Aufgabenorientierung in viele Führungsmodelle integriert sind. Wo nicht auf diese beiden Führungsstile zurückgegriffen wird, geht es verstärkt um den Grad der Partizipation. So berücksichtigt die Weg-Ziel-Theorie, vor allem aber das normative Entscheidungsmodell solche Führungsstile, die als Erweiterung der klassischen Arbeiten von Lewin et al. (1939) zu betrachten sind (vgl. auch Abbildung 6, S. 38). Erst in den neo-charismatischen Führungstheorien wurde die Bandbreite des Führungsverhaltens stark über diese beiden klassischen Führungsstil-Konzeptionen hinaus erweitert, so dass die bis dahin entwickelten Theorien als „traditionelle Führungstheorien“ bezeichnet werden können. Zur traditionellen Führung sind auch die nicht-transformationalen Führungsbausteine aus dem Modell von Bass (1985) zu zählen. Weniger explizit wurde in den meisten neo-charismatischen Konzepten hingegen der Kontingenzgedanke fortgesetzt.

Tabelle 6. Ähnlichkeit zwischen den verschiedenen, in dieser Arbeit (Kapitel 2.4 und 2.5) vorgestellten Führungstheorien (modifiziert nach Tosi et al., 2000, S. 472)

Führungsmodell	Bausteine des Modells		
Führungsstile von Lewin et al.[1]	Laissez Faire	demokratisch	autoritär
Ohio State-Studien (Fleishman; Blake & Mouton)	Consideration		Initiating Structure
Michigan-Studien (Likert)	employee centered		production centered
		Situative Faktoren	
Kontingenztheorie (Fiedler)	mitarbeiterorientiert (hoher LPC)[2]	▪ Positionsmacht ▪ Aufgabenstruktur ▪ Führer-Mitarbeiter- ▪ Beziehung	aufgabenorientiert (niedriger LPC)
Reifegradmodell (Hersey & Blanchard)	mitarbeiterorientiert	Reifegrad (Fähigkeit und Motivation)	aufgabenorientiert
3-D-Modell (Reddin)	mitarbeiterorientiert	▪ Arbeitsanforderungen ▪ Führungsstil des Vg[3] ▪ Kollegen ▪ Unterstellte ▪ Organisation	aufgabenorientiert
Weg-Ziel-Theorie (House)	Unterstützende und partizipative Führung	Klarheit der Weg-Ziel-Zusammenhänge	Direktive und leistungsorientierte Führung
Normatives Entscheidungsmodell (Vroom & Yetton)	beratend (CI, CII) Gruppe (GII)[4]	▪ Akzeptanz ▪ Qualität ▪ Information ▪ Konflikt ▪ Zielkongruenz	Autokratisch (AI, AII)[4]
Führungsmodell von Bass	Transformationale Führung	▪ Krise i.w.S. ▪ Andere moderierende Situationsvariablen untersucht (Kap. 2.5)	Transaktionale Führung

Anmerkungen. [1]von Tannenbaum und Schmidt (1958) zu einem Kontinuum erweitert, [2] LPC: „Least Preferred Coworker“ als umstrittenes Maß für die Mitarbeiter-/Aufgabenorientierung einer Führungskraft, [3]Vg: direkter Vorgesetzter. [4]CI, CII: konsultative Führung/Entscheidungsfindung, GII: Gruppenentscheidung, AI, AII: autoritäre Führung/Entscheidungsfindung.
Nicht in die Übersicht aufgenommen sind die neocharismatischen Führungstheorien mit Ausnahme von Bass (1985) sowie reine Attributions- und Substitutionsansätze.

Vielleicht ist das größte Verdienst der Konzeptentwicklung zur transformationalen Führung gar nicht das Konzept selbst, sondern die Tatsache, dass dadurch das Feld der Führungsforschung transformiert wurde, das bis dahin eher fragmentarisch, statisch und irrelevant für den angewandten Führungsprozess war (Hunt, 1999; Lombardo & McCall, 1978).[51] In dem Zusammenhang sind nämlich viele neue Konzepte entstanden, die später als „new leadership" betitelt wurden (Bryman, 1992). So wurden Führungskräfte zu „managers of meaning" (Bryman, 1996), indem sie Visionen vermitteln, symbolische Aspekte der Organisation und die Organisationskultur als Ganze stärker betrachten sollten. Als Antwort auf diese Führung wurden auf Geführtenseite zunehmend emotionale Gesichtspunkte betont (Chemers, 1997; Yukl, 1999b) und kollektive Prozesse betrachtet (Shamir, 1995). In Untersuchungsdesigns fällt auf, dass die neo-charismatischen Führungsmodelle im Unterschied zu den traditionellen Modellen auch durch qualitative Forschung und an hochrangigen Führungspersonen untersucht werden (statt quantitativ und ausschließlich an face-to-face-Führungsdyaden unterer Hierarchieebenen) (Bryman, 1996). Aus dieser Perspektive wird die Popularität der transformationalen Führung sogar fast unabhängig von der empirischen Evidenz, die inzwischen ja zusätzlich erbracht ist, nachvollziehbar. Wie ist die transformationale Führung aber in das Führungsspektrum traditioneller Führungstheorien einzuordnen? Im Folgenden wird dies unter Rückgriff auf Abbildung 3 (S. 34) diskutiert.

Für die *personalen Theorien* sprechen diejenigen Überlegungen, die insbesondere den charismatischen Anteil der transformationalen Führung auch im Sinne von Personenvariablen konzipieren (führerzentrierter Ansatz). Dass aber all denen zu widersprechen ist, die mit der transformationalen Führung die Wiedergeburt der Trait-Theorien herauf beschworen haben, zeigt sich auch in den unbedeutenden Korrelationen zu verschiedenen Persönlichkeitsvariablen. So korrelieren beispielsweise die Big Five mit der transformationalen Führung deutlich geringer als mit anderen Führungsverhaltensweisen. Am stärksten, dennoch nur sehr gering, fällt die Korrelation zum charismatischen Faktor aus (Bono & Judge, 2004). Dazu passt, dass manche *charismatischen* Führungstheorien (z.B. House, 1977) stärker auf Persönlichkeitsmerkmale fokussieren als andere new-leadership-Theorien (vgl. S. 56). Bei charismatischer Führung steht auch eher die Abhängigkeit von der Führungsperson im Mittelpunkt, während bei der transformationalen Führung die verfolgte Mission betont wird und kein einseitiger, sondern ein gegenseitiger Beeinflussungsprozess angestrebt wird (Barbuto, 1997). Möglicherweise betreiben charismatische Führungskräfte auch besonders geschicktes Impression Management. Dass dies wichtig ist, wurde mehrfach hervorgehoben (Bass, 1985; Conger & Kanungo, 1987; Gardner & Avolio, 1998; House, 1977).

[51] Lombardo und McCall (1978, p. 3) charakterisieren die Führungsforschung entsprechend: "If leadership is bright orange, leadership research is slate gray."

Dass die Existenz und Wirkung transformationaler Führung auch davon abhängt, ob die Geführten ihrer Führungskraft Charisma zusprechen, wird sogar im MLQ verdeutlicht, indem der erste Faktor als „Idealized Influence – *attributed*“ bezeichnet wird (geführtenzentrierter Ansatz).[52] Da transformationale Führung generell darauf abzielt, dass die Mitarbeiter ihre Werte und verschiedene Annahmen verändern, ist zweifellos deren Wahrnehmung von Führung wichtiger als die tatsächliche Ausübung von Führung. Mit der Weg-Ziel-Theorie verbindet die transformationale Führung vor allem, dass beide auf Erwartungs*Wert-Theorien rekurrieren und deshalb auch als Prozesstheorien bezeichnet werden (z.B. Tosi et al., 2000).

Als *situativer Ansatz* kann die transformationale Führung insofern betrachtet werden, als verschiedene moderierende Kontextvariablen bestehen – für das Auftreten oder die Attribution sowie für die Wirksamkeit von transformationaler Führung. Denkt man an die Krise, dann ist eher von einem Makroansatz auszugehen; denkt man an die ausreichende Emergenz transaktionaler Führung, dann wäre das als Mikroansatz zu verstehen. Hersey et al. (2001) entwickelten analog zum ursprünglichen Reifegradmodell ein situatives Modell zur Transformation in Organisationen. Die Achsen, die darin die Führungsquadranten bestimmen, sind inspirierendes und strukturierendes Führungsverhalten. Generell geht es dabei eher um vorbereitende Prozesse für eine Transformation, die zu durchlaufen sind, bis im „vierten Quadranten“ dann der Status Quo zerstört und die neue, transformierte Struktur implementiert werden können. Auch wenn das Abstraktheitsniveau im Vergleich zu Bass deutlich höher ist, kann das Fazit gezogen, werden, dass nicht in jeder Organisationsphase (situative Bedingung) eine gleich starke, transformationale Führung angebracht und zielführend ist. Nach dem Modell muss sie an die „organizational readiness“ angepasst werden, worunter unterschiedlich flexible Ausprägungen der Organisationskultur zu verstehen sind.

Dass die transformationale Führung in unsicheren Zeiten mit erhöhter Wahrscheinlichkeit auftritt, steht im Einklang mit der Substitutionstheorie für Führung von Kerr und Jermier (1978), da in stabilen Zeiten klare Strukturen und bekannte, messbare Ziele das Schiff auf Kurs halten und Sicherheit geben können. Diese Ziele sind durch transaktionales Führen erreichbar (House, 1996; Yukl & Howell, 1999), während in unsicheren Zeiten klare Strukturen ins Wanken geraten und durch transformationale Führung und erstrebenswerte Visionen ersetzt werden müssen.

Was die *Positionstheorien* betrifft, so ist davon auszugehen, dass sehr innovative Transformationsideen eher von Personen mit starker Machtbasis ansteckend wirken (Machttheorien). Beruht die Macht der Führungskraft auf Identifikation („referent power“ nach French & Raven, 1959; Raven, 1992) – einem wichtigen Ziel transformationaler Führung –, dann ist eher mit dem erwünschten hochengagierten und selbstbestimmten Verhalten der

[52] Zur genaueren Darstellung des MLQ vgl. Kapitel 2.7.4 und Kapitel 5.2.2.

Mitarbeiter zu rechnen als bei anderen Machtbasen (vgl. auch Conger & Kanungo, 1987). Betrachtet man transformationales Führungsverhalten aus der Perspektive der Rollentheorien, dann ist es als Extrarollenverhalten zu charakterisieren. Ganz explizit ist dies im Idiosynkrasie-Kredit-Modell von Hollander (1995) so konzipiert, wonach gesammelte Kredit-Punkte gegen rollenüberschreitendes, transformationales Verhalten getauscht werden können.

Zum Prinzip der *Interaktionstheorien* bestehen viele Gemeinsamkeiten. So ist transformationale Führung sogar darüber definiert, dass sich Führungskraft und Geführte in einer Art Pflichtgemeinschaft gegenseitig weiterentwickeln. Wechselseitige und dynamische Einflussprozesse zwischen Führungskraft und Mitarbeiter sind also wie auch im Idiosynkrasie-Kredit-Modell, zu dem viele Gemeinsamkeiten bestehen, immanent wichtig. Generell lässt die transformationale Führung extrem kooperatives Führungsverhalten zu. Die Individuelle Berücksichtigung, die ja immer auch eine Art Abstimmung zwischen Führungskraft und Mitarbeiter bedeutet, ist sogar ein explizit benannter Baustein transformationaler Führung (vgl. Abbildung 15). Weiter ist eher als interaktionistisch denn als personal einzuordnen, dass Personen, die sich selbst ähnlich extravertiert wie die Führungskraft wahrnehmen, dieser eher transformationale Führung zuschreiben (Felfe & Schyns, 2004).

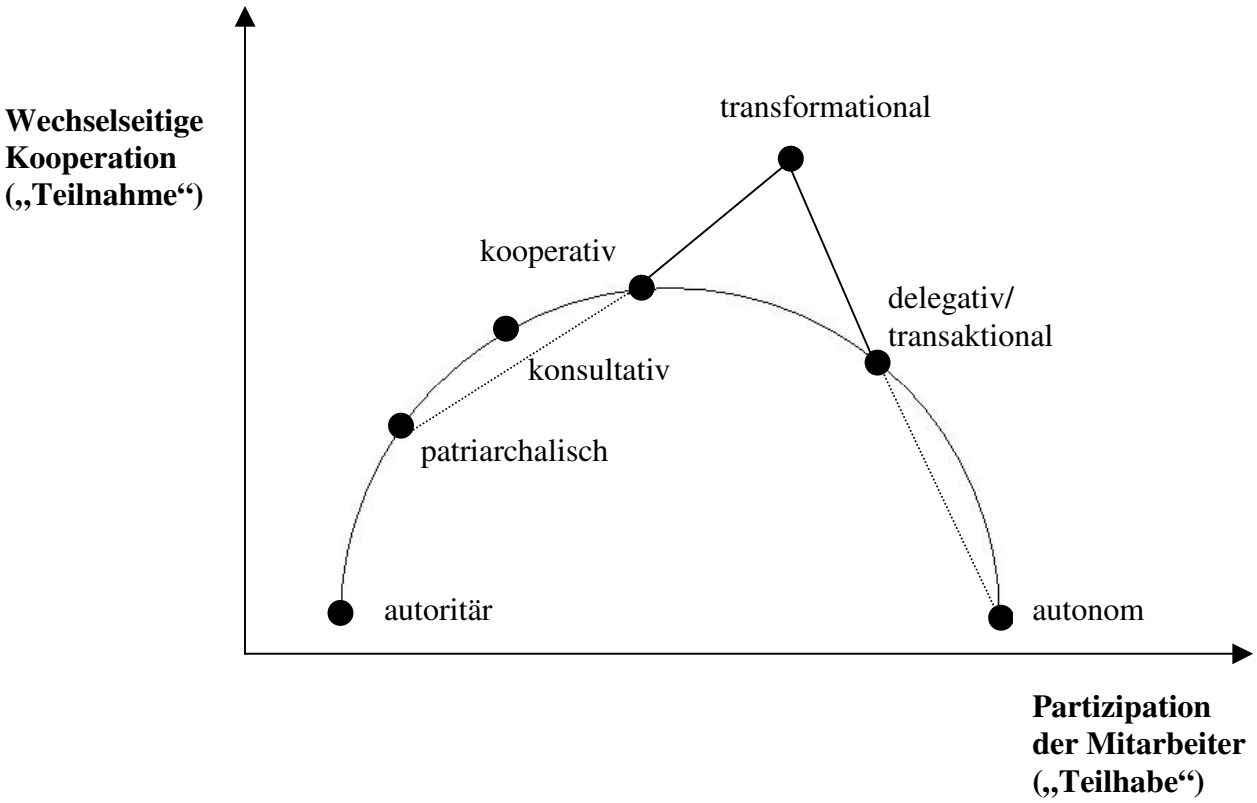


Abbildung 15. Transaktionale und transformationale Führung im Führungsstilkontinuum (nach Wunderer, 2001, S. 248)

Schließlich kann festgehalten werden, dass die transformationale Führung zu allen vier großen Theoriegruppen (personale, situative, Positions- und Interaktionstheorien) Affinität aufweist. Deshalb wäre sie innerhalb der herangezogenen Systematik von Abbildung 3 genau an deren Schnittstelle anzusiedeln.

Andere neo-charismatische Theorien (neben der transformationalen Führung nach Bass) sind deshalb eindeutiger einer der vier Theoriengruppen zuzuordnen, weil sie weniger vielfältige Führungsverhaltensweisen fordern und deshalb in der Regel als Ausschnitte der transformationalen Führungstheorie konzipiert werden können. In den entsprechenden Abschnitten in Kapitel 2.5.2 wurde darauf bereits verwiesen. So ist die charismatische Führung nach House (1977) sowie nach Conger und Kanungo (1987, 1998) stärker auf die Person der Führungskraft fokussiert, die charismatische Theorie von Shamir et al. (1993) auf die Person der Geführten und die Idiosynkrasie-Kredit-Theorie auf die Interaktion. Dies verdeutlicht weiter, dass die transformationale Führung nach Bass mehr umfasst als charismatische oder visionäre Führung alleine. Während Vertrauensbildung und visionäre Führung in vielen neo-charismatischen Führungstheorien enthalten sind, gehen individuell berücksichtigende und intellektuell stimulierende Verhaltensweisen über andere Theorien hinaus (vgl. House & Shamir, 1993).

Transformationale Führung ist also nicht mit charismatischer Führung gleichzusetzen (Barbuto, 1997; Bass, 1997), auch wenn sie den stärksten Faktor der transformationalen Führung darstellt und mit anderen transformationalen Faktoren korreliert. Dass es sich um trennbare Konstrukte handelt, wird in der täglichen Praxis deutlich. So kann eine Führungskraft ihre Mitarbeiter beispielsweise individuell berücksichtigen, ohne charismatisch oder transformational auf den anderen Dimensionen zu sein (Bass & Avolio, 1993).

Im Folgenden sollen drei Punkte dargestellt werden, in denen sich die transformationale Führung nach Bass (1985) konzeptionell von anderen Führungstheorien unterscheidet (zu inhaltlichen Unterschieden vgl. Tabelle 3).

Erstens kann festgehalten werden, dass im Laufe der Zeit mehrdimensionale Modelle die Oberhand gewonnen haben, um der Komplexität von Personalführung gerecht zu werden. Auch einzelne Führungsaufgaben bleiben nicht isoliert stehen, sondern werden zu Katalogen an Führungsaufgaben zusammengefasst. Indem das komplette Führungsspektrum zur Verfügung gestellt wird, kann das Führungsmodell von Bass (1985) als Gipfel dieser Strömung bezeichnet werden.

Unstreitbar ist inzwischen auch, dass der Führungs-Erfolgs-Zusammenhang von situativen Merkmalen moderiert wird. Die Logik der damit angesprochenen Kontingenzforschung besteht darin, die Erfolgsprognose eines spezifischen Führungsverhaltens dadurch zu erhöhen, dass *immer mehr* situative Randbedingungen berücksichtigt werden. Auf der anderen Seite ist die Ausdifferenzierung, wie sie im Entscheidungsbaum von Vroom

und Yetton (1973) vorgenommen wird, bereits zu komplex, um in der Praxis unmittelbar umgesetzt werden zu können. Von einer pragmatischen Perspektive aus betrachtet ist die Kontingenzforschung also bereits „zu weit“ fortgeschritten, während sie aus der Forschungsperspektive noch lange keine zufriedenstellenden Ergebnisse erbringt.

Bass (1985) unternahm mit dem Konzept der transformationalen Führung einen Versuch, diese enge Grundstruktur der situativen Theorien, die einerseits nicht weit genug gehen können und andererseits immer zu weit gehen, aufzubrechen. Diese Erweiterung besteht darin, dass Führungskräfte aus einem Pool an Führungsverhalten schöpfen können. Dabei scheint die transformationale Führung fast universell wirksam zu sein, wenngleich auch bereits situative Moderatorbedingungen untersucht und extrahiert werden. Eine zusätzliche Erweiterung im Modell von Bass regt dazu an, über Basisvariablen von Führung nachzudenken.[53] So besagt die Augmentationshypothese, dass transaktionale Führungsverhaltensweisen die Basis dafür darstellen, dass transformationale Führung ihre Wirkungsweise voll entfalten kann. Die hierarchische Strukturierung von Führungsverhaltensweisen geht auch über die traditionellen Führungsmodelle hinaus. Eine ähnliche Annahme findet sich lediglich im Idiosynkrasie-Kredit-Modell von Hollander (1995), denn auch dort kann transformationale Führung nur auf der Basis von erfolgreicher transaktionaler Führung entstehen. Ein großer Unterschied besteht jedoch darin, dass transformationale Führung bei Hollander (1995) einen Spezialfall der Transaktionen darstellt. Erst wenn die Führungskraft nämlich genug Kredits gesammelt hat, kann sie – quasi im Austausch dafür – nicht normgemäßes Verhalten zeigen und dadurch Transformationsprozesse einleiten. Bei Bass (1985) hingegen ist transformationale Führung von Austauschgedanken losgelöst.

Wenn klassische Führungsaufgaben aber eher der transaktionalen Führung zuzusprechen sind (siehe unten), dann gehören sie möglicherweise mehrheitlich zum Basisverhalten einer Führungskraft. Zu untersuchen wäre auch, ob aufgaben- und mitarbeiterorientierte Führung solche Basisvariablen guter Führung darstellen. Dafür sprechen erste Ergebnisse, die für eine darüber hinausgehende Wirkung transformationaler Führung sprechen (Seltzer & Bass, 1990).

Unter diesem Blickwinkel stellt sich aber die Frage, ob der "Full Range of Leadership" Führung annähernd vollständig aufgreift. Dabei geht es Bass (1985) nicht um eine vollständige Aufzählung aller effektiven Führungsverhaltensweisen, sondern darum, ein potentiell wirksames Führungsspektrum aufzuspannen. Dass alle „großen Perspektiven“ (Person, Situation, Interaktion, Position) eingenommen werden können, wurde oben bereits diskutiert. Aber auch im Vergleich zum Führungsverhalten, das andere Führungstheorien beinhalten, zeigt sich, dass keine wesentlichen Führungsverhaltensweisen fehlen – vorausgesetzt man beschränkt sich auf die dyadische Führung und vernachlässigt organisa-

[53] Auch Basisvariablen können im Sinne der Kontingenzmodelle als situative Faktoren betrachtet werden.

tionale Ebenen, auf denen beispielsweise zusätzlich strategische Führungselemente bedeutsam wären. So findet sich das Ausmaß an Partizipation und Aktivität, wie es die klassischen Führungsstile von Lewin et al. (1939) variieren, im „Full Range of Leadership“ genauso wieder[54] wie die Aufgaben- und Mitarbeiterorientierung. Die Aufgabenorientierung ist in Form der transaktionalen Verhaltensweisen integriert (Klarheit, Rollenstrukturierung). Consideration geht hingegen in eine transformationale Führungsdimension, die Individualisierte Berücksichtigung (Individualized *Consideration*) ein.[55] Diese umfasst allerdings mehr als das, was das traditionelle Konzept von Consideration ausmacht (Bass & Avolio, 1993). Während unter Consideration insbesondere beziehungsorientierte Führungsverhaltensweisen zu verstehen sind, die im Sinne eines Tauschgeschäfts bei den Mitarbeitern Akzeptanz der Führungskraft und Zufriedenheit mit der Führungskraft hervorrufen, ist Individualisierte Berücksichtigung funktional für eine Transformation konzipiert: Die individuellen Ziele und Potentiale der Mitarbeiter werden so unterstützt, dass diese sich fortentwickeln können.

In einem letzten Schritt sollen nun die in Kapitel 2.3 angerissenen Führungsaufgaben dahingehend überprüft werden, ob sie im "Full Range of Leadership" Modell einzuordnen oder einzelnen Führungstheorien zuzuordnen sind. So kann die Leistungsbeurteilung als erstgenannte Führungsaufgabe im transaktionalen Führungsspektrum verankert werden, da jede kontingente Belohnung beziehungsweise jede Führungsintervention im Ausnahmefall („Management by exception“) als Beurteilung betrachtet werden kann.

Zielsetzung ist in Abhängigkeit von den Qualitäten des Ziels entweder im Sinne von klarer Rollenverteilung innerhalb einer Austauschbeziehung (transaktional) oder im Sinne von Visionen (Inspirierende Motivierung) zu sehen.

Die Personalentwicklung kann durch viele verschiedene Führungsverhaltensweisen unterstützt werden. So kommt es durch Transformation genauso zu Weiterentwicklungen wie durch operante Verstärkung (transaktionale Führung). Auf einer konkreteren Ebene soll vor allem die Individualisierte Berücksichtigung die Einwicklung einzelner Mitarbeiter fördern. Zu diesem Zwecke können nicht nur Aufgaben und Verantwortungsbereiche delegiert werden, sondern es kann generell partizipativ geführt werden. Insofern berühren auch alle Modelle, die an der Partizipationsdimension ausgerichtet sind, die Aufgabe der Personalentwicklung. Auch im Reifegradmodell endet der anvisierte Entwicklungszyklus damit, dass hochreife Mitarbeiter delegativ geführt werden können. Delegieren kann somit im Sinne von individuellem Berücksichtigen an der Grenze zur transformationalen Führung gesehen werden. Auch die intellektuelle Stimulierung wird stärker initiiert als durch Nicht-

[54] Zur Erinnerung: Der „Full Range of Leadership“ wird sogar an der Dimension Aktivität ausgerichtet (vgl. Abbildung ***9***).

[55] Die Korrelation zwischen Initiating Structure und Consideration ist entsprechend dieser Zuordnung auch geringer als die zwischen Consideration und Charisma als wichtigstem transformationalem Faktor (r = .60 vs. r = .69; Seltzer & Bass, 1990).

oder falsche Delegation. Delegieren kann aber auch im Sinne des Austauschgedankens, also als transaktionale Führung formuliert werden: Die Führungskraft vertraut ihrem Mitarbeiter und geht davon aus, dass dieser die zu delegierende Aufgabe erfüllt. Der Mitarbeiter vertraut seiner Führungskraft insofern, dass sie Rückendeckung für den Fall von Fehlschlägen erwartet und nimmt die Delegation an. Er bringt also die Leistung, die von ihm erwartet wird – beispielsweise im Austausch für das entgegen gebrachte Vertrauen (vgl. Bauer & Green, 1996).

Wie stark organisierend Führungskräfte intervenieren müssen, hängt von situativen Faktoren ab, zum Beispiel der Aufgabenstruktur. Prinzipiell können viele Führungsverhaltensweisen in dem Sinne organisierend wirken, dass die Erwartungs- und Wert-Komponenten der gesteckten Ziele gestärkt werden. Insbesondere die Wert-Komponente ist durch transformationale Führung positiv zu beeinflussen. Im Speziellen zielt vor allem die Weg-Ziel-Theorie auf die Erwartungs- und Wert-Komponenten ab.

Letztlich spiegelt sich die Aufgabe des Entscheidens und Entscheidenlassens im Aktivitäts- beziehungsweise Partizipationsgrad wieder. Sie ist daher mit vielen Theorien vereinbar (z.B. Modell von Bass oder von Vroom & Yetton). Keine der zentralen Führungsaufgaben liegt also außerhalb des "Full Range of Leadership" Modells. Eine weitere Übersetzung von Führungstheorien in Führungsaufgaben könnte jedoch dazu beitragen, theoretische Erkenntnisse für die Praxis besser nutzbar machen.

Die klassischen Führungsaufgaben sind also durch transaktionale Führung zu bewerkstelligen, auch wenn teilweise transformationales Führen unterstützend wirken können. Hier gelingt dann auch die Integration von Management in Führung, ohne disjunkte Aufgabenbereiche beschreiben zu müssen. Jede Führungskraft muss managen, also transaktional konzipierbare Führungsaufgaben erfüllen. Wenn sie zudem aber wirklich führt, sind zusätzliche Transformationen erreichbar. Tabelle 7 gibt einen Überblick über den prinzipiellen Unterschied zwischen Führung und Management und verdeutlicht dadurch die Ähnlichkeit zur Unterscheidung zwischen transformationaler und transaktionaler Führung.

Tabelle 7. Unterschied zwischen Führung und Management (nach Rost, 1998)

	Führung	Management
Beziehung	Beeinflussung	Autorität
Rollen	Führer/Führungskräfte (FK) und Geführte	Manager und unterstellte Mitarbeiter
Ziel	FK streben echte Veränderungen an	produzieren/verkaufen Güter und Dienstleistungen
Prozess	angestrebte Veränderungen durch *gegenseitigen* Prozess	Güter/Dienstleistungen als Ergebnis koordinierter Aktivitäten

Das Führungsmodell von Bass (1985) kann also als übergreifendes Konzept verstanden werden, in das sowohl verschiedene Führungsverhaltensweisen als auch konkrete Führungsaufgaben integrierbar sind. Lediglich die situativen Bedingungen, die in die verschiedenen Kontingenzmodellen integriert sind, werden nicht in dieser expliziten Form berücksichtigt. Vor dem Hintergrund dieses hohen Veränderungs- und Integrationspotentials, das in diesem Modell steckt, prognostiziert Wunderer (2001, S. 538), dass sich die transformationale Führung zum zentralen Führungsstil entwickeln wird. Dafür muss jedoch eine bessere Nutzbarkeit für die Praxis erarbeitet werden.

Bislang weist die transformationale Führung eine Nähe zu vielen wirtschaftspolitischen Konzepten auf, was ihre Bedeutung unterstreicht. So wird auf der Basis eines wirkungsvollen Selbstkonzepts vor allem durch intellektuell stimulierende Verhaltensweisen Intrapreneurship hervorgerufen (vgl. auch Organizational Citizenship Behavior), wie es zunehmend von Mitarbeitern, nicht mehr nur von Führungskräften gefordert wird. Auch Konzepte wie die Ich-AG oder weitläufig geforderte Selbständigkeiten sind ohne ein solches Unternehmertum zum Scheitern verurteilt. Auch lernende Organisationen können von inspirierten Mitarbeitern nur profitieren. Letztlich ist Empowerment ein Prozess, der von transformationalen Führungskräften ausdrücklich unterstützt wird. Empowerment moderiert die Beziehung zwischen transformationaler Führung und Arbeitszufriedenheit (Fuller, Morrison, Jones, Bridger & Brown, 1999). Insofern ist es im Sinne der Substitutionstheorie der Führung (Howell, Bowen, Dorfman, Kerr & Podsakoff, 1990; Kerr & Jermier, 1978) als „Enhancer“ zu betrachten (Fuller et al., 1999).

Dennoch muss sich die Führungsforschung inzwischen wohl eingestehen, dass sie für die unmittelbare praktische Umsetzung erst ansatzweise taugliche Modelle erbracht hat und möglicherweise auf einer sehr konkret-praktikablen Ebene auch gar nicht erbringen kann. Um jedem Einzelfall gerecht werden zu können, sind flexiblere Modelle notwendig. Entsprechend wird immer wieder kritisiert, dass die Führungsforschung generell für den Praktiker keine Relevanz besitze: „If you’re not serving Bill and Barbara, then you’re not serving leadership”, lautet entsprechend der Titel von Mintzberg (1982). Doch nicht allein inhaltlich definiertes Führungsverhalten, sondern eine Heuristik, die in Abhängigkeit vom Einzelfall ausgestaltet werden muss, wird der Komplexität des praktischen Führungsprozesses am ehesten gerecht (Beispiel in Tabelle 8). Ob dies langfristig in eine Führungstheorie integriert werden kann, bleibt offen.

Tabelle 8. Beispiel für eine pragmatische Führungsheuristik (nach Gebert, 1992)

1)	Zu welchem Ziel soll der Mitarbeiter einen Beitrag leisten?
2)	Welches Verhalten muss der Mitarbeiter - vor dem Hintergrund der Erfahrungen und theoretischen Erkenntnisse und - bei Vorliegen welcher situativer Rahmenbedingungen zeigen, um dem Ziel näher zu kommen?
3)	Welches Führungsverhalten ist - vor dem Hintergrund der Erfahrungen und theoretischen Erkenntnisse - bei Vorliegen bestehender Realisierungsbedingungen (z.B. Reifegrad des Mitarbeiters) besonders geeignet, um dieses Verhalten des Mitarbeiters hervor zu rufen oder zu unterstützen?

Relevant ist der Rahmen, vor dessen Hintergrund Führung stattfindet. Lange Zeit stand mindestens implizit die Frage im Raum, wie ideale Führungskräfte aussehen. Allein deshalb, weil kaum Genies verfügbar sind, muss jedoch die Frage umformuliert werden. Wichtig zu erforschen ist demnach, wie gewöhnliche Menschen wirksam führen können (Malik, 2000, S. 18). Vor diesem Hintergrund ist der Aufschwung charismatischer Führungsmodelle dann erneut nüchtern zu betrachten, wenn man sie als personale und weniger als verhaltensbezogene Ansätze konzipiert. Vorteilhaft wären also verhaltensbezogene, transformationale Führung oder das Ziel, die Selbstwirksamkeit der Mitarbeiter zu erhöhen. Doch auch diese Antworten sind für Praktiker auf zu abstraktem Niveau, weil sie keine generelle Anweisung für konkretes Handeln geben. Dennoch lässt ein solch offenes Modell Raum für die Forschung nach konkreten (situativen) Rahmenbedingungen, die die Effektivität von Führung erhöhen können und gleichzeitig erlernbar oder anderweitig hervorzurufen sind. So wird die Suche nach den Wirkungsmechanismen von Führung in der vorliegenden Arbeit auf eine interaktionelle Perspektive ausgedehnt. In diesem Sinne ist die Fokussierung dieser Arbeit auf die multiperspektivische Übereinstimmung von Führungsbeurteilungen zu betrachten. Es ist zu untersuchen, ob sie wirklich eine wichtige Bedingung für erfolgreiche Führung darstellt. Deshalb beschäftigt sich der zweite große Theorieteil (Kapitel 3) mit der multiperspektivischen Führungsbeschreibung, nachdem im folgenden Kapitel ein grundlegender Überblick über die Messung von Führungsverhalten gegeben wird.

2.7 Messung von Führungsverhalten

Seit langer Zeit gehört die Leistungs- und Ergebnisbeurteilung von Mitarbeitern zur normalen Aufgabe von Führungspersonen. Sie zieht sich als zentrale Führungsaufgabe durch Forschung und Praxis (vgl. Kapitel 2.3.1), da sie die wesentliche Aufgabe der Mitarbeiter einer Organisation kontrolliert, nämlich ob Leistungen gemäß der Organisationsziele auch erbracht werden. Doch warum sollten nur Mitarbeiter in ihrem Arbeits- und Leistungsverhalten beurteilt werden? Aus welchem Grund sollte das Führungsverhalten von Führungskräften von regelmäßigen Evaluationen ausgenommen werden – gerade da diesem so weit reichende Wirkungskraft zugesprochen wird? Es erscheint ein wenig blauäugig, Führungskompetenz als gegeben vorauszusetzen, wo doch immer noch häufig die besten Fachkräfte zu Führungskräften avancieren.[56]

Wenn Führungsverhalten gezielt eingesetzt, entwickelt oder honoriert werden soll, dann muss es messbar sein. Dabei kann "Messung" von Führungsverhalten gleichgesetzt werden mit Beurteilung von Führungsverhalten. Es leuchtet ein, dass ein so unstrukturierter Aufgabenbereich und ein so facettenreiches Konstrukt wie die Mitarbeiterführung schwierig zu erfassen sind. Der tatsächliche Messgegenstand muss also genau bestimmt werden. Erschwert wird die Messung außerdem dadurch, dass unter den Führungskräften, die in ihrer zentralen Aufgabe, dem Führen von Mitarbeitern, beurteilt werden, häufig verschiedene Vorbehalte herrschen. Sie betreffen die Glaubwürdigkeit der Fremdbeurteiler genauso wie Vermutungen über langfristig negative Konsequenzen (Prämienkürzung etc.). Diese Problematik, mit der sich jede forschungs- und betriebsbedingte Messung arrangieren muss, sei hier nur am Rande erwähnt und muss an gegebener Stelle bei der Ergebnisinterpretation mitbedacht werden.

2.7.1 Methodenüberblick

Zum Einsatz können verschiedene Methoden kommen: Prinzipiell sind die Verfahren, in denen die Führungskraft sich selbst beschreibt (Selbstbeschreibung) von denen zu unterscheiden, in denen andere Personen diese aus ihrer Perspektive beschreiben (Fremdbeschreibung). Zu Letzteren zählt die *Beobachtung* von alltäglichem Führungsverhalten (in Echtzeit oder in Ausschnitten) genauso wie die *Befragung* Dritter zum erlebten Führungs-

[56] Dass Fachkräfte (z.B. Ingenieure in der Industrie, Wissenschaftler an Universitäten) zu Führungskräften werden, zeigt die Unternehmensrealität. Eine Aus- oder Weiterbildung in Personalführung wird allenfalls durch unternehmensinterne Programme quasi nebenberuflich verfolgt. Generell wird vorausgesetzt, dass mit der Beförderung die Person auch Führungskompetenz erwirbt. Das Peter-Prinzip (Peter & Hull, 1969/2001) fasst anekdotisch zusammen, dass Beförderungen, die nicht auf der Grundlage entsprechender (Führungs-) Potentiale beruhen, zu einer „Hierarchie der Unfähigen" führt.

verhalten. Weiter kann aus der Fremdperspektive dasjenige Führungsverhalten erfasst werden, das beispielsweise in Personalselektionsverfahren durch bestimmte Aufgaben fernab der echten Führungsbeziehung provoziert wird (z.B. durch einzelne Übungen aus Assessment Centern oder die Management-Fallstudien von Fennekels & D'Souza, 1999). Ist man hingegen eher an der Selbstbeschreibung interessiert, dann kann neben der Befragung von Führungskräften die „Tagebuch-Methode" verwendet werden. Das bedeutet, dass die Führungskraft im Protokollstil notiert, was sie wann wie mit wem wozu tut.

Es fällt auf, dass nur subjektive Quellen für die Messung von Führung existieren, da Kriteriumsrelevanz objektiver Maße zu gering ist. Allenfalls können einzelne harte und objektiv messbare Variablen, zum Beispiel Fehltage der Mitarbeiter, als Führungs*erfolg* definiert werden und als Annäherung für den Führungserfolg, nicht für Führung als solche erhoben werden. Solche Maße sind also immer von einer hohen Kriteriumsdefizienz begleitet (erfassen relevante Kriteriumsanteile nicht). Da es mit zunehmender Komplexität und mit zunehmendem zwischenmenschlichen Interaktionsgrad der zu erfassenden Leistung immer schwieriger wird, objektive Kriterien mit hoher Relevanz zu finden, sind subjektive Beurteilungsquellen unverzichtbar.

Mit Abstand am häufigsten werden *Fragebogen* zur Messung von Führungsverhalten eingesetzt.[57] Der wohl gewichtigste Vorteil dabei ist, dass auf ökonomische Art und Weise Fremd- und Selbstbeschreibungen kombiniert werden können. Dazu muss der Fragebogen lediglich von der Führungskraft selbst und zusätzlich von weiteren Personen bearbeitet werden. Ziel kann sein, einzelne Führungskräfte in ihrem Führungsverhalten zu beurteilen (Individualanalyse) oder einen Trend quer über das Unternehmen zu gewinnen (Klimaanalyse). Für diesen letztgenannten Fall ist die Vorgesetztenbeurteilung in Form einer Mitarbeiterbefragung zum Thema Führung zu verstehen. Analog zu den unterschiedlichen Zielebenen kann auch die Analyseebene mehr oder weniger summarisch sein: Ein allumfassender Führungsscore, einzelne Führungsdimensionen oder einzelne Iteminhalte können im Vordergrund stehen. Die verfügbaren Instrumente unterscheiden sich insbesondere nach ihren Inhalten. Insgesamt erscheint der Fragebogenmarkt für Führungsfragebogen sehr unübersichtlich, da eine Vielzahl an Instrumenten unveröffentlicht blieb. Einige Fragebogeninstrumente werden in der Folge exemplarisch skizziert. Ausgewählt wurden

- der Fragebogen zur Vorgesetzten-Verhaltens-Beschreibung (Fittkau & Fittkau-Garthe, 1971), der die deutschsprachige Führungsforschung lange Zeit prägte,
- die Qualitative Führungsstilanalyse (Fennekels, 2000), ein aktuelles Instrument der Praxis, und
- der Multifactor Leadership Questionnaire (Bass, 1985), der ein breites Spektrum an Führungsverhalten erfasst („Full Range of Leadership") und in einer aktuellen Version von Felfe und Goihl (2003) in der vorliegenden Studie eingesetzt wurde.

[57] Wenn hier von Fragebogen die Rede ist, sind Einstufungsverfahren gemeint (vgl. Kapitel 2.3.1).

Es soll nicht unerwähnt bleiben, dass auch Fragebögen zur Erfassung von Führungsverhalten nicht nur kriteriumsrelevante, sondern zusätzlich irrelevante Information enthalten. Diese sogenannte „Kriteriumskontamination“ bildet die Grundlage für die problematische Interpretation später betrachteter interperspektivischer Beurteilungsdiskrepanzen (Kapitel 3). Für eine ausführlichere Diskussion der Problematik von Kriteriumsdefizienz und -kontamination in der Leistungsbeurteilung wird auf Marcus und Schuler (2001) verwiesen (vgl. auch S. 149f.).

2.7.2 Fragebogen zur Vorgesetzten-Verhaltens-Beschreibung (FVVB)

Der FVVB (Fittkau & Fittkau-Garthe, 1971) basiert auf den Zwei-Faktoren-Modellen, die Mitarbeiter- und Aufgabenorientierung als zentrale Führungsprinzipien betrachten (vgl. Kapitel 2.4). Genauer ziehen die Autoren als Konstruktionsgrundlage den Supervisory Behavior Description Questionnaire von Fleishman (1953) heran. Im Gegensatz dazu postulieren Fittkau und Fittkau-Garthe jedoch eine Fünf-Faktoren-Struktur, die folgende Dimensionen umfasst (Reihenfolge nach absteigender Varianzaufklärung durch den Faktor): Freundliche Zuwendung und Respektierung (F), Mitreißende, zur Arbeit stimulierende Aktivität (A), Ermöglichung von Mitbestimmung und Beteiligung (M), Kontrolle versus Laissez Faire (K) sowie Freundliche Zuwendung und stimulierende Aktivierung (F/A). Das strenge Zusammenspiel von Theorie- und Fragebogenentwicklung führte zu einer Auswahl an betrachteten Führungsverhaltensweisen, ohne letztlich plausibel begründen zu können, dass kein wichtiger Führungsaspekt vergessen wurde (vgl. Neuberger, 2002).

Bis heute wird die Faktorstruktur in Frage gestellt, von Kritikern wird die übliche Zwei-Faktoren-Struktur favorisiert (z.B. Allerbeck, 1978; Nachreiner, 1978; Titscher & Titscher, 1977). Weiter wird bemängelt, dass die Polung (z.B. sehr stark bis überhaupt nicht und umgekehrt) und die Verankerung der Items häufig variieren (in ständigem Wechsel unter anderem: oft bis sehr selten, oft bis fast nie, sehr stark bis überhaupt nicht). Außerdem fehlen Angaben zur Interpretation der Ergebnisse. Insbesondere die veraltete Normstichprobe von 1970 aber, die im Übrigen gleichzeitig Konstruktions- und Validierungsstichprobe darstellte, lässt einen heutigen Einsatz mangels Vergleichbarkeit problematisch erscheinen. Auch wenn der FVVB heute aufgrund verschiedener Probleme und des inzwischen antiquiert anmutenden Vokabulars, wenn überhaupt, dann nur noch aus Mangel an Alternativen eingesetzt wird, muss ihm ein historisches Verdienst zugesprochen werden. Dadurch, dass der Fragebogen als einer der ersten auf die Aufwärtsbeurteilung durch direkt unterstellte Mitarbeiter (Mitarbeiterbeurteilung) ausgerichtet war, wurde zunehmend auch die Perspektive der Mitarbeiter in ein erweitertes Führungsverständnis integriert (wie es auch

in der vorliegenden Arbeit postuliert wird). Nie mehr erlangte ein anderer deutschsprachiger Führungsfragebogen auch nur annähernd so viel Interesse.

2.7.3 Qualitative Führungsstilanalyse (QFA) und Modifikation

Die QFA (Fennekels, 2000) wurde von Beginn an so konzipiert, dass die Selbstbeschreibung der Führungskraft den Fremdbeschreibungen der Mitarbeiter gegenüber gestellt werden kann. Um den inhaltlichen Bereich abzugrenzen, wurden in Führungstrainings solche Problembereiche erarbeitet, die eine Zusammenarbeit im Unternehmen behindern. Folgende wurden extrahiert: Planung & Organisation, Entscheidungsverhalten, Soziale Kompetenz, Anerkennung & Mitwirkung, Leistungs- & Führungsverhalten. Die "kritischen Ereignisse" (im Sinne von Flanagan, 1954) dieser Dimensionen lagen der Itemkonstruktion des Fragebogens zugrunde. Zwar konnten die Faktoren empirisch bestätigt werden, doch besteht keine theoretische Grundlage für deren Auswahl. Dem gegenüber steht die Anwenderfreundlichkeit des Instruments, die mit dessen Erweiterung zum Multidirektionalen Feedback (MDF 360°, Fennekels, 2003) zusätzlich verbessert werden konnte. Im MDF 360° sind drei weitere Beurteilungsperspektiven erfassbar: die der Kollegen (Peers), die der Vorgesetzten und die der Kunden. Zum Ziel der praktischen Anschlussfähigkeit passt auch, dass die Inhaltsbereiche starke Überlappung mit den oben dargestellten klassischen Führungsaufgaben erkennen lassen, jedoch auf keiner Führungstheorie aufbauen.

Bei einer solchen Auswahl an Inhaltsbereichen besteht die Gefahr, dass diese lediglich ein an Willkür erinnerndes, zusammenhangsloses Sammelsurium an mehr oder weniger wichtigen Führungsaspekten darstellen, die adressatenspezifisch operationalisiert sind. Die Ausrichtung auf eine bestimmte Population ist allerdings zugleich von hoher Relevanz. Denn die Messung von Konstrukten, die im jeweiligen Kontext irrelevant sind, wird keine Akzeptanz erlangen und letztlich ihre Funktion nicht erfüllen können. Gängig – mit allen Vor- und Nachteilen – ist die adressatenspezifische Konstruktion von Fragebögen jedenfalls (z.B. für Führung in Militär, Schule, Polizei etc.). Die vielen Tools, die zum Kauf angeboten werden, sind im Bedarfsfall hinsichtlich ihrer Eignung für die bearbeitete Fragestellung genau zu überprüfen. Einige weitere Praxis-Instrumente sind beispielsweise im Anhang bei Neuberger (2000) nachzulesen.

2.7.4 Multifactor Leadership Questionnaire (MLQ)

Der MLQ basiert auf der Theorie zur transformationalen und transaktionalen Führung von Bass (1985) und erfasst in der eingesetzten Version alle in Kapitel 2.5.3.2 vorgestellten Bausteine des "Full Ranges of Leadership" Modells (transformationale und transaktionale Führung sowie Laissez Faire). Ursprünglich wurden bei der Itemsammlung durch verschiedene Arten von Interviews und qualitativen Methoden in unabhängigen Stichproben Charakteristika von transformationaler und transaktionaler Führung exploriert. Aus dem dadurch gewonnenen Material wurde eine erste MLQ-Version mit 73 Items konstruiert, die drei transformationale (Charisma, Intellektuelle Stimulierung und Individualisierte Berücksichtigung) und zwei transaktionale Subskalen (Kontingente Belohnung und „Management by exception“) sowie Laissez Faire erfasste (Bass, 1985). Ziel war es, deren differenzierten Einfluss auf die postulierten Erfolgsmaße (zunächst Extra Effort, Effektivität, Zufriedenheit mit Führung) zu messen. Diese wurden dann folgerichtig in eigenen Subskalen operationalisiert. Die faktorenanalytische Überprüfung dieses sogenannten „MLQ 1“ führte dazu, dass „Management by exception“ (MBE) in einen passiven und einen aktiven MBE-Faktor getrennt wurde (Hater & Bass, 1988). Obwohl in einzelnen neueren Studien diese Fünf-Faktoren-Struktur (Laissez Faire blieb unbeachtet) konfirmatorisch bestätigt werden konnte, wird aufgrund der hohen Interkorrelationen und geringen differentiellen Einflussgrößen auf Erfolgsvariablen immer wieder ein Modell mit weniger Faktoren favorisiert (Bycio et al., 1995).

Inzwischen wurde der Fragebogen mehrfach mit den Zielen modifiziert, die Trennung der Faktoren zu verbessern, neu ermittelte Faktoren aufzunehmen und stichprobenspezifische Versionen zu entwickeln. So variieren die vorliegenden Fassungen nicht nur in Länge und Anzahl ihrer Subskalen, sondern auch in ihrer Passung für bestimmte Tätigkeitsbereiche (z.B. Militärversion „Multifactor Officer Questionnaire“ von Yammarino & Bass, 1990) oder Themenschwerpunkte (z.B. zu Arbeitssicherheit von Barling, Loughlin & Kelloway, 2002).

In den meisten MLQ-Analysen (konfirmatorische und explorative Faktorenanalysen) können die transformationale und die transaktionale Führung getrennt werden. Häufig weisen MLQ-Analysen aber aufgrund der ähnlichen Kovariationen von „Management by exception passive“ (MBP) und Laissez Faire (LF) mit Drittvariablen sowie der hohen Interkorrelationen der übrigen Subskalen (z.B. Meta-Analyse von Lowe et al., 1996; Felfe & Goihl, 2003) darauf hin (vgl. Kapitel 6.1.1.2), MBP und LF als „passiven“ Faktor zusammenzufassen und der „aktiven Führung“ (bestehend aus den anderen Faktoren) gegenüber zu stellen (Den Hartog, Van Muijen & Koopman, 1997). Auf diese Faktoren zweiter Ordnung weist bereits Bass (1985) selbst hin. Zumindest mehren sich die Vorschläge für unterschiedliche MLQ-Versionen – auch für die in dieser Arbeit eingesetzte Version – lediglich zwei bis vier Faktoren (zum Beispiel aktive und passive oder transformationale,

transaktionale und passive Führung) zu unterscheiden (für MLQ 5x: Avolio, Bass & Jung, 1999; für andere Versionen: Den Hartog et al., 1997; Geyer & Steyrer, 1998; Lievens, Van Geit & Coetsier, 1997; Tejeda, Scandura & Pillai, 2001). Dann wären die oben angesprochenen Probleme, also die hohen Skaleninterkorrelationen und die mangelnden spezifischen Wirkungen auf Erfolg, aufgehoben.

Nach Bass (1998) sind die angesprochenen methodischen Probleme indes mit dem Modell vereinbar: Zum einen steigt beispielsweise die Wahrscheinlichkeit, dass eine Führungskraft inspirierend führt, wenn sie bereits charismatisch geführt hat. Da jedoch kein zwingender Zusammenhang besteht, wird die konzeptuelle Differenzierung nicht aufgegeben und statt dessen die Interkorrelation in Kauf genommen. Zum anderen greift transformationale Führung auf transaktionale Führung zurück. Die Korrelation zwischen den transformationalen Skalen und „Contingent Reward“ wird erklärt und akzeptiert, indem "Contingent Reward" als Schnittstelle zwischen transaktionaler und transformationaler Führung wahrgenommen wird (Tejeda et al., 2001). Gibt eine Führungskraft beispielsweise individuell Anerkennung für gute Leistung, dann kann dieses Verhalten auf der einen Seite als Belohnung im Sinne transaktionaler Führung gewertet werden, auf der anderen Seite aber auch als individualisierte Berücksichtigung.

Die in der vorliegenden Studie eingesetzte deutschsprachige Version basiert auf der Version MLQ 5x Short (Bass & Avolio, 1995, 2000), wurde aber um eine zehnte Subskala (Ausstrahlung) erweitert (Felfe & Goihl, 2003). Diese wird in Kapitel 5.2.2 eingeführt, in Anhang C finden sich die Items des gesamten Fragebogens. Bass und Avolio (1997) testeten mit konfirmatorischen Faktorenanalysen einige Modelle gegeneinander, wobei die neun-faktorielle Lösung des MLQ 5x die beste Modellpassung aufwies. Auf der anderen Seite entwickelten auch Tejeda et al. (2001) eine Kurzversion, nachdem sie andere Versionen nicht bestätigen konnten. Sie berichten dazu psychometrische Indizes, die die des MLQ 5x übertreffen.

Die Faktorstruktur und psychometrische Güte kann also auch für den MLQ 5x nur annäherungsweise repliziert werden, so dass alle Ergebnisse immer auch vor dem Hintergrund zu betrachten sind, dass der MLQ einer kontinuierlichen Evaluation und Verbesserung unterliegt. Insbesondere die Trennung der transformationalen Führung in verschiedene Faktoren gestaltet sich als schwierig. Letztlich ist jede Modifikation ein Versuch mehr, die Gratwanderung zwischen Ökonomie und Testgüte ohne „Beinbruch“ auf der einen oder anderen Seite zu bewältigen, also ein Versuch, aus (in vorliegender Version von Felfe & Goihl, 2003) 48 Items 13 voneinander unabhängige Faktoren[58] zu extrahieren. Die inkonsistente Ergebnislage aus unabhängigen Faktorenanalysen sollte aber meines

[58] sechs transformationale, drei transaktionale, eine Laissez Faire und drei Erfolgsskalen (vgl. Kapitel 5.2.2)

Erachtens nicht voreilig zur Rückkehr zur Zwei-Dimensionalität führen, wie sie auch für den MLQ 5x gefordert wird (z.B. Vandenberghe et al., 2002).

Im Gegenteil, es sollte eher der eingeschlagene Weg verstärkt verfolgt werden, um spezifische Messinstrumente zur Erfassung spezifischer und modellgerecht plausibler Führungsfacetten zu erarbeiten. Goodwin, Wofford und Whittington (2001) schlagen aufgrund dieser Unzulänglichkeiten vor, das Modell theoretisch und empirisch zu erweitern. Vor allem schlagen sie vor, "Contingent Reward" in einen impliziten und einen expliziten Faktor zu trennen. Während Ersterer eher Erwartungen für gute Leistung beinhaltet und mit den transformationalen Skalen stark positiv einhergeht, beschäftigt sich Letzterer mit konkreten Verhandlungs- und Austauschprozessen und ist enger mit anderen transaktionalen Skalen verknüpft.

Möglicherweise muss aber auch der Kontext stärker als bisher berücksichtigt werden. Für homogene Stichproben konnte die Neun-Faktorenstruktur des MLQ 5x (36 Items) nämlich konfirmatorisch bestätigt werden. Da in unterschiedlichen Kontexten Führung von den Führenden unterschiedlich operationalisiert und von den Geführten unterschiedlich wahrgenommen wird, wird empfohlen, zur Prüfung der Konstruktvalidität des Fragebogens keine inhomogenen Stichproben zu verwenden (Antonakis, Avolio & Sivasubramaniam, 2003).

Da der MLQ wie die meisten Fragebögen nicht situationsbezogen aufgebaut ist, basiert sein Einsatz auf der impliziten Annahme, dass Führungsverhalten transsituativ wirksam ist.[59] Lediglich die Häufigkeitsanker lassen dem Bearbeiter Spielraum, von einer universellen Information abzuweichen. Außerdem gibt der Aufbau des Fragebogens vor, dass jede Führungskraft durch ein charakteristisches Verhaltensprofil beschrieben und dieses durch den Fragebogen erfasst werden kann (Neuberger, 2002, S. 420).

Einige gewichtige Vorteile des MLQ (gegenüber anderen Führungsfragebögen) sollen an dieser Stelle nicht unerwähnt bleiben:

- Alle erfassten Dimensionen, die konzeptuell und empirisch erarbeitet wurden, spiegeln unmittelbar das theoretische Modell zur transformationalen Führung (Bass, 1998) wider. Die Basis einer guten Theorie erscheint der Arbeitsgruppe wichtiger als ein gutes Messinstrument, da sie letzteres – zum Beispiel in folgender Analogie – eindeutig als zweiten Schritt charakterisiert: „Atoms, X rays, and quarks were

[59] Dies spiegelt auf der einen Seite die von Bass vertretene Meinung wider, dass die Universalität des Modells einer kontingenztheoretischen Sichtweise überlegen sei (z.B. Interview mit Bass: Hooijberg & Choi, 2000, S. 297). Doch ergeben sich auf der anderen Seite Reibungspunkte mit der situationsabhängigen Wahrscheinlichkeit der Emergenz transformationaler Führung und Einsatzhäufigkeit einzelner Führungsfacetten sowie mit einem situationsmitbedingten Führungserfolg.

discussed and conceptualized long before they could be observed and accurately measured.“ (Bass & Avolio, 1993, S. 61)

- Da die Empfehlung, nicht nur transaktional, sondern zusätzlich transformational zu führen, von Effektivitätsuntersuchungen abgeleitet ist, erscheint es nur folgerichtig, dass entsprechende Erfolgsskalen konstruiert wurden. Damit unterstreicht die Arbeitsgruppe außerdem den Entwicklungscharakter, den sie mit allen Bausteinen ihrer Arbeit verfolgt. Es geht nicht darum, aus normativen Gründen den einen oder anderen Führungsstil zu zeigen, sondern darum, dass sich Führungskräfte zu effektiven Führungskräften entwickeln und damit zusätzlich Modelle für unterstellte Führungskräfte darstellen (vgl. Domino-Effekt, S. 82).
- Der MLQ ist das einzige Instrument, das ein derart breites Spektrum an Führungsverhaltensweisen erfasst (Lowe et al., 1996). Folglich ist sein Einsatzgebiet weniger eingeschränkt als bei anderen spezialisierteren Fragebögen. Der MLQ erwies sich auf verschiedenen Organisationsebenen sowie in unterschiedlichen Organisationstypen als brauchbar. Für die Konstruktvalidierung neu konstruierter Fragebögen eignet er sich deshalb besonders. Die hier eingesetzte Version deckt im Vergleich zu früheren Versionen des MLQ noch mehr Dimensionen ab.
- Durch unabhängige Faktorenanalysen wird das Instrument kontinuierlich überprüft und verbessert (vgl. S. 107f.). Auch wenn die Trennung der Subfaktoren von transformationaler und transaktionaler Führung bisher nur annäherungsweise repliziert werden konnte, kann zumindest die Trennung in passive und aktive Führung durchgehend nachvollzogen werden.
- Metaanalysen konnten sowohl dem Konzept der „neuen Führung“ (vgl. Kapitel 2.5) als auch dem MLQ selbst empirischen Nutzen bestätigen (DeGroot, Kiker & Cross, 2000; Fuller, Patterson, Hester & Stringer, 1996; Lowe, Kroeck, & Sivasubramaniam, 1996). Die Kriteriumsvalidität kann als gesichert angesehen werden. Problematisch erscheint allerdings, dass transformationale Führung nahezu ausschließlich mit dem MLQ gemessen wird, so dass jede Interpretation immer an die Gütekriterien des MLQ gebunden ist (vgl. Lowe et al., 1996). Erste entkräftende Hinweise gibt die Meta-Analyse von DeGroot et al. (2000). Sie stellt den charismatischen Baustein des MLQ anders erhobener charismatischer Führung gegenüber und findet keinen Unterschied. Auch Messungen der charismatischen Führung mit der „Conger-Kanungo—Scale“ korrelieren hochpositiv mit der Charisma-Skala des MLQ (Conger & Kanungo, 1994).
- Die Iteminhalte sind durchgängig leicht verständlich und – mit Ausnahme des Faktors „Idealized Influence (attributed)“ sowie der hinzugefügten Skala „Ausstrahlung“ – verhaltensbezogen formuliert. Da zudem die zugrunde liegende Theorie

plausibel und einfach ist, kann die Akzeptanz des Fragebogeneinsatzes sowie dessen Interpretationsobjektivität als gesichert angesehen werden.

- Bei Bedarf kann auf die verschiedenen Versionen (zum Beispiel Selbst- und Aufwärtsbeschreibung) des MLQ zurückgegriffen werden, was die Nützlichkeit für die Praxis stark erhöht. Vor allem durch die Kurzversionen verfolgen die Autoren auch das Kriterium der Ökonomie.

So kann geschlossen werden, dass für eine umfassende Messung von Führungsverhalten, die neben den traditionellen Führungskonzepten die „neue Führung“ abdeckt, der MLQ als anwenderfreundliches und theoretisch fundiertes Instrument eingesetzt werden kann. Wird hingegen darauf abgezielt, nur einen speziellen Führungsbereich zu erfassen oder normative Modelle mit dem "Full Range of Leadership" abzugleichen, muss (zusätzlich) ein weiteres, auf die Fragestellung ausgerichtetes Instrument eingesetzt beziehungsweise konstruiert werden – so auch in der vorliegenden Studie. Neuberger (2000, S. 38ff) warnt regelrecht vor dem universellen Einsatz vorhandener Instrumente.

3 Multiperspektivische Führungsbeschreibung

In Abbildung 16 ist das Themengebiet der multiperspektivischen Führungsbeschreibung (syn. Führungsbeurteilung) stichpunktartig zusammen gefasst. Gleichzeitig gibt die Abbildung einen Überblick über die einzelnen Gliederungspunkte dieses Kapitels.

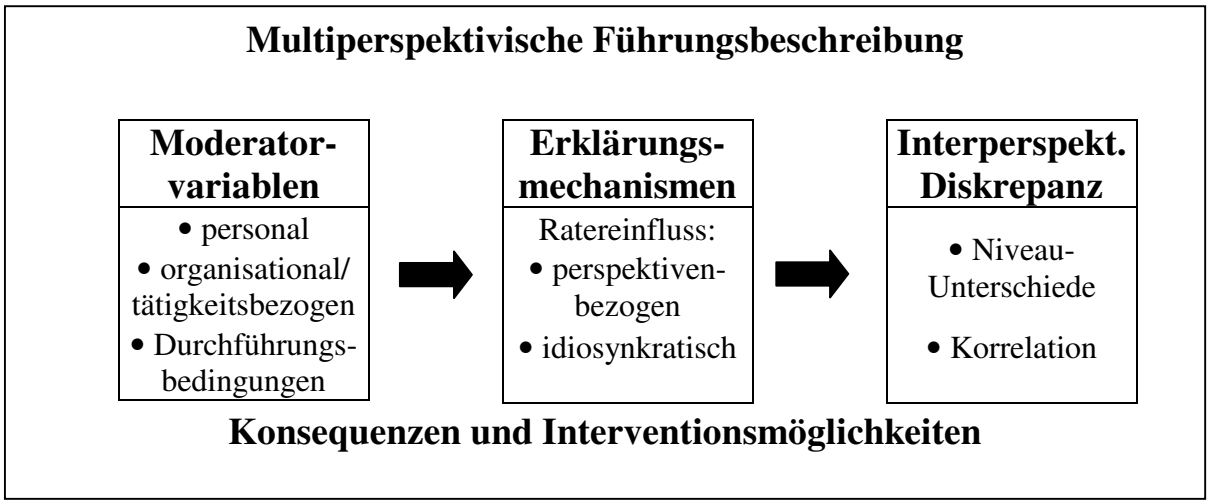


Abbildung 16. Multiperspektivische Führungsbeschreibung und ihre theoretischen Grundlagen – Überblick

In Kapitel 3.1 wird beschrieben, wie eine multiperspektivische Führungsbeschreibung durchgeführt wird und welche Vorteile sie gegenüber einer gewöhnlichen Beurteilung mit sich bringt. Es folgen in Kapitel 3.2 aktuelle Forschungsergebnisse zur Passung von Selbst- und Fremdbeurteilung. Dabei wird der Rahmen über Personalführung hinaus erweitert, um Befunde zu Führung überhaupt einordnen zu können, aber auch um die Ergebnisse einer breiten Forschungstradition nutzen zu können. Vor allem in Meta-Analysen bestätigte Moderatorvariablen, durch die eine bessere Passung zwischen Selbst- und Fremdbildern erzielt werden kann, sind in Kapitel 3.3 dargestellt. Im darauf folgenden Kapitel 3.4 werden Erklärungsmechanismen für das Zustandekommen der mangelhaften Passung zwischen Selbst- und Fremdbildern aufgezeigt. Hier werden auch die Moderatorvariablen erneut diskutiert. Die Konsequenzen einer mangelhaften beziehungsweise guten Passung werden in Kapitel 3.5 diskutiert, bevor abschließend in Kapitel 3.6 die Feedbackintervention beschrieben und ihre angenommene Wirkweise dargestellt werden.

Aufgrund eines Mangels an empirischen Befunden oder theoretischen Modellannahmen wird auf die Fremd-Fremd-Übereinstimmung (Abwärts-Aufwärts) nur am Rande eingegangen. Sie wird jedoch deshalb nicht gänzlich vernachlässigt, weil sie in der vorliegenden Studie ein ähnliches Gewicht erhält wie die Selbst-Fremd-Übereinstimmung.

3.1 Einführung und Begriffsbestimmung

„Das ‚Personnel Journal' gab sich optimistisch. Auf die Frage, ob Mitarbeiter künftig ihre Vorgesetzten beurteilen sollten, antwortete ein Autor des US-Fachblatts ohne Umschweife: ‚Es wird Zeit.' ... ‚Die Methode mag noch nicht verbreitet sein', so sein Resümee, ‚aber sie wird kommen'. Das war im Winter 1944". (Rieker, 1994, S. 181).

Will man Führungskräfte in ihrem zentralen Arbeitsverhalten, dem Führen von Mitarbeitern, beurteilen, dann zeigt sich hier wie anderswo die *Abwärtsbeurteilung* als traditionelle Beurteilungsperspektive. Über 90% aller leistungsbezogenen Ratings werden offenbar von Vorgesetzten abgegeben (Bernardin & Beatty, 1984). Ohne deren Wert mindern zu wollen, lässt sich jedoch berechtigtermaßen anzweifeln, ob eine Führungskraft von einer ihr unterstellten Führungskraft wirklich weiß, wie diese führt. Häufig lässt eine große Führungsspanne gar nicht zu, die mannigfaltigen Führungsinteraktionen, die jede Führungskraft mit ihren Mitarbeitern unterhält, umfassend zu beobachten oder über Informationen aus dritter Hand genau zu erfassen.

3.1.1 Die Überwindung der „Abwärts-Vorherrschaft"

"Nachdem [...] aus Arbeitern Mitarbeiter und aus Vorgesetzten Führungskräfte wurden, denen niemand unterstellt oder gar untergeben ist, und Mitarbeiter keinen Weisungen folgen, sondern ihren Führungskräften berichten, wird zunehmend interessant, was Mitarbeiter zu sagen haben." (Wiendieck, 1997).

Mit der stetig zunehmenden Verflachung der Hierarchien – also mit steigenden Führungsspannen – sinken die *Gelegenheiten* der potentiellen Abwärtsbeurteiler, das relevante Verhalten (z.B. Führungsverhalten) aller unterstellen Mitarbeiter (mit Führungsverantwortung) zu beobachten. Mit der zunehmenden Spezialisierung der Arbeitstätigkeiten sinken ihre *Fähigkeiten* zur adäquaten Beurteilung. Zudem erschweren mikropolitische Strategien eine genaue Beurteilung, da Vorgesetzte beispielsweise ihre Mitarbeiter weder durch schlechte Beurteilungen demotivieren noch durch gute Beurteilungen an andere Abteilungen verlieren wollen.

Meyer (1991) plädiert für die Hinzunahme zumindest der *Selbstbeschreibung*, nachdem er die pure Abwärtsbeurteilung als anachronistische, autoritäre Vater-Kind-Beziehung charakterisierte. Partizipation (und damit eine gleichberechtigtere/wertschätzendere Beziehung) bei Beurteilungen kann also nur erreicht werden, indem der zu Beurteilende um die Bewertung seiner eigenen Leistung, die im Übrigen nur ihm selbst direkt zugänglich ist, gebeten wird. Die Teilnahme am eigenen Beurteilungsverfahren führt dazu, dass Kritik von anderen eher akzeptiert werden kann, dass man sich des Leistungsverhaltens eher

bewusst wird und insgesamt zu Introspektion und Selbstevaluation angeregt wird (vgl. auch Fletcher & Baldry, 2001).

Im Folgenden werden drei argumentative Linien verfolgt, die (zusätzlich zur Abwärtsbeschreibung) außerdem die *Aufwärtsbeschreibung* fordern. Zunächst erscheint es unbestreitbar, dass eine Abwärtsbeurteilung den Führungsprozess nur einseitig darzustellen vermag. Denn gerade hier, wo es um die Führung der Mitarbeiter geht, sollte deren Meinung eine der wichtigsten sein (Aufwärtsbeurteilung). Den Mitarbeitern eröffnet sich nämlich aufgrund ihrer Geführtenrolle eine unmittelbare und dadurch einzigartige Außen-Sicht auf das Führungsverhalten. Am Rande sei bemerkt, dass gemäß der Definition von Weibler (2001, vgl. S. 16) gar keine Führung stattfände, wenn die Mitarbeiter, auf die die Führung abzielt, die Beeinflussung nicht akzeptierten.

Zweitens spricht die Reliabilität dafür, die Mitarbeiterperspektive in den Prozess der Führungsbeurteilung zu integrieren: Da die meisten Führungskräfte mehrere Mitarbeiter haben, stellen diese im Vergleich zu anderen Beurteilungsgruppen die potentiell größte Ratergruppe dar. Dank dieses Vorteils kann verschmerzt werden, dass die Reliabilität von Einzelratings in der Aufwärtsperspektive minimal geringer als die der Peerperspektive und deutlich geringer als die der Abwärtsperspektive ausfällt (Meta-Analyse von Conway & Huffcutt, 1997; Greguras & Robie, 1998).[60] Denn mit der Anzahl aggregierter (gemittelter) Aufwärtsbeurteilungen (die im Unterschied zu anderen Perspektiven leicht erhöht werden kann) steigt die Reliabilität und übersteigt sogar die Reliabilität einzelner Abwärtsbeurteilungen (Greguras & Robie, 1998; Scullen, Mount & Goff, 2000) sowie Peerbeurteilungen (Greguras, Robie, Schleicher & Goff, 2003 [nur für Ratingprozesse, die auf die Mitarbeiterentwicklung abzielen]). Um beispielsweise eine Reliabilität von .70 zu erreichen,[61] sind nach den Berechnungen von Conway und Huffcutt (1997) sechs Mitarbeiterratings nötig. Für einen vergleichbaren Reliabilitätsindex wären für die Beurteilung von Management-Tätigkeiten drei Abwärtsbeurteiler oder fünf Peerbeurteiler notwendig, was de facto in kaum einer Organisationsstruktur ermöglicht.[62] Zur Wirkung von Moderatorvariablen auf die Höhe der Reliabilität sei auf Conway und Huffcutt (1997) verwiesen. Zum Beispiel steigt die Reliabilität der Aufwärtsbeurteilungen, wenn es um die Beurteilung interpersoneller (versus kognitiver) Bereiche geht (beispielsweise Kommunikationsverhalten oder

[60] Die Reliabilitätsschätzungen von Conway und Huffcutt (1997) erbringen ohne Berücksichtigung der Skalenlänge im Durchschnitt: Aufwärts: .30, Abwärts: .50, Peer: .37.
Ähnlich fallen die Angaben von Greguras und Robie (1998) aus, die sich nur auf *Führungs*beschreibungen beziehen (N = 153). Sie sind insgesamt (für Skalen bis zu 20 Items) etwas niedriger und weisen vernachlässigbare Unterschiede zwischen Peer- und Aufwärtsbeurteilung auf. Die im Vergleich zu anderen Perspektiven höhere Reliabilität der Abwärtsbeurteilungen könnte einen Übungseffekt darstellen. Denn es ist Aufgabe von Vorgesetzten, ihre unterstellten Führungskräfte/Mitarbeiter kontinuierlich zu beurteilen (vgl. Kapitel 2.3.1).

[61] Reliabilitätskriterium nach Nunnally, 1978

[62] Greguras, Robie et al. (2003), die neben der Anzahl an Ratern auch die Anzahl an Items einbeziehen, extrahieren aus ihrer Management-Stichprobe (Ratingzweck = Entwicklung) folgende Rateranzahl, um mit 10 Items eine Reliabilität von .70 zu erreichen: sechs Aufwärts- beziehungsweise neun Peerrater.

mitarbeiterorientierte Führung). Erinnert wird daran, dass eine Steigerung der Reliabiliät auch in erhöhter Validität der Aufwärtsbeschreibungen (Korrelation mit Außenkriterien) resultieren kann.

Genau darin besteht ein drittes Argument, das die Wichtigkeit der Aufwärtsbeschreibung unterstreicht: Mehrfach konnte bisher belegt werden, dass ein signifikanter Zusammenhang zwischen der Aufwärtsbeschreibung von Führung und Erfolgsvariablen besteht (Kriteriumsvalidität). Dieser ist deutlich stärker als der Zusammenhang zwischen Selbstbeschreibung von Führung und Erfolgsvariablen (z.B. Bass & Yammarino, 1991; Furnham & Stringfield, 1994; Sala & Dwight, 2002). Wenn mehrere Aufwärtsrater aggregiert werden, erklären diese auch mehr Varianz als einzelne Abwärtsrater, deren Anzahl in der Praxis nur in Ausnahmefällen erhöht werden kann (Mount & Scullen, 2001). Die Mitarbeitersicht des Führungsverhaltens kann somit als Indikator für Erfolg betrachtet werden.

Tabelle 9 gibt einen Überblick über die Vielzahl und Höhe der Zusammenhänge. Dabei sind die Erfolgsvariablen entsprechend dem Einstellungs-Verhaltens-Modell von Fishbein (1967) in Meinungen, Einstellungen, Verhaltensintentionen und Verhalten getrennt, so dass mit aufsteigender Reihenfolge die unternehmensweite Relevanz steigen sollte.

Tabelle 9. Übersicht über den Zusammenhang zwischen Aufwärtsbeschreibung von Führung und Erfolgsvariablen

Erfolgskriterium	**Führungs-prädiktor**	***r****	**Branche**	**Autoren (Beispiel)**
Einstellungen				
Commitment[1]	Management-Verhalten	.47 ‡	Krankenhaus	Shipper & Wilson, 1992
Affektives Commitment[1]	Tf	.41 ‡	Krankenhaus	Bycio, Hackett & Allen, 1995
	Ta	.36		
Commitment zu bleiben[1]	Tf	-.04 ns		
	Ta	-.06 ns		
Normatives Commitment[1]	Tf	.16 ‡		
	Ta	.20		
Zufriedenheit mit Führung[1]	Tf	.78 ‡	verschiedene	Bass, 1985
	Ta	.37 ‡		
Arbeitszufriedenheit[∞]	IS	.31	Einzelhandel	Koene et al., 1992
	C	.39		
	Ch	.49		
Verhaltensintention				
Kündigungsabsicht[1]	Tf	-.29 ‡	Krankenhaus	Bycio et al., 1995
	Ta	-.21		
Arbeitsmotivation[1]	IS	.49	Einzelhandel	Koene et al., 1992
	C	.68		
	Ch	.72		
Extra Effort[1]	Tf	.84	Bankgewerbe	Geyer & Steyrer, 1998
	Ta	.82		

Verhalten & Verhaltensergebnisse

Stress/Anspannung[1]	Management-Verhalten	-.44 ‡	Krankenhaus	Shipper & Wilson, 1992
Fehltage	M	-.21	Automobilzulieferer	Schmidt, 1996
	F, A, K, F/A ($\bar{r}$)	.03 ns		
Fluktuation (2-Jahres-Zeitraum)	F	-.21	Automobilzulieferer	Schmidt, 1996
	A	-.19		
	M	-.30 #		
	K	.06 ns		
	F/A	-.23		
Innovatives Verhalten (Einreicherquote BVW) [2]	F, A, M, F/A K		Automobil	Nickel & Krems, 1998
Projektqualität (u.a. technisch und Unternehmenswert)	Tf	.31 ‡	F&E-Abteilungen	Keller, 1992
	IS	.25		
	C	.26		
Leistungsbeurteilung	Tf	.30 †	Marine-Offiziere	Atwater & Yammarino, 1992
		.29 ‡		
	Ta	.20 ‡		Bass & Yammarino, 1991
	LF	-.31		
Empfehlung zu vorgezogener Beförderung			Marine-Offiziere	
▪ kürzlichst	Tf	.21 †		Atwater & Yammarino, 1992
▪ früher/insgesamt		.32 †		
		.31 ‡		
	Ta	.23 ‡		Bass & Yammarino, 1991
	LF	-.31		
Beförderung nach 7 Jahren	Management-Verhalten	.22	US-Polizeieinheit	McEvoy & Beatty, 1989
Effektivität der Arbeitsgruppe[1]	Tf	.53 ‡	Lieferfirma	Hater & Bass, 1988
	Ta	.08 ‡		
Effektivität der Führungskraft[1]	Tf	.32 ‡	Lieferfirma	Hater & Bass, 1988
	Ta	-.20 ‡		
Schwund-Satz an der Brutto-Verkaufssumme (Fehlbetrag)	IS	ns	Einzelhandel	Koene et al., 1992
	C	ns		
	Ch	-.29		
Jahresleistung	Management-Verhalten	.33 ‡	Krankenhaus	Shipper & Wilson, 1992
Produktvolumen (Konten etc.) (kurz- und langfristig)	Tf	.27	Bankgewerbe	Geyer & Steyrer, 1998
	Ta	.19		

Anmerkungen.

*Alle Korrelationen zwischen Führungsprädiktor und Kriterium sind mindestens auf dem 5%-Niveau signifikant, wenn nicht anders vermerkt, ns: nicht signifikant (wenn in der Tabelle kein Korrelationsindex angegeben ist, dann wurde auch im entsprechenden Artikel keiner berichtet).

[1] Das Erfolgskriterium wurde genauso wie das Führungsverhalten von den Mitarbeitern erfasst. Zur Problematik des Single-Source-Bias, die diese Studien betrifft, wird auf Kapitel 2.5.4.1 verwiesen.

[2] Die Mittelwertsvergleiche zwischen zwei Extremgruppen (Abteilungen mit hoher bzw. niedriger Einreichungsquote) verweisen auf Unterschiede auf den Dimensionen F, A, M und F/A.

‡Angegeben sind jeweils die selbst berechneten *mittleren* Korrelationen aus den Korrelationen zwischen den einzelnen Führungs-Subskalen und dem Erfolgskriterium.

C: Consideration, IS: Initiating Structure, Ch: Charismatische Führung, Tf: Transformationale Führung, Ta: Transaktionale Führung (ohne passives Management by exception (MBE) bzw. ohne MBE für den Fall, dass aktives und passives MBE nicht getrennt wurden), LF: Laissez Faire, F, A, M, K & F/A: 5 Faktoren des FVVB (vgl. Kapitel 2.7.2).

Nur M hält weiteren Kontrollen stand.

† In die Korrelationen gehen die gesamten Aufwärtsbeurteilungen ein, ohne deren Diskrepanz zur Selbstbeurteilung zu betrachten (dafür vgl. Kapitel 3.5).

Wie Mitarbeiter das Führungsverhalten ihres Vorgesetzten wahrnehmen, geht also mit verschiedenen Erfolgsvariablen einher: angefangen bei Commitment- und Zufriedenheitskonstrukten über Motivationsvariablen bis hin zu beobachtbarem Verhalten und Verhaltensergebnissen aus den Bereichen Gesundheit, Innovation und Leistung. Insofern ist die Sichtweise der Mitarbeiter nicht nur für die Organisation als Ganzes von Bedeutung, sondern auch für die jeweiligen Führungskräfte. Denn deren Aufgabe ist es, genau diese Ergebnisse zu beeinflussen.

Die dargestellten Befunde in Tabelle 9 zeigen, dass die Führungs-Erfolgs-Zusammenhänge von den Einstellungs- bis zu den Verhaltens-Kriterien nicht kontinuierlich schwächer werden. Wer die Mitarbeitersicht der eigenen Führung ignoriert, verschenkt also tatsächlich auch Indikatoren für die aktuelle Ergebnislage seines Teams. Dies gilt insbesondere vor dem Hintergrund, dass die Selbstbeschreibung der Führungskräfte kaum mit Führungserfolg korreliert.

Die starken Zusammenhänge zwischen der Aufwärtsbeurteilung und den Erfolgskriterien ist nicht auf die Aggregation der Aufwärtsbeurteiler zurückzuführen, da die Analysen, die mit individuellen (nicht-aggregierten) Aufwärtsratings arbeiten (z.B. Bycio et al., 1995; Geyer & Steyrer, 1998; Schmidt, 1996), etwa gleich starke Erfolgszusammenhänge erbringen wie die Analysen, die mit aggregierten Ratings rechnen (z.B. Bass & Yammarino, 1991; Hater & Bass, 1988). Folglich ist der Aufwärtsbeurteilung per se Kriteriumsvalidität zuzusprechen.

Meta-Analysen bestätigen, dass auch die Aufwärtsbeurteilung von beruflicher Leistung generell (nicht auf Führungsverhalten beschränkt) in ähnlicher Höhe mit objektiven Leistungsmaßen korreliert (ρ = .25 bei Conway, Lombardo & Sanders, 2001). Entsprechende Ergebnisse finden sich dort auch für Peer-Beurteilungen (ρ = .29). Damit ist die Kriteriumsvalidität von Aufwärtsbeurteilungen zwar etwas geringer als die der Abwärtsbeurteilung beruflicher Leistung (ρ = .35 bei Bommer, Johnson, Rich, Podsakoff & MacKenzie 1995, siehe unten).[63] Sie ist aber nicht vernachlässigbar, weil sie – wie bereits oben erwähnt – einen inkrementellen Beitrag über die Abwärtsbeurteilung beruflicher Leistung hinaus leistet (Conway et al., 2001).

Vergleicht man nicht verschiedene Beurteilungsperspektiven in ihrer Vorhersagekraft für Erfolgskriterien miteinander, sondern zieht man andere Leistungsprädiktoren heran, gilt die Zusammenfassung von McEvoy und Beatty (1989, p. 50): “The predictive validity places subordinate ratings in the upper echelon of predictors of managerial performance, along with cognitive ability tests, biodata, peer evaluations, and assessment centers.”

Eine Überprüfung der Validität uniperspektivischer Beurteilungen ist Voraussetzung für eine adäquate Erforschung der interperspektivischen Diskrepanzen. Wenn auch vieles da-

[63] Da die Vertrauensintervalle für Aufwärts- und Peerbeurteilungen bei Conway, Lombardo und Sanders (2001) sehr hoch sind, ist der Unterschied zur Abwärtsbeurteilung nicht deutlich.

für spricht, ist bislang noch nicht eindeutig geklärt, ob die Diskrepanz zwischen Selbst- und Aufwärtsbeschreibung einen besseren Beitrag zum Erfolg leisten kann als die Aufwärtsbeschreibung allein (vgl. Kapitel 3.5.1). Diskrepanzen zur Abwärtsbeschreibung oder gar Abwärts-Aufwärts-Diskrepanzen sind hinsichtlich ihrer Kriteriumsvalidität bisher kaum untersucht. Eine weitere Aufklärung dessen bildet eine Hauptfragestellung der vorliegenden Untersuchung.

3.1.2 Definition und Begriffsklärung

Wird die Aufwärtsbeschreibung mit der Selbstbeschreibung der Führungskraft und der traditionellen Abwärtsbeschreibung durch den direkten Vorgesetzten kombiniert, dann kann von einem "Sandwich-Modell" gesprochen werden. Werden zusätzlich noch die Urteile von Kollegen oder externen Partnern (z.B. von Kunden) eingeholt, dann wird dieses Vorgehen als "360°-Feedback" bezeichnet (siehe Abbildung 18).[64]

360°-Feedback ist das Schlagwort, unter dem die multiperspektivische Führungsbetrachtung populär wurde (z.B. Neuberger, 2000; Pollack & Pollack, 1996; Runde, Kirschbaum & Wübbelmann, 2001). Eine solche Rundumbetrachtung trägt dem Ziel Rechnung, das Führungsverhalten möglichst genau erfassen zu wollen und verfolgt die grundlegende Idee, die Führungseffektivität über den Weg der Selbsterkenntnis der Führenden zu verbessern, die sich aus der Gegenüberstellung verschiedener Beurteilungen ergeben soll.

Allgemein kann jede Beurteilung von mindestens zwei Beurteilern als "multiperspektivische Führungserfassung, -beurteilung, -beschreibung oder -messung" bezeichnet werden. Zugunsten dieser Terminologie wird in der vorliegenden Studie auf die uneinheitlich gebrauchte, eher metaphorisch zu verstehende Bezeichnung "360°-Feedback" aus zwei Gründen verzichtet. Erstens stellt das in der vorliegenden Studie verwendete Design ein Sandwich-Modell dar und entspricht in der Winkel-Nomenklatur folglich nur einem *270°*-Modell (Scherm & Sarges, 2002, S. 2).[65] Diese 270°-Methode ist im Übrigen viel gängiger als das allseits diskutierte 360°-Feedback, da es trotz seiner relativen Ökonomie vielfältige Informationen erbringt. Zweitens liegt der Schwerpunkt auf der Leistungsmessung, nicht auf dem *Feedback*prozess, welcher in der Regel in Entwicklungsmaßnahmen integriert wird (vgl. Kapitel 3.6).

[64] Die Definition ist nicht einheitlich. Während beispielsweise Neuberger (2000) erst dann von einem 360°-Feedback spricht, wenn auch unternehmensexterne Personengruppen (Kunden und/oder Familienmitglieder) befragt werden, liegt die Definitionsschwelle bei anderen niedriger (z.B. Tornow, 1993). Hier genügen die vier organisationsinternen Perspektiven der Selbst-, Abwärts- (durch Vorgesetzten), Aufwärts- (durch Mitarbeiter) und Seitwärtsbeschreibung (durch Kollegen/Peers). Scherm und Sarges (2002) bezeichnen beides als 360°-Feedback – mehr oder weniger idealtypisch.

[65] Für Zusammenfassungen von Kunden- beziehungsweise Peerratings wird auf Paradise-Tornow (1998) und Jochum (1991) beziehungsweise auf die Meta-Analyse von Conway und Huffcutt (1997) verwiesen. In der vorliegenden Arbeit werden sie nur sehr randständig behandelt.

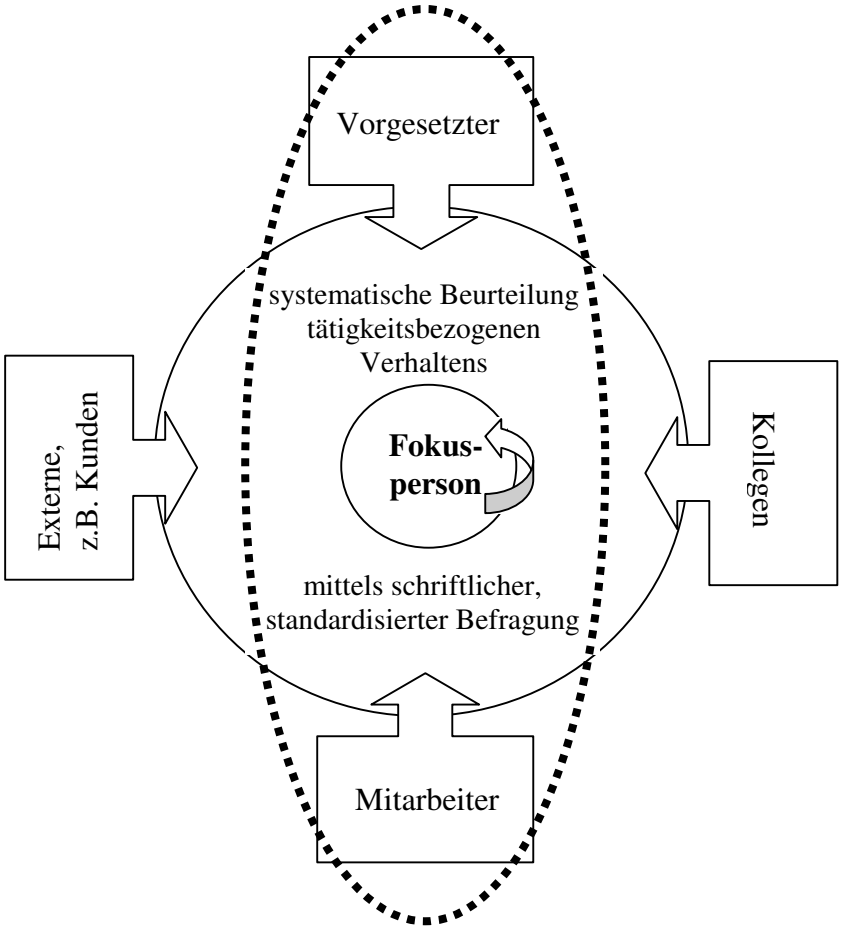


Abbildung 18. Prototyp eines 360°-Feedbacks (gesamte Abbildung) und eines Sandwich-Modells (gepunktete Ellipse) (nach Scherm & Sarges, 2002)

Unabhängig von der uneinheitlichen Begriffsverwendung verbirgt sich hinter den Konzepten das gleiche (nach Scherm & Sarges, 2002), nämlich eine

- systematische Beurteilung (in vorliegender Studie: vgl. Kapitel 5.3)
- der Fokusperson (in vorliegender Studie: Führungskräfte)
- auf tätigkeitsbezogenen Dimensionen (in vorliegender Studie: Führungsverhalten)
- mittels schriftlicher, standardisierter Befragung (in vorliegender Studie: Fragebogen vgl. Kapitel 5.2 und 5.5)
- von verschiedenen Beurteilergruppen (in vorliegender Studie: Abwärts-, Selbst- und Aufwärtsbeurteiler).

Terminologisch gilt für die vorliegende Arbeit, dass jeweils die Perspektive des Beurteilers (Rater), nicht die des Beurteilten (Ratee), notiert ist. *Mitarbeiterratings* sind also Aufwärtsratings von Mitarbeitern, die ihre direkte Führungskraft beurteilen und nicht Abwärtsratings für Mitarbeiter als Beurteilte.

3.1.3 Ablauf und Ziele

Wie sieht eine multiperspektivische Führungsbeschreibung aber nun konkret aus? Fast schon als Standard kann folgender Ablauf festgeschrieben werden: Erstens Fragebogenbearbeitung, zweitens externe Auswertung und drittens Feedbackgespräche mit Führungskraft und Team:

1. Alle Beteiligten erhalten einen Fragebogen, der inhaltlich genau die Dimensionen abdeckt, die das entsprechende Unternehmen von seinen Führungskräften fordert. Dabei spielen Verhaltensmaße die größte Rolle. An dieser Stelle sei ausdrücklich darauf verwiesen, dass in vorliegender Studie aus den verschiedenen Perspektiven die gleichen Verhaltensweisen (Items) beurteilt werden und daher der in der Praxis typischere Fall in dieser Arbeit unbeachtet bleibt: Häufig werden nämlich aus jeder Perspektive nur diejenigen Dimensionen erfasst, die aus der jeweiligen Perspektive besonders gut beobachtbar sind, zum Beispiel Führung aus der Mitarbeiterperspektive, Verwaltung aus der Vorgesetztenperspektive (Brutus, Fleenor & London, 1998) oder Kooperation aus der Peerperspektive (Pollack & Pollack, 1996).

2. Die Fragebögen werden in der Regel extern ausgewertet, die beteiligten Mitarbeiter anonymisiert. Anonymität wird von Mitarbeitern nicht nur präferiert, sondern führt auch dazu, dass diese eher zur Teilnahme bereit sind und ihre Urteile weniger stark "schönen" als unter identifizierbaren Bedingungen (Antonioni, 1994; London, Wohlers & Gallagher, 1990; Mabe & West, 1982).[66] Dort, wo Anonymität nicht gegeben oder aufgrund schlechter Erfahrungen angezweifelt wird, lebt die Angst der Mitarbeiter vor einer "Retourkutsche" bei ehrlicher Kritik. So finden viele Autoren die Anonymitätsfrage nicht einmal eine Diskussion wert (z.B. Kiessling & Spannagl, 1996). Die Kehrseite dessen impliziert jedoch die Annahme, dass die Mitarbeiter durch Anonymitätszusicherung wirklich vor genau solchen Retourkutschen ihrer Chefs beschützt werden müssen, während dem Chef die alleinige Verantwortung für eine Veränderung zugesprochen wird. Damit ist das Dilemma perfekt: Die Verantwortung für den Beurteilungs- und anschließenden Entwicklungsprozess kann nur dadurch allen Beteiligten, auch den Mitarbeitern, übertragen werden, wenn diese identifizierbar sind. Die Identifizierbarkeit der beteiligten Fremdbeurteiler führt allerdings dazu, dass sie deutlich mildere Ratings abgeben als unter anonymen Bedingungen oder überhaupt nicht mehr teilnehmen. Identifizierbarkeit scheint also mit ehrlicher Verantwortungsbereitschaft (noch) nicht vereinbar zu sein; eine gemeinsame Verpflichtung zur Verbesserung von Führung scheint also schwierig. Nicht umsonst wird eine Art Re-

[66] Während Mitarbeiter unter identifizierbaren Bedingungen also signifikant positivere Urteile abgeben als unter anonymen Bedingungen, bewerten die beurteilten Führungskräfte den Beurteilungsprozess positiver, wenn ihre Mitarbeiter für sie identifizierbar sind (Antonioni, 1994). Allerdings ist bislang nicht bekannt, ob sich Führungskräfte dann (bei Identifizierbarkeit ihrer Mitarbeiter) auch intensiver mit dem erhaltenen Feedback auseinandersetzen und eher zu Veränderungen motiviert sind.

chenschaftspflicht oder Verantwortungsbereitschaft als "Achillesferse" des multiperspektivischen Feedbackprozesses beschrieben (London, Smither & Adsit, 1997; vgl. S. 203).[67] „Accountability“ erfährt derzeit insgesamt eine hohe Befürwortung (z.B. Jay, 2003; Lebow & Spitzer, 2002).

Zumindest ist es wichtig klarzustellen, dass Anonymität weder bedeutet, dass die Ergebnisse abstrakt bleiben und verpuffen, noch, dass Tendenzen unterstützt werden, sich bei der Führungskraft für angesammelte Unzufriedenheiten mit einer extrem negativen Bewertung zu rächen. Um die Anonymität der Mitarbeiter zu gewährleisten, schlagen Bass und Avolio (1997, S. 11) vor, mindestens drei Mitarbeiter zu befragen. Andere Autoren setzen die Mindestgröße auf fünf Mitarbeiter. Diese willkürlich erscheinenden Mindestangaben sollten aus Gründen der Reliabilität erhöht werden (vgl. S. 114). Da es in der Regel nur einen Abwärtsbeurteiler gibt, können Vorgesetzte nie anonym bleiben.

3. Die Feedbackgespräche verfolgen das Ziel, gemeinsam die Befunde aus der Fragebogenerhebung zu erklären, um letztlich Abmachungen zwischen der Führungskraft und ihren Mitarbeitern zu treffen, wie Problembereiche aus der Vergangenheit konkret in effektive Zusammenarbeit münden sollen. Moderiert oder nicht: Solche Gespräche zwischen Führungskraft und Mitarbeitern erhöhen die Chance, die angestrebte Veränderung/Verbesserung auch zu erreichen und wirken deshalb für einen *Entwicklungsprozess* sehr unterstützend (vgl. Kapitel 3.6).[68]

So kann das Procedere prinzipiell also mehr bieten als eine Momentaufnahme der Führung. Es kann eine Steuerungsfunktion für die nächsten Schritte übernehmen. Für eine weniger individuums-, sondern eher *organisationsbezogene Fragestellung* – wie sie hier vorliegt – kann aber auch eine solche Momentaufnahme gewinnbringend sein: Das Führungsbild des gesamten Unternehmens wird greifbarer, die Personalsysteme werden evaluiert und ein falscher Kurs wird rechtzeitig signalisiert. Insofern ist eine multiperspektivische Befragung immer Teil eines Organisationsprozesses, da sie – ob gesteuert oder nicht – immer Konsequenzen nach sich zieht. Wird im Zuge dessen eine Entwicklungsmaßnahme angestrebt, dann muss klar sein, dass aggregierte Daten auch nur zu aggregierten Veränderungsmaßnahmen führen können. In diesem Fall hat man bei negativen Befragungsergebnissen oft damit zu kämpfen, dass “nicht ich, sondern die anderen dafür verantwortlich sind" und deshalb zunächst Änderungsmotivation aufgebaut werden

[67] Gerade die Anonymität wird von Befürwortern einer offenen Feedbackkultur (z.B. Sprenger, 1999) abgelehnt. In der Regel besteht jedoch Einstimmigkeit darüber, langfristig das Ideal einer solchen erreichen zu wollen. Doch zum Zeitpunkt der multiperspektivischen Beurteilung müssen Vor- und Nachteile zielbezogen abgewägt werden.

[68] Für ausführlichere Praxis-Anleitungen kann beispielsweise auf Borg (2000) für Mitarbeiterbefragungen im Allgemeinen beziehungsweise auf Bungard und Jöns (1997), Hofmann, Köhler und Steinhoff (1995), Tornow und London (1998) oder Scherm und Sarges (2002) für multiperspektivische und Vorgesetztenbeurteilung im Speziellen verwiesen werden.

muss. Es empfiehlt sich, Befragungen mit Personalentwicklungszielen getrennt von solchen mit *administrativen Zielen* durchzuführen (z.B. Meyer, 1991 bzw. S. 146). Letztere bilden eine sehr heterogene Gruppe an Zielen, die personalpolitische Entscheidungen der Beförderung und Karriereplanung genauso umfassen wie die Festlegung von Gehaltsanteilen oder die Ausschüttung von Prämien. Solche Entscheidungen basieren letztlich auf einem interindividuellen Vergleich der Führungsbeurteilungen, während entwicklungsorientierte Ratings lediglich den Beurteilten betrachten. Brutus und Derayeh (2002) plädieren für ein phasenweises Vorgehen: Erst wenn multiperspektivische Ratings mit Entwicklungszweck akzeptiert sind, sollten sie für administrative Ziele eingesetzt werden.

Multiperspektivische Beurteilungen können aber nicht nur dazu eingesetzt werden, den Entwicklungsbedarf zu messen, sondern auch dazu, eingetretene Entwicklungen (beispielsweise nach entsprechenden Maßnahmen) festzuhalten (für einen Überblick Martineau, 1998). Weiter werden sie in der Führungsforschung häufig zur Messung der unabhängigen Variable oder für Forschungsarbeiten im Grundlagenbereich verwendet, beispielsweise zur Ermittlung der Gütekriterien von Ratings (z.B. Hater & Bass, 1988; Scullen, Mount & Goff, 2000).

Dalton (1998) nennt fünf Gründe ganz anderer – konkret praktischer – Art für multiperspektivische Beurteilung von Führungskräften. Man könne dadurch den Bedürfnissen von strategisch wichtigen Mitarbeitern (high potentials) entgegen kommen, das Selbstmanagement von Führungskräften stärken, alle Führungskräfte zumindest bis zu einem Mindeststandard entwickeln, zu einer angestrebten Kulturänderung beitragen (hin zu einer Lernkultur, auch O'Connor-Wilson, McCauley & Kelly-Radford, 1998) und die Normen im engen Arbeitsumfeld der Feedbackempfänger und –geber verändern. Weitere Argumente finden sich beispielsweise bei Scherm und Sarges (2002). Generell gilt, dass Unternehmen mit Beurteilungen meist mehrere Ziele verfolgen anstatt auf ein Ziel zu fokussieren (Cleveland, Murphy & Williams, 1989).

3.1.4 Verbreitung

> "Perhaps the fact that multi-source feedback takes place at all, and that it has been adopted so widely and so quickly, is an indicator of the very considerable shift in organisational culture that has already taken place – feedback systems are more a symptom or expression of change than a factor in its causation."
> (Fletcher & Baldry, 2001, S. 136)

Unter anderem aus den oben genannten Gründen erlebt die Vorgesetztenbeurteilung seit den 90er Jahren einen wahrhaften Boom (Hedge, Borman & Birkeland, 2001). Sie lässt sich im generellen Aufschwung von Mitarbeiterbefragungen (z.B. Borg, 2000, 2001) und

Total Quality Management sowie von Beurteilungen[69] und der Tradition von Mehrfachbewertungen in Assessment-Centern verankern. Die Befragten werden als interne beziehungsweise externe Kunden betrachtet, auf deren Bedürfnisse eingegangen werden muss, um von ihnen Kundenbindung etc. erwarten zu können. 1989 nennt Schuler in seinem Enzyklopädie-Kapitel die drei Beurteilergruppen Vorgesetzte, Kollegen und Selbst, während die Aufwärtsbeurteilung durch Mitarbeiter noch nicht einmal erwähnt ist. Im Jahre 1996 gaben jedoch bereits circa 75% der 100 umsatzstärksten deutschen Unternehmen an, Mitarbeiterbefragungen zum Thema Vorgesetztenverhalten durchzuführen (Bungard, Fettel & Jöns, 1997). Multiperspektivische Bewertung ist inzwischen weit verbreitet (Edwards & Ewen, 2000; London & Smither, 1995) und wird von Jahr zu Jahr häufiger eingesetzt (Brutus & Derayeh, 2002). Mit überwiegender Mehrzahl wird sie erstens in großen Unternehmen, zweitens mit dem Ziel der Entwicklung (versus Verwaltung/Gehalt), drittens für die Zielgruppe der Führungskräfte, viertens von externen Beratern, fünftens mit Papier und Bleistift und sechstens im jährlichen Zyklus durchgeführt (Brutus & Derayeh, 2002: Befragung kanadischer Unternehmen).

Mit der Annahme, dass Feedback aus verschiedenen Perspektiven valider ist als nur aus einer Perspektive (Conway, Lombardo & Sanders, 2001), geht auch der Befund von Bernardin, Dahmus und Redmon (1993) plausibel einher: Demnach haben erstens diejenigen Führungskräfte mehr Bedenken, dass Mitarbeiterratings ihre Autorität untergraben könnten, die ausschließlich von Mitarbeitern (nicht von Vorgesetzten) Feedback erhalten. Zweitens sind Führungskräfte den Mitarbeiterratings gegenüber auch insgesamt positiver eingestellt, wenn sie von oben und unten anstatt nur von oben Feedback erhalten. Beide Perspektiven aufzunehmen erhöht also nicht nur die Validität, sondern auch die Akzeptanz der Rückmeldung durch die Führungskräfte, was sich auf eine darauf folgende Zielvereinbarung positiv auswirken sollte.

Letztlich hat sich die einstige Modeerscheinung fast zu einem Standard innerhalb des Human Resource Managements entwickelt. Unter welchen Bedingungen dies gerechtfertigt ist, wird aus den folgenden Kapiteln zu ersehen sein. Zu kritisieren ist der derzeitige, kaum hinterfragte, geschweige denn überprüfte Umgang mit multiperspektivischer Beurteilung und anschließendem Feedback. Der Gefahr, dass dies unbegründet zu einem Selbstläufer wird, muss die Forschung durch das Hervorbringen besserer Erkenntnisse entgegentreten.

[69] Folgendes Zitat kann als Tenor einer Strömung verstanden werden, die im Sinne des Tichy-Modells Beurteilungen als zentrales Element jeglichen Human Resource Managements betrachtet: „Es kommt nicht darauf an, was beurteilt wird, sondern dass beurteilt wird.“ (Neuberger, 2000, S. 2)

3.2 Interperspektivische Passung der Beurteilungen von beruflicher Leistung

"The greatest magnifying glasses in the world are a man's own eyes
when they look upon his own person."
(Alexander Pope, 1688-1744,
zit. nach Sedikides & Strube, 1997)

Was aber findet man vor in Unternehmen, wenn Selbst- und Fremdbilder miteinander verglichen werden? Man kann sagen: viele Wirklichkeiten – je nach Perspektive und individuellen Motiven fällt die Bewertung des Leistungsverhaltens einer Person anders aus.

3.2.1 Niveau-Unterschiede

Führungskräfte beschreiben sich selbst (unter anderem ihr Führungsverhalten) im Durchschnitt deutlich *besser* als sie von anderen beschrieben werden (vgl. Tabelle 10). Das bedeutet nicht automatisch, dass die Selbstbeschreibung immer fehlerhaft und die Fremdbeschreibung immer richtig ist – oder umgekehrt. Es handelt sich zunächst um verschiedene Sichtweisen des gleichen Verhaltens. Dies berücksichtigend werden derartige Niveau-Unterschiede in der vorliegenden Arbeit der Einfachheit zuliebe dennoch als „Überschätzung“ bezeichnet.

Verschiedene Meta-Analysen haben das Ausmaß der Diskrepanz untersucht. Es muss darauf verwiesen werden, dass in alle Meta-Analysen Studien eingehen, die die multiperspektivische Beurteilung *verschiedenster* Verhaltensbereiche erfassen. So sind die Fokuspersonen der in die Meta-Analysen eingehenden Studien nicht nur Führungskräfte aus Industrie, Dienstleistung und Militär, sondern auch technische Angestellte, Arbeiter, Pfarrer, Lehrer und andere (vgl. Tabelle 10). Mabe und West (1982) legen sich nicht einmal auf berufliche Leistung fest, sondern integrieren beispielsweise auch schulische Leistung oder im Labor provoziertes Leistungsverhalten unterschiedlicher Art. Das bedeutet, dass die Analysen zwar die für die vorliegende Studie relevante Stichprobe der Führungskräfte umfassen, sich aber bei weitem nicht auf sie beschränken. Wo die Ergebnisdarstellung der Meta-Analysen eine getrennte Betrachtung erlaubt, wird das Management im vorliegenden Kapitel gesondert dargestellt.

Tabelle 10. Übereinstimmung von Selbst- und Fremdratings (Abwärts, Aufwärts) zu (beruflicher) Leistung: Meta-Analytische Befunde

	Mabe & West, 1982[a]	Harris & Schaubroeck, 1988	Conway & Huffcutt, 1997	Heidemeier & Moser, 2002
ρ Selbst – Ab	.29 k = 55	.22 (.35) k = 36	.22 (.31) k = 50	.22 (.35) k = 76
	(.08)[b] k unbekannt[c]	.27[b] k = 11	.19 (.29)[b] k = 19	.19 (.32)[b] k = 21
ρ Selbst – Auf	-	-	.14 (.26) k = 50	-
ρ Ab – Auf	-	-	.22 (.57) k = 22	-
d Selbst – Ab	-	.70 k = 18	-	.35 k = 54
				.34[b] k = 17
Stichprobe	81% Studierende (also nicht nur berufliche Leistung!), Rest: u.a. Management	verschiedene Bereiche, z.B. Management, Service.	verschiedene Dimensionen der Arbeitsleistung	Feldstudien zu Arbeitsleistung in aktuellem Berufsverhältnis, verschiedene Sektoren, 72% Industrie
Außenkriterien	verschiedene, z.B. Abwärtsratings (16%) oder objektive Tests (50%)	Abwärts- und Peerratings[d]	Abwärts-, Aufwärts- und Peerratings[d]	Abwärtsratings

Anmerkungen.
k = Anzahl der unabhängigen Stichproben
In Klammern sind die korrigierten Indizes angegeben. Bei allen korrigierten Indizes wird die Reliabilität der Ratings berücksichtigt. Bei Harris und Schaubroeck (1988) wurden sie zusätzlich um die Einschränkung des Ranges bei Selbstbeurteilungen (aufgrund der Mildetendenz) korrigiert.
[a] Die berichteten Korrelationen beschreiben die Zusammenhänge zwischen der Selbstbeschreibung und Außenkriterien, die nur zu 16% Abwärtsbeurteilungen darstellen (vgl. entsprechende Spalte der Tabelle).
[b] Ergebnis für Teilstichprobe "Management/Führung". Bei allen Angaben, die Aufwärtsbeurteilungen enthalten, ist naturgemäß die Beurteilung von Führungskräften betroffen, da andere Personenstichproben keine Mitarbeiter haben.
[c] Da nur 4 der 55 Studien der Leistungskategorie Management/Führung im Feld zuzuordnen sind, gingen in diese Analyse maximal vier Studien ein.
[d] Die Ergebnisse der Peerratings sind in der Tabelle nicht berücksichtigt.

Da in der vorliegenden Studie lediglich das *Führungsverhalten* der Führungskräfte beurteilt wird, ist wissenswert, dass sich die Beurteilung von deren gesamtem Leistungsverhalten (Management) kaum von der Beurteilung speziell ihres Führungsverhaltens unterscheidet (Conway & Huffcutt, 1997, S. 338).[70] Aufgrund ihrer Komplexität nehmen

[70] Eine klare Trennung von Management und Personalführung ist ohnehin ein schwieriges Unterfangen (vgl. Kapitel 2.5.2.2 und 2.6), das aufgrund der ähnlichen Befunde bei Conway und Huffcutt (1997) in der vorliegenden Arbeit nicht weiter verfolgt wird.

Managementtätigkeiten - im Vergleich zu anderen Tätigkeiten - eine Sonderrolle ein, die sich auch als Moderatorvariable empirisch zeigt (vgl. Kapitel 3.3.2). Um einen Überblick über die aktuelle Ergebnislage zu geben, werden auf den folgenden Seiten deshalb sowohl Ergebnisse der Meta-Analysen (vgl. auch Tabelle 10) als auch Ergebnisse von Einzelstudien, die die multiperspektivische Beurteilung von Management/Führungsverhalten untersuchen, dargestellt.

Besonders hoch erweisen sich die Niveau-Unterschiede zwischen Selbst- und Fremdbildern: So fällt die Selbstbeurteilung deutlich besser aus als die Abwärtsbeurteilung (d = .70 bei Harris & Schaubroeck, 1988; d = .35 bei Heidemeier & Moser, 2002). Heidemeier und Moser (2002) geben für die Management-Stichprobe einen Unterschied von einer Drittel Standardabweichung an. Weniger überhöht erscheint das Selbstbild im Vergleich zur Kollegenbeschreibung (d = .28 bei Harris & Schaubroeck, 1988).[71] Da in keine Meta-Analyse Aufwärtsbeschreibungen einbezogen wurden, muss für entsprechende Aussagen gänzlich auf andere Studien zurückgegriffen werden. Mount und Scullen (2001) zitieren zwei nicht publizierte Studien, die an großen Manager-Stichproben alle vier Perspektiven hinsichtlich der Niveau-Unterschiede verglichen, und halten fest, dass sich Aufwärtsbeschreibungen generell kaum im Niveau von Kollegen- und Vorgesetztenbeschreibungen unterscheiden (auf einzelnen Dimensionen leicht über oder unter diesen liegen) und sich folglich ebenfalls deutlich von Selbstratings unterscheiden. Wird zur Führungsbeschreibung der MLQ (vgl. S. 107) herangezogen, dann liegt die Selbstbeschreibung um etwa einen halben Skalenpunkt (bei einer 5er Skala) höher als die Aufwärtsbeschreibung (Atwater & Yammarino, 1992; Bass & Yammarino, 1991). Insgesamt fallen Selbstbeschreibungen also deutlich besser aus als Fremdbeschreibungen von Vorgesetzten, Kollegen und Mitarbeitern, während sich die letzten drei kaum unterscheiden (Fleenor, McCauley & Brutus, 1996). Auch im Vergleich zu externen Beratern überschätzen Manager ihre Management-Fähigkeiten (Furnham & Stringfield, 1998).

Der wiederholt berichtete diskrepante Befund für den Selbst-Fremd-Vergleich wäre weniger bedeutend, wenn er durch Addition einer Konstanten aufgelöst werden könnte. Dann würden die Profile der Selbst- und Fremdbeschreibung trotz der Niveau-Unterschiede übereinstimmen, also hoch korrelieren. Die Mittelwertsdifferenzen könnten beispielsweise mit einem Ankereffekt (Tversky & Kahneman, 1974) erklärt werden. Doch Führungskräfte beschreiben das Führungsverhalten *anders* als Außenstehende (vgl. Tabelle 10 und Kapitel 3.2.2).

[71] Einschränkend anzumerken ist, dass die Meta-Analyse lediglich 18 Studien für die Niveau-Unterschiede zwischen Selbst und Abwärts beziehungsweise vier Studien für die Niveau-Unterschiede zwischen Selbst und Peers einbeziehen konnte. Die Niveau-Unterschiede erregten insgesamt weniger Forschungsinteresse als die Korrelationen, zu denen aktuellere Meta-Analysen mit größeren Stichproben durchgeführt wurden (vgl. Kapitel 3.2.2).

Nicht zu vergessen ist selbstverständlich, dass es auch Führungskräfte gibt, die sich in Übereinstimmung mit Fremdbeurteilern befinden oder sich negativer als diese bewerten. Tornow (1993) berichtet in seinem Diskussionspapier folgende Zahlen: Nur zehn Prozent der Manager schätzen sich im Einklang mit ihren Mitarbeitern ein, die anderen weichen mindestens um eine halbe Standardabweichung vom Fremdbild ab. Circa zwei Drittel der diskrepanten Einschätzungen sind dergestalt, dass sich die Führungskräfte besser beschreiben als ihr Fremdbild. Erstaunlich sind andere Zahlen, die annähernd auf eine Drittelung in Unterschätzer, Übereinstimmer und Überschätzer (zur Kategorisierung vgl. Fußnote 112) hinweisen (Atwater, Roush & Fischthal, 1995 [Studenten der Marineakademie]; Van Velsor, Taylor & Leslie, 1993 [Manager]). Möglicherweise ist ein Teil dieses offensichtlichen Widerspruchs dadurch aufzuklären, dass das Ausmaß der Überschätzung das der Unterschätzung deutlich übersteigt, so dass im Durchschnitt eine Überschätzung resultiert.

Wichtig sind erste Hinweise darauf, dass die Diskrepanz nicht durch eine Perspektive allein verursacht wird, sondern sowohl durch die Selbst- als auch Fremdperspektive. Bei einer hohen Diskrepanz scheint also nicht allein die Selbstperspektive nach oben zu tendieren, sondern auch die Fremdperspektive nach unten. Für die Unterschätzer zeichnet sich ein überraschendes Bild ab: Sie schätzen sich selbst im Vergleich zu anderen am schlechtesten ein, werden von anderen jedoch am besten beurteilt (Atwater & Yammarino, 1992; Van Velsor, Taylor & Leslie, 1993).

Wie die Selbst-Fremd-Kongruenz bei der Beurteilung von Führungsdimensionen im Vergleich zu der bei der Beurteilung anderer Leistungsbereiche einzuordnen ist, wird am Ende des folgenden Kapitels dargestellt.

3.2.2 Korrelativer Zusammenhang

Auch auf korrelativer Ebene finden sich nur geringe Zusammenhänge zwischen dem Selbstbild der Führungskraft und den Fremdbildern (Tabelle 10). Für allgemeine berufliche Leistungsbeurteilungen schätzen die Meta-Analysen folgende Zusammenhänge mit dem Abwärts-/Kollegenbild: ρ = .31/.31 (Conway & Huffcutt, 1997), ρ = .35/.36 (Harris & Schaubroeck, 1988) beziehungsweise ρ = .35/-[72] (Heidemeier & Moser, 2002). Noch geringere Korrelationen ergeben sich für die Substichprobe der Führungskräfte: ρ = .29 (Conway & Huffcutt, 1997) beziehungsweise ρ = .08 (Mabe & West, 1982; nicht nur Abwärtsbeurteilung, sondern verschiedene Außenkriterien). Von allen 12 Leistungsbereichen, die Mabe und West (1982) untersuchten, fallen die Korrelationen zwischen Selbstbewertung und Kriterium im Management-Bereich damit am geringsten aus.

[72] - bedeutet: nicht berechnet.

In der großen Management-Stichprobe (N = 2.056) von Fleenor, McCauley und Brutus (1996) korreliert die Selbstbeschreibung der Management-Kompetenzen mit der Abwärtsbeschreibung etwas geringer (r = .19) als mit der Aufwärtsbeschreibung (r = .24), wobei festzuhalten ist, dass die Ergebnisse für Selbst-Abwärts-Zusammenhänge stark schwanken und in der Regel höher liegen als die für Selbst-Aufwärts-Zusammenhänge. Diese erweisen sich im Unterschied zu den Selbst-Abwärts-Zusammenhängen als recht stabil. Auch Atwater, Ostroff, Yammarino und Fleenor (1998) finden für die Führungsbeschreibung von 1.326 Managern eine Korrelation von r = .25 zwischen Selbst- und Aufwärtsbeurteilung. In diese Größenordnung reihen sich andere Ergebnisse widerspruchslos ein: Selbst-Aufwärts (z.B. Atwater & Yammarino, 1992 [r = .19 für Marinestudenten bzw. r = .12 für Marineoffiziere]; Baril, Ayman & Palmiter, 1994 [r = .16 bis .34 für unterschiedliche Beurteilungsdimensionen]; Church, 1997a [r = .25]; London & Wohlers, 1991 [r = .24]). Ebenso verhält es sich in der – über den Führungsbereich hinausreichenden – Meta-Analyse von Conway und Huffcutt (1997; ρ = .14, korrigiert ρ = .26). Im Unterschied zu Abwärts- und Aufwärtsratings korreliert die Selbstbeurteilung auch mit der Beurteilung durch externe Kunden bzw. Berater sehr gering (Church, 1997b bzw. Furnham & Stringfield, 1998).

Insgesamt gilt, dass die Korrelationen steigen, wenn das Mittel der Fremdbeurteilungen (auch über die Perspektiven hinweg) anstatt individuelle Werte in die Korrelation mit der Selbstbewertung eingehen (Wohlers & London, 1989). Außerdem bestehen dann höhere korrelative Zusammenhänge, wenn Fremdbilder miteinander verglichen werden und das Selbstbild außen vor bleibt. Diesen Befund erbringen sowohl die Meta-Analysen für die Führungsbeurteilung von Führungskräften (Conway & Huffcutt, 1997/Harris & Schaubroeck, 1988: Aufwärts-Abwärts: ρ = .57/-, Aufwärts-Peer: ρ = .66/-, Abwärts-Peer: ρ = .79/.62) als auch Einzelstudien (Fleenor et al. [1996]: r = .41 für Aufwärts-Abwärts-Korrelation bzw. r = .24/.19 für Selbst-Aufwärts/Abwärts-Korrelation; Furnham & Stringfield [1998]: r = .58 als mittlere Korrelation aus 72 Fremd-Fremd-Korrelationen bzw. r = .13 aus Selbst-Fremd-Korrelationen[73]).

Das kann als Zeichen dafür gewertet werden, dass die Selbstbewertung einen großen Anteil an der mangelhaften Übereinstimmung trägt. Außerdem gibt dies einen Hinweis darauf, dass Fremdbeobachter Ähnlichkeiten zueinander aufweisen, dass aber die Selbstwahrnehmung der Führungskräfte nahezu unabhängig von den Fremdurteilen ist. Unklar bleibt, ob die Ähnlichkeit der Fremdrater nicht nur darin besteht, ähnlichen Stereotypen von Führung zu „erliegen“. Aufgrund stereotyp-konformer Ratingantworten könnten diese dann hoch miteinander, nicht jedoch mit der Selbstbeurteilung der Führungskraft korrelieren. Dafür sprechen Ansätze wie der der „romantisierenden Führung“ von Meindl (1993,

[73] Die Fremdbeurteilungen verteilen sich auf andere Manager, Peers und Berater, wobei Peers mit den Beratern am geringsten korrelieren, während sich andere Fremd-Fremd-Korrelationen nicht unterscheiden.

1998, vgl. S. 21). Danach könnte Führung von Dritten sehr ähnlich beschrieben werden, selbst wenn die zu beschreibenden Führungspersonen oder –dimensionen nahezu unbekannt sind. Gegen eine stereotyp-bedingte Korrelation spricht allerdings, dass Fremdbeurteilungen und Selbst-Fremd-Diskrepanzen höher mit Erfolgskriterien korrelieren als Selbstbeurteilungen (vgl. Kapitel 3.1.1 und 3.5.1).

Verschiedene Studien zeigen, dass Selbstbeurteilungen kaum mit anderen Außenkriterien, z.B. mit unternehmensinterner Beförderung korrelieren (z.B. Bass & Yammarino, 1991; Mabe & West, 1982). Daraus folgt für die Praxis, dass Führungskräfte durch die eigene Einschätzung ihrer Führung erstens keinen erfolgsrelevanten Indikator gewinnen und zweitens keinen Hinweis auf die Einschätzung anderer gewinnen und damit auch auf indirektem Weg keinen Hinweis auf deren Leistungsbereitschaft, Wohlbefinden oder ähnliches (vgl. Tabelle 9). Konsequenterweise müsste also jedes Unternehmen danach streben, Fremdbeurteilungen als wichtige Indikatoren für Erfolg heranzuziehen.

Generell wurde in der Forschung deutlich mehr Wert auf korrelative Zusammenhänge zwischen den Perspektiven als auf Niveau-Unterschied gelegt. Methodische Fragestellungen, insbesondere zur Reliabilität von Selbstbeurteilungen, standen dabei im Vordergrund. Dies gilt nicht nur für berufliche Leistungsbeurteilung und Beurteilung von Führung im Allgemeinen, sondern auch für spezifische Anforderungen, die an Führungskräfte gestellt werden (z.B. eigenverantwortliches Handeln, Koch, 2001).

Es stellt sich die Frage, wie stabil solch diskrepante Beurteilungen über die Zeit hinweg sind. Hier stehen die Studien, die Veränderungsmöglichkeiten durch Feedback aufzeigen (vgl. Kapitel 3.6) denjenigen gegenüber, die eine mittlere Retest-Reliabilität von bis zu r = .81 erbringen (Nilsen & Campbell, 1993; Smither, London, Vasilopoulos, Reilly, Millsap & Salvemini, 1995). Damit liegt der Stabilitätsindex für Führungsdifferenzen durchschnittlich nur leicht unter dem für Abwärtsbeschreibungen von genereller Arbeitsleistung (ρ = .81, Meta-Analyse Viswesvaran, Ones & Schmidt, 1996). Dies erstaunt, wenn man die Kritik an der Reliabilität von Differenzscores ernst nimmt (Johns, 1981, vgl. Kapitel 5.2.1).

Geringe interperspektivische Korrelationen, aber auch hohe Niveau-Unterschiede (Selbst > Fremd) zwischen Selbst- und Fremdbeschreibungen wurden neben dem Führungsbereich in vielen anderen Bereichen gezeigt. Dazu gehören die Beurteilung von Leistungen in Assessment-Centern (Drees, 1994; Stempfle, Hagmayer, Hübner, Iwanoff & Kaufmann, 2004: geringe Korrelation zwischen Selbst- und Beobachter-Beurteilung), klerikale Arbeit (z.B. Parker, Taylor, Barrett & Martens, 1959: hohe Niveau-Unterschiede und geringe Korrelationen), Pflegefertigkeiten (z.B. Klimoski & London, 1974: geringe Selbst-

Fremd-Korrelationen) oder Depressivität und Ängstlichkeit (Okazaki, 2002: von nahestehenden Personen unterschätzt und lediglich mittlere Selbst-Fremd-Korrelationen).[74]

Hervorgehoben werden soll die Studie von Malloy, Yarlas, Montvilo und Sugarman (1996), auch wenn sie sich selbst eher dem Forschungsbereich psychosozialer Entwicklung von Kindern verschreibt. Hier zeigt sich, dass für die Beurteilung verschiedener Verhaltens-, Leistungs- und Persönlichkeitsdimensionen (z.B. Klassenverhalten, Lesefertigkeit, Attraktivität) von Schülern der ersten bis sechsten Jahrgangsstufe die Selbst-Fremd-Korrelationen ($\bar{r}$ = .20 bzw. $\bar{r}$ = .33) deutlich geringer ausfallen als die Fremd-Fremd-Korrelationen ($\bar{r}$ = .70). Fremdbeurteilungen wurden von Lehrern und Peers eingeholt. Genauso wie es für die berufliche Leistung berichtet wurde, korreliert die Selbsteinschätzung in anderen Bereichen nur gering mit objektiven Leistungsparametern (z.B. Bailey & Bailey, 1974 bezüglich Schulleistungen).

Doch auch das als sicher geglaubte Wissen um die Normalität der „Selbstüberschätzung“ wurde durch eine Meta-Analyse erschüttert, die regelrecht davor warnt, Selbstbeurteilungen grundsätzlich als inflationär abzustrafen. So finden Crampton und Wagner (1994), dass die Selbstbeurteilungen nur in 38 von 143 untersuchten Fällen statistisch bedeutsam positiver ausfallen als andere (multiperspektivische) Beurteilungen (t-Test, $p <$ .05). Da sie Befunde aus den verschiedensten Bereichen der Organisationsforschung einbezogen, plädieren sie dafür, bereichsspezifische Untersuchungen durchzuführen, um letztlich die Bereiche zu identifizieren, die besonders anfällig für eine Erhöhung der Selbstbeurteilungen sind.

Lässt man die vereinzelten Hinweise auf eine Drittelung der Führungskräfte in Unterschätzer, Übereinstimmer und Überschätzer außer Acht (vgl. S. 128), dann ist aus den berichteten Ergebnissen von Crampton und Wagner (1994) zu schließen, dass im Führungsbereich mit verhältnismäßig milden Selbstratings zu rechnen ist. Auch wenn man einen Vergleich zwischen den oben berichteten Lebensbereichen und dem Führungsbereich zieht, sticht letztgenannter negativ hervor. Hier scheinen die Selbst-Fremd-Korrelationen nämlich besonders gering auszufallen (z.B. Mabe & West, 1982). Ob bzw. welche Konsequenzen dies mit sich bringt, wird in Kapitel 3.5 diskutiert.

[74] Nur in Randgruppen scheint sich diese Tendenz zu hohen Niveau-Unterschieden und geringen korrelativen Zusammenhängen umzukehren. So bewerten – störungskonform und deshalb wenig überraschend – soziale Phobiker ihre Leistung, eine kurze Rede zu halten, im Vergleich zu Fremdratings negativer (Rapee & Lim, 1992).

3.2.3 Above-Average-Effekt

Stellt man nicht die Einschätzungen von verschiedenen Personen einander gegenüber, sondern bittet man die Probanden, sich relativ zum Durchschnitt selbst zu beschreiben, dann zeigt sich – wiederum in den verschiedensten Bereichen – ähnliches: Die große Mehrzahl der Befragten gibt an, glücklicher (Freedman, 1978) und intelligenter (Larwood & Whittaker, 1977) als die meisten anderen zu sein, besser, sicherer und langsamer Auto zu fahren als der Durchschnitt (McCormick, Walkey & Green, 1986; Svenson, 1981; Walton & Bathurst, 1998), mit höherer Wahrscheinlichkeit positive Ereignisse (z.B. guten Arbeitsplatz bekommen) und mit geringerer Wahrscheinlichkeit negative Ereignisse (z.B. aggressives Verhalten oder Kontrollverlust nach Alkoholgenuss, Alkoholismusproblem) (Paglia & Room, 1999; Weinstein, 1980; für eine Übersicht Taylor & Brown, 1988) oder Krankheiten (Perloff & Fetzer, 1986) zu erleben. Insgesamt konnte dieser so genannte "Above-Average-Effekt" (auch: "better-than-average effect") sowohl für Eigenschaftsratings (siehe Intelligenz oder auch Dunning, Meyerowitz & Holtzberg, 1989 [verschiedene Traits]) als auch Verhaltensratings (siehe Autofahren oder Allison, Messick & Goethals, 1989 [gute/schlechte Verhaltensweisen]; Messick, Bloom, Boldizar & Samuelson, 1985 [faire/unfaire Verhaltensweisen]) gezeigt werden. Was berufliche Dimensionen betrifft, schätzen sich 94% der College-Lehrer so ein, als leisteten sie überdurchschnittlich gute Lehre (Cross, 1977), als hätten sie bessere Führungsfertigkeiten und soziale Kompetenzen (College Board, 1976/77, zitiert nach Dauenheimer, Stahlberg, Frey & Petersen, 2002) und als seien sie überdurchschnittlich gut für Management-Tätigkeiten geeignet (Larwood, 1978). Sowohl Management-Studenten im Planspiel als auch Vorstandsvorsitzende von Produktionsbetrieben schreiben ihrem eigenen Unternehmen überdurchschnittliche Wachstumsraten zu (Larwood & Whittaker, 1977). In der inzwischen legendär gewordenen Studie von Meyer (1980) schätzten Führungskräfte ihre Leistung durchschnittlich besser ein als die von 78% der anderen mit vergleichbaren Aufgaben. Nur 2 von 92 Teilnehmern schätzten sich (leicht) unterdurchschnittlich ein. Generell zählen sich mindestens 40% der Angestellten aller Art bezüglich ihrer Arbeitsleistung zu den "Top Ten". Von einer Management- und Experten-Stichprobe (hohe Hierarchieebenen) ordnen sich sogar 80% dieser Kategorie zu (Meyer, 1980). Der Above-Average-Effekt ist inzwischen also für einen breiten Bereich an inhaltlichen Dimensionen bestätigt und keineswegs nur auf den Führungsbereich beschränkt. Für diesen zeichnet sich aber ein extrem starker Effekt ab.

Um den Above-Average-Effekt zu erklären, kann man zwei prinzipiell unterschiedliche Mechanismen heranziehen. Folgt man einem motivationalen Mechanismus, dann konzeptualisiert man die überdurchschnittlich gute Selbstbeschreibung so, als diene sie der Befriedigung der Bedürfnisse nach Selbstwerterhöhung und –schutz (vgl. Kapitel 3.4.2.3). Dabei lassen vage Vergleichsgruppen mehr Spielraum für selbstwertdienliche Abwärtsvergleiche als konkrete Vergleichspersonen (Perloff & Fetzer, 1986). Dem gegenüber kann

der Effekt auf kognitivem Wege über die Verfügbarkeitsheuristik erklärt werden, da selbstbezogene negative Ereignisse weniger verfügbar sind (vgl. Kapitel 3.4.2.1). Gegen eine ausschließliche Verfügbarkeits-Erklärung spricht jedoch, dass der Above-Average-Effekt für positive Dimensionen (wie sie auch in der vorliegenden Studie betrachtet werden) stärker ausfällt als in umgekehrter Ausprägung für negative (z.B. Alicke, Klotz, Breitenbecher, Yurak & Vredenburg, 1995; vgl. auch Asymmetrie bei Ingroup-Favorisierung, S. 139). Für eine detaillierte Abhandlung potentieller Erklärungsansätze wird auf Kapitel 3.4 verwiesen, in dem der Above-Average-Effekt hinsichtlich einzelner Erklärungsmechanismen weiter diskutiert wird.

In einer Reihe von sieben Laborexperimenten (studentische Stichproben) zeigte sich, dass der Effekt reduziert werden kann, indem der Vergleich nicht zum durchschnittlichen Studierenden, sondern zu einem individuellen Studierenden gezogen wird. Eine weitere Reduktion gelingt dann, wenn zu diesem persönlicher Kontakt besteht (Alicke, Klotz, Breitenbecher, Yurak & Vredenburg, 1995). Der Effekt scheint also vom Abstraktionsniveau der Vergleichsperson(en) abzuhängen: Je vager die Vergleichsperson oder –gruppe charakterisiert ist, umso mehr Möglichkeiten bleiben, um selbsterhöhende Information abzurufen. Auch das Bedürfnis, sich von einer unbekannten Durchschnittsperson positiv abzuheben, sollte größer sein als das Bedürfnis, sich gegenüber einer Person, zu der Kontakt besteht, abzuheben. Mit dieser teilt man dann nämlich die Zugehörigkeit zu mindestens einer sozialen Gruppe.

Man kann also aus empirischen Befunden festhalten,

- dass Führungskräfte ihr eigenes Leistungsverhalten im Durchschnitt besser einschätzen als dies von anderen eingeschätzt wird (Niveau-Unterschiede),
- dass die Mehrzahl der Führungskräfte ihr eigenes Leistungsverhalten besser einschätzt als das anderer (Above-Average-Effekt),
- dass ihre Selbstbeschreibungen weniger mit Fremdbeschreibungen korrelieren als diese untereinander und
- dass Fremdbeschreibungen stärker mit (organisationalen) Erfolgskriterien korrelieren als Selbstbeschreibungen.

Zu den einzelnen Perspektiven ist zusammenfassend festzuhalten, dass die Selbstbeurteilung die geringsten Korrelationen zu anderen Perspektiven aufweist. Möglicherweise ist sie deshalb am intensivsten untersucht worden, was sich daran zeigt, dass sie in fast alle multiperspektivischen Beurteilungsstudien integriert ist. So konnte sie auch in alle vier beschriebenen Meta-Analysen aufgenommen werden. In der Praxis bewertete man lange Zeit die Abwärtsbeurteilung als Perspektive mit höchster Relevanz. Da dies die akzeptierteste Form der Leistungsbeurteilung darstellte und darstellt, wurde auch sie in vielen Studien

und letztlich ebenso in allen vier Meta-Analysen berücksichtigt. Im Unterschied dazu setzte sich die Aufwärtsbeurteilung erst in den vergangenen Jahren verstärkt durch.

Obwohl alle Perspektiven für sich genommen – mehr (z.B. Abwärtsbeurteilung) oder weniger (Selbstbeurteilung) – valide sind, gewinnt die multiperspektivische Leistungsbeschreibung deutlich durch die inkrementelle Validität der verschiedenen Perspektiven. Die Validität wird dadurch ermöglicht, dass die Perspektiven untereinander nur gering korrelieren und folglich wenig redundante, sondern spezifische, kriteriumsrelevante Information beisteuern (Meta-Analyse von Conway, Lombardo & Sanders, 2001; Scullen, Mount & Goff, 2000).

3.3 Moderatorvariablen für die Passung zwischen Selbst- und Fremdbild

Nachdem gerade die *geringe* Übereinstimmung zwischen Selbst- und Fremdbildern oben herausgestellt wurde, stellt sich die Frage, unter welchen Umständen entweder die korrelative Übereinstimmung zwischen zwei Perspektiven steigen beziehungsweise die Niveau-Unterschiede zwischen diesen sinken kann. Die Frage zielt also auf potentielle Moderatorvariablen. Da erst wenige Studien durchgeführt wurden, die auf solche moderierende Variablen zur Güte der Selbstbeschreibungen des *Führungsverhaltens* fokussieren, werden zunächst auch für diesen Forschungsbereich alle Variablen beschrieben, die bereits in Meta-Analysen zur Selbsteinschätzung von Leistung insgesamt untersucht wurden. Wo möglich, wird auch hier getrennt auf Ergebnisse für die Ratee-Stichprobe der Manager eingegangen. Die meta-analytischen Ergebnisse sind in Tabelle 11 überblicksartig zusammengestellt, bevor sie in den Kapiteln 3.3.1 bis 3.3.3 genauer beschrieben werden. Diese drei Kapitel beschäftigen sich jeweils mit Moderatorvariablen aus einem der folgenden Bereiche (Einteilung in Anlehnung an Moser, Donat, Schuler, Funke & Roloff, 1994):

- personale Faktoren
- tätigkeitsbezogene und organisationale Faktoren
- Faktoren, die die Durchführungsbedingungen betreffen

Einzelstudien, die gezielt den Inhaltsbereich des Führungsverhaltens betrachten, werden immer dort angeführt, wo solche bekannt sind. Im Übrigen bieten die Ergebnisse ein Grundgerüst für den Transfer auf die Dimension "Führungsverhalten". Da Hinweise darauf bestehen, dass kein Zusammenhang zwischen den Niveau-Unterschieden und den korrelativen Zusammenhängen besteht (vgl. Moser, 1999), werden diese getrennt voneinander betrachtet. Dafür spricht auch die unterschiedliche Wirkung von Moderatorvariablen auf Niveau- und korrelative Übereinstimmung, wie sie Heidemeier und Moser (2002) finden. Sie sind bislang auch die einzigen, die in einer Meta-Analyse die Moderatorwirkung verschiedener Variablen nicht nur hinsichtlich des korrelativen Zusammenhangs zweier Perspektiven, sondern auch hinsichtlich der Niveau-Unterschiede überprüften. Allerdings wurde lediglich die Übereinstimmung zwischen Selbst- und Abwärtseinschätzung in die Analyse aufgenommen.

Tabelle 11. Meta-analytisch untersuchte Moderatorvariablen für die Übereinstimmung zwischen Selbstbeurteilung von Leistung und Außenkriterien (vgl. Tabelle 10)

	Mabe & West, 1982[b]	Harris & Schaubroeck, 1988	Conway & Huffcutt, 1997[c]	Heidemeier & Moser, 2002
signifikante Moderatorvariablen für *r* bzw. *D*[a]	▪ Anonymitätsinstruktion ▪ Validierungserwartung ▪ Bewertungserfahrung ▪ Relative Selbstbewertung im Vergleich zu anderen vs. absolutes Rating (z.B. Likertskala) ▪ Einschätzung vergangener vs. zukünftiger Leistung ▪ Leistung vs. Fähigkeit	▪ Tätigkeit: Management/ Experten vs. Arbeiter/Service[d]	▪ Tätigkeit: Management/ Offiziere vs. Nicht-Management ▪ Komplexität der Arbeitsaufgabe: hoch – mittel – gering	▪ Tätigkeit: Management vs. Arbeiter *nur für D:* ▪ Tätigkeit: Management/Experten vs. Arbeiter ▪ Zweck: Forschung vs. Entwicklung ▪ Validierungserwartung ▪ aufgabenbezogene vs. umfeldbezogene Leistung ▪ Kultur: westlich vs. asiatisch ▪ Verifizierbarkeit
nicht signifikante Moderatorvariablen für *r*	▪ Instruktion mit Bezug zu bestimmter Vergleichsgruppe: klare Nennung der Referenzgruppe ▪ Wissen über Fähigkeiten der Referenzgruppe ▪ hohe Passung von Bewertungsgegenstand und Kriterium	▪ dimensionales vs. globales Ratingformat ▪ trait- vs. verhaltensbezogener Ratinggegenstand		▪ Instruktion zum sozialen Vergleich ▪ trait- vs. verhaltensbezogen ▪ Einzelitem (global vs. dimensional) vs. Aggregation ▪ Anonymität ▪ Multiperspektivische Erfassung (>2 Fremdperspektiven) ▪ lange vs. kurze Organisationszugehörigkeit ▪ Bildungsniveau ▪ Geschlecht (weibliche Selbstbeurteiler)

Anmerkungen. [a] *r:* korrelativer Zusammenhang zwischen Selbstbeurteilung und Kriterium, *D:* Niveau-Unterschied zwischen Selbstbeurteilung und Abwärtsbeurteilung (nur letzte Spalte). [b] Die Ergebnisse beruhen darauf, ob die Moderatorvariablen signifikant mit der Korrelation (z-transformiert) aus Selbst*Kriterium korrelieren. [c] Da die Dimension interpersonelle vs. kognitive Tätigkeit nicht für Selbstbeurteilungen getestet wurde, erscheint sie nicht in der Tabelle. [d] Unterstrichen sind diejenigen Ausprägungen der Moderatorvariablen, die mit einer höheren Validität/geringeren Niveau-Unterschieden einhergehen (obere Zeile) beziehungsweise nach Hypothese einhergehen sollten (untere Zeile). Wo nur eine Ausprägung aufgeführt ist, wurde sie dem Gegenteil dieser Ausprägung gegenüber gestellt, z.B. Anonymitätsinstruktion vs. keine Anonymitätsinstruktion. Für die genannte Ausprägung wird eine höhere Kriteriumsvalidität angenommen.

Alle vier Meta-Analysen (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002; Mabe & West, 1982) untersuchen Moderatorvariablen für die Validität der Selbstbeurteilungen (Korrelation mit Außenkriterien/Fremdbeurteilungen), keine Meta-Analyse untersuchte bisher potentielle Moderatoreffekte auf die Fremd-Fremd-Übereinstimmung. Um die im Folgenden dargestellten Ergebnisse besser bewerten zu können, wird vorab auf einige Unterschiede hingewiesen: Während Mabe und West (1982) ein sehr breites Verständnis von Leistung sowie von Validierungskriterien zugrunde legen (Ratings aller Art sowie objektive Tests, Schulnoten und anderes), beschränken sich alle anderen auf berufliche Leistung beziehungsweise Fremdratings als Außenkriterien. Harris und Schaubroeck (1988) lassen Studien mit Abwärts- und Peerbeurteilungen zu. Studien mit Mitarbeiterratings kommen bei Conway und Huffcutt (1997) sowie in geringerem Umfang (16% der Kriterienmaße) auch bei Mabe und West (1982) hinzu. Dagegen arbeiten Heidemeier und Moser (2002), wie oben erwähnt, ausschließlich mit Abwärtsbeurteilungen (vgl. Tabelle 10). Auf diese Unterschiede können einige unerwartete Ergebnisse zurückgeführt werden. An den entsprechenden Stellen wird darauf Bezug genommen.

Insgesamt kann bezweifelt werden, ob die gängige Praxis, Fremdbeschreibungen als Außenkriterium für die Validierung von Selbstbeschreibungen heranzuziehen, zielführend ist. Letztlich werden verschiedene Rater ja gerade wegen ihrer unterschiedlichen Perspektiven engagiert, da man sich von diesen spezifische Information erwartet. Wenn in der vorliegenden Arbeit in diesem Sinne von „Validität“ gesprochen wird, dann geschieht das in Anlehnung an das Gros der Literatur.

3.3.1 Moderatorvariablen im personalen Bereich

Validität. Auch wenn Mabe und West (1982) aus verschiedenen Überlegungen das Resumée ziehen, dass hohe *Intelligenz*, hohes *Qualifikationsniveau* und interne *Kontrollüberzeugung* persönliche Variablen des Selbstraters sind, die die Validität von Selbstbeschreibungen erhöhen können,[75] verfolgen sie deren Moderatorwirkung in ihrer Meta-Analyse nicht weiter. Erst Heidemeier und Moser (2002) erfassen das *Bildungsniveau* der Selbstrater, finden jedoch weder für die Validität noch für den Niveau-Unterschied eine signifikante moderierende Wirkung.

[75] Erklärt wird dies mit der Annahme, dass intelligentere/fähigere Personen zu besserer Wahrnehmung der eigenen Leistung und ihrer Bedeutung im Vergleich zu anderen imstande seien. Entgegen dieser Annahme steht, dass Personen mit höherer Qualifikation eine kleinere Vergleichsgruppe haben, so dass sie sich aufgrund eines mangelhaften Maßstabs weniger akkurat beurteilen können.
Interne Kontrollüberzeugung ist mit erhöhtem Wunsch und erhöhter Anstrengung verbunden, Information über das eigenen Leistungsverhalten zu erhalten. Dies soll in der Folge zu genauerer Selbsteinschätzung führen.

Auch andere Persönlichkeitsvariablen der Fokusperson wurden in letzter Zeit als Einflussfaktoren für die Selbst- und Fremdbewertung sowie deren Übereinstimmung untersucht (Brutus et al., 1998). Denkbar ist beispielsweise, dass das Verhalten einer extravertierten Person valider eingeschätzt werden kann als das einer introvertierten Person, da erstere mehr Beobachtungsgelegenheiten bietet.
In der vorliegenden Arbeit werden sowohl Charakteristika der Rater als auch der Ratees berücksichtigt (z.B. Alter), jedoch keine Persönlichkeitseigenschaften.

Niveau. Entgegen der Annahme, dass sich Frauen genauer einschätzen als Männer (vgl. London & Wohlers, 1991) und sich in männlichen Aufgaben sogar eher unterschätzen (Beyer & Bowden, 1997),[76] ergibt sich kein signifikanter Effekt für *Geschlecht* (weder für Korrelation noch für Niveau, Heidemeier & Moser, 2002). Auch im Führungsbereich kann der Unterschied nicht generell gefunden werden (Van Velsor, Taylor & Leslie, 1993). Dies könnte daran liegen, dass sich Frauen zwar weniger überschätzen als Männer, dass sie aber auch von Fremdbeurteilern schlechter bewertet werden als Männer, da weibliche Führungskräfte prinzipiell negativer betrachtet werden als männliche (Heilman, Block & Martell, 1995) und *erfolgreichen* Führungskräften eher stereotyp männliche Verhaltensweisen und Charkteristika zugeschrieben werden (Heilman, Block, Martell & Simon, 1989; Schein, Mueller, Lituchy & Liu, 1996).[77] Dementsprechend können für weibliche Führungskräfte in Settings, die traditionell Männern zugesprochen werden (z.B. Produktionsbetriebe) auch höhere Diskrepanzen als für männliche Führungskräfte oder weibliche Führungskräfte in Bildungseinrichtungen (traditionell weibliche Settings) gemessen werden (Becker, Ayman & Korabik, 2002).

Im Gegensatz dazu erbringt die Studie von Heidemeier und Moser (2002) ein „positives“ Ergebnis für *Kultur* (westlich versus asiatisch) in der Hinsicht, dass die Niveau-Unterschiede für asiatische Stichproben deutlich geringer ausfallen als für westliche. Erklärt wird dies damit, dass in kollektivistischen Kulturen (Hofstede, 1983) aufgrund des Gruppenprimats die Selbst-Aufwertung an Wichtigkeit verliert. Dies wiederum hängt möglicherweise damit zusammen, dass die Equity-Norm schwächer ist (Kim, Park & Suzuki, 1990), also die Erwartung von individuellem Gewinn nicht so stark von einer im Vergleich zu anderen sehr positiven Beurteilung abhängt. Alternativ kann die moderierende Wirkung von Kultur auch mit den unterschiedlichen Kommunikationsstilen, die zwischen östlichen (China, Südkorea) und westlichen Kulturen (USA) existieren, begründet werden. Im Unterschied zu Ersteren lässt sich für die USA vor allem in Arbeits-Settings eine geringere

[76] Im Experiment von Beyer und Bowden (1997) wurde als "männliche Aufgabe" ein Multiple-Choice-Test zu Sportaufgaben herangezogen. Da auch Managementverhalten als „männlich“ charakterisiert wird (Schein, Mueller, Lituchy & Liu, 1996), kann für Frauen auch für diesen Bereich von einer Selbstwahrnehmung ausgegangen werden, die im Vergleich zu Männern weniger stark überzogen ist.
[77] Dieser Effekt ist bei männlichen (versus weiblichen) Beurteilern besonders stark (Schein et al., 1996) und er gilt umgekehrt, wenn es um weibliche Rollen geht (Eagly, Karau & Makhijani, 1995).

Sensibilität/Aufmerksamkeit für indirekte Kommunikation („zwischen den Zeilen“) festhalten (Sanchez-Burks, Lee, Choi, Nisbett, Zhao & Koo, 2003), was in höheren Diskrepanzen resultieren kann (vgl. Kapitel 3.5.2). Der Einfluss von Kultur auf *Fremd-Fremd*-Diskrepanzen ist bisher nicht untersucht; sie müssten gemäß dieser Erklärung für östliche Kulturen ebenfalls höher ausfallen als für westliche.

Da selbst-bezogene Prozesse in Einzelstudien immer wieder als wichtige Moderatorvariablen gehandelt werden, sollen sie hier kurz erwähnt werden: *Selbst-Monitoring* (Snyder, 1976), also die Fähigkeit, Hinweisreize aus der Umwelt zu erkennen und sein Verhalten adäquat darauf auszurichten, hängt mit überhöhter Beurteilung fremder (fiktiver) Leistung zusammen (Jawahar, 2001). Die Selbstbeschreibung hingegen scheint mit steigendem Selbst-Monitoring genauer auszufallen. Denn je stärker das Selbst-Monitoring von Führungskräften (N = 504) ausgeprägt ist, umso geringer fällt die Diskrepanz zwischen Selbst- und Fremdbeurteilung aus (Church, 1997a). Die unterschiedliche Wirkung von Selbst-Monitoring bei Selbstbeurteilung (bei Validierungserwartung) und Fremdbeurteilung könnte dadurch erklärt werden, dass hohe Selbst-Monitorer ihr Ratingverhalten an den erwarteten Konsequenzen ausrichten: Prinzipiell werden gute Beurteilungen eher belohnt; für den Fall einer Selbst-Fremd-Gegenüberstellung werden eher „richtige“ (übereinstimmende) Beurteilungen belohnt. Alternativ werden die Ergebnisse besser vereinbar, wenn man den beurteilten Inhaltsbereich betrachtet. Selbst-Monitoring scheint nämlich nur dann positiv auf die Genauigkeit von Ratings zu wirken, wenn es um die Beurteilung von interpersonellem Verhalten (z.B. Teammanagement) geht und nicht um die Beurteilung von purer Arbeitskompetenz (z.B. Entscheidungen treffen, Planung) (Warech, Smither, Reilly, Millsap & Reilly, 1998). Interpersonelles Verhalten ist aber bei einer Management-Stichprobe im Feld (wie bei Church, 1997a) bedeutender als bei der Beurteilung fiktiver Leistung (wie bei Jawahar, 2001).

Weiter zeigt sich im Labor, dass Personen mit einem hohen *Selbstwertgefühl* positive (negative) Attribute eher sich selbst (anderen) als anderen (sich selbst) zuschreiben (Brown, 1986). Allerdings zeigt die Arbeitsgruppe von A. Mummendey immer wieder, dass eine Asymmetrie zwischen positiven und negativen Stimuli insofern besteht, dass die Ingroup viel stärker durch die gesteigerte Zuschreibung positiver als durch die geminderte Zuschreibung negativer Attribute favorisiert wird (Übersicht in Otten & Mummendey, 2000).

Führungskräfte, die sich selbst besser kennen und zudem bereit sind, sich zu verändern (also hohe *Self-Awareness* haben), gehören weniger der Gruppe der Über- als vielmehr der Gruppe der Unterschätzer oder Übereinstimmer an (Van Velsor et al., 1993). Fehlende Self-Awareness wird von verschiedenen Autoren als Erklärung für die mangelhafte Selbst-Fremd-Korrelation herangezogen; letztere wird sogar als Indikator für Self-Awareness

betrachtet (Bass & Yammarino, 1991; Church, 1997a; Wohlers & London, 1989). Von dieser Sichtweise wird in der vorliegenden Arbeit Abstand genommen, da bisher die Entstehung der Diskrepanzen ungeklärt ist und mit hoher Wahrscheinlichkeit verschiedene Prozesse daran beteiligt sind (vgl. auch Kapitel 3.4). Dass das Konzept hier als Moderatorvariable, nicht als Erklärungsmechanismus dargestellt wird, liegt an den bisherigen Untersuchungsdesigns. Wer sich selbst gut kennt, bringt letztlich elaborierte Selbstkonzepte hervor und kann sodann leichter auf selbstbezogenes Wissen zurückgreifen (Markus, 1983). Doch hohe Self-Awareness verbessert nicht nur diesen Zugriff, sondern hängt auch negativ mit Impression Management zusammen (Pryor et al., 1977). Wenn aber auch selbstwertdienliches Impression Management zu überhöhten Selbstratings führt (vgl. Kapitel 3.4.2.3), dann kann Self-Awareness auch indirekt – über weniger überzogene Selbstdarstellung – mit einer genaueren Selbsteinschätzung zusammenhängen.

In einer militärischen Stichprobe zeigte sich eine hohe interperspektivische Übereinstimmung vor allem dann, wenn die Führungskräfte hohe Werte auf der *Introversions*- und Sensitivitätsskala (MBTI) hatten (Roush & Atwater, 1992). Wenn aufgrund solch eher dispositionaler Persönlichkeitsmerkmale manche Personen dazu prädispositioniert sind, sich selbst zu überschätzen, könnte deren Erfassung dabei helfen, genaue und ungenaue Selbstbeurteiler zu trennen.

3.3.2 Moderatorvariablen im tätigkeitsbezogenen und organisationalen Bereich

Validität. Wird die Selbsteinschätzung von Personen mit *Management-Tätigkeiten* (oder Fachexperten) mit der von Personen mit einfacheren Tätigkeiten verglichen, dann zeigt sich eine höhere Validität für letztere. Der Tätigkeitsbereich erweist sich als sehr stabiler Moderator (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002). Auch wenn dies der oben dargestellten Bildungsniveau-Hypothese widerspricht (da davon auszugehen ist, dass Manager/Experten ein höheres Qualifikations- und Bildungsniveau haben als Arbeiter), sind die Ergebnisse plausibel erklärbar. Zum einen kann angenommen werden, dass für Führungskräfte weniger relevante Vergleichspersonen zur Verfügung stehen als für Personen, die in einer Unternehmenshierarchie weiter unten angesiedelt sind. Nach der Theorie des sozialen Vergleichs (siehe Kapitel 3.4.1.7) kann diese mangelnde Gelegenheit zum Vergleich von Fertigkeiten oder Meinungen zu ungenaueren Selbsteinschätzungen führen. Zum anderen kann auf die Komplexitäts-Annahme zurückgegriffen werden. Sie besagt, dass mit Zunahme der Komplexität einer Tätigkeit diese auch vielfältiger zu interpretieren, weniger vollständig zu beobachten sowie schwieriger zu beurteilen ist (da die Integration der Einzelbeobachtungen schwieriger ist). Aufgrund dieser

unsystematischen Verzerrungen stimmen mehrere Beobachter weniger gut in ihrem Urteil überein als bei einfachen, klar geregelten Arbeitsgängen. *Aufgabenkomplexität* nimmt auch als eigenständige Variable (nicht nur verdeckt im Tätigkeitstyp) einen signifikanten moderierenden Einfluss auf die Validität der Selbstbeschreibung (Conway & Huffcutt, 1997). Auch Wherry und Bartlett (1982) schließen aus ihrer komplexen Ratingtheorie, dass unter anderem über die Fokuspersonen hinweg geltende, konstante Arbeitsbedingungen zur Güte der Ratings beitragen. Aufgabenkomplexität ist mit einer solchen Konstanz aber nicht vereinbar.

Weiter ist anzunehmen, dass die *inhaltliche Dimension* der zu beurteilenden Leistung einen Einfluss auf die Beurteilungsgüte nimmt. So sollten kognitive Aufgaben wegen der schlechteren Beobachtbarkeit weniger valide zu beurteilen sein als interpersonelle Tätigkeiten, was jedoch empirisch nicht für Selbstbeschreibungen untersucht wurde. Bei Conway und Huffcutt (1997) findet sich immerhin ein erster Hinweis darauf, dass verschiedene Fremdbeurteiler dann höher übereinstimmen, wenn es um die Beurteilung interpersoneller Fertigkeiten geht (im Gegensatz zu allen inhaltlichen Dimensionen im Ganzen). Für das Führungsverhalten ist insbesondere die interpersonelle Dimension charakteristisch. Jedoch gehen den Handlungen oft wichtige Phasen voraus, die kognitiver Natur sind oder in Interaktion mit Personen stattfinden, die später nicht als Rater befragt werden.

Niveau. Bei der Beurteilung von *umfeldbezogener Leistung* ergeben sich geringere Niveau-Unterschiede als bei der Beurteilung von aufgabenbezogener Leistung, das heißt von konkreten Aufgaben mit festsetzbaren Zielen (Heidemeier & Moser, 2002).[78] Da umfeldbezogene Leistungsbereiche über die formalen Rollenerwartungen hinausgehen, nur indirekt ergebnisunterstützend und nicht jobspezifisch sind, charakterisieren sie das Führungsverhalten eher als aufgabenbezogene Leistungsbereiche. Folglich geht der Befund damit einher, dass sich auch für *Manager/Experten* geringere Diskrepanzen zeigen als für Arbeiter (Heidemeier & Moser, 2002). Die Wirkung beider Moderatorvariablen – des Leistungsbereichs und der Arbeitstätigkeit – wirkt zunächst überraschend, weil konkrete Aufgaben von Arbeitern genauer beurteilbar sein müssten. Ihre Wirkung könnte aber zum einen auf dem höheren Bildungsgrad und der besseren Selbstreflexion der Führungskräfte beruhen, zum anderen darauf, dass Abwärtsbeurteiler von Führungskräften sich ebenfalls als Führungskraft kategorisieren und deshalb keine Distinktheit schaffen „müssen“ (Heidemeier und Moser betrachten lediglich Selbst-*Abwärts*-Unterschiede; vgl. Kapitel

[78] Die Einteilung beruflicher Leistung in aufgabenbezogene (task performance) und umfeldbezogene Leistung (contextual performance) erfolgte bei Heidemeier und Moser (2002) im Sinne von Borman und Motowidlo (1993). Die umfeldbezogene Leistung umfasst für Führungskräfte folgende Inhaltsbereiche: die Aufrechterhaltung von Arbeitsbeziehungen, die Repräsentation der Organisation nach außen, die Ausdauer und Anstrengung zur Zielerreichung sowie organisationales Commitment. Sie spielt sich also deutlicher als aufgabenbezogene Leistung im interpersonellen Kontext ab und betrifft weniger den Kernauftrag des Unternehmens (z.B. Produktion von Autos) als aufgabenbezogene Leistung.

3.4.1.7). Die Autoren selbst bemühen Erklärungen, die die überhöhten Ratings der Arbeiter auf deren Motivation zurückführen, in eine andere "job community" aufgenommen werden zu wollen. Dieser Punkt muss meines Erachtens genauer untersucht werden, zumal für Manager eine geringere korrelative Übereinstimmung berichtet wurde als für Arbeiter (vgl. S. 140). Er soll hier deshalb nicht überstrapaziert werden, weil sich die Diskrepanz für Manager alleine (d = .34, nicht Experten) – also die für die vorliegende Studie relevante Stichprobe – nicht von der durchschnittlichen Diskrepanz der Gesamtstichprobe (d = .35) unterscheidet.

Weder für die Validität noch für die Niveau-Unterschiede erweist sich die Dauer der *Organisationszugehörigkeit* als Moderator, was aufgrund besserer Kenntnisse der Anforderungen, Ziele und ungeschriebenen Normen zu erwarten wäre (Heidemeier & Moser, 2002). Anzunehmen wäre allerdings, dass genau dieses Wissen für die Beschreibung von Führung an Relevanz gewinnt, da sie der Dreh- und Angelpunkt zu dessen Verbreitung ist. Diesbezügliche Ergebnisse für den Führungsbereich stehen aber noch aus.

3.3.3 Moderatorvariablen im Bereich der Durchführungsbedingungen

Die Durchführungsbedingungen lassen sich in zwei Gruppen trennen. Zum einen werden Variablen betrachtet, die speziell das Format des Befragungsinstruments betreffen, zum anderen geht es um allgemeine Rahmenbedingungen.

Validität. Innerhalb der ersten Variablengruppe (Format des Befragungsinstruments) zeigten sich bisher lediglich zwei Moderatorvariablen (Mabe & West, 1982): Wird *vergangene* Leistung eingeschätzt (versus zukünftige prognostiziert), dann steigt die Validität der Selbstbeurteilung. Der gleiche Effekt ergibt sich, wenn gezeigte *Leistung* beurteilt wird anstatt zugrunde liegende Fähigkeiten. Beide Variablen präzisieren den zu beurteilenden Gegenstand, so dass weniger unsystematische Beurteilungsfehler auftreten können. Je präziser die Operationalisierung des zu beurteilenden Gegenstands in der Itemformulierung gelingt, umso besser können verschiedene Urteiler übereinstimmen und umso valider können die Antworten ausfallen.

Da eine spezifische Operationalisierung für abgegrenzte, spezifische Inhaltsbereiche besser möglich ist als für globale, ist anzunehmen, dass *dimensionale Skalen* zu einer höheren Validität führen als globale Skalen (Harris & Schaubroeck, 1988). Auf der anderen Seite könnten sich Abweichungen auf einzelnen Dimensionen in globalen Ratings wieder ausgleichen, so dass die Übereinstimmung für letztere höher sein müsste (Heidemeier & Moser, 2002). Doch weder in die eine noch in die andere Richtung konnte bisher ein moderierender Einfluss gezeigt werden. Weiter konnte lediglich tendenziell (nicht signifikant) bestätigt werden, dass für *verhaltensbezogene* Ratings eine höhere Übereinstimmung

erreichbar ist als für eigenschaftsorientierte Ratings (Harris & Schaubroeck, 1988). Dabei dürfte die Beantwortung von Items, die durch konkrete Verhaltensbeschreibungen operationalisiert sind, insgesamt weniger Inferenzen voraussetzen und insbesondere weniger Inferenzen, die aus außerberuflichen Lebensbereichen gezogen werden und nur dem Selbstbeurteiler zugänglich sind. Dieser uneindeutigere Interpretationsprozess bei der Beantwortung eigenschaftsorientierter Items, der ja bereits bei der Auswahl der heranzuziehenden Indikatoren beginnt, sollte zu stärkeren Beurteilerdifferenzen führen. Die Annahme eines solchen „störungsresistenteren" Beurteilungsprozesses bei Verhaltensratings entbehrt indes seiner realistischen Grundlage; denn Beurteiler geben Verhaltensratings häufig auf der Basis von vorher gebildeten Eigenschaftsurteilen ab und können ohne Training Verhalten von Eigenschaften nicht trennen (vor allem, wenn die Formulierung des Verhaltens „verkappte" Eigenschaftsdimensionen enthält) (Diskussion dazu vgl. Marcus & Schuler, 2001). Da Führungsverhalten per se jedoch schwierig zu beobachten ist, sollte hier besonders auf klare Operationalisierung geachtet werden.

Aus psychometrischen Gründen sollte die Validität bei Skalen mit *mehreren Items* höher sein als bei einem einzelnen Item. Dies konnte in der Meta-Analyse von Heidemeier und Moser (2002) jedoch nicht gefunden werden. Wider Erwarten beeinflusst auch der Grad der *Ähnlichkeit* zwischen dem Gegenstand der Selbstbeurteilung und des Kriteriumsmaßes deren Interkorrelation nicht (Mabe & West, 1982). Dies entkräftet ansatzweise die Kritik an der Verwendung weicher Erfolgsmaße sowie am Schweregrad eines potentiellen Single-Source-Bias (siehe oben). In der Regel besteht nämlich eine stärkere Ähnlichkeit zwischen Prädiktor- und Kriteriumsvariablen, wenn weiche (versus harte) Kriterien betrachtet werden beziehungsweise wenn beide Variablen von der selben Person erhoben werden. Ähnlichkeit scheint aber wirkungslos zu bleiben und die Validität nicht ungerechtfertigterweise in die Höhe zu treiben.

Im Folgenden wird die zweite Gruppe an Moderatorvariablen im Bereich der Durchführungsbedingungen betrachtet: die Rahmenbedingungen. Anzunehmen ist, dass Personen, die darin geübt sind, ihre Fertigkeiten selbst einzuschätzen (sowie Personen, die häufig eingeschätzt werden), mehr und realistischeres Wissen über ihre eigenen Leistungsbereiche ansammeln. Zumindest in der Meta-Analyse von Mabe und West (1982), in der Leistungsbeschreibung in einem sehr breiten Verständnis (über berufliche Leistung hinaus) untersucht wurde, konnte eine deutliche Korrelation zwischen *Selbstbewertungs-Erfahrung* und der Validität von Selbstbeurteilungen festgestellt werden. Eine Übertragung auf den Führungskontext steht noch aus. Neben solchen intraindividuellen Referenzpunkten früherer Erfahrungen kann davon ausgegangen werden, dass auch interindividuelle Bezugspunkte die Selbsteinschätzung beeinflussen. Da Personen dazu tendieren, ihre eigene Leistung mit

Hilfe sozialer Vergleichsprozesse einzuordnen (Festinger, 1954),[79] müsste die Validität steigen, wenn Vergleichsinformation verfügbar ist beziehungsweise wenn der Selbstbeurteilung „standardisierte“ Referenzpersonen zugrunde gelegt werden. Dies wird in der Laborstudie von Farh und Dobbins (1989) bestätigt: Die Validität der Selbstbeurteilungen ist für diejenigen Versuchspersonen höher, denen als Vergleichsinformation die individuellen Leistungen von mehreren angeblichen „Kollegen“ vorgelegt werden. Eine derartige Leistungsevaluation hat also eher relativen (Vergleich zur Referenzpersonen) als absoluten Charakter. Ankerskalen, die eine Relation zu einer Bezugsgruppe festlegen (besser/schlechter als Durchschnitt), geben explizit eine *Instruktion zum sozialen Vergleich* (z.B.: "as compared to your fellow workers"). Im Unterschied zu gewöhnlichen Likert-Skalen (absolute Ratings) erwies sich diese als signifikante Moderatorvariable (Mabe & West, 1982, vgl. auch Reduktion des Above-Average-Effekt, Kapitel 3.2.3).[80] Es ergaben sich keine weiteren moderierenden Effekte von Variablen, die den sozialen Vergleich konkretisieren könnten: Die Korrelation zwischen Selbstbild und Kriterium steigt weder dann, wenn die Referenzgruppe besser definiert wird, noch dann, wenn mehr Wissen über ihre Kompetenzen gegeben wird (Mabe & West, 1982). Likert-Skalen ohne Instruktion zum sozialen Vergleich überlassen es den Ratern, ob/welche Referenzgruppe sie bei der Beurteilung heranziehen. Dadurch bleibt mehr Spielraum für selbstwertdienliche Antworten, was die Beurteilerübereinstimmung mindert. Zu vermuten ist, dass gerade Führungsverhalten soziale Vergleichsprozesse erfordert, da aufgrund seiner Komplexität mehr Unsicherheit besteht und aus Mangel eines objektiven Standards kein anderer Referenzpunkt vorhanden ist.

Widersprüchliche Ergebnisse bestehen hinsichtlich der zugesicherten *Anonymität* (der Ratingantworten) und der *Validierungserwartung*, also der Erwartung, dass die abzugebende Selbstbeschreibung an anderen Kriterien (z.B. Fremdratings) validiert wird: Während bei Mabe und West (1982) beide Variablen unter die vier wichtigsten Moderatoren für die Validität von Selbstbeschreibungen fallen, konnten Heidemeier und Moser (2002) keine moderierenden Effekte auf die korrelative Übereinstimmung zwischen Selbst- und Abwärtsbeurteilungen nachweisen. Dazu passt ihr Ergebnis, dass sich hier auch bei *360°-Befragungen* (Selbst-, Abwärts-, Aufwärts- *und* Peerbeurteilung), bei denen eine Validie-

[79] Festinger (1954) postuliert, dass Individuen dazu motiviert sind, ihre Fertigkeiten (und Meinungen, was für die vorliegende aber Arbeit irrelevant ist) akkurat beurteilen zu können. Für den Fall, dass absolute Standards nicht gegeben sind, müssen sie sich deshalb mit anderen vergleichen, um ihr Leistungsniveau zu bestimmen. Genaues Wissen um ihr eigenes Leistungsniveau wird von Individuen deshalb angestrebt, um ihren Erfolg vorherzusagen beziehungsweise um Misserfolgssituationen zu vermeiden. Dem gegenüber steht der in Kapitel 3.2.3 beschriebenen Above-Average-Effekt, der in der Regel durch Verzerrungen bei *relativen* Ratings gefunden wird.

[80] "Relatives Ratingformat" erweist sich auch in der Meta-Analyse von Heneman (1986) als Moderatorvariable für den Zusammenhang zwischen Abwärtsbeurteilungen und objektiven Leistungskriterien. Dies konnte bei Bommer et al. (1995) nur tendenziell bestätigt werden. Dennoch können die Ergebnisse als Hinweis dafür gewertet werden, dass relative Ratings leichter und genauer zu beantworten sind als absolute.

rung an anderen Ratings nahe liegt, keine höhere Selbst-Abwärts-Korrelation und keine geringeren Niveau-Unterschiede einstellen als bei gewöhnlichen Befragungen von weniger Perspektiven. Möglicherweise finden Heidemeier und Moser deshalb keine moderierenden Effekte durch Anonymität und Validierungserwartung, weil die beiden Variablen für den Vergleich zu Beurteilungen aus der Abwärtsperspektive keine wichtige Rolle spielen; die Abwärtsbeurteilung könnte für die Selbstbeurteiler zu gewohnt sein, da sie lange Zeit die einzige Beurteilungsrichtung darstellte. Insgesamt erscheint es entsprechend der Ergebnisse von Mabe und West (1982) durchaus plausibel, dass die Beurteilerübereinstimmung steigt, wenn die Selbstbeurteiler davon ausgehen, dass ihre Einschätzung anhand anderer Kriterien validiert wird (Validierungserwartung)[81] beziehungsweise wenn sich die Risiken aus einer Veröffentlichung minimieren (Anonymität). Im Vergleich zu anderen beruflichen Leistungen wird die Güte des Führungsverhaltens im organisatorischen Alltag selten unmittelbar durch andere Kriterien rückgekoppelt. Insofern könnte man mutmaßen, dass die Moderatorwirkung von Validierungserwartung auf die Selbst-Fremd-Korrelation stärker wird, wenn noch keine Gewöhnung an Beurteilung eingesetzt hat und eine Validierung eigener Aussagen bisher selten ist. Auch Anonymität wirkt sich entsprechend der Erwartungen zumindest auf Mitarbeiterebene stark aus (siehe Kap. 3.1.3), da es bei der Beurteilung des Führungsverhaltens um den zwischenmenschlichsten Bereich zwischen Führungskraft und Mitarbeiter geht, der die tägliche Zusammenarbeit bestimmt.

Niveau. Mit der Identifizierbarkeit beziehungsweise mit den erwarteten Folgen der Bewertung verbunden ist der *Beurteilungszweck*. Diesbezüglich zeigen sich geringere Niveau-Unterschiede, wenn es um Ziele der Personalentwicklung geht (im Vergleich zu reinen Forschungszielen), die von allen Seiten gleichermaßen Zustimmung finden (Heidemeier & Moser, 2002). Für administrative Ratings (Zweck: Gehaltserhöhung und Beförderung) ist der größte Niveau-Unterschied zu erwarten, da es um „harte“, folgenschwere Konsequenzen geht – und diese in der Regel auf gute Leistung folgen. Eine negative Selbstbeurteilung bliebe also erfolglos. Entsprechend zeigt eine Meta-Analyse zu genau dieser Unterscheidung, dass bei administrativen Ratings mit positiveren Abwärtsratings von Führungskräften gerechnet werden muss als bei Forschungs- oder Entwicklungsratings. Für Aufwärtsratings ergab sich der Effekt nur in sehr geringem Ausmaß (Jawahar & Williams, 1997). Dieser Unterschied bezüglich der Perspektiven kann daran liegen, dass sich Vorgesetzte selbst mit dem Erfolg ihrer unterstellten Führungskräfte schmücken können, nicht jedoch deren Mitarbeiter.

[81] Zu bedenken gilt, dass eine Selbstbeschreibung in diesem Fall nicht mehr als reine Beschreibung der Selbstwahrnehmung zu interpretieren ist, sondern als Vorhersage der Einschätzung durch andere und damit als gezielte Selbstdarstellung. In der Literatur zur Selbstpräsentation gilt, dass die Selbstdarstellung positiver ausfällt, je schwieriger es für ein Publikum ist, sie zu überprüfen (Schlenker, 1980).

Im Gegensatz zu ihrem Ergebnis für die Validität finden Heidemeier und Moser (2002) den erwarteten Effekt auf Niveau-Unterschiede, wenn die Selbstbeurteiler eine *Validierung* ihrer Angaben durch andere erwarten. Ähnlich ist die *Verifizierbarkeit* zu betrachten: Geht es um die Beurteilung von Leistungen, die anhand objektiver Kennwerte überprüfbar sind, verringern sich die Niveau-Unterschiede im Vergleich zu Beurteilungen von Leistungen, die lediglich interpersonell (zum Beispiel Führungsverhalten) gut validierbar sind (kein Effekt für Korrelation, Heidemeier & Moser, 2002). Hingegen hat die *Anonymität* in den Daten von Heidemeier und Moser (2002) keinen Einfluss auf die Selbst-Abwärts-Niveau-Unterschiede. Möglicherweise ist eine Anonymitätsinstruktion unglaubwürdig, wenn es um die Geheimhaltung vor dem eigenen Vorgesetzten (Abwärtsbeurteiler) geht. Eine solche Anonymitätsinstruktion wäre auch dann wirkungslos, wenn der Selbstbeurteiler davon ausgeht, dass sein Vorgesetzter seine Leistung ohnehin gut kennt.

Von den Variablen, die die Gestaltung des Befragungsinstruments betreffen, wurde bisher lediglich eine (eigenschafts- versus *verhaltensorientierte Ratings*) hinsichtlich ihres Effekts auf Niveau-Unterschiede untersucht. Sie erwies sich als nicht signifikant (Erklärung vgl. S. 143). Unklar ist, ob sich für die Beurteilung von Führungsverhalten ein anderes Bild ergibt.

Einen Eindruck davon, wie stark der moderierende Effekt sein kann, vermitteln Mabe und West (1982, S. 293) für die in ihrer Meta-Analyse vier bedeutsamsten Moderatorvariablen *Anonymität*, *Validierungserwartung*, *Bewertungserfahrung* und *relative* statt absolute Ankerskalen (Instruktion zum sozialen Vergleich). Die Korrelation zwischen der Selbstbeurteilung und den Außenkriterien (Validität) steigt von $r = .00$ für den Fall, dass keine der vier Bedingungen erfüllt ist, auf $r = .63$ für den Fall, dass alle vier Bedingungen erfüllt sind.[82] Als Bestätigung für deren moderierende Wirkung auch im beruflichen Kontext[83] kann eine eben dort angesiedelte Studie herangezogen werden, in der alle vier Bedingungen eingeführt wurden (Moser et al., 1994): Hier konnte eine vergleichsweise hohe Korrelation von $r = .42$ zwischen Selbst- und Abwärtsbeurteilungen gefunden werden. Gerade die Moderatorvariablen, die an den Durchführungsbedingungen ansetzen, lassen sich unmittelbar in Durchführungsempfehlungen umsetzen (vgl. Kapitel 5.3.1).

[82] Einschränkende Anmerkung: Nur in vier Studien waren alle vier Bedingungen erfüllt.
[83] Die höchsten Validitäten ergaben sich bei Mabe und West (1982) für schulische und sportliche Leistungen, so dass der allgemeine Validitätsindex den für berufliche Leistung überschätzt.

Zusammenfassend gilt, dass die Forschungsarbeiten, die das Zustandekommen der diskrepanten Beurteilungen untersuchten, sehr lange nur auf Moderatorvariablen abzielten. Dadurch konnten verschiedenste moderierende Effekte auf Persönlichkeits-, Tätigkeits- und Durchführungsebene aufgedeckt werden. Führungstätigkeit scheint eine moderierende Wirkung auf die Selbst-Fremd-Übereinstimmung zu haben. So ist für das *Führungsverhalten* festzuhalten, dass eher eine geringe Korrelation zwischen Selbst- und Fremdbild sowie ein geringer Niveau-Unterschied zwischen Selbst- und Abwärtsbild zu erwarten sind – verglichen mit der Übereinstimmung hinsichtlich anderer Leistungsbereiche. Welche Moderatorvariablen die Selbst-Fremd- oder die Fremd-Fremd-Übereinstimmung bei Führungsbeurteilungen beeinflussen, wurde bislang kaum untersucht.

Im Folgenden werden nun umfassendere Erklärungen für die oben beschriebene geringe Validität beziehungsweise hohe niveaubezogene Diskrepanz dargestellt. Jede dieser Argumentationslinien kann auch die Wirkung verschiedener Moderatorvariablen erklären. Deshalb werden diese in den entsprechenden Abschnitten des folgenden Kapitels erneut aufgegriffen.

3.4 Potentielle Erklärungsmechanismen für die mangelnde Passung zwischen Selbst- und Fremdbild

Es drängt sich die Frage auf, *warum* die Selbstbewertung in aller Regel besser (Niveau) beziehungsweise anders (Korrelation) ausfällt als die Fremdbewertung. Menschliche Beurteilung ist nie eine "1:1-Abbildung" einer Wirklichkeit, sondern immer eine individuelle Beschreibung, die auch genährt ist von persönlichen Interessen, Motiven, Erwartungen und mentalen Verarbeitungsprozessen der Rater. Verkürzt bedeutet dies, dass sich eine multiperspektivische Beschreibung aus so vielen Wirklichkeiten wie Beurteilern zusammensetzen muss und deshalb keine totale Übereinstimmung anvisiert werden kann. Doch warum differieren diese Wirklichkeiten so stark? Und warum unterscheiden sich die Wirklichkeiten der sich selbst beurteilenden Führungskräfte so stark von denen anderer Beurteiler? Vieles deutet darauf hin, dass ein Zusammenspiel verschiedener Ursachen für die Ergebnisse verantwortlich ist. Auch im vorliegenden Kapitel steht entsprechend des aktuellen Erkenntnisstandes die Selbst-Fremd-Übereinstimmung im Mittelpunkt. Einzelne Erklärungsansätze werden auf die Fremd-Fremd-Übereinstimmung übertragen.

Auf verschiedenen Ebenen können Faktoren die Verzerrung von Selbst- oder Fremdbild beeinflussen. Bei deren Darstellung wird zurückgegriffen auf allgemeine Beschreibungen des Beurteilungsprozesses, dessen Ergebnis sich aus drei Klassen von Faktoren ergibt (nach Wherry & Bartlett, 1982; erweitertes Begriffsverständnis bei Scullen et al., 2000): erstens aus dem Leistungsverhalten der Fokusperson (*Ratee-Leistung*) selbst, zweitens aus verschiedenen Prozessen der Informationsverarbeitung im weitesten Sinne, die der Aussage des Raters (Selbst- oder Fremdrater) zugrunde liegen (*Rater-Bias*) sowie drittens aus dem *Messfehler*. Die **Ratee-Leistung** ergibt sich aus dem **generellen Leistungsniveau** der Fokusperson sowie ihren **speziellen Leistungen** auf einzelnen Dimensionen.[84] Idealiter geben Leistungsratings nur dieses aktuelle Leistungsverhalten der Fokusperson wieder. Doch die Korrelationen zwischen Leistungsratings und objektiven Leistungsmaßen sind nach meta-analytischen Schätzungen allenfalls mittelmäßig (maximal r = .39; Bommer et al., 1995; Conway et al., 2001), da Ratingantworten kontaminiert und defizient sind. Während Bommer et al. (1995) Abwärtsratings als subjektive Maße heranzogen, waren es bei Conway et al. (2001) Aufwärts- und Peerratings. Dadurch gewinnen die Ergebnisse für die vorliegende Untersuchung an Relevanz, weil sie alle auch hier einbezogenen Perspektiven abdecken. Außerdem gibt es keinen Hinweis auf deutlich andere Zusammenhänge in der Beurteilung von *Führungsverhalten,* auch wenn die zitierten Meta-Analysen berufliche Leistung im Allgemeinen betrachten. Diese Annahme wird durch die Befunde zum

[84] Da manche Leistungsvoraussetzungen für verschiedene Leistungsfacetten sehr ähnlich sind, ist von einem „wahren Halo“ auszugehen, der folglich nicht als Fehler missinterpretiert werden darf.

Zusammenhang zwischen Aufwärtsratings und Erfolgskriterien, die in Kapitel 3.1.1 berichtet wurden, unterstützt.

Der nur mittelmäßige Zusammenhang zwischen Ratings und objektiven Kriterien darf aber nicht als Limit, das nicht übertroffen werden kann, bewertet werden. Es zeigt sich nämlich, dass subjektive Ratings und objektive Maße sehr stark korrelieren (r = .71), wenn sie so konstruiert sind, dass sie genau die gleiche Leistungsdimension erfassen (Meta-Analyse von Conway et al., 2001). Zu bedenken gilt weiter, dass auch objektive Maße defizient sind, also auch diese zu den geringen Korrelationen beitragen (vgl. Murphy, Cleveland & Mohler, 2001). So fasst beispielsweise das objektive Maß der Fehltage eher die Führungsleistung, die sich im Bereich der Motivierung und Mitarbeiterorientierung abspielt, vernachlässigt aber die Führungsleistung, die sich mit der Außendarstellung des Unternehmens befasst. Für komplexe Bereiche wie das Führungsverhalten ist es schwierig, ein relevantes, objektives Kriterium zu finden.

Wenn aber Ratings mit objektiven Maßen in aller Regel nur gering korrelieren, die Ratee-Leistung also nur ansatzweise abgebildet wird, dann bedeutet das, dass ein Teil der Ursachen für die mangelnde Übereinstimmung von Selbst- und Fremdbild auf der Raterseite, am **Rater-Bias**, liegt: Zum einen kann sie in den unterschiedlichen organisationalen Perspektiven von Selbst- und Fremdratern begründet sein (**perspektivenbezogener Bias**). Dies liegt deshalb sehr nahe, weil die Übereinstimmung von Ratern aus derselben Perspektive höher ist als von Ratern aus verschiedenen Perspektiven (Borman, 1974; siehe oben). Konkrete Möglichkeiten, wie eine geringe Selbst-Fremd-Übereinstimmung in der Führungsbeurteilung durch Perspektiven-Unterschiede bedingt sein kann, sind in Kapitel 3.4.1 beschrieben. Zum anderen kann eine mangelnde Selbst-Fremd-Übereinstimmung aber auch dadurch zustande kommen, dass die Beurteiler das zu beurteilende Führungsverhalten aufgrund individueller Faktoren verzerren (**idiosynkratischer Bias**, Kapitel 3.4.2). Dies stellt zugegebenermaßen eine Kategorie mit nur sehr vagen Grenzen dar: "Consequently, we use the term idiosyncratic rater effects to include all of the effects associated with individual raters." (Mount & Scullen, 2001, p. 165). Diese Kategorie umfasst also – im Gegensatz zum unsystematischen Fehleranteil – die systematische Varianz, die mit einzelnen Ratern, nicht aber mit dem Ratee-Verhalten oder den Ratings anderer verbunden ist (nach Scullen, Mount & Goff, 2000).[85] Darunter fallen individuelle Beurteilertendenzen, die beispielsweise auf dem Motiv, sich positiv darzustellen, gründen.

Verzerrungen, die letztlich zu den geringen Übereinstimmungen zwischen Selbst- und Fremdbild führen, können die Selbstbeschreibung genauso betreffen (in der Regel nach

[85] Scullen, Mount und Goff (2000) nennen beispielhaft für idiosynkratische Verzerrungen lediglich den Halo- und den Milde-Effekt, beschränken sich also auf klassische Beurteilungsfehler. Wenn man die mangelnde Passung nicht nur aus der methodischen Perspektive betrachtet, lassen sich die Verzerrungen auf verschiedenste plausible Motive zurückführen. In der konfirmatorischen Faktorenanalyse werden idiosynkratische von perspektivenbezogenen Effekten getrennt, indem jeder Rater als einzelnes Maß beziehungsweise alle Variablen aus einer Perspektive als ein Maß modelliert wurden.

oben verzerrt) wie die Fremdbeschreibung oder beide. Dass die geringe Übereinstimmung von Selbst- und Fremdbild von einer ungerechtfertigten Selbstüberschätzung der Führungskräfte zumindest mitbedingt ist (vgl. Nilsen & Campbell, 1993), lässt sich aus verschiedenen Befunden folgern: Ein erster schwacher Hinweis ist, dass bei Selbstratings die Verteilung stark zum besseren Pol hin verschoben ist. Deutlicher zeigt sich zweitens die Selbstüberschätzung in relativen Urteilen, denn in der Tat schätzt sich die Mehrzahl der Führungskräfte im Vergleich zu ihren Kollegen überdurchschnittlich gut ein (vgl. Above-Average-Effekt, Kapitel 3.2.3). Weiter korreliert das Selbstbild drittens (im Unterschied zu anderen Perspektiven) mit keiner anderen Beurteilergruppe in ähnlichem Ausmaß wie Fremdbeurteiler untereinander. So sind die Korrelationen zwischen den verschiedensten Fremdbeurteilern höher sowie die Niveau-Unterschiede zwischen ihnen niedriger als bei einem beliebigen Selbst-Fremd-Vergleich. Hinzu kommt, dass für Manager besonders die Selbst-*Fremd*-Korrelationen (Selbst-Peer sowie Selbst-Abwärts) deutlich geringer ausfallen als für Arbeiter, während die Fremd-Fremd-Korrelationen (Peer-Abwärts) von der Moderatorvariable *Tätigkeit* (Manager/Experten vs. Arbeiter) kaum beeinflusst sind (Harris & Schaubroeck, 1988). Das bedeutet, dass insbesondere *Selbstbeurteiler in Führungspositionen* in einer geringeren Übereinstimmung zu Fremdbeurteilern stehen. Da also Verzerrungen bei der sich selbst beschreibenden Führungskraft wahrscheinlich sind, wird bei der Darstellung potentieller idiosynkratischer Verzerrungen in Kapitel 3.4.2 auf die Selbstrater fokussiert.

Umgekehrt ist dadurch ganz und gar nicht ausgeschlossen, dass auch Fremdratings positiver (z.B. um gute Beziehungen zu halten) oder negativer als gerechtfertigt (z.B. aufgrund von Rachegefühlen oder Konkurrenzdenken) ausfallen. Abwärtsbeurteiler können aufgabenbezogene Ziele (z.B. Mitarbeiter motivieren, Karriere fördern), interpersonale Ziele (z.B. gute Zusammenarbeit belohnen, Schwierigkeiten mit Mitarbeiter vermeiden), strategische Ziele (z.B. zur positiven Darstellung des Unternehmens nach außen beitragen) und internalisierte Ziele (z.B. Selbstbild durch viele gute Mitarbeiter aufwerten, den Unternehmensnormen entsprechen) verfolgen (Murphy & Cleveland, 1995).

In Übereinstimmung mit anderen Autoren (z.B. Becker et al., 2002; Borman, 1997; Moser, 1999; Murphy & Cleveland, 1995) muss eine unbefriedigende Forschungslage festgehalten werden, was die Erklärung der geringen Selbst-Fremd-Übereinstimmung von Leistungsratings angeht. Trotz der dünnen und uneindeutigen Ergebnislage bestehen für viele Erklärungen unterstützende Argumente. Obwohl sie teilweise schon sehr lange diskutiert werden, wurden sie jedoch selten gezielt untersucht. Borman (1997) beschränkt sich beispielsweise in seinem Review auf zwei perspektivenbezogene Erklärungsmechanismen (unterschiedliche [Gewichtungen von] Dimensionen sowie unterschiedliche Beobachtungsgrundlage) sowie zwei Ratingperspektiven (Abwärts- und Peerbeschreibungen), da für andere Erklärungen und Perspektiven die Datenlage noch schwächer ausfällt. Diese in

den Kapiteln 3.4.1.1, 3.4.1.3 und 3.4.1.4 vorgestellten Erklärungsansätze stellen alsdann auch eine Erweiterung der grundlegenden Ideen von Borman (1997) dar. Andere Autoren beschränken sich darauf, die Diskrepanz auf die Mildetendenz in den Selbstbeurteilungen zurückzuführen,[86] was genauso verkürzt erscheint und allenfalls als deskriptive Notiz verstanden werden sollte wie ihre alleinige Rückführung auf illusorisch positive Selbstwahrnehmung (z.B. bei Taylor & Brown, 1988).

Im Folgenden wird eine Vielzahl an potentiellen perspektivenbezogenen sowie idiosynkratischen Erklärungen für die geringe Selbst-Fremd-Übereinstimmung nebeneinander gestellt, da zum aktuellen Zeitpunkt keine der Hypothesen ausreichend Unterstützung durch empirische Befunde findet, um sie allein zu verfolgen. Wo einzelne Studien zu Selbst-Fremd-Übereinstimmung gezielt einzelne Erklärungshypothesen testeten, werden diese angeführt. Im Anschluss an die einzelnen Erklärungen werden jeweils diejenigen Moderatorvariablen aus Kapitel 3.3 erneut aufgegriffen, die plausibel damit einhergehen.

3.4.1 Perspektivenbedingte Diskrepanzen

Wie bereits erwähnt, wird die Darstellung potentieller Erklärungsmechanismen danach getrennt, ob der Raterbias eher perspektivenbezogene oder eher idiosynkratische Ursachen aufgreifen. Das vorliegende Kapitel beschäftigt sich mit erstgenannten.

3.4.1.1 Unterschiedliche Verhaltensstichproben

Gegner der multiperspektivischen Führungsbeschreibung argumentieren, dass sich Führungsverhalten ohnehin nicht von verschiedenen Personen ähnlich beschreiben lässt, weil es im Sinne von LMX-Modellen (Graen & Scandura, 1987a, 1987b) auf den einzelnen Mitarbeiter ausgerichtet sei und sich folglich individuell gestalte (und gestalten müsse). Jeder Mitarbeiter nehme demzufolge einen Ausschnitt des Führungsverhaltens wahr, der nur ihm in dieser Weise zugänglich sei und nicht für das gesamte Führungsverhalten der Führungskraft repräsentativ sei. Entsprechend müsste hier jedem Rating per se Validität zugesprochen werden (vgl. Borman, 1997; Brutus et al., 1998). Hohe Korrelationen sind dann gar nicht zu erwarten, Niveau-Unterschiede hingegen schon. Denn während Führungskräfte ihr Führungsverhalten in Interaktion mit einer Vielzahl an Personen erleben und immer, wenn Führungsverhalten gezeigt wird, direkten Zugang dazu haben, können Fremd-

[86] Eine Mildetendenz allein könnte allenfalls die Niveau-Unterschiede erklären, nicht jedoch die geringen Korrelationen. Entsprechend wird "leniency" häufig synonym für Niveau-Unterschiede verwendet (z.B. Jawahar, 2001).

beurteiler lediglich Ausschnitte beobachten. Unmittelbar können nur die Verhaltensweisen erlebt werden, die in der Interaktion mit der Führungskraft geschehen. Allein die daraus entstehende unterschiedliche Größe der Verhaltensstichprobe führt dazu, dass für Selbstbeurteiler die gesuchten Ereignisse leichter zugänglich sind (Verfügbarkeitsheuristik).

Wenn man die LMX-Theorien zugrunde legt, müsste die Übereinstimmung zwischen Ratings stark vom Abstraktionsniveau der Ratinginhalte abhängen: Auf einer sehr konkreten Ebene (z.B. Fortbildung), auf der eine Führungskraft die unterschiedlichsten Bedürfnisse ihrer Mitarbeiter zu befriedigen hat, machen alle Mitarbeiter entsprechend unterschiedliche Erfahrungen/Verhaltensbeobachtungen, so dass auch ihre Ratings sehr unterschiedlich ausfallen und nicht korrelieren sollten. Auf einer abstrakteren Ebene wird jede Führungskraft aufgrund eines begrenzten Verhaltensrepertoires die Art und Weise, wie sie die individuellen Mitarbeiterbedürfnisse eingeht, nicht drastisch ändern. Wenn die Ratinginhalte also insofern abstrakt formuliert sind, dass sie auf verschiedene Führungsinhalte anwendbar sind (z.B. Partizipation bei Entscheidungen in verschiedenen Inhaltsbereichen), dann ist auch vor dem Hintergrund der LMX-Theorien nicht mehr von einer Null-Korrelation auszugehen.

Da man jedoch davon ausgehen kann, dass die Verhaltensstichproben von Personen der gleichen Perspektive ähnlicher sind als die von Personen unterschiedlicher Perspektiven und Positionen, wird die Erklärung unter die perspektivenbezogenen Diskrepanzen gefasst. Dies ist deshalb plausibel, weil Bedürfnisse an die Führung aus gleichen Positionen heraus ähnlicher sind als aus unterschiedlichen. Entsprechend findet man höhere Korrelationen zwischen Ratings aus der gleichen Perspektive als zwischen Ratings aus verschiedenen Perspektiven (Tsui & Ohlott, 1988). Auch erklärbar wird dann, weshalb Abwärts- und Peerratings höher miteinander korrelieren (r = .34, r_{korr} = .79, Conway & Huffcutt, 1997) als sie jeweils mit Aufwärtsratings korrelieren (r = .22). Denn den Mitarbeitern bleibt ein Teil des Managementverhaltens, den andere Führungskräfte (also Peer- oder Abwärtsbeurteiler) häufig zu Gesicht bekommen, eher verschlossen, z.B. Verhandlungen mit Kunden (Borman, 1997; Murphy & Cleveland, 1995). Die mangelnde Passung zwischen den Perspektiven ist hier jedoch nicht mehr als echter Bias zu bewerten; stattdessen spiegelt sie einen Teil des wahren Leistungsverhaltens wider, das nur der bestimmten Perspektive zugänglich ist.

Den LMX-Modellen muss entgegnet werden, dass verschiedene Außenstehende (auch aus verschiedenen Positionen) oft ähnlicher beurteilen als Selbst- und Aufwärtsbeurteiler, obwohl sie im Unterschied zu diesen in verschiedenen Führungsdyaden agieren. Die Selbstbeurteiler müssten sogar die genauesten Beurteilungen abgeben, da ihnen per se jede gezeigte Verhaltensweise zugänglich sein sollte.

Mit dieser Erklärung geht die oben berichtete Moderatorwirkung der *Arbeitstätigkeit* (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002) plausibel einher. Denn während das Führungsverhalten in viele Facetten zerfallen kann, können weniger *komplexe* Tätigkeiten von vielen Beurteilern nahezu vollständig beobachtet werden. Jedoch ist nicht anzunehmen, dass die unterschiedlichen Verhaltensstichproben den Großteil der fehlenden Übereinstimmung erklären, da verschiedene Fremdbeurteiler (auch aus verschiedenen Positionen) oft ähnlichere Urteile abgeben als die Führungskraft selbst und Fremdbeurteiler. Dies wäre hingegen nicht zu erwarten, wenn man unterschiedlichen Beurteilergruppen oder gar einzelnen Beurteilern extrem differierende Beobachtungen zusprechen würde. Einen speziellen Fall unterschiedlicher Beobachtungsstichproben stellt folgender Punkt dar.

3.4.1.2 Mangelnde Beobachtbarkeit und Transparenz von Entscheidungsprozessen

Wenn die zu beurteilenden Verhaltensweisen mit Leichtigkeit beobachtet werden können, dann steigt die Güte der Ratings (Ratingtheorie von Wherry & Bartlett, 1982). Doch die Beobachtung von Führung ist für Fremdbeurteiler häufig alles andere als leicht. Deshalb wird mangelnde Gelegenheit zur Beobachtung des zu beurteilenden Verhaltens seit langem als Faktor diskutiert, der die Reliabilität von Ratings mindert (Dunnette, 1966; Nagle, 1953). Empirische Bestätigung findet dieser Punkt erstmals bei Rothstein (1990), die einen starken Zusammenhang zwischen der Interrater-Reliabilität und der Dauer der Führungstätigkeit[87] feststellte. Möglicherweise liegt dies auch daran, dass weniger Beobachtungsmöglichkeiten zu systematischen Beurteilungsfehlern führen, die die Varianz reduzieren. Das wiederum mindert die Reliabilität (Rothstein, 1990). Vorausgesetzt, dass Führungskräfte ihr eigenes Verhalten beobachten können, bleibt durch diesen Erklärungsansatz unerklärt, weshalb Fremdbeurteilungen untereinander höher korrelieren als Selbst- mit Fremdbeurteilungen.

Die Beobachtung von Führung ist allerdings per se, also auch bei ausreichender Beobachtungsdauer, ein schwieriges Unterfangen. So erfahren Mitarbeiter beispielsweise aufgrund der fehlenden Transparenz von Führungsentscheidungen oft nur das Ergebnis, nicht aber den Prozess und die Argumente einer Entscheidungsfindung. Erwartungen der Mitarbeiter werden dann ohne Begründung erfüllt oder – schlimmer – bleiben ohne Begründung unerfüllt. Ähnlich ergeht es den Vorgesetzten der Führungskräfte, die vorwiegend an der Zielerreichung, nicht am detaillierten Weg zum Ziel interessiert sind. Da bei der Beurtei-

[87] strittige Operationalisierung der potentiellen *Beobachtungsdauer*

lung dann aber unweigerlich Attributionsprozesse (incl. Attributionsfehler, z.B. Korrespondenzverzerrung, für einen Überblick Gilbert, 1995) einsetzen, die den tatsächlichen Entscheidungsprozess nicht richtig und vollständig rekonstruieren, scheint eine mangelnde Korrelation zwischen Selbst- und Fremdbild unvermeidlich. Die Führungskraft wird – insbesondere bei unzufriedenstellenden Ergebnissen – schlechter bewertet, denn nur sie selbst kennt in einem solchen Fall den Hintergrund und weiß, dass sie unter den vorliegenden Umständen vielleicht sogar "das Beste herausgeholt" hat.

Plausibel erscheint, dass vor allem *komplexe Tätigkeiten* beziehungsweise *Management-Tätigkeiten* für Außenstehende teilweise intransparent bleiben, so dass hier Selbst- und Fremdbild am wenigsten übereinstimmen (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002). Die Personen, die schon länger zusammenarbeiten, sollten besser übereinstimmen, da sie bereits mehr Gelegenheiten zum Austausch – auch über Entscheidungsprozesse – hatten. Weiter sollte in den Bereichen, die umfassender beobachtet werden können (z.B. *umfeldbezogene/interpersonelle* Dimension) sowie in *verhaltensbezogenen* Ratings höhere Übereinstimmung erfolgen. Dies erfuhr bisher tendenzielle Bestätigung (Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002).

3.4.1.3 Unterschiedliche Definitionen von Führung

Es ist unbestreitbar, dass Führungskräfte – quasi per definitionem – vielfältige Erfahrungen mit dem zu beurteilenden Gegenstand – dem Führungsverhalten – machen; anders als die Mitarbeiter, die zwar auch viele, aber weniger qualitativ unterschiedliche Führungssituationen erleben. Dies liegt vor allem darin begründet, dass Führungskräfte in ihrer bisherigen Laufbahn als Führungskraft bereits verschiedene Mitarbeiter mit verschiedenen Anliegen und Zielen in verschiedenen Situationen geführt, während ein Mitarbeiter nur die Führung in seinem eigenen Handlungsraum erfährt. Zudem haben Führungskräfte als Akteure der Führung eine andere, komplexere Perspektive auf Führungsprobleme als Mitarbeiter als Zielpersonen der Führung. Diese unterschiedlichen Erlebniswelten könnten unterschiedliche Definitionen von Führung allgemein oder von einzelnen Führungsdimensionen provozieren. Während für einen Mitarbeiter beispielsweise auch umfassende Einarbeitung in ein neues Arbeitsgebiet inklusive der Einweisung in entsprechende informelle Wege zu Führung gehört, versteht eine Führungskraft darunter möglicherweise nur die klare Aufgabenverteilung. Klimoski und London (1974) führen es auf unterschiedliche, implizite Leistungsdimensionen zurück, dass in ihrer Studie Vorgesetzte von Krankenschwestern Items zu deren Kompetenz nicht von Items zu deren Anstrengung

trennen können – im Unterschied zu Peers und den Krankenschwestern selbst. Diese Erklärung kann auf erste Überlegungen von Guion (1965) zurückgeführt werden.

Vor allem bei nicht-spezifischen, eher abstrakten Ratings gründen Selbst- und Fremdbeurteiler ihre Leistungsbeschreibung auf unterschiedliche Facetten (Borman, 1974). Unterschiedliche implizite Leistungstheorien, aufgrund derer unterschiedliche Verhaltensmuster zur Beurteilung einer allgemeinen Leistung herangezogen werden, begünstigen dies. Zwangsläufig ergibt sich nur eine geringe Korrelation zwischen Selbst- und Fremdbeschreibung. Analog gilt, dass Fremdbeurteilungen umso ähnlicher ausfallen sollten, je ähnlicher die Erlebniswelten von Führung, aus denen die jeweiligen Definitionen generiert werden, sind. Insgesamt besteht eine unklare Ergebnislage (Borman, 1997).

Anzunehmen ist, dass komplexe *Tätigkeiten* (z.B. Führung) eine größere Bandbreite an Definitionen zulassen, so dass deren Moderatorwirkung (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002) erklärbar ist. Außerdem ist denkbar, dass im Zuge mehrfacher Beurteilungen eine Annäherung der Definitionen erfolgt, so dass auch die moderierende Wirkung von *Bewertungserfahrung* (Mabe & West, 1982) plausibel wird. Da Peers sowie Vorgesetzte selbst auch Erfahrung mit Führungsverhalten haben, müssten deren Urteile aber – entgegen der meta-analytischen Befunde (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988) – eng mit dem Selbstrating der Führungskraft verbunden sein. Weiter sollte der Spielraum für unterschiedliche Definitionen durch konkrete *verhaltensbezogene* Operationalisierungen der Items genauso reduziert werden wie durch die Vermeidung *globaler* Items. Dies erfuhr tendenzielle Bestätigung in Meta-Analysen (Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002). Eine Angleichung der Definitionen sollte auch mit der Dauer der Zusammenarbeit und Zugehörigkeit zur Organisation steigen. Weiter sollte sie bei ähnlichen Ideal-Vorstellungen der Beurteiler und bei Kenntnis der Führungsgrundsätze höher sein. Niveau-Unterschiede zwischen Selbst- und Fremdratern sind nicht plausibel erklärbar.

3.4.1.4 Unterschiedliche Gewichtung von Führungsverhaltensweisen

Auch wenn man davon ausgeht, dass Führungskräfte und Mitarbeiter (beziehungsweise Vorgesetzte) prinzipiell die gleichen Aspekte heranziehen oder aufgrund der Itemvorgaben heranziehen müssen, besteht die Möglichkeit, dass sie diese unterschiedlich stark gewichten (vgl. Borman, 1997). Persönlich Wichtiges wird mit höherer Wahrscheinlichkeit in die Itembeantwortung aufgenommen als Unwichtiges (vgl. Sudman, Bradburn & Schwarz, 1996). Den Führungskräften selbst ist die Personalführung eine Aufgabe unter vielen, neben Kundenkontakten oder Kontakt nach oben. Selbst wenn sie der Personalführung

höhere Priorität als ihren anderen Aufgaben einräumen, kann vor allem bei steigenden Führungsspannen, die die Verschlankung der Hierarchien mit sich bringt, für die Führung jedes einzelnen Mitarbeiter nicht viel Zeit aufgebracht werden. Die Führungskräfte haben entsprechend reduzierte Erwartungen an ihr Führungsverhalten und bewerten sich auch bei mittelmäßiger Leistung schon sehr positiv.

Den Mitarbeitern ist hingegen aus ihrer individuellen Perspektive heraus die Art, wie sie geführt, gefördert oder beurteilt werden, so wichtig, dass sie gute Noten nur für den verteilen, der den hohen Ansprüchen auch wirklich genügt. Wenn Führungskräfte einer Hierarchieebene aufgrund ihrer ähnlichen Führungs- und Ergebnisverantwortung andere Prioritäten herausbilden als Personen über- oder untergeordneter Hierarchien (Tsui & Ohlott, 1988), dann sollte die Korrelation zwischen Selbst- und Peerbeurteilungen deutlich höher ausfallen als andere. Dies ist jedoch nicht der Fall (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988); möglicherweise greifen auf Peer-Ebene aber motivationale Faktoren besonders stark, Kollegen nicht „zu gut“ bewerten zu dürfen. Die Selbst-Fremd Korrelation sollte zudem für globale Ratings deutlich niedriger sein als für differenzierte Ratings, da erstere auf mehreren Facetten beruhen und diese dann mit den jeweils unterschiedlichen Gewichten versehen werden können. Es zeigt sich jedoch kein Unterschied zwischen den Korrelationen hinsichtlich globaler versus differenzierter Ratings (Harris & Schaubroeck, 1988 untersuchten in ihrer Meta-Analyse die Selbst-Abwärts-Korrelation).

Betrachtet man anstatt der perspektiven- und zeitbedingten Gewichtungen eher solche, die auf individuellen Vorlieben basieren, dann sind die geringen Selbst-Fremd-Korrelationen besser erklärbar. Es ist bekannt, dass positives Feedback dazu führt, dass der Feedbackgegenstand in seiner Wichtigkeit aufgewertet wird (Rost, Frey & Rost-Schaude, 1975) beziehungsweise dass konforme Informationen ("gute Nachrichten") für wichtiger gehalten werden als andere (Kahneman, Slovic & Tversky, 1982). Wenn eine Führungskraft beispielsweise immer wieder positive Rückmeldung zur Art und Weise ihrer Mitarbeitergespräche bekommt, wird sie dieser allmählich immer mehr Wichtigkeit zusprechen. Entsprechend könnten Führungskräfte die Dimensionen für wichtig erachten, auf denen sie sich besonders gut bewerten oder sie annehmen, dass sie von anderen darin gut bewertet würden (konsistenzerhaltend und selbstwerterhöhend). Mitarbeiter sprechen hingegen eher den Dimensionen besondere Wichtigkeit zu, die sie sich besonders wünschen (z.B. weil ein bestimmtes Führungsverhalten als fehlend/mangelhaft wahrgenommen wird). Items von hoher Wichtigkeit führen aber aus der Fremdperspektive zu hohen Erwartungen,[88] die nur schwer erfüllt werden können, da die Leistung an einem hohen impliziten Referenzwert gemessen wird. So fällt die Selbstbeurteilung auf wichtigen Dimensionen eher positiv, die Fremdbeurteilung eher negativ aus. Wenn nicht einmal die gleichen Dimensionen als wichtig erachtet werden, verwundert die geringe Selbst-Fremd-Korrelation nicht.

[88] Aus der Selbstperspektive sind hohe Erwartungen mit der angenommenen hohen Selbstbewertung erfüllt.

Diese Erklärung sagt voraus, dass sowohl für *globale* als auch für *dimensionale* Skalen eine Diskrepanz entstehen kann. Entsprechend finden die Meta-Analysen auch keine Moderatorwirkung (Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002). Außerdem erklärt die Wichtigkeitshypothese, weshalb gerade für *Management*-Beurteilungen nur geringe Zusammenhänge zu erzielen ist (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002). Denn die *Komplexität* des Gegenstands lässt einen großen Spielraum für unterschiedliche Gewichtungen, die durch unterschiedliche soziale Kategorisierung (vgl. Kapitel 3.4.1.7) zudem gefördert werden. Dass weiter bei *relativen* Selbstbewertungen höhere Übereinstimmungen erzielt werden als bei absoluten (Mabe & West, 1982), kann daran liegen, dass dann weniger die Wichtigkeit als Referenzpunkt herangezogen wird, sondern vielmehr die gegebene Vergleichsgruppe. Bei *Leistungen* (versus Fähigkeiten) sowie bei *vergangenen* (versus zukünftigen) Leistungen, die genauer eingeschätzt werden (Mabe & West, 1982), gelingt möglicherweise auch die Einschätzung der Wichtigkeit besser. Da außerdem *umfeldbezogene* Leistungen im interpersonellen Kontext stattfinden, könnte ein vermehrter Austausch auch (implizit oder explizit) zu einer besseren Einigung über die Wichtigkeit führen – und schließlich zur gefundenen geringeren Diskrepanz zwischen Selbst- und Abwärtsbild (Heidemeier & Moser, 2002).

3.4.1.5 Unterschiedliche Differenzierungen von Selbst- und Fremdbild

Manager greifen bei ihrer Selbstbeurteilung insgesamt stärker auf spezielles als globales Leistungsverhalten zurück (Scullen, Mount & Goff, 2000, siehe S. 178). Da jedes Selbstkonzept differenzierter ist als ein Fremdbild, können Führungskräfte ihre Selbstbeurteilung auf spezielleren Aspekten gründen als Fremdbeurteiler. Zur Beurteilung mancher Führungsdimensionen – zum Beispiel Kommunikationsverhalten, Kritikfähigkeit oder Durchsetzungsvermögen – können zudem arbeits-irrelevante Selbstkonzepte (multiple Selbstkonzepte, Markus & Wurf, 1987) relevant werden. So ist die Selbstbeschreibung aber möglicherweise von einem allgemeinen Selbstbild gestützt, während es den Fremdbeobachtern besser gelingt, sich lediglich auf arbeitsbezogene Situationen zu beziehen. Wenn jedoch auf unterschiedliche Information zurückgegriffen wird, korrelieren Selbst- und Fremdbild in der Folge kaum. Durch die Aktivierung mehrerer Selbstkonzepte steht dem Selbstbeurteiler auch ein größerer Pool an Informationen zur Verfügung, woraus er – im Sinne der Selbstwerttheorie – zunächst diejenigen erinnert und abruft, die positiv sind. Und: Es wird nicht jede relevante Information zur Urteilsfindung bzw. zur Beantwortung von Items herangezogen. Indes geht es vielmehr um einen Suchprozess, der vorzeitig abgebrochen wird, sobald ausreichend Information zur Itembeantwortung verfügbar ist (vgl. Bodenhausen & Wyer, 1987). So ist auch der Niveau-Unterschied erklärbar.

Nur durch Items, deren Formulierung ausschließlich berufsbezogene Inhalte aktiviert, also durch sehr *spezifische Operationalisierungen*, lässt sich diese Verzerrung abschwächen (vgl. Fragestellung 4). Doch selbst dann kennt eine Führungskraft von sich selbst mehr Facetten (in Interaktion mit einer Vielzahl an Kollegen etc.) als andere, die hauptsächlich nur ihre eigene Interaktion mit der Fokusperson beobachten.[89] Je länger die Fremdbeurteiler bereits mit der Führungskraft zusammenarbeiten und je mehr Kontakt sie zu ihr haben oder Führungsverhalten aus eigener Führungserfahrung kennen, umso stärker sollte die Selbst-Fremd-Übereinstimmung sein. Gegen diese Hypothese spricht, dass Selbstbeschreibungen eine geringere Streuung aufweisen als Fremdbeschreibungen (z.B. Fleenor, McCauley & Brutus, 1996).

3.4.1.6 Akteur-Beobachter-Verzerrung

Der "Actor-Observer-Bias" besagt, dass die Handelnden dazu tendieren, ihr Verhalten auf situationale Bedingungen zu attribuieren, während Beobachter dazu tendieren, das Verhalten (des Handelnden) auf die Person des Handelnden zu attribuieren (Jones & Nisbett, 1972). Jones und Nisbett führen die diskrepante Verhaltensbeschreibung darauf zurück, dass den Handelnden andere und detailliertere Informationen (z.B. über die Umstände und die handlungsleitenden Motive) verfügbar sind als den Beobachtern und dass ihnen unterschiedliche Informationen salient sind. Während für den Akteur eher situationale Charakteristika und Anforderungen (also auch Rollenanforderungen) salient und damit handlungsauslösend erscheinen, nimmt der Beobachter das Verhalten des Handelnden selbst als "Figur" vor dem "Grund" der Situation wahr und stützt seine Kausalattribution folglich eher auf saliente Charakteristika der handlungsausführenden Person. Informationen über Handlungsabsichten und deren Zusammenhang zu Umweltstimuli kennt er in der Regel gar nicht. So attribuieren Akteur (Führungskraft) und Beobachter (Vorgesetzte oder Mitarbeiter) unterschiedlich, was zur Folge hat, dass ihre Urteile gering korrelieren. Denn ob beobachtete Verhaltensweisen beziehungsweise ihre Ergebnisse eher auf die Situation oder Person attribuiert werden, wirkt sich beispielsweise darauf aus, in welchem Ausmaß der Führungskraft bestimmtes Verhalten überhaupt zugesprochen wird oder darauf, wie differenziert der zur Beurteilung herangezogene Verhaltenspool ist (siehe oben, Kapitel 3.4.1.5). Verschiedene Beobachter sollten in ihren Urteilen jedenfalls besser übereinstimmen als Akteur und Beobachter.

[89] Analoge Information gewinnt man aus der Literatur zu Assessment Centern: Shore, Tetrick und Shore (1998) stellen fest, dass Assessees (im Gegensatz zu Beobachtern und Peers) für die Selbstbeurteilung ihrer Management-Potentiale auf Information zurückgreifen, die nicht im Assessment Center generiert wurde.

Legen die salienten Merkmale des Handelnden nahe, dass die Beurteilung schlechter als gerechtfertigt ausfällt, kommt es zu einem Niveau-Unterschied. Exemplarisch dafür können die Ergebnisse von Becker et al. (2002) angeführt werden, die Selbst-Aufwärts-Diskrepanzen von Führungsbeurteilungen untersuchen: *Weibliche* Führungskräfte weisen in männlichen Domänen (z.B. Industrie vs. Bildungseinrichtung) höhere Niveau-Unterschiede[90] zu Mitarbeiterratings auf als männliche Führungskräfte (oder als weibliche Führungskräfte in rollenkonformen Einrichtungen, z.B: Bildungseinrichtung). Die Autoren begründen dies damit, dass den weiblichen Führungskräften selbst eher ihr Managementverhalten salient ist, während ihren Mitarbeitern eher das Geschlechtsmerkmal "Frau" salient ist. Da Letzteres Erwartungen weckt, die von der Rolle der Frau gespeist werden und die sich von den Erwartungen an die Rolle eines Managers unterscheiden, wird eine weibliche Führungskraft in ihrem Führungsverhalten nicht als "angemessen weiblich" beurteilt. Dass im Unterschied dazu in der Meta-Analyse von Heidemeier und Moser (2002) für Geschlecht kein moderierender Effekt gefunden werden konnte, liegt an der Bandbreite an Settings, in denen die fokussierte Leistung erbracht wurde. Denn nicht überall sticht das Geschlecht als salientes Merkmal hervor.

Vor dem Hintergrund dieses Erklärungsansatzes erscheint auch die Moderatorwirkung von *Kultur* sehr plausibel: Denn in asiatischen, kollektivistischen Kulturen findet man sowohl eine geringere Selbst-Fremd-Diskrepanz (Heidemeier & Moser, 2002) als auch eine geringere Anfälligkeit für die Korrespondenz- und Akteur-Beobachter-Verzerrung. Situationale Information wird stärker berücksichtigt (Choi & Nisbett, 1998; Choi, Nisbett & Norenzayan, 1999). Insgesamt ist jedoch auch für diesen Erklärungsmechanismus (Akteur-Beobachter-Verzerrung) die aktuelle Befundlage hinsichtlich Selbst-Fremd-Übereinstimmung von Leistungsbeurteilungen unzureichend.

3.4.1.7 Sozialer Vergleich und soziale Identität

Die soziale Vergleichstheorie (Festinger, 1954) und Weiterentwicklungen zur Theorie der sozialen Identität (Tajfel & Turner, 1986; Turner, 2000) können einen Erklärungsrahmen für die Diskrepanzen zwischen Selbst- und Fremdbild bieten, der über die oben skizzierten Moderatoreffekte hinausgeht. Mabe und West (1982) argumentieren als eine der ersten mit sozialen Vergleichsprozessen, um die geringe Validität von Selbstbeschreibungen zu erklären.

Möglicherweise ergibt sich die geringe Kongruenz zwischen Selbst- und Fremdbild daraus, dass diese auf der Basis verschiedener Vergleichsprozesse generiert werden. Im

[90] Das in der Studie verwendete Differenzmaß lässt es nicht zu, die Richtung der Niveau-Unterschiede zu bestimmen.

Unterschied zu Festingers ursprünglicher Annahme, dass soziale Vergleichsprozesse dem Motiv dienen, sich akkurater einschätzen zu wollen, können sie auch dem Motiv nach Selbstwertschutz beziehungsweise –erhöhung entspringen. So kann auf Strategien des Abwärtsvergleichs (Wills, 1981) oder der Abwertung von Vergleichspersonen zurückgegriffen werden beziehungsweise der Vergleich komplett vermieden werden (für eine Übersicht Dauenheimer et al., 2002; Hoog, 2000). Versucht man, diese Vergleichsmöglichkeiten auf multiperspektivische Beurteilungen zu übertragen, dann ist die Vermeidung eines Vergleichs am ehesten denen zuzuschreiben, die nicht an der Befragung teilnahmen. Die teilnehmenden Führungskräfte (Selbstbeschreiber), die sich als Führungskräfte (Ingroup) kategorisieren und sich nach unten vergleichen,[91] werden sich selbst deutlich abheben, indem sie sich besonders positiv beschreiben (soziale Distinktheit). Wenn sie andere in ihren Fremdbeschreibungen abwerten, dann resultiert daraus letztlich das Above-Average-Symptom (vgl. Kapitel 3.2.3).

Selbstbeurteiler wählen ihre Vergleichsperson also in Abhängigkeit von selbstwertrelevanter Motivation und vergleichen sich nicht immer mit relevanten Anderen. Außenstehende vergleichen das zu beurteilende Verhalten hingegen eher mit dem relevanter Personen. Die Varianz in den bemühten Referenzpunkten erklärt, weshalb Selbst- und Fremdbild kaum, verschiedene Fremdbilder jedoch beachtlich korrelieren. Denn abhängig vom Vergleichsniveau fällt die Selbsteinschätzung besser oder schlechter aus. Plausibel erscheint dies vor dem Hintergrund der potentiellen Vergleichsmöglichkeiten: Während Führungskräfte in der Regel nur wenige relevante interindividuelle Vergleichsmöglichkeiten haben (denn welche Führungskraft lässt sich schon gerne ins Handwerk schauen?), um ihr eigenes Führungsverhalten im Sinne der sozialen Vergleichstheorie adäquat einzuordnen, verfügen beispielsweise ihre Vorgesetzten gewöhnlich über mehr Referenzpunkte, auf die ihre Abwärtsbeschreibung gestützt sein kann. Zumindest in Linienorganisationen haben sie mehrere Mitarbeiter mit Führungsverantwortung.

Zudem zeigt sich experimentell, dass Personen mit hohem Selbstwert soziale Vergleiche insbesondere nach Misserfolgen anstreben, während Personen mit niedrigem Selbstwert eher nach Erfolgsmeldungen Vergleiche zu anderen ziehen (Wood, Giordano-Beech, Taylor, Michela & Gaus, 1994). Geht man aber davon aus, dass Führungskräfte mehrheitlich einen hohen Selbstwert haben und dass ihnen selten Misserfolge (im Führungsverhalten) zurück gemeldet werden, dann stellen sie vergleichsweise wenig soziale Vergleichsprozesse an.

Weiter kann anhand solcher Modelle, die auf der Theorie des sozialen Vergleichs oder der sozialen Identität basieren, die Wirkung verschiedener Moderatorvariablen erklärt wer-

[91] Auf einen Abwärtsvergleich könnten sie ausweichen, weil sie sich erstmalig mit der Beurteilung durch ihre Mitarbeiter konfrontiert sehen. Dies könnte als Bedrohung des Status Quo und damit als Selbstwertbedrohung wahrgenommen werden. Letztere gilt als Voraussetzung für einen Abwärtsvergleich (vgl. Hoog, 2000).

den: Unmittelbar angesprochen werden *relative Selbstratings im Vergleich zu anderen* (Mabe & West, 1982), bei denen eine annähernde Standardisierung der Vergleichsgruppe besteht (und so beispielsweise Abwärtsvergleiche ausgeschlossen sind). Eine Beurteilung vor dem gleichen Referenzrahmen (gleiche Vergleichsinformation über die Leistung anderer) zeigt auch im Labor bei der Leistungsbeurteilung hinsichtlich Korrekturlesens bessere Übereinstimmungen zwischen Selbst- und Abwärtsbeurteilungen (durch den Strohmann in der Rolle des Chefredakteurs) sowie zwischen Selbstbeurteilung und objektiver Leistung (Farh & Dobbins, 1989). Für *Leistungen* (versus Fähigkeiten) bieten sich insgesamt mehr Gelegenheiten zum Vergleich, so dass die Referenzgruppe repräsentativer und dadurch die Einschätzung genauer wird. Durch *Anonymitäts-* sowie *Validierungserwartungen* oder *Verifizierbarkeit* werden alle selbstwertmotivierten Vergleichsstrategien eingeschränkt, so dass die Diskrepanzen für alle genannten Variablen sinken (teilweise Heidemeier & Moser, 2002; Mabe & West, 1982). Außerdem ist die Übereinstimmung geringer bei der Einschätzung *zukünftiger* (versus vergangener) Leistung, was damit erklärt werden kann, dass auch Fremdbeurteilern für zukünftige Leistung keine besseren Vergleichsmöglichkeiten zugänglich sind.

Begibt man sich auf Gruppenebene, können in einer Organisation die zwei großen Gruppen der Führungskräfte und Mitarbeiter unterschieden werden. Liegt eine derartige soziale Kategorisierung und Identität vor, dann müssten Vorgesetzte ihre unterstellten Mitarbeiter mit Führungsverantwortung (deren Förderer sie zudem sind) in ihrem Leistungsverhalten besser bewerten als ihre unterstellten Mitarbeiter ohne Führungsverantwortung (Ingroup-Favorisierung zum Erhalt positiver sozialer Identität[92]). Geringere Niveau-Unterschiede zwischen Selbst- und Abwärtsbeurteilung für die *Management*-Stichprobe (versus Arbeiter-Stichprobe) wären die Folge – genauso wie es Heidemeier und Moser (2002) berichten. Während in der Management-Stichprobe nämlich sowohl Abwärts- als auch Selbstbeurteiler der Gruppe der Führungskräfte angehören, sind in der Arbeiter-Stichprobe lediglich die Abwärtsbeurteiler Führungskräfte. Die sich selbst beurteilenden Arbeiter tragen keine Führungsverantwortung und werden von den abwärts beurteilenden Führungskräften anders sozial kategorisiert. Um soziale Distinktheit zu schaffen, erzeugen die Abwärtsbeurteiler allein aufgrund der Kategorisierung eine höhere Distanz zu Arbeitern (ohne Führungsverantwortung) als zu Mitarbeitern mit Führungsverantwortung (Managern). In der vorliegenden Untersuchung sind alle Fokuspersonen von Abwärtsbeurteilungen Führungskräfte.

Nach dem gleichen Prinzip müssten allerdings auch Aufwärtsbeurteiler mit eigener Führungsverantwortung (versus Aufwärtsbeurteiler ohne Führungsverantwortung) ihre

[92] Eine positive soziale Identität wird deshalb angestrebt, weil dadurch auch das personale Selbstkonzept aufgewertet werden kann. Gruppenidentität und persönliche Identität sind nicht voneinander unabhängig.

Vorgesetzten besser bewerten. Zudem wird erklärbar, weshalb Abwärts- und Peerratings zum Führungsverhalten einer Führungskraft höher miteinander korrelieren als sie jeweils mit Aufwärtsratings korrelieren (Scullen, Mount & Goff, 2000): Während Abwärts- und Peer-Rater der selben sozialen Kategorie angehören (Führungskräfte) und deshalb keine Distinktheit schaffen wollen, tragen Aufwärtsbeurteiler häufig keine Führungsverantwortung und vergrößern durch ihr Rating die Distanz. Außerdem könnte aus einer negativen sozialen Identität (z.B. *nur* Führungskraft der vierten Ebene) sogar die Unterschätzung eigener Fertigkeiten resultieren. Auch wenn die abgeleiteten Hypothesen plausibel erscheinen, stehen Studien aus, die die Selbst-Fremd-Diskrepanzen von Führungsbeurteilung auf der Grundlage dieser Theorien untersuchen.

3.4.2 Idiosynkratisch bedingte Diskrepanzen

Bevor im Folgenden potentielle idiosynkratische Verzerrungen dargestellt werden, soll angemerkt sein, dass diese nicht gänzlich losgelöst von der Perspektive zu betrachten sind.

3.4.2.1 Antworttendenzen und Urteilsfehler

Große Anteile der idiosynkratischen Verzerrungen basieren auf den **Antworttendenzen**, die seit langem in der Literatur beschrieben werden (vgl. Brandstätter, 1970). Die wichtigsten werden nun kurz angerissen.

Erstens sind *Mittelwertstendenzen* zu nennen, was vereinfacht bedeutet, dass die Urteile besser (leniency) oder schlechter (severity) ausfallen als es die "wahre" Leistung des zu Beurteilenden erlaubt. Leniency scheint bei Selbstbeurteilern fast regelhaft zu sein und ist mit nahezu allen skizzierten Erklärungsmechanismen vereinbar. Möglicherweise vergrößert sich die Diskrepanz durch Severity-Ratings der Fremdbeurteiler.

Eine zweite Gruppe an Urteilstendenzen sind die *Streuungstendenzen* (auch Häufigkeitstendenzen), die besagen, dass sich Ratingantworten bei bestimmten Skalenwerten häufen und dadurch den Range verengen. So gibt es beispielsweise die Tendenz zur Mitte oder die extreme Verlagerung der Werte zu einem Skalenpol. Möglicherweise führt bei solchen Ratings, wie sie in der vorliegenden Studie eingesetzt wurden, eine Ja-Sage-Tendenz (Akquieszenz) zu einer Verlagerung der Antworten zum positiven Skalenende. Geringe Informationsbasis, ungenügende Differenzierungsleistung oder besondere Wichtigkeitszuschreibungen können solche Streuungstendenzen beispielsweise bei den Selbstbeurteilern (Führungskräften) unterstützen und dadurch die Korrelation mit Fremdbeurteilern mindern.

Als Beispiel für die dritte Verzerrungsgruppe – die *Korrelationstendenzen* – sei der Halo-Effekt angeführt. Die unbegründet hohe Korrelation verschiedener Beurteilungsdimensionen kommt dadurch zustande, dass ein Eindruck bestimmter Merkmale die anderen sozusagen "überstrahlt". Zur Rechtfertigung "wahrer" Zusammenhänge wird auf Fußnote 84 verwiesen. Korrelationstendenzen bei Selbst- oder Fremdbeurteilern wirken sich auch auf deren Interkorrelation und die Validität der Ratings aus.

Aus der kognitiven Sozialpsychologie sind **Urteilsfehler** bekannt (für einen Überblick Strack & Deutsch, 2002), die auch bei Ratingantworten auftreten können. So kann ein *Anker* (Startwert einer Schätzung) *mit anschließender ungenügender Adjustierung* die Ratingantworten in Richtung Anker verzerren. Bei der vorliegenden Untersuchung könnte also die Urteilsverzerrung durch implizite Verankerung in Ideal- oder Wichtigkeitswerten begünstigt werden.

Die *Verfügbarkeitsheuristik*, nach der unter anderem Häufigkeiten aufgrund guter kognitiver Verfügbarkeit von Beispielen nach oben verzerrt werden, kann die Unterschiede zwischen Selbst- und Fremdbeurteilung dadurch erklären, dass selbstbezogene Information besser verfügbar ist als andere Information. Den Führungskräften sind also ihre eigenen Führungsverhaltensweisen (selbstbezogene Information) relativ gut verfügbar. Fremdbeurteiler müssen indes Information über das Verhalten anderer – der fokussierten Führungskraft – abrufen, was weniger verfügbar als selbstbezogenes Verhalten ist. Besser verfügbare Information erscheint dem Beurteiler aber häufiger, so dass Selbstbeurteiler ihr Führungsverhalten besser bewerten als Fremdbeurteiler das Führungsverhalten der Fokusperson. Ross und Sicoly (1982) nennen vier Faktoren, die deren Verfügbarkeit erhöhen. Dies sind die selektive Enkodierung und Speicherung von selbstbezogener Information, der selektive Erinnerungs-/Abrufprozess, die höhere Quantität an Wissen (die eine Führungskraft im Vergleich zu ihren Fremdratern über ihr Führungsverhalten hat) sowie motivationale Einflussfaktoren wie das Motiv zur Selbstwerterhöhung.

Der *Rückschaufehler* ("hindsight bias", Hawkins & Hastie, 1990) beschreibt, dass im Rückblick, nachdem die Lösung für eine Aufgabe präsentiert wurde, die Aufgabe so erscheint, als hätte man sie mit hoher Wahrscheinlichkeit lösen können. Für Fragebogenuntersuchungen, beispielsweise in der vorliegenden Arbeit, kann der Rückschaufehler insofern eine Rolle spielen, dass Iteminhalte, die das erwünschte Führungsverhalten beschreiben (analog zum Lösungsvorschlag für eine Aufgabe), eher eine überhöht positive Beantwortung im Sinne von "Aufgabe leicht lösbar" provozieren.

Derartige Verzerrungen sind zweifellos nicht auf Selbstbeschreiber beschränkt. Je nach Richtung der Antworttendenz bei Fremdbeurteilern wird dadurch die Diskrepanz erhöht oder gemindert. Aus verschiedenen Gründen (siehe Seite 151) kann angenommen werden, dass die diskrepanten Einschätzungen stark von der sich selbst beschreibenden Führungs-

kraft bestimmt werden. Daher werden überwiegend potentielle Verzerrungen dieser Beurteiler vorgestellt, die die Diskrepanz zu Fremdbeschreibungen hervorrufen/erhöhen können. Diese können sowohl idiosynkratisch als auch perspektivenbedingt sein. Sie sind hier im Kapitel zu idiosynkratischen Erklärungen aufgeführt, weil sie für einzelne Führungskräfte vermutlich stark variieren. Im Übrigen sind idiosynkratische von perspektivenbezogenen Einflussfaktoren für die Ratergruppe der Selbstbeurteiler (um die es im Folgenden geht) auch methodisch nicht zu trennen (Scullen, Mount & Goff, 2000).

3.4.2.2 Mangelndes Feedback

Häufig lässt sich beklagen, dass Führungskräfte von anderen kein Feedback über ihr Führungsverhalten bekommen (Longenecker & Gioia, 1992). Während der „Normalbürger“ immer wieder zu spüren bekommt, wie angemessen oder unangemessen er sich verhalten hat, machen Führungskräfte diese Erfahrungen seltener. Das bedeutet umgekehrt, dass sie seltener dazu „gezwungen“ werden, ihr Selbstbild zu reflektieren und zu aktualisieren. Dies gilt vor allem vor dem Hintergrund, dass negatives Feedback noch seltener gegeben wird als positives (z.B. Tesser & Rosen, 1975) und dass es für weniger informativ gehalten wird (Ilgen, Fisher & Taylor, 1979). Dass Menschen aber überwiegend positive Rückmeldung erhalten und suchen (Brown, 1986), wird nicht nur durch soziale Normen gesteuert, die Rückmeldungsprozesse reglementieren, sondern auch durch selbst initiierte Interaktionsstrategien, die das Gegenüber dazu auffordern, das bestehende positive Selbstkonzept zu bestätigen. So haben beispielsweise viele Außenstehende die Erfahrung gemacht, dass das Geben von positivem Feedback mit positiven Konsequenzen (Lob, gute Beziehung etc.) durch den Feedbacknehmer belohnt wird, was gemäß der Lerntheorien die Auftretenswahrscheinlichkeit positiven Feedbacks erhöht (und umgekehrt für negatives). Ohne kritische Aufwärtskommunikation bleibt dezentral vorhandenes Wissen aber ungenutzt. Dies betrifft nicht nur Wissen über Führung, sondern auch Lösungsvorschläge, die direkt Arbeitserfolg oder -zufriedenheit verbessern könnten. Insofern kann eine fehlende (kritische) Aufwärtskommunikation unmittelbar die Leistung mindern beziehungsweise vorhandene Potentiale unausgeschöpft lassen.

Weiter kann vermutet werden, dass sie auch die in der vorliegenden Studie betrachtete Diskrepanz herbeiführen kann. Alle Feedbackmodelle – vom plakativ-plausiblen Johari-Fenster (Luft, 1961, 1963)[93] bis hin zu etwas komplexeren Modellen (z.B. London &

[93] Das Johari-Fenster ist in vier Quadranten geteilt, die sich durch die Kombination aus Aktivitäts-Bereichen ergeben, die der Fokusperson selbst beziehungsweise anderen bekannt beziehungsweise unbekannt sind. Die Verhaltensweisen und Motivationen, die andere kennen, der Fokusperson selbst allerdings unbewusst sind, charakterisieren den so genannten "Blinden Fleck". Durch eine Gegenüberstellung von Selbst- und Fremdbild soll dieser verkleinert werden. Eine Weiterentwicklung stellt – wenn es auch nicht als solche bezeichnet

Smither, 1995)[94] – fokussieren darauf, dass Feedback zu einer höheren Übereinstimmung zwischen Selbst- und Fremdbild führt. Interventionsstudien legen dies ebenfalls sehr nahe (siehe auch Kapitel 3.6). Deshalb beklagen Longenecker und Gioia (1992), dass Führungskräfte immer weniger Feedback erhalten, je höher sie in der Hierarchie steigen. Dies betrifft also gerade diejenigen, die die unstrukturiertesten und schwierigsten Tätigkeiten eines Unternehmens übernehmen.

Doch selbst wenn negatives Feedback geäußert wird, unterliegt es weiteren verzerrenden Adaptationsprozessen bei der Enkodierung oder dem nachträglichen Abruf. Denn am besten wird diejenige Information wahrgenommen, enkodiert, erinnert und abgerufen, die selbstwertdienlich beziehungsweise konsistent mit vorher bestehenden Selbstkonzepten ist (vgl. Modell von London, 1994). So kann Feedback – insbesondere wenn es zweideutig formuliert ist – umgedeutet (Jacobs, Berscheid & Walster, 1971) oder sogar ignoriert werden. Entsprechend finden Shrauger und Schoeneman (1979) in ihrer Übersichtsarbeit einen deutlichen Zusammenhang zwischen dem Selbstbild und der eigenen Einschätzung der Fremdurteile.

Fehlt Feedback gänzlich oder wird es umgedeutet, so bleibt derjenige Verhaltensbereich groß, den nur Außenstehende (nicht die Führungskraft selbst) wahrnehmen. Dieser sogenannte "Blinde Fleck" (Luft, 1961, 1963) macht niedrigere Selbst-Fremd- als Fremd-Fremd-Korrelationen wahrscheinlich, da unterschiedliche Aspekte in die Beurteilung einfließen. Gleichzeitig begünstigt er Niveau-Unterschiede, wenn man bedenkt, dass er überwiegend blind vor negativer Kritik macht (Schlenker, 1980). Führungskräfte überschätzen sich dann also aufgrund der Einseitigkeit an Informationsverteilung. Da sie bei anderen aber durchaus auch negativ zu bewertende Verhaltensweisen wahrnehmen, kann der Above-Average-Effekt entstehen. Der umgekehrte Fall, dass eine Führungskraft unerwartet viel negatives Feedback bekommt, kann dazu führen, dass sie sich in ihrer Führungsleistung *unterschätzt*.

Mit dieser Feedback-Hypothese ist erklärbar, dass sich die Beurteilten mit *Bewertungserfahrung* – also Geübte im Umgang mit Selbstbeurteilungen – in höherer Übereinstimmung mit anderen einschätzen (Mabe & West, 1982). Das gleiche gilt für *Arbeiter* (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002), denen (im Vergleich zu Management-Positionen) Feedback eher gegeben werden darf (soziale Norm). Mit dem Alter, der Dauer der Zusammenarbeit, einer geringen Führungsspanne

---

wird – das Modell von Waldman und Atwater (1998) dar, die aus der besseren Kenntnis des Selbst eine Leistungsverbesserung ableiten.

[94] Durch eine Wahrnehmung der Selbst-Fremd-Diskrepanzen kommt es nach dem Modell von London und Smither (1995) – unter dem Einfluss verschiedener Variablen (z.B. Selbstbild) – auch zur Wahrnehmung der Ist-Ziel-Diskrepanz. Dadurch kann eine Aktualisierung des Selbstbilds angestoßen werden, was wiederum in neuen Zielsetzungen und Veränderungsverhalten resultieren kann (in Abhängigkeit von z.B. Impression Management). Da sehr komplexe Konstrukte in das Modell aufgenommen sind, ist es schwer überprüfbar.

und mit zunehmender Erfahrung als Beurteilter sollte über den hier fokussierten Feedback-Mechanismus die Selbst-Fremd-Übereinstimmung außerdem steigen.

3.4.2.3 Selbsttäuschung und Eindrucksmanagement

Die Problematik der sozialen Erwünschtheit (Social Desirability) nimmt in der Literatur einen breiten Raum ein (Bortz & Döring, 2003; Mummendey, 2003). Dabei kann die Angst vor negativen Konsequenzen genauso eine Rolle spielen wie beispielsweise der Versuch, durch eine positive Selbstdarstellung einen guten Eindruck zu machen. Paulhus (1984) identifizierte zwei unterschiedliche Prozesse, Fragebogenantworten sozial erwünscht abzugeben. Dies ist zum einen das Eindrucksmanagement (Impression Management)[95], zum anderen die Selbsttäuschung (Self deception). Während Ersteres adressatenorientiert ist, indem Menschen das Bild, das andere von ihnen haben, mitzugestalten und zu kontrollieren versuchen,[96] wird Letzteres als generelles Phänomen der Selbstwahrnehmung gehandelt. Hierbei stellt sich die Person nämlich so dar, wie sie sich selbst sieht, ohne sich vor anderen bewusst in besserem Licht darzustellen. Dennoch wird unterstellt, dass diese Darstellung besser ausfällt als ihre „'tatsächliche' Persönlichkeit" (Mummendey, 2003, S. 202) und die Person folglich nicht nur andere, sondern auch sich selbst täuscht. Dieser Positivitätsbias wurde bereits als Erklärung für geringe Selbst-Fremd-Korrelationen diskutiert (Harris und Schaubroeck, 1988). Jeserich (1996) führt Selbst-Fremd-Diskrepanzen ausschließlich auf das Motiv nach Selbstwertschutz zurück und stellt – wie häufig in der entsprechenden Literatur ohne ausreichende Befundgrundlage – heraus, dass Führungskräfte dafür besonders anfällig sind.

Selbsttäuschung ist die Tendenz, vorteilhafte (aber ehrlich gemeinte) Selbstbeschreibungen abzugeben. Dies ist zumindest kurzfristig mit einem höheren Selbstwert sowie mit höherem psychischen Wohlbefinden verbunden (Paulhus & Reid, 1991; Taylor & Brown, 1988; für langfristige negative Konsequenzen: Paulhus, 1998). Ohne Zweifel sind Erfolge – bei Führungsratings also der jeweils positiv verankerte Pol einer Skala – selbstwertdienlicher als Misserfolge. Entsprechend ist die Kernaussage der Theorie der Selbstwerterhöhung und des Selbstwertschutzes (für einen Überblick Dauenheimer et al., 2002), dass Menschen grundsätzlich nach Erhöhung und Schutz ihres Selbstwerts streben. Dazu stehen ihnen zwei Techniken der Selbsttäuschung zur

[95] Impression Management und Selbstdarstellung werden in der vorliegenden Arbeit synonym verwendet, da die Ansätze der verschiedenen Selbstdarstellungsforscher nicht unterschieden werden. Autoren, die diesen Themenbereich genauer betrachten, verwenden „Selbstdarstellungstheorie" als Sammelbegriff für Impression Management- und verwandte Ansätze (z.B. Mummendey, 2002).

[96] Zur Diskussion über den Grad an zugrunde liegender Absicht: siehe Mummendey (2002). Schlenker (1980) integriert bewusste und unbewusste Versuche, das Selbstbild zu steuern, in seine Definition.

ihnen zwei Techniken der Selbsttäuschung zur Verfügung: Erstens können sie sich übermäßig positive Attribute zuschreiben, also ihr Können übertreiben (Enhancement), zweitens können sie negative Attribute verleugnen, also "ihre Sünden kleinhalten" (Denial).[97] Dabei scheint Enhancement eher mit Selbstwerterhöhung und Denial eher mit Selbstwertschutz verbunden zu sein (Paulhus & John, 1998). Wenn aber eine solch vorteilhafte Selbstbeschreibung genauso wie ein unrealistischer Optimismus oder wie Kontrollillusionen (bzgl. des Selbst) charakteristischer für Menschen ist als eine akkurate Selbstbeschreibung (für einen Überblick Dauenheimer et al., 2002; Taylor & Brown, 1988), dann ergeben sich ein positiveres Selbst- als Fremdbild (Niveau-Unterschied) und der in Kapitel 3.2.3 beschriebene Above-Average-Effekt. Hinzu kommt, dass positive selbstbezogene Informationen besser abrufbar sind als negative (Kuiper & Derry, 1982). Da ein großer Teil des Selbstwerts durch berufsbezogene Ereignisse geprägt ist, lassen sich die Kernaussagen dieser Selbstwert-Theorie zweifelsohne auf Arbeitsbeziehungen übertragen.

Daneben könnte eine geringe Selbst-Fremd-Korrelation dadurch erklärt werden, dass es (durch die Erhöhung der Selbstbeschreibung) zu einer Verringerung der Varianz der Selbstbeschreibung kommt und folglich zu einer geringeren Korrelation. Da eine Korrektur der Streubreite den Effekt aber nicht aufheben kann (Harris & Schaubroeck, 1988), müssen andere Hypothesen im Sinne der Selbstwert-Theorien generiert werden, um die geringen Korrelationen zwischen Selbst- und Fremdbewertung zu erklären.

Eine potentielle Ursache dafür kann in der Attributionsasymmetrie (Miller & Ross, 1975) liegen: Der Beurteilte selbst attribuiert Misserfolge extern und Erfolge intern (self-serving bias, z.B. Kernis & Grannemann, 1990), während Beobachter (Fremdbeurteiler) unter bestimmten Umständen umgekehrt vorgehen (Boski, 1983). Eine geringe Interkorrelation zwischen Selbst- und Fremdbild wäre die Folge, während Fremdbeobachter hoch miteinander korrelieren sollten. Zusätzlich sollten Erfolge – wenn sie hauptsächlich intern attribuiert werden – auch eher kognitiv verfügbar sein, zur Selbstbeschreibung herangezogen werden und diese insgesamt positiv verzerren.[98]

Neben selbstwertdienlichen Attributionsprozessen stehen dem Individuum weitere Strategien zur Verfügung, mit deren Hilfe das Bedürfnis nach Selbstwertschutz und –erhöhung befriedigt werden kann: die selektive Informationssuche, –verarbeitung und wiedergabe von selbstbezogener Information (vgl. London, 1994) oder verschiedene Strategien der Selbstdarstellung sowie des sozialen Vergleichs. Wenn aber auf diesem Wege unterschiedliche Informationen zur Selbst- und Fremdbeurteilung herangezogen werden, dann sinkt auch die Korrelation zwischen den Beurteilungen.

---

[97] Paulhus und Reid (1991) zeigen, dass diese beiden Taktiken nur für Selbsttäuschung getrennt werden müssen, während sie sich für Impression Management nicht trennen lassen.

[98] Die Erhöhung der kognitiven Verfügbarkeit ist allein aufgrund der Selbstwertdienlichkeit der Information zu erwarten. Sie sollte jedoch erhöht werden, wenn selbstwertdienliche Information zudem internal attribuiert wird.

Dies kann durch selbstwertschützende Motive der Fremdrater verstärkt werden. So können Mitarbeiter oder Vorgesetzte ihre Wahrnehmung und Beschreibung der fokussierten Führungskraft zu deren Vorteil verzerren, um ihr eigenes Selbstwertgefühl zu schützen. Denn wer gesteht sich beispielsweise gerne ein, seit Jahren bei einer Führungskraft zu arbeiten, die er als inkompetent beurteilt (Gefährdung der Selbstkonsistenz, vgl. Kapitel 3.4.2.4) beziehungsweise führungsverantwortlich für Führungskräfte zu sein, die schlechte Leistungen erbringen (Ingroup-Favorisierung). Dennoch geht es für Fremdbeurteiler nicht in erster Linie um den Selbstwert der Fokusperson, so dass andere (nicht dem Selbstwert der Fokusperson dienende) Informationen in der Regel ebenso leicht zugänglich sind und Fremdbeurteiler untereinander stärker übereinstimmen. Verschiedene zugrunde gelegte Informationen erklären dann die geringe Selbst-Fremd-Korrelation und – wenn die Information unterschiedlich valent ist – auch die Niveau-Unterschiede.

Die Selbstwerthypothesen erklären erstens die Wirkung der *kulturellen* Moderatorvariable (Heidemeier & Moser, 2002), da in kollektivistischen Kulturen im Unterschied zu individualistischen weniger auf Erhöhung und Schutz des Selbst als vielmehr der Gruppe geachtet werden sollte. Zweitens ist stimmig, dass zukünftige Verhaltensweisen ungenauer eingeschätzt werden als *vergangene* (Mabe & West, 1982), denn der unrealistische Optimismus besagt, dass die eigene Zukunft positiver prognostiziert wird als die Vergangenheit und für sich selbst besser als für andere (für einen Überblick, Taylor & Brown, 1988). Drittens wird plausibel, weshalb sich Menschen auf den unterschiedlichsten Dimensionen unrealistisch positiv beschreiben und dies nicht auf Führungskräfte beziehungsweise berufliche Leistung beschränkt bleibt. Geht man viertens davon aus, dass der Selbstwert von Arbeitern tendenziell niedriger (also schützenswerter und mit mehr Spielraum für eine Erhöhung des Selbstwerts) ist als der von Führungskräften,[99] dann wird auch erklärbar, weshalb Heidemeier und Moser (2002) finden, dass sich die Selbstbeschreibungen von *Führungskräften* im Niveau weniger stark von Abwärtsbeurteilungen unterscheiden als die Selbstbeschreibungen von Arbeitern.

Auch wenn feststeht, dass die Selbstbewertung in aller Regel stark zum positiven Ende einer Skala hin tendiert, kann ihr allein deshalb noch nicht das Attribut "unrealistisch" oder "Täuschung" zugesprochen werden. Dies wird allerdings durch relative Angaben evident. Erinnert sei hier an den "Above-Average-Effekt" (Kapitel 3.2.3), der de facto keine betriebliche Wirklichkeit wiederspiegeln kann.

Die Ergebnisse könnten jedoch auch aus folgenden Gründen erwachsen: Das Bedürfnis nach Selbstwertschutz ist eng verbunden mit entsprechendem **Eindrucksmanagement**, um sowohl positive Eindrücke (z.B. Erfolge) zu generieren als auch negative Eindrücke (z.B. Blamage) zu vermeiden (Tedeschi, 1981; für einen Überblick Mummendey, 2002). Eine

[99] Hinweise darauf geben beispielsweise die Meta-Analysen von Scholz und Schuler (1993) sowie von Twenge und Campbell (2002).

derartige strategisch positive Selbstdarstellung wird in Unternehmen häufig mit positiven Konsequenzen belohnt (vgl. Murphy & Cleveland, 1995). Es gehört bereits zum allgemeinen Erfahrungswissen, dass der Erfolg einer Person im Unternehmen sehr stark davon abhängt, wie sie von anderen beurteilt und gefördert wird, ob sie also einen "Mentor" hat (vgl. auch anekdotische Darstellung von „Protektion & Beförderung“ bei Peter & Hull, 2001). Deshalb scheint es erfolgsversprechend, wenn andere Personen einen guten Eindruck von ihr haben. So erreichen positive Selbstdarsteller (auch Manager) beispielsweise mit höherer Wahrscheinlichkeit höhere Gehälter oder erleiden mit geringerer Wahrscheinlichkeit einen Karrierestopp (Gould & Penley, 1984). Verschiedene Selbstdarstellungsstrategien, insbesondere sich kompetent darzustellen, führen in Bewerbungsgesprächen zu besserer Wahrnehmung durch den Interviewer (Kristof-Brown, Barrick & Franke, 2002) und letztlich zu höherem Erfolg/mehr Jobangeboten (Stevens & Kristof, 1995). Auch für Assessment Center schließt Fleenor (1996) in der Diskussion seiner Studie, dass „good actors“ von den Beurteilern als die effektiveren Manager betrachtet werden (zusammenfassend siehe auch Baumeister, 1989; Schlenker, 1980).[100]

In verschiedenen Führungstheorien, vor allem in Theorien charismatischer Führung, wird die Effektivität einer Führungskraft sogar ganz explizit auf deren Selbstpräsentation zurückgeführt (z.B. Sosik, Avolio & Jung, 2002: Impression Management bei charismatischer Führung, vgl. Kapitel 2.5.2 und 2.6). Rosenfeld, Giacalone und Riordan (1995) – also die Arbeitsgruppe, die Impression Management ausführlich auf den organisationalen Kontext übertrug[101] – zählen es zum täglichen Leben im Unternehmen (S. 11) und empfehlen, es als Ressource, nicht als Defizit zu betrachten (S. 140).[102] Fragebogen werden dabei insgesamt als "hervorragende Mittel der Selbstdarstellung" (Mummendey, 2003, S. 198) charakterisiert. Diskrepante Ratings, die dadurch zustande kommen, dass die Selbstbeurteiler ein Interesse an der eigenen positiven Selbstdarstellung haben, können demzufolge auch als Indikator für zielorientiertes Impression Management verstanden werden. Die Selbstdarstellungstheorie kann neben den Niveau-Unterschieden aber auch erklären, weshalb Selbstratings mit Fremdratings so gering korrelieren. Die Verzerrung aufgrund von Impression Management variiert nämlich nicht nur über verschiedene Inhaltsdimensionen hinweg, sondern auch über verschiedene Selbstbeurteiler hinweg, da die Personen unterschiedliche Darstellungs-Ziele verfolgen (Wohlers & London, 1989). Denn auch wenn nahezu jeder versucht, sich kompetent darzustellen, variieren die Operationalisierungen. Während der eine sich beispielsweise vor allem vor seinen Vorgesetzten im Bereich Perso-

[100] Im Labor zeigt sich, dass Studierende, die die Rolle von Managern übernehmen, auf eine eher übertriebene Erfolgsdarstellung mit besseren Konsequenzen für Beruf und Karriere reagieren würden als Studierende in der Rolle von Mitarbeitern (Wosinska, Dabul, Whetstone-Dion & Cialdini, 1996). Generell zeigt die gleiche Studie, dass Frauen im Vergleich zu Männern mit einer bescheideneren Selbstdarstellung erfolgreicher sind.
[101] erstmals Giacalone und Rosenfeld, 1989, 1991
[102] Da die Definition eines "wahren" Selbst sowieso unmöglich erscheint, steht hier die Frage nach dem Selbstkonzept der Führungskraft im Vordergrund, weniger die Frage nach einer "verfälschten" Darstellung.

nalauswahl kompetent darstellen will, stellt ein anderer eher sein fachliches Wissen vor seinen Mitarbeitern heraus. Wenn das Selbstdarstellungsziel nicht beinhaltet, sich mittelmäßig/durchschnittlich darzustellen, sondern sich zumindest etwas besser als der Durchschnitt darzustellen, dann resultiert daraus der *Above-Average-Effekt*.

Die Wirkung verschiedener Moderatorvariablen wird erklärbar: Unter einer *Anonymitätsinstruktion* fehlt der Adressat der Selbstdarstellung, da keiner seine individuellen Ergebnisse (Selbstdarstellung) erfährt.[103] Es kann also bei niemandem ein positiver Eindruck hinterlassen werden. Weiter sollte bei einer *Validierungserwartung* oder *Verifizierbarkeit* als Selbstdarstellungsziel eher Genauigkeit und weniger Glorifizierung[104] überwiegen. Also sagt die Theorie des Impression Managements in beiden Fällen die gefundenen geringeren Niveau-Unterschiede (für Anonymität bisher nur tendenziell belegt, Heidemeier & Moser, 2002) und höheren Korrelationen (Mabe & West, 1982) voraus. Da in Management-Ebenen generell weniger starre Anforderungs- und Belohnungsrollen vorliegen als in anderen Unternehmensbereichen, haben Führungskräfte mehr Spielraum, durch Impression Management Belohnungen zu erzielen. Wenn in Management-Ebenen aber mehr Eindrucksmanagement vollführt wird, dann ist auch die moderierende Wirkung der Tätigkeit plausibel. Weiter zeigt sich, dass Fremdratings (Führungskräfte als Ratees), die einen administrativen *Zweck* haben (versus Forschungszweck), milder ausfallen (Harris, Smith & Champagne, 1995). Eine Meta-Analyse (Jawahar & Williams, 1997) schätzt weiter, dass bei administrativen Ratings die Mildetendenz für die Ratergruppe der Manager (im Feld) um eine Drittel Standardabweichung höher liegt als für Studierende (im Labor[105]). Ebenso stark erhöht war der Effekt für Abwärtsratings im Vergleich zu Aufwärtsratings. Auch wenn keine Selbst-Fremd-*Diskrepanzen* untersucht wurden, ist die Mildetendenz bei administrativen Ratings über die Selbstdarstellung der Fremdbeurteiler erklärbar; die Konsequenzen im Feld (nicht im Labor) können weit reichend sein – und nur bei administrativem Verwendungszweck (nicht bei Forschungsfragen) gibt es etwas zu gewinnen oder zu verlieren. Auch auf abwärts beurteilende Manager selbst wirft es ein gutes Licht, wenn ihre Mitarbeiter belohnt werden. Bei einem alleinigen Motiv nach Selbstwertschutz und -erhöhung (ohne Selbstdarstellungstendenzen) sollten diese drei Unterschiede (für Anonymität, Validierungserwartung/Verifizierbarkeit und Ratingzweck) nicht auftreten.

Sowohl mit der Selbstdarstellungstheorie als auch der Theorie der Selbstwerterhöhung und des Selbstwertschutzes geht stimmig einher, dass in *unterstützenden Arbeitssettings* Selbst- und Fremdratings besser übereinstimmen als in Settings, die als nicht unterstützend

[103] Impression Management ist zwar prinzipiell auch vor vorgestelltem Publikum oder in Interaktion mit sich selbst möglich. Es sollte dann aber weniger stark sein.

[104] In der Persönlichkeitspsychologie wird die Selbstdarstellung häufig im Spannungsfeld zwischen „Glaubwürdigkeit“ und „Glorifizierung“ eingeordnet (z.B. Laux & Schütz, 1996).

[105] Im Labor wird häufig mit Szenarien gearbeitet, da es andernfalls schwierig ist, einen administrativen Zweck einzuführen.

wahrgenommen werden (Yammarino & Dubinsky, 1990). Es ist plausibel, dass in wohlgesinnter Atmosphäre auch eine nicht makellose Selbstdarstellung erlaubt ist beziehungsweise der Selbstwert weniger stark schützenswert ist. Ähnlich kann argumentiert werden, wenn es um die *Wichtigkeit* der Skaleninhalte geht: Je wichtiger der zu beurteilende Gegenstand ist, umso wichtiger ist auch ein guter Eindruck in diesem Bereich. Damit wird erklärbar, weshalb für aufgabenbezogene *Inhalte* größere Diskrepanzen als für umfeldbezogene/interpersonelle Inhalte bestehen. Denn Ersteren wird im Unternehmen mehr Erfolgsvarianz zugesprochen. So sind nicht nur einzelne Führungssysteme (z.B. Anreizsysteme) aufgabenorientiert ausgerichtet, sondern die Aufgabenorientierung als solche ist den Führungskräften generell wichtiger (z.B. bei der Beschreibung ihrer Mitarbeiter) (Conway, 1999).

3.4.2.4 Motiv zu Selbstkonsistenz

Das Motiv zur Selbstkonsistenz wird an dieser Stelle nur insofern beschrieben, wie es Unterschiede zum bereits beschriebenen Motiv zur Selbstwerterhöhung vorhersagt. Konsistenztheoretische Ansätze gehen auf Heiders Balancetheorie (1958) und Festingers Theorie der kognitiven Dissonanz (1957) zurück. Später wurden sie in Form neuer Begriffe (zum Beispiel self-verification, Swann, 1983) aktualisiert.

Das Bedürfnis nach Selbstkonsistenz motiviert Personen dazu, ihr bestehendes Selbstkonzept zu bestätigen. Sie streben also danach, selbstbezogene Informationen zu suchen und solche Informationen abzurufen, die mit ihrem Selbstkonzept konsistent sind. Zieht man die Above-Average-Studien (vgl. S. 132) heran, dann ist anzunehmen, dass die Mehrzahl der Führungskräfte positive Selbstkonzepte bezüglich ihrer Führung hat. Unabhängig also von ihrem tatsächlichen Leistungsniveau ist allein ihr überhöhtes Selbstbild für die selektive Verarbeitung positiver (konsistenter) Information verantwortlich. Genauso wie der Above-Average-Effekt insgesamt fußt auch diese Selektion nicht auf betrieblichen Wirklichkeiten, so dass von einem Positivitätsbias und einer Verzerrung der Rangfolge (Niveau-Unterschied und geringe Korrelation) auszugehen ist.Ein Vorteil der Selbstkonsistenztheorie ist (gegenüber Selbsttäuschung und Eindrucksmanagement), dass sie auch erklären kann, weshalb sich manche Führungskräfte im Vergleich zu Fremdurteilen unterschätzen oder richtig einschätzen (Atwater & Yammarino, 1992; Van Velsor et al., 1993). Es sollen wieder die Above-Average-Studien herangezogen werden, denn auch hier findet man (wenige) Führungskräfte, die ein unterdurchschnittliches Führungsselbstbild haben. Genau für diese Führungskräfte würde die Konsistenztheorie postulieren, dass sie ihrer Selbstbeurteilung auch negative, selbstbezogene Information zugrunde legen (entsprechend ihres führungsbezogenen Selbstkonzepts). Tatsächlich wei-

sen empirische Ergebnisse darauf hin, dass sich Personen mit niedrigem Selbstwert schlechter einschätzen als Personen mit hohem Selbstwert (z.B. Laborstudie von Aitkenhead, 1984). Dass hier das Motiv nach Selbstkonsistenz und weniger das nach Selbstwerterhöhung vorherrscht (das gerade bei niedrigem Selbstwert zu selbstwerterhöhender/ positiver Informationsverarbeitung führen müsste), geht einher mit dem Integrativen Selbstschemaansatz (Petersen & Stahlberg, 1995). Dieser besagt, dass für Bereiche mit umfangreichem Selbstwissen (schematische Selbstkonzeptbereiche) selbstbezogene Information eher im Sinne der Selbstkonsistenz als der Selbstwerterhöhung verarbeitet wird. Da Führungskräfte zum Bereich der Personalführung in der Regel viel selbstbezogenes Wissen besitzen, sollte das Motiv nach Selbstkonsistenz also überwiegen. Für den häufigeren Fall, dass Personen in Führungspositionen einen höheren Selbstwert haben als andere, kommen die Selbstkonsistenz- und die Selbstwerterhöhung zu gleichen Annahmen. Nach beiden Theorien streben sie nach möglichst positivem Feedback. Der Streitpunkt zwischen den Theorien entzündet sich nur, wenn es um Personen mit niedrigem Selbstwert geht.[106]

Verschiedene moderierende Effekte erscheinen vor dem Hintergrund der Selbstkonsistenztheorie plausibel. So ist es umso unwahrscheinlicher, dass auch Fremdbeurteiler alle Tätigkeitsfacetten konsistent beurteilen, je komplexer eine Tätigkeit ist. Weiter ist es wahrscheinlich, dass Beurteilungen einer Person mit hohem Konsistenzbedürfnis dann von Dritten konsistenter ausfallen, wenn sie auch konsistenter kommuniziert werden können. Das wiederum ist bei verhaltensbezogenen Skalen, leistungsorientierten Skalen und bei vergangenen Beurteilungsinhalten eher der Fall.

3.4.3 Bewertung der potentiellen Erklärungsmechanismen

In einem abschließenden Kapitel zu den potentiellen Erklärungsmechanismen wird dargestellt, welche Erklärungsmechanismen sich bislang als die tauglichsten herausgestellt haben und welche Information die Ratings aus den jeweiligen Perspektiven enthalten. Dazu wird zunächst gegenüberstellend diskutiert, welche Erklärungsmechanismen erstens die geringen korrelativen Selbst-Fremd-Übereinstimmungen und die hohen Niveau-

[106] Es geht um folgende Frage: Suchen Menschen Rückmeldung, die ihr eigenes Selbstbild bestätigt (Konsistenztheorie), oder suchen sie möglichst positive Rückmeldung (Theorie der Selbstwerterhöhung)? Menschen mit niedrigem Selbstwertgefühl sollten folglich nach negativer Rückmeldung suchen, wenn ihre Konsistenz im Vordergrund steht, beziehungsweise ausschließlich nach maximal positiver, wenn das Motiv nach Selbstwerterhöhung im Vordergrund steht. Für Menschen mit niedrigem Selbstwertgefühl gehen aus den beiden Theorien also zwei unterschiedliche Antworten hervor. Für eine Integration wird auf Shrauger (1975) sowie Petersen und Stahlberg (1995) verwiesen. Ersterer versucht die Integration, indem zwar kognitiv auch negativ-konsistente Information gesucht wird, affektiv aber immer nur positive. Letztere nehmen im bereits erwähnten Integrativen Selbstschema-Ansatz an, dass die Tendenz zur Selbstkonsistenztheorie steigt, je stärker die Elaborationstiefe des betroffenen Bereichs ist.

Unterschiede in der Beurteilung von Führungsverhalten plausibel erklären können. Zweitens werden die Erklärungen dahingehend diskutiert, ob beziehungsweise wieviele und welche Moderatorvariablen plausibel damit einhergehen.

Da lediglich die Erklärung über unterschiedliche Definitionen von Führung nicht potent genug ist, auch Niveau-Unterschiede plausibel herzuleiten, kann diese als weniger tauglich eingestuft werden als die anderen, die sich auf den ersten Blick hinsichtlich ihrer Brauchbarkeit nicht unterscheiden. Dennoch liefern meines Erachtens Selbsttäuschung und Eindrucksmanagement einen besonderen Beitrag, da das Motiv nach selbstwertdienlicher Darstellung in verschiedene andere Ansätze eingeht. Sie sind es auch, die neben zwei anderen Ansätzen (mangelndes Feedback, soziale Identität) den Above-Average-Effekt plausibel erklären können. Mangelndes Feedback und das Motiv nach Selbstkonsistenz sagen zudem unter bestimmten Bedingungen voraus, dass es zu einer Unterschätzung der eigenen Fertigkeiten kommen kann. Diese Ansätze (Selbsttäuschung, Eindrucksmanagement, Selbstkonsistenz, Feedback, soziale Identität), allen voran das mangelnde Feedback, bergen insgesamt eine hohe Plausibilität – zum einen, weil sie auf bekannte psychologische Theorien zurückgreifen, zum anderen, weil sie auch für Praktiker eingängiger sind als andere Erklärungsmechanismen. Letzteres führt dazu, dass entsprechende Ansatzpunkte in der praktischen Anwendung eher Akzeptanz finden würden. Auch der Actor-Observer-Bias wirkt zunächst plausibel, was die geringe Korrelation zwischen Selbst- und Fremdbeurteilern betrifft, da er direkt bei den befragten Perspektiven (Selbst und Fremd) ansetzt. Um die tatsächliche Brauchbarkeit dieses sowie der anderen Ansätze für die Erklärung der mangelnden Übereinstimmung zwischen Selbst- und Fremdbild besser bewerten zu können, müssen langfristig experimentelle Studien durchgeführt werden, die in der Lage sind, kausale Mechanismen zu überprüfen.

Was die Erklärung verschiedener Moderatorwirkungen betrifft (Tabelle 12), so fällt auf, dass fast alle Ansätze die am besten bestätigte Moderatorvariable – die Tätigkeit – integrieren können. Um auch die Wirkung von Validierungserwartung und Verifizierbarkeit, die sich ebenfalls als zentrale Moderatorvariablen für Selbst-Fremd-Korrelationen beziehungsweise Niveau-Unterschiede erwiesen haben, plausibel mit einem Erklärungsansatz fassen zu können, muss auf Soziale Vergleichsprozesse oder auf die Selbstdarstellungstheorie zurückgegriffen werden. Dies spricht also weiter für diese beiden Ansätze. Gleiches gilt für die Wirkung der Anonymitätsinstruktion. Betrachtet man die Anzahl der Moderatorvariablen, die die Erklärungsmechanismen aufnehmen können, so sticht – wiederum neben den sozialen Vergleichsprozessen – die Wichtigkeitshypothese hervor. Möglicherweise ist sie ein zentraler Ansatz, um die perspektivenbedingte Anteile der Ratingantworten zu erklären.

Tabelle 12. Gegenüberstellung von Erklärungsmechanismen und Moderatorvariablen: Plausibilitäts-Passung

	Erklärungsmechanismen (vgl. Kapitel 3.4.1 und 3.4.2)										
	perspektivenbedingt							idiosynkratisch			
Moderatorvariablen (vgl. Kapitel 3.3)	unterschiedliche Verhaltens-Stichproben	Mangelnde Transparenz	unterschiedliche Definitionen	unterschiedliche Wichtigkeit	unterschiedliche Differenzierungen	Akteur-Beobachter-Bias	Sozialer Vergleich/ Soziale Identität	Mangelndes Feedback	Selbsttäuschung	Impression Management	Selbstkonsistenz
Tätigkeit	x	x	x	x			x	x	x	x	
Komplexität der Tätigkeit		x	x	x				x			x
Validierungserwartung							x			x	
Verifizierbarkeit							x			x	
Anonymitätsinstruktion							x			x	
interpersonelle Dimension		x							x	x	
umfeldbezogene Dimension		x		x					x	x	
verhaltensorientierte Skala		x	x		x						x
Bewertungserfahrung			x								x
dimensionale Skala			x	x							
relative (vs. absolute) Skala				x			x				
Leistung (vs. Fähigkeiten)				x	x		x				x
vergangene (vs. zukünftige) Leistung				x	x		x		x		x
weibliche FK						x					
Kultur (kollekiv vs. individualistisch)						x			x		
Entwicklungszweck: Forschung vs. Administration										x	
unterstützendes Arbeitssetting									x	x	

Anmerkungen. Die Moderatorvariablen werden in diesem Kapitel nicht danach getrennt, ob sie die Selbst-Fremd-Korrelation oder deren Niveau-Unterschied beeinflussen.
kursiv: nur für Niveau-Unterschiede (Kultur, Zweck, Verifizierbarkeit, umfeldbezogene Iteminhalte), nur tendenziell für korrelativen Zusammenhang oder nur in Einzelstudien bestätigt, allerdings unter der jeweiligen Erklärungsperspektive plausibel.
Die Antworttendenzen und Urteilsfehler werden an dieser Stelle nicht mehr angeführt, weil davon auszugehen ist, dass sie – wie auch bei allen anderen Befragungen nicht gänzlich auszuschließen sind. Außerdem sind sie in viele andere Erklärungsmechanismen integriert.

Abschließend ist die Frage um die Effektivität der Erklärungen noch lange nicht zu klären. Dass erst verschiedene Prozesse gemeinsam die Diskrepanzen aufklären, scheint allein deshalb plausibel, weil sowohl aus der Abwärts- als auch aus der Aufwärtsperspektive idiosynkratische und perspektivenbezogene Anteile in den Ratingantworten enthalten sind

Wie man bei vielen der letztgenannten Begründungsversuche erkennen konnte, entstehen idiosynkratische Verzerrungen nicht immer losgelöst von der Perspektive, aus der beurteilt wird. Dennoch sollen hier die Ergebnisse einer Studie mit einer großen Management-Stichoprobe dargestellt werden (Scullen, Mount & Goff, 2000; vgl. Abbildung 19), in der versucht wurde, die Ratingantworten in die auf Seite 149f. dargestellten Anteile zu trennen (Ratee-Leistung [generell, dimensional], Rater-Bias [perspektivenbezogen, idiosynkratisch], Messfehler). Erkenntnisse darüber können langfristig die Auswahl geeigneter Perspektiven unterstützen.

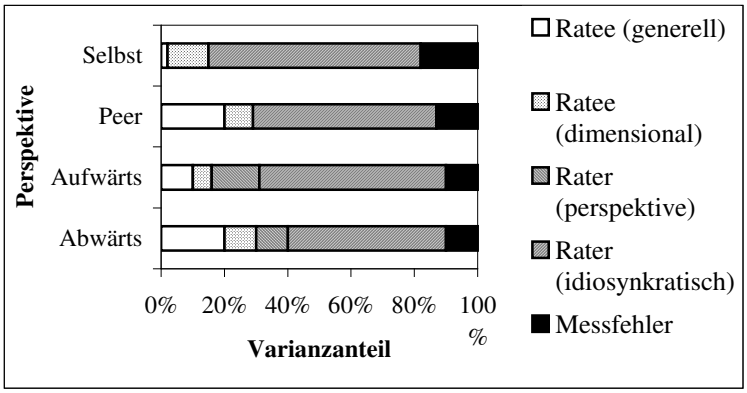


Abbildung 19. Varianzanteile, die durch Ratee-Leistung, Rater-Bias bzw. Fehler aufgeklärt werden (unterschiedliche Perspektiven) (nach Scullen et al., 2000)

Insgesamt mehren sich die Hinweise, dass idiosynkratische Verzerrungen den größten Varianzanteil an Leistungsratings für Manager beschreiben (auch Greguras, Robie et al., 2003). Er ist nicht nur höher als der Anteil perspektivenbezogener Verzerrungen, sondern auch höher als der, der von der gezeigten Leistung der Fokusperson erklärt wird.[107] Das bedeutet, dass Faktoren, die außerhalb des zu beurteilenden Gegenstands liegen, deutlich mehr Anteile an Ratingantworten tragen als dieser selbst.

In Abwärtsratings wird – im Vergleich zu den anderen Perspektiven – die Ratee-Leistung am stärksten berücksichtigt (ca. 30% Varianzaufklärung). Dennoch liegt dieser Anteil deutlich unter der Varianz, die durch den Rater-Bias aufgeklärt wird. 10% können allein auf die Abwärtsperspektive zurückgeführt werden, während circa 50% der beobachteten Varianz auf idiosynkratischen Verzerrungen basieren.

[107] Um die beiden Rater-Komponenten methodisch trennen zu können, wurden sie in der konfirmatorischen Faktorenanalyse unterschiedliche modelliert: So wurde entweder jeder Rater in einer eigenen Variable konzeptualisiert (idiosynkratischer Anteil) oder es wurden alle Ratings, die aus einer Perspektive abgegeben wurde, so behandelt, als wären sie gemeinsam gemessen worden (perspektivenbezogener Anteil).

Der größte Unterschied zwischen Abwärts- und Peerratings liegt nicht in der Verteilung auf Ratee-Leistung und Rater-Bias an sich, sondern darin, dass bei Kollegenurteilen letzterer komplett durch idiosynkratische Verzerrungen bestimmt ist, während die Perspektive selbst überhaupt keine Varianz an den Ratingantworten aufklärt.

Bei Aufwärtsbeurteilungen nimmt die perspektivenbezogene Varianz hingegen eine wichtigere Stellung als bei allen anderen Perspektiven ein. Sie entspricht nämlich mit einem Anteil von circa 15% dem Anteil, der von der Ratee-Leistung selbst bestimmt wird. Allerdings schlägt auch hier der idiosynkratische Bias stark (fast 60% Varianzaufklärung) zu Buche. Da ein perspektivenbezogener Bias immer auch spezifische wahre Anteile des zu beurteilenden Gegenstands repräsentiert, der sich ausschließlich aus einer bestimmten Perspektive beobachten lässt, kann dieser auch zum Anteil des Leistungsverhaltens der Fokusperson gezählt werden (statt zu den unerwünschten Faktoren). Denn beim multiperspektivischen Feedback geht es ausdrücklich auch darum, die inkrementelle Validität, die sich aus den einzigartigen Sichtweisen der einbezogenen Perspektiven ergibt, zu nutzen (Meta-Analyse von Conway et al., 2001).

Im Vergleich zu den Außenperspektiven zeigt sich für die Selbstbeurteilung eine Umkehrung der Anteile, die auf die Ratee-Leistung selbst zurückgehen: Während bei Fremdbeurteilungen die generelle Leistung etwa zwei Drittel der Ratee-Komponente ausmacht, umfasst sie bei Selbstbeurteilungen lediglich ein Fünftel. Das bedeutet, dass Selbstbeurteiler im Gegensatz zu Fremdbeurteilern vor allem spezielles Leistungsverhalten zur Beurteilung heranziehen. Dies stützt die oben skizzierte Hypothese, dass die geringen Selbst-Fremd-Korrelationen auf unterschiedlichen Spezifizierungen beruhen (vgl. Kapitel 3.4.1.5).

Für alle Perspektiven gilt, dass der (unerwünschte) idiosynkratische Bias mehr Varianz aufklärt als die perspektivenbezogenen Faktoren mit dem Ratee-Verhalten zusammen. Dies gilt insbesondere für die Selbstbeschreiber; jedoch können hier perspektivenbezogene und idiosynkratische Anteile nicht getrennt werden (Mount, Judge, Scullen, Sytsma & Hezlett, 1998; Scullen, Mount & Goff, 2000). Eine aktuelle Studie schätzt die Ratereffekte für Aufwärts- und Peerbeschreibungen noch stärker ein und platziert die Ratee-Effekte erst an dritter Stelle – noch hinter dem unsystematischen Fehler (Greguras, Robie et al., 2003). Dies ist deshalb jedoch nicht verwunderlich, weil der Ratereffekt nicht vom Interaktionseffekt (Rater x Ratee) getrennt wurde. Zudem zeigt sich, dass der Ratee-Effekt für Aufwärtsratings (nicht für Peerratings) deutlich steigt, wenn die Beurteilung ausschließlich für den Zweck durchgeführt wird, aus den Ergebnissen einen Entwicklungsplan für den Ratee zu entwickeln (reiner Entwicklungszweck).

Dies stimmt damit überein, dass auch die Interrater-Reliabiliät für Aufwärtsbeurteiler (nicht für Peers) dann höher ist, wenn Ratings für einen Entwicklungs- versus administrativen Zweck (Bonuszahlung oder Beförderung) erhoben werden. Der geschätzte Ratee-

Effekt für Entwicklungsziele (17%) geht dann besser mit den oben berichteten Ergebnissen von Scullen et al. (2000) einher, die übrigens ausschließlich Ratings für Entwicklungsziele betrachten. Im Einklang mit der bereits in Kapitel 3.3.3 angesprochenen Moderatorvariable *Zweck* steigt die Güte der Ratings also, wenn sie für Entwicklungsziele eingesetzt werden.

Auch wenn die berichteten Ergebnisse letztlich darauf hinweisen, dass die Varianz durch die Perspektive gering ausfällt, werden die verschiedenen Perspektiven nicht belanglos und irrelevant. Verschiedene Perspektiven sind deshalb wichtig, weil sie inkrementell zur Vorhersage objektiver Leistung beitragen (Conway, Lombardo & Sanders, 2001). Aus keiner Perspektive werden per se falsche Aussagen gemacht; es handelt sich um verschiedene „Wahr-Nehmungen". Aber es ist nur eine Möglichkeit, die Validität von Ratings durch Hinzunahme mehrerer Perspektiven zu erhöhen. Eine andere Chance bietet sich in der Erhöhung der Raterzahl. Während die Valität auf dem erstgenannten Weg dadurch erhöht wird, dass mehr Anteile des Kriteriums abgedeckt werden, gelingt dies auf dem anderen Weg dadurch, dass die Reliabilität erhöht wird. Aus folgendem Grund ist dieser Weg ebenfalls zu empfehlen: Da die idiosynkratischen Verzerrungen sehr starkes Gewicht haben, müssen diese stärker kontrolliert werden. Neben Beurteilertrainings (Hoyt & Kerns, 1999) kann ihr Einfluss aber, wie bereits erwähnt, vor allem durch die Erhöhung der Raterzahl beziehungsweise die Aggregation von Ratings erzielt werden (Conway & Huffcutt, 1997; Hoyt & Kerns, 1999; Scullen, Mount & Goff, 2000).

Bisher gibt es kaum Versuche, verschiedene der skizzierten Ansätze zu einem theoretischen Modell zu integrieren, das die in einem multiperspektivischen Beurteilungsprozess entstehenden Diskrepanzen erklären kann. Erste Modellversuche bleiben sehr abstrakt (Ashford, 1989; London, 1994) oder gehen nicht über eine Ansammlung an Moderatorvariablen hinaus (Atwater & Yammarino, 1997), so dass sie in dieser Arbeit nicht weiter beschrieben werden. Im folgenden Kapitel wird begründet, weshalb es für Führungskräfte wichtig ist, auch die Fremdbilder, die andere über ihre Führung haben, kennen zu lernen.

3.5 Die Rolle einer hohen und geringen Übereinstimmung zwischen Selbst- und Fremdbild

"Selbsterkenntnis ist der beste Weg zur Besserung."
(Volksmund)

Bisher wurde lediglich festgehalten, dass Selbst- und Fremdbild in der Regel nicht übereinstimmen und eine positiv verzerrte Wahrnehmung eher gewöhnlich als ungewöhnlich ist. Weiter wurde versucht zu erklären, weshalb diese Ergebnisse regelmäßig in dieser Art gefunden werden. Wichtig erscheint es nun, der sich folgerichtig anschließenden Frage nachzugehen, was kongruente oder diskrepante Selbst- und Fremdbilder von Führung eigentlich für Konsequenzen haben.

Die Volksmundweisheit „Selbsterkenntnis ist der beste Weg zur Besserung“ basiert auf der Annahme, dass die Wahrheit über sich selbst einen heilsamen Charakter habe und sich daher jeder möglichst oft Spiegelbilder seiner selbst einholen sollte. Dem gegenüber stehen jedoch oben angerissene Ergebnisse (siehe S. 168), dass sich Selbsttäuschungen und nicht Selbsterkenntnis positiv auf verschiedene Lebensbereiche auswirken, zum Beispiel auf die psychische Gesundheit (Taylor & Brown, 1988). Diese für den Menschen so charakteristische rosarote Brille, durch die er sich selbst betrachtet, stellt also auf der einen Seite eine Art Antidepressivum dar. Denn eine positive Betrachtung der eigenen Fertigkeiten wirkt als Motivationsfaktor, der die Arbeitsleistung erhöhen kann. Führungskräfte, die sich überschätzen, werden eher als karrieretauglich wahrgenommen. Die Mitarbeiterbeschreibung ihrer Führung korreliert signifikant mit der Anzahl an Empfehlungen zu vorgezogener Beförderung (Atwater & Yammarino, 1992). Die Teilnehmer an Assessment Centern stellen eine selektive Gruppe dar, in der sich nur noch Selbstüberschätzer befinden (Stempfle et al., 2004; N = 38). Eine entsprechende Selbstdarstellung im Vorfeld kann den Erfolg der ersten Auswahlverfahren, die letztlich in der Einladung zum Assessment Center resultieren, mitbedingen. Bei Assessment Centern erhalten diejenigen Führungskräfte die bessere Bewertung, die ihre Management-Fähigkeiten im Vergleich zum Urteil ihres Vorgesetzten überschätzen oder mit diesem im Urteil überdurchschnittlicher Leistung übereinstimmen (Nowack, 1997).[108]

Doch ist wirklich jede rosarote Sicht mit positiven Konsequenzen verbunden; kann sie auf der anderen Seite nicht auch negative Konsequenzen mit sich bringen? Obwohl Taylor und Brown (1988) herausarbeiten, dass positive Illusionen dem Wohlbefinden zuträglich sind, prophezeien sie in ihrem Resumée auch negative Folgen einer Selbstüberschätzung: Menschen könnten Karrieren und Interessen verfolgen, für die sie ungeeignet sind; sie

[108] Letzteres verwundert nicht, denn dies sind gewissermaßen die gesuchten Personen, die „high potentials“. Dennoch zeigt auch dies, dass im Unterschied zu Überschätzern für Unterschätzer keine positiven Konsequenzen (gute Bewertung im Assessment Center) folgen.

könnten Ressourcen in die Lösung von Aufgaben stecken, obwohl diese für sie unlösbar oder unkontrollierbar sind; sie könnten Risiken nicht als solche wahrnehmen und entsprechende Vorkehrungen unterlassen.

Auch Colvin und Block (1994) ziehen in ihrer Diskussion des Artikels von Taylor und Brown das Fazit, dass nur suboptimale beziehungsweise schlecht angepasste Verhaltensweisen aus positiven Verzerrungen hervorgehen können. Kurzfristige positive Konsequenzen, die sich in Laborstudien auf positive Illusion zurückführen ließen, seien mit den Bedingungen des täglichen Lebens nicht vergleichbar und müssten in Feldstudien erst noch bestätigt werden. Jeserich (1996) warnt gerade erfolgreiche Führungskräfte, die gewohnt seien, selbstwertdienliche Strategien einzusetzen, vor deren inflationärem Gebrauch, da er langfristig negative Konsequenzen prognostiziert. Gerade bei Assessment Centern *mit hohen Anforderungen* scheint der Erfolg auch stark von der Selbstreflexionsfähigkeit (Selbst-Beobachter-*Übereinstimmung*) abzuhängen (Stempfle et al., 2004).

3.5.1 Zusammenhang zwischen diskrepanter Führungsbeurteilung und Führungserfolg

In diesem Kapitel soll eine Nützlichkeitsperspektive eingenommen werden. Die Rolle der Diskrepanz zwischen Selbstbild von Führungskräften und den entsprechenden Fremdbildern wird hinsichtlich der Konsequenzen beurteilt, die sich daraus für die Mitarbeiter, das Team, das Unternehmen und/oder die Führungskraft selbst ergeben (unabhängig davon, worauf die Diskrepanz gründet). So ist nicht nur ein Streit darüber entfacht, ob Selbstüberschätzung für die Fokusperson selbst nützlich oder schädlich ist, sondern – und dies soll hier vor allem verfolgt werden – auch darüber, ob sich der Grad an Übereinstimmung zwischen Selbst- und Fremdbild von Führungsverhalten[109] auf unternehmensrelevante Erfolgsvariablen auswirkt. Verschiedene Studien sprechen dafür (vgl. Tabelle 13), andere dagegen (siehe unten).

Die widersprüchlichen Ergebnisse könnten darauf beruhen, dass in der Forschung sehr verschiedene methodische Indizes für die Selbst-Fremd-Diskrepanz üblich sind (vgl. Brutus, Fleenor & Tisak, 1999). So werden Korrelationen zwischen Selbst- und Fremdbeurteilung (z.B. Church, 1997b) oder Distanzmaße (z.B. Hazucha, Hezlett & Schneider 1993; Wexley, Alexander, Greenawalt & Couch, 1980) mit den betrachteten Erfolgsmaßen in Beziehung gesetzt. Weiter existieren beispielsweise regressionsanalytische Vorgehensweisen (z.B. Nowack, 1997), die den Einfluss von Selbst- und Fremdbeurteilung genauso wie

[109] Während Forscher zu Selbst- und Fremdbild von Führung in der Regel auf die sich daraus ergebenden Konsequenzen fokussieren, achten sie in anderen Inhaltsbereichen eher auf Moderatorvariablen, so dass zwar oben allgemeine Meta-Analysen herangezogen wurden, in diesem Kapitel aber lediglich Studien aus dem Management-Bereich zitiert werden.

den von deren Interaktion auf Erfolgsvariablen betrachten können. Wieder andere Studien (z.B. Atwater & Yammarino, 1992; Fleenor et al., 1996) kategorisieren die Führungskräfte nach dem Ausmaß ihrer Diskrepanz und dem Niveau der Beurteilung, um dann Unterschiede in den Erfolgsmaßen über die Kategorien hinweg zu untersuchen. Atwater et al. (1998) plädieren in Folge dessen dafür, sich nicht auf ein einzelnes Passungsmaß zu verlassen. Aus diesen Gründen wird bei den im Folgenden dargestellten Untersuchungen jeweils auf die Methode verwiesen. Zusätzlich werden die Erfolgsoperationalisierung und die Stichprobe genannt, die ihrerseits natürlich auch Ergebnis-Unterschiede zwischen Studien hervorrufen können. Tabelle 13 gibt einen Überblick über empirische Ergebnisse.

Tabelle 13. Zusammenhänge zwischen Übereinstimmung von Selbst- und Fremdbeschreibung der Führung und Erfolgsmaßen

Prädiktor: Übereinstimmung zwischen Selbst- und	**Erfolgskriterium**	Art des **Zusammenhangs** zwischen Prädiktor und Kriterium	**Stichprobe**	**Quelle**
Aufwärts-beschreibung	▪ Leistungsbeurteilung von oben (mehrere Beurteilungszeitpunkte) ▪ Empfehlung zu vorgezogener Beförderung (frühere Jahre) ▪ Empfehlung zu vorgezogener Beförderung (unmittelbar verg. Beurteilung)	sign. Korrelation zwischen Aufwärts und Erfolg für Übereinstimmer: $r = .39$ $r = .41$ $r = .22$	158 Marine-offiziere	Atwater & Yammarino, 1992
Aufwärts-beschreibung (vergleichbare Ergebnisse für Peers)	▪ Effektivität (Vorgesetzten-Einschätzung)	Regressionsmodell, das zusätzlich zu Selbst- und Aufwärtsbeschreibung interaktive Prädiktoren enthält, klärt mehr Varianz auf	1.326 FK[1] (öffentlicher und privater Sektor	Atwater et al., 1998
Aufwärts-beschreibung	▪ Vorgesetztenbeurteilung ▪ Empfehlung (von Vorgesetztem) zu vorgezogener Beförderung	1. Differenz (Selbst minus Aufwärts[aggregiert]) auf einigen Führungsskalen korreliert negativ mit Erfolgskriterien. 2. Hoch Erfolgreiche auf beiden Erfolgskriterien (vs. mittelmäßig und niedrig Erfolgreiche) zeigen geringere Diskrepanz zu Aufwärts.	155 Marine-offiziere	Bass & Yammarino, 1991
Aufwärts-beschreibung	▪ Starke bzw. durchschnittliche Leistungsträger (SL, DL) werden nach Kennwerten der Organisation identifiziert (von externer Beurteilung bis Vorgesetztenbeurteilung)	Für SL (versus DL) signifikant höhere Übereinstimmung zwischen Selbst und Aufwärts (Korrelation, Differenz, Kategorisierung[1])	504 FK (drei Industriezweige)	Church, 1997a
Aufwärts-, Abwärts-, Peer- und Kunden-beschreibung	▪ Servicequalität ▪ Servicekreativität	Selbst-Fremd-Korrelation korreliert positiv mit Erfolgskriterien (bis $r = .38$, Unterschiede für Perspektive und Kriterium)	152 Dienstleistungsanbieter (mit Führungsverantwortung)	Church, 1997b

Aufwärts-beschreibung	▪ jährliche Vorgesetztenbeurteilung	Regression: Kreuzprodukte (Interaktionseffekte Selbst-Aufwärts) sind signifikante Prädiktoren für Erfolgskriterium	396 FK (internationale Fluglinie in Hong Kong)	Furnham & Stringfield, 1994
Aufwärts-beschreibung (MLQ)	▪ Aufwärtsbeurteilung (Effektivität, Extra Effort und Zufriedenheit mit Führung)	Für Überschätzer[2] geringste Leistung[3]	63 FK (Informations-technologie)	Sosik & Megerian, 1999
Aufwärts-beschreibung (MLQ)	▪ Aufwärtsbeurteilung (Effektivität, Extra Effort und Zufriedenheit mit Führung und Arbeitszufriedenheit)	Nicht näher beschrieben	17 FK (öffentlicher Dienst)	Tartler, Goihl et al., 2003
Aufwärts-beschreibung	▪ Leistungsratings von unten, u.a. Derailment	Für Überschätzer[2] geringste Leistung	79 Krankenhaus-verwalter	VanVelsor et al., 1993
Aufwärts-beschreibung (durch Studierende)	▪ Verschiedene Zufriedenheitsdimensionen	Signifikante Korrelation zwischen Differenz-Score und Zufriedenheitsratings	194 FK	Wexley et al., 1980
Aufwärts-beschreibung	▪ Über- und unterdurchschnittliche Betriebe werden nach Kennzahlen aus dem Arbeits- und Gesundheitsschutz klassifiziert	Selbst-Fremd-Diskrepanz signifkant höher bei unterdurchschnittlichen Betrieben (versus überdurchschnittliche)	vier über- und fünf unterdurchschnittliche Betriebe	Zimolong, 2001

Anmerkungen. [1]FK = Führungskräfte, [2]vgl. Fußnote 112, [3]Es ergeben sich keine signifikanten Gruppenunterschiede wenn die Leistungskriterien von oben/abwärts beurteilt werden.

Es überrascht keineswegs, dass die Führungskräfte, die ihre eigenen Führungsfertigkeiten überschätzen, die schlechtesten Aufwärtsbeurteilungen ihrer Führung bekommen (Atwater et al., 1995; Van Velsor et al., 1993). Obwohl dies häufig als Beleg dafür angeführt wird, dass die Selbst-Fremd-Diskrepanz für das Ausmaß an Erfolg ausschlaggebend ist, ist die Argumentation zirkulär. Denn es ist ein und dasselbe Maß, das zum einen in die Kategorisierung als „Überschätzer“ eingeht und zum anderen das Erfolgsmaß (bei Atwater et al. immerhin zu einem späteren Messzeitpunkt erneut erhoben) darstellt.

Validere Aussagen lassen sich aus den Ergebnissen von Bass und Yammarino (1991) ableiten, weil hier Führungs- und Erfolgsbeurteilungen aus verschiedenen Quellen stammen. Höhere Kongruenz zwischen Selbst- und Aufwärtsbeurteilung von Führungsverhalten[110] von Marineoffizieren hängt mit besseren Vorgesetztenbeurteilungen sowie häufigeren Empfehlungen zu frühzeitiger Beförderung zusammen. Beide Erfolgsmaße nehmen auf die Militärlaufbahn sowie den zukünftigen Rang und Status des Offiziers starken Einfluss. Analoge Zusammenhänge ergeben sich bei Church (1997b) zwischen Selbst-

[110] Signifikante Ergebnisse ergeben sich für charismatische Führung, intellektuelle Stimulierung und Bedingte Belohnung. Für Laissez Faire hängt die Selbstunterschätzung im Vergleich zur Aufwärtsbeschreibung erwartungsgemäß mit negativen Konsequenzen (seltene Empfehlungen zur Beförderung) zusammen.

Fremd-Diskrepanz bezüglich Management-Verhaltensratings (Fremd = Aufwärts-, Abwärts-, Peer- oder Kundenbeschreibung) und Leistung im Service-Bereich (Qualität und Kreative Problemlösung).

Regressionsanalytisch finden Furnham und Stringfield (1994), dass die Übereinstimmung zwischen Selbst- und Aufwärtsbewertung von Führung[111] (sowie auch von Commitment) einen prognostischen Wert für die jährliche Vorgesetztenbeurteilung der Führungskräfte (Selbstbeurteiler) hat. Hier zeigt sich auch, dass es lohnt, die Diskrepanz in der Führungsbeurteilung besonders zu betrachten, da die Übereinstimmung in anderen Management-Tätigkeiten (z.B. Planung, Entscheidungsfindung) keinen signifikanten Beitrag zum Erfolg leistet.

Kategorisiert man die Führungskräfte gemäß ihrer Übereinstimmung mit ihren Mitarbeitern,[112] dann zeigt sich, dass die Mitarbeiter von Überschätzern am wenigsten zusätzliche Anstrengung (Extra Effort) aufbringen, am wenigsten mit der Führung zufrieden sind, dieser die geringsten Effektivitätswerte zusprechen (Sosik & Megerian, 1999) und außerdem am wenigsten arbeitszufrieden sind (Tartler, Goihl et al., 2003[113]). Umgekehrt kann für die Führungskräfte, die mit der Einschätzung ihrer Mitarbeiter annähernd übereinstimmen, auch der höchste Zusammenhang zwischen Führungs- und Erfolgsvariablen (Vorgesetztenbeurteilung der vergangenen Jahre und Empfehlungen zur Beförderung) festgehalten werden (Atwater & Yammarino, 1992). Doch auch für diese Studie müssen Einschränkungen hingenommen werden, denn die Fremdratings zur transformationalen Führung gehen sowohl in die Korrelation mit den Erfolgsvariablen ein als auch in die kategoriale Einordnung der Führungskräfte als Übereinstimmer.

Die Arbeitsgruppe von Atwater greift in weiteren Forschungsarbeiten mehrfach auf diese Kategorien zurück (vgl. Fußnote 112) und unterteilt die Übereinstimmer weiter in gute und schlechte Leistungsträger (über- und unterdurchschnittliche Beurteilung) (z.B. Atwater & Yammarino, 1997; Yammarino & Atwater, 1997). Diese unterscheiden sich hinsichtlich ihrer Effektivität und Veränderungsbereitschaft. Während den Überschätzern die insgesamt negativsten Folgen auf individueller (z.B. Karrierestopp, vgl. auch Van Velsor et al., 1993) und organisationaler Ebene (z.B. geringe Effektivität, viele Konflikte oder geringes Commitment) zuzusprechen sind, kann für die Unterschätzer ein gemischtes Fazit gezogen werden, da diese auf der einen Seite die negativen Konsequenzen des geringen Selbstwerts zu tragen hätten, aber auf der anderen Seite durch kontinuierliches Feedback Verände-

[111] Die Dimension umfasst Führung und speziell auch Motivierung.

[112] Als Überschätzer/Unterschätzer gelten diejenigen Führungskräfte, deren Selbstbewertung mehr als eine halbe Standardabweichung über/unter den Fremdratings liegt (mittlerer Skalenwert der Ratings für diese Führungskraft aus einer Perspektive). Übereinstimmung ist operationalisiert als maximal um eine halbe Standardabweichung different von den entsprechenden Fremdratings.

[113] Da in der Arbeit nicht angegeben ist, wie die Differenz berechnet wurde bzw. ob in die Kategorie der hoch diskrepanten Führungskräfte auch Unterschätzer eingehen, gilt diese Aussage nur aufgrund der Annahme, dass dies nicht der Fall ist. Die Stichprobe ist mit $N = 17$ Selbstbeschreibern (Führungskräfte) sehr klein.

rungspotential genutzt werden könne. Übereinstimmung auf niedrigem Leistungsniveau resultiert quasi einstimmig in negativen individuellen und organisationalen Konsequenzen, während Übereinstimmung auf hohem Leistungsniveau die erfolgreichste Subgruppe kennzeichnet. Anstatt mit kategorialer Einordnung arbeiten Atwater, Ostroff, Yammarino und Fleenor (1998) mit Regressionsanalysen und anschließenden dreidimensionalen Oberflächenanalysen in Anlehnung an Edwards (1994a, 1994b, 2002). Letztere dienen insbesondere der Interpretation der Ergebnisse, die die Regressionsanalyse erbrachte. Die Daten dieser Studie verweisen darauf, dass die Führungseffektivität (gemessen durch Vorgesetzten-Abwärtsratings) umso höher ist, je besser Selbst- und Fremdbeurteilung von Management-Verhalten ausfallen und je weniger die Selbstbeschreibung die Aufwärtsbeschreibung übersteigt (besser unterschreitet).

Für eine Management-Stichprobe aus verschiedenen Industriezweigen zeigt sich, dass diejenigen Manager mit ihren Mitarbeitern besser in der Beurteilung ihres Führungsverhaltens übereinstimmen, die von ihrem Unternehmen als überdurchschnittliche Leistungsträger identifiziert wurden (Church, 1997a). Da das Ergebnis in dieser Studie erstens mit verschiedenen Kongruenzmaßen nachvollzogen wurde (Differenzmaß, Korrelation und kategoriale Einteilung entsprechend Fußnote 112) und zweitens für ein Erfolgsmaß gefunden wurde, das unabhängig von Selbst- und Fremdratings ist, kann es im Vergleich zu vorab dargestellten Ergebnissen als besonders valide hervorgehoben werden. Allerdings untersucht auch Church (1997a) nicht, ob die Unterschiede zwischen den Leistungsgruppen genauso gut durch Unterschiede in einer Beurteilungsperspektive allein erklärt werden können.

Ebenso kategorisiert die Arbeitsgruppe von Zimolong (2001)[114] die Untersuchungseinheit nach dem Erfolgsmaß, bevor sie untersucht, ob sich die Differenz zwischen Selbst- und Fremdwahrnehmung von Führung für überdurchschnittlich erfolgreiche Betriebe (hinsichtlich verschiedener Indikatoren zum Arbeits- und Gesundheitsschutzes) von der für unterdurchschnittlich erfolgreiche Betriebe unterscheidet. Die Untersuchungseinheit geht hier also weit über die einzelne Führungskraft (und ihr Team) hinaus. Abgestimmt auf den Bereich des Erfolgsmaßes ist "Führung" als Partizipation, Motivation und Zielsetzung nur im Bereich des Arbeits- und Gesundheitsschutzes zu verstehen. Es zeigt sich, dass die Führungskraft-Mitarbeiter-Diskrepanzen in unterdurchschnittlich erfolgreichen Betrieben bis zum Vierfachen der Abweichungen in überdurchschnittlich erfolgreichen Betrieben betragen. Als Erklärung wird die Kommunikation zwischen Führungskraft und Mitarbeiter herangezogen (vgl. Kapitel 3.5.2).

[114] Die Ergebnisse entstammen dem Verbundprojekt „Prävention im betrieblichen Arbeits- und Gesundheitsschutz – Ganzheitliches Management des betrieblichen Arbeits- und Gesundheitsschutzes (GAMAGS)“.

Letztlich birgt eine hohe Diskrepanz zwischen Selbst- und Fremdbild des Führungsverhaltens aber auch für die Karriere der Führungskraft selbst negative Seiten. Denn mit steigender Insensitivität für Mitarbeiter,[115] was häufig als Diskrepanz interpretiert wurde, steigt die Wahrscheinlichkeit für ein sogenanntes „career derailment“ der Führungskraft (Interviewstudie von McCall & Lombardo, 1983; Übersichtsarbeit von Van Velsor & Leslie, 1995): Die Führung entgleist, die Führungskraft steigt entgegen aller Erwartungen nicht weiter auf, sondern bleibt auf einer Position unter den Erwartungen stehen, wird degradiert, versetzt oder entlassen. Im Einklang mit diesem Ergebnis zeigt sich für Führungskräfte, die an einem Plateau angekommen sind, eine höhere Selbst-Fremd-Diskrepanz als für "High Potentials", denen ein weiterer Aufstieg vorgezeichnet wird (nicht publizierte Arbeit von Nilsen, 1991, ausführlich dokumentiert in Tornow, 1993).

Verschiedene Untersuchungen tragen also ein breites Spektrum an unerwünschten Kennwerten für die Organisation, das Team und die Führungskraft selbst zusammen – für den Fall, dass Führungskraft und Mitarbeiter in ihrer Einschätzung weit auseinander liegen, sind dies: geringe (Arbeits-)Zufriedenheit, geringe Leistungsbereitschaft über das erwartete Mindestmaß hinaus und geringes Commitment der Mitarbeiter, schlechte Kennzahlen im Arbeits- und Gesundheitsschutz, vermehrt Konflikte, insgesamt geringe Effektivität der Führungskraft sowie eine höhere Wahrscheinlichkeit für einen Karriereknick. Es kann also ein deutliches Übergewicht der Studien festgehalten werden, die zeigen, dass sich erfolgreiche Führungskräfte in höherer Kongruenz zu Fremdbeurteilungen einschätzen können als weniger erfolgreiche Führungskräfte. Die überwiegende Mehrzahl an Studien untersucht im Zusammenhang mit Erfolgsvariablen lediglich die Diskrepanz zwischen Selbst- und Aufwärtswahrnehmung (vgl. Tabelle 13). Ergebnisse aus Studien mit anderen Fremdperspektiven, die der Selbstbeurteilung gegenüber gestellt werden, sind erst vereinzelt publiziert, weisen aber immerhin in die gleiche Richtung. Untersuchungen zur Validität der Fremd-Fremd-Übereinstimmung liegen nicht vor. Auch die vorliegende Studie legt den Schwerpunkt auf die Selbst-Aufwärts-Übereinstimmung, wenn es um die Untersuchung der Rolle der Diskrepanz geht. Dies hat verschiedene Gründe, die später dargelegt werden.

Wie oben bereits angerissen, gibt es vereinzelt Studienergebnisse, die durch eine Fremdbeschreibung allein die abhängige Erfolgsvariable besser prognostizieren können als durch ein Übereinstimmungsmaß von Selbst- und Fremdbeschreibung.

So finden beispielsweise Fleenor et al. (1996), dass von den Ratings nur die Mitarbeiterperspektive darauf Einfluss nimmt, welche Leistung – gemessen als Vorgesetzten-Beurteilung – die Fokus-Führungskraft erbringt. Die Autoren kategorisieren die Führungs-

[115] In einem Überblicksartikel stellen Van Velsor und Leslie (1995) die Insensitivität für andere sowie schlechte Arbeitsbeziehungen als zentral für Derailment heraus.

kräfte in gute und schlechte Leistungsträger (über- und unterdurchschnittliche Aufwärtsbeurteilung), die zusätzlich jeweils die Ausprägung Überschätzer, Unterschätzer und Übereinstimmer annehmen kann (vgl. Fußnote 112). Hier ergibt sich, dass gute Leistungsträger unabhängig von der Passung zwischen Selbst- und Aufwärtsratings auch von ihren Vorgesetzten besser eingeschätzt werden als schlechte Leistungsträger. Da jedoch keine unabhängigen Erfolgsvariablen verwendet wurden, sondern lediglich Abwärtsratings auf dem gleichen Instrument, das auch für die Selbst- und Aufwärtsratings herangezogen wurde, spiegeln die Ergebnisse meines Erachtens nur die höheren Fremd-Fremd- im Vergleich zu Selbst-Fremd-Korrelationen wider (vgl. Kapitel 3.2.2). Dennoch verweist das prinzipielle Vorgehen auf einen häufigen Forschungsfehler. Denn nur selten wird tatsächlich untersucht, ob Führungserfolg, wie er durch das Ausmaß der Kongruenz/Diskrepanz prognostiziert wird, nicht auch durch das fremdeingeschätzte Leistungsniveau allein vorhersagbar wäre. Hinweise darauf ergab bereits die regressionsanalytische Auswertung von Bass und Yammarino (1991), bei der die Selbstbeschreibung keine zusätzliche Varianz – über die Aufwärtsbeschreibung hinaus – aufklären konnte. Dies müsste der Fall sein, wenn man annimmt, dass Diskrepanzen durch Selbst- und Fremdbeurteilungen entstehen und kein explizites Diskrepanzmaß als Prädiktor aufgenommen wird.

Integrierend ist zu betonen, dass möglicherweise erst durch die gleichzeitige Betrachtung erstens der Übereinstimmung, zweitens der Richtung der potentiellen Abweichung sowie drittens des Leistungsniveaus an sich der Führungserfolg besser erklärt werden kann als durch die alleinige Betrachtung einer dieser Perspektiven. Die einzelnen Anteile sind in zukünftigen Forschungsarbeiten erst noch zu ermitteln.

Für die Praxis bedeutet dies, dass konsequenterweise jedes Unternehmen danach streben müsste, in einem ersten Schritt Aufwärts- und andere Fremdbeurteilungen als Indikatoren für Erfolg ernst zu nehmen und die Führungsleistung in den entsprechenden Punkten stetig zu verbessern (also das Niveau zu heben). In einem nicht zu vernachlässigenden zweiten Schritt müssen aber die Arbeitsteams auch dabei unterstützt werden, die Kongruenz zwischen der Selbstwahrnehmung der Führungskräfte und der Fremdwahrnehmung ihrer Mitarbeiter (anderer) zu verbessern (vgl. Kapitel 3.6). Es ginge dann nicht nur darum, Fremdbeurteilungen ernstzunehmen, sondern auch darum, die Diskrepanz zu verringern.

Nachdem oben bereits beschrieben wurde, dass in der Literatur die Niveau-Unterschiede zwischen Selbst- und Fremdbild als Indikator für eine Mildetendenz und deren Korrelation als Indikator für die Validität der Selbstratings herangezogen werden (z.B. Heidemeier & Moser, 2002), erweist sich die mangelnde Übereinstimmung zwischen Selbst- und Fremdeinschätzung (unabhängig vom verwendeten Index) neben diesen methodischen Gesichtspunkten also zunehmend als Indikator für mangelnden Erfolg. Doch wie kann dies erklärt werden?

3.5.2 Potentielle Erklärungsmechanismen für den Zusammenhang zwischen diskrepanter Führungsbeurteilung und Führungsmisserfolg

Typischerweise wird die Schlussfolgerung gezogen, dass zwischen den Befragten, die nicht übereinstimmen, *Kommunikationsprobleme* und eine hohe Anfälligkeit für Missverständnisse bestehen (z.B. Atwater et al., 1998, S. 585). Bereits Luft (1963) stellt den allgemeinen Zusammenhang auf, dass die Kommunikation umso schlechter ist, je kleiner „Quadrant 1“ im Johari-Fenster ausfällt. Dieser Quadrant ist der „Bereich der freien Aktivität“ und umfasst Verhaltensweisen und Motivationen, die anderen und der Person selbst bekannt sind. In diesem Fall ist auch ihre Beurteiler-Übereinstimmung verhältnismäßig hoch. Entsprechend findet man auch, dass die Selbst-Aufwärts-Kongruenz hinsichtlich eines breiten Spektrums an Führungsverhaltensweisen (MLQ) mit einer besseren Kommunikation einher geht (Tartler, Goihl et al., 2003[116]). Bei Church (1997b; siehe Tabelle 13) ist es vor allem die Kongruenz bezüglich des interpersonellen, nicht nur des kommunikativen Verhaltens, die die Korrelation zwischen kongruenter Beurteilung und Erfolg bestimmt. Auch was die Selbst-Aufwärts-Übereinstimmung von Führungsverhalten im Rahmen des Arbeits- und Gesundheitsschutzes (AGS) betrifft, ist für Betriebe, die in diesen Bereichen besonders erfolgreich sind, nicht nur eine höhere Kongruenz festzuhalten, sondern auch eine verstärkte Einbindung und Partizipation der Mitarbeiter. Die Schlüsselrolle wird der Information und Kommunikation zugeschrieben: Denn je besser die innerbetriebliche Kommunikation funktioniert, umso näher liegen die Einschätzungen der Führungskraft selbst mit denen ihrer Mitarbeiter beisammen (Zimolong, 2001).

Zudem zeigten Hatfield und Huseman (1982), dass kongruente Wahrnehmungen der Kommunikation mit höherer Zufriedenheit einhergehen. Mit höherer Zufriedenheit geht aber auch eine höhere Selbst-Aufwärts-Übereinstimmung auf verschiedenen Einstellungs- und Persönlichkeitsdimensionen einher (Wexley, Alexander, Greenawalt & Couch, 1980), was häufig als Indikator für eine bessere Beziehung zwischen Führungskraft und Mitarbeiter interpretiert wird (Tartler, Goihl et al., 2003). Die Beziehung sowie Konflikthäufigkeit selbst wird in der Untersuchung von Wexley et al. (1980) nur hinsichtlich der Übereinstimmung zwischen Führungskraft und Mitarbeiter für die Fokusperson des Mitarbeiters untersucht, wobei sich ein positiver Zusammenhang ergibt: Schätzt eine Führungskraft ihre Mitarbeiter also ähnlich ein wie diese sich selbst, dann kommt die Führungskraft nach Aussagen ihres Vorgesetzten besser mit Kollegen zurecht, bereitet weniger Probleme und ist insgesamt erfolgreicher.

Plausibel erscheint die Erklärung über die Güte der Beziehung und Kommunikation zweifellos. Denn es ist anzunehmen, dass dann auf der einen Seite viele der oben genann-

[116] Leider bleibt unbekannt, welche Art der Kommunikation die drei konstruierten Items messen. Das zitierte Beispielitem deutet auf effektives Besprechungsmanagement hin. Insofern könnte man spekulieren, dass klare Zielabsprachen und regelmäßige Bewertungen der Teilschritte zu einer höheren Übereinstimmung führen.

ten Erklärungsmechanismen für die Entstehung der Diskrepanz (Kapitel 3.4) außer Kraft gesetzt werden. Zum Beispiel wird in guten Beziehungen eher Feedback akzeptiert, was einen kontinuierlicheren Abstimmungsprozess ermöglicht, oder es wird weniger Impression Management betrieben oder es werden weniger selbstwertdienliche Vergleiche durchgeführt, so dass nicht nur die Fragebogenratings weniger stark verzerrt werden, sondern zur Erfüllung dieser Selbstdarstellungen auch weniger Kapazitäten von der eigentlichen Aufgabe abgezogen werden (vgl. Baumeister, 1989). Durch eine klare Kommunikation sollten auf der anderen Seite weniger Reibungsverluste im Sinne von Steiner (1972, 1976) auftreten, da mehr über Ziele, Wichtigkeiten und Entscheidungen kommuniziert wird. So gelingt es eher, dass alle das gleiche Ziel anstreben und keine Verluste entstehen, wenn Ziele und Aufgaben vorab klar kommuniziert werden: "Zu große Abweichungen in der Sicht und der Einschätzung der Abläufe oder in den Wertorientierungen verhindern, werden sie nicht als Anlass für die Entwicklung neuer Problemlösungen genutzt, die Ausrichtung und Abstimmung des Leistungshandelns auf ein Ziel und damit eine Optimierung des AGS" (Elke, 2001, S. 173). Entsprechendes kann auch für Ziele außerhalb des AGS gelten.

Bislang ungeklärt ist jedoch, ob eine schlechte Kommunikation ursächlich für die Diskrepanz ist, die wiederum in geringem Erfolg resultiert, oder ob die Erfolglosigkeit genauso auf schlechter Kommunikation beruht wie die Diskrepanz selbst. Plausibel erscheint, dass mangelhafte Kommunikation mit den Fremdbeurteilern sowohl zu diskrepanten Wahrnehmungen führt als auch eine zielgerichtete und erfolgsversprechende Führungstätigkeit unterbindet. Als dritte Möglichkeit wäre aber sogar denkbar, dass durch die Erfolglosigkeit mehr Konflikte oder schlechtere Kommunikationsstile bedingt werden, die dann wiederum zu einer diskrepanten Führungsbeschreibung führen. Dies sollte beispielsweise durch Langzeiterhebungen näher beleuchtet werden.

Neben der Kommunikation kann auch die *Veränderungsmotivation* ins Zentrum der Erklärung gerückt werden. Denn es ist davon auszugehen, dass Überschätzer – ohne einen stetigen Anstoß von außen – kaum Motivation zur kontinuierlichen Verbesserung aufbringen, da sie ihre Leistung bereits sehr gut bewerten. Sie strengen sich bei Aufgaben weniger an, als deren adäquate Erfüllung erfordern würde. Außerdem erliegen sie auch eher der Illusion, besser wahrgenommen zu werden und mehr Vertrauen zu genießen als dies der Fall ist. Da daraus allerdings eine Missachtung von Feedback und anderen Meinungen entstehen kann, werden die entsprechenden notwendigen Veränderungsprozesse nicht vollzogen, so dass kein optimaler Arbeitsablauf mehr gewährleistet ist. Aufgrund der Selbstüberschätzung wird die Notwendigkeit zur Veränderung nicht wahrgenommen, werden also keine Veränderungsziele gesteckt, so dass keine Veränderungsmotivation aufgebaut wird und Anpassungen unterlassen werden. In der Folge kann ein Teufelskreislauf in Kraft treten, da mangelnde Anpassung zu größeren Fehlern, zu geringerem Erfolg, Vertrauensverlust und zu noch größeren Diskrepanzen führen kann.

An dieser Stelle laufen die beiden angesprochenen Erklärungen zusammen. Denn wenn Führungskräfte die Erwartungen und Rollen der Mitarbeiter, Vorgesetzten, Peers und Kunden kennen, dann können sie sich besser an die immer schneller werdenden Veränderungszyklen anpassen (London, 1995b). Dies zu forcieren ist Ziel der multiperspektivischen Führungsbeschreibung mit anschließendem Feedback. Dyadische Führung (LMX) oder die transformationale Führung versuchen dies ebenfalls zu unterstützen.

Dass die Ergebnisse für Unterschätzer gemischt ausfallen, überrascht nicht. Denn sie können auf der einen Seite versuchen, ihre vermeintlich geringen Fertigkeiten durch vermehrte Anstrengung auszubügeln und hohe Erfolge erzielen. Sie können auf der anderen Seite aber auch bei geringer Erfolgserwartung nur mit geringer Motivation arbeiten, was letztlich zu schlechten Ergebnissen führt. Möglicherweise sind diejenigen Führungskräfte, die in ihrer Führungsbeschreibung mit anderen übereinstimmen, insgesamt sensibler dafür, was in bestimmten Situationen gebraucht wird und sind deshalb erfolgreicher.

Eine interessante Hypothese stellen Kruger und Dunning (1999) auf. Sie finden im Labor, dass den Überschätzern die Kapazität oder die metakognitiven Fertigkeiten dazu fehlen, ihre eigene Kompetenz oder die anderer zu erkennen. Dass dies ausgerechnet bei den Überschätzern der Fall ist, wird damit erklärt, dass für sie die Aufgaben (z.B. Grammatik-Übungen) schwierig sind (da die gezeigte Leistung unter ihrer Erwartung liegt) und daher viel Kapazität zur Aufgabenbewältigung gebraucht wird.

Basierend auf der Annahme, dass die Beschreibungen die jeweiligen Wirklichkeiten darstellen und nicht Inszenierungen bestimmter Ziele, kann Folgendes resümiert werden: Stimmen die Urteile von Führungskraft und Mitarbeiter annähernd überein, spricht das eher für einen reibungslosen, aufeinander abgestimmten und dadurch erfolgreichen Arbeitsablauf. Prozessverluste im Sinne von Steiner sollten dann seltener sein. Weichen ihre Urteile voneinander ab, dann gibt das Hinweise auf Unklarheiten, die aufgrund derer die Arbeitsergebnisse nicht optimal ausfallen können. Diese Prozessverluste erklären die negativen Erfolgskennwerte der Überschätzer. Für Unterschätzer müssten diese dann ähnlich stark sein, worauf die bisherigen Daten nicht hindeuten. Sie sprechen für weniger negative Resultate als für Überschätzer. Dies ist deshalb auch plausibel, weil sich Unterschätzer riskante Projekte nicht zutrauen und deshalb nicht daran scheitern und in der Folge auch weniger negative Konsequenzen für das Unternehmen provozieren. Die Ergebnislage für Unterschätzer ist jedoch unzureichend.

Betrachtet man nicht den Beurteilungsbereich des Führungsverhaltens, sondern eine Bandbreite an arbeitsbezogenen *Einstellungen und Persönlichkeitsvariablen*, dann findet sich, dass Mitarbeiter zufriedener sind, wenn ihre Fremdwahrnehmung mit der Selbstwahrnehmung ihrer Führungskraft auch auf diesen Dimensionen besser übereinstimmt

(Wexley et al., 1980). Weniger eindeutig hängt die Zufriedenheit der Mitarbeiter mit der Diskrepanz zwischen der Selbstbeschreibung des Mitarbeiters (Mitarbeiter als Fokusperson) und der Fremdbeschreibung durch die Führungskraft (Rater) zusammen. Dies deutet darauf hin, dass eine (zwischen Führungskraft und Mitarbeiter) kongruente Wahrnehmung der Einstellungen der Führungskraft, möglicherweise also der Führung an sich, eine wichtigere Schlüsselposition für die Zufriedenheit der Mitarbeiter darstellt als eine kongruente Wahrnehmung eigener Einstellungen. Das ist deshalb alles andere als überraschend, weil der Führung die Erfolgsgenerierung zugeschrieben wird – und dies kann nur geschehen, wenn alle Beteiligte sie „richtig“ verstehen.

In keine der in Kapitel 2 dargestellten Führungstheorien ist explizit die Selbst-Fremd-Übereinstimmung als vermittelnde Variable für Führungserfolg einbezogen. Wenn sich die Ergebnisse, auf die Kapitel 3.5.1 hinweist, aber weiter bestätigen, sollte die Diskrepanz sowie die vermittelnden Variablen zum Erfolg auch in der Theorienbildung berücksichtigt werden – und zwar über angrenzende Begriffe wie Leader-Member-Exchange, Mitarbeiterorientierung, Individualisierte Berücksichtigung oder Beziehungsstruktur (Situationsvariable im Kontingenzmodell bei Fiedler) hinaus. Konkreter werden bislang normative Modelle aus der Praxis, in denen sich regelmäßig die Dimension „Zusammenarbeit“ oder „Kommunikation“ findet und in denen für alle Arten an Mitarbeitergesprächen (z.B. Zielsetzungs-, Beurteilungs-, Rückkehrgespräche) proklamiert wird, dass spätestens am Gesprächsende das Verständnis und besser noch die Akzeptanz des Mitarbeiters sicher gestellt sein muss.

Kennt eine Führungskraft nämlich die Fremdwahrnehmung anderer, schafft sie sich dadurch die Möglichkeit, darauf zu reagieren. Für eine Führungskraft ist es sowohl relevant, das Fremdbild ihres Vorgesetzten – also desjenigen, der Belohnungsmacht hat – zu kennen als auch die Fremdbilder ihrer Mitarbeiter, denn diese erschaffen durch ihre Arbeit letztlich den Teamerfolg, an dem auch die Führungskraft selbst gemessen wird. Das folgende Kapitel hinterfragt, ob ein solches Kennenlernen der Fremdeinschätzungen auch zu den entsprechenden Veränderungen führt.

Zuvor soll angemerkt werden, dass Diskrepanzen zwischen Selbst- und Fremdbild nicht nur im Führungsbereich negative Konsequenzen nach sich ziehen können. So prognostiziert beispielsweise ein hoher Niveau-Unterschied zwischen Selbst- und Fremdeinschätzung (durch Versuchsleiter und Freunde) von Persönlichkeitsattributen kurzfristig schädigendes Verhalten in sozialen Beziehungen und langfristig (5 Jahre) schlechte soziale Kompetenzen sowie schlechte psychische Anpassung (Colvin, Block & Funder, 1995). Insbesondere bei langanhaltendem Missverhältnis bzw. Missverständnis zwischen Selbst- und Fremdeinschätzung überwiegen die Anpassungsschwierigkeiten den Motivationsschub bei optimistischer Selbstsicht, so dass verschiedenste Misserfolge entstehen.

3.6 Feedback-Intervention nach multiperspektivischer Führungsbeurteilung

„Feedback selbst stellt immer ein Geschenk dar und wie mit allen Geschenken, entscheidet der Empfänger, ob er es annimmt oder nicht, d.h. in diesem Fall ob er die Ratschläge aufnehmen und sein Verhalten ändern möchte."
(Steinert, 2003)

In der Literatur viel zitiert ist der Konfliktpunkt, den Tornow (1993) einbrachte, indem er das wissenschaftliche und praktische Interesse an multiperspektivischen Beurteilungen nahezu unvereinbar nebeneinander stellte. Für Erstere sei die Messung an sich Zentrum des Interesses (z.B. die Ratinggenauigkeit), so dass eine Reduzierung der mangelnden Übereinstimmung zwischen den Ratern nur deshalb angestrebt würde, um den Ratingfehler zu minimieren. Die Praktiker hingegen wollten die Methode nutzen, um die individuelle und organisationale Effektivität zu steigern und setzten auf Lernprozesse, die durch die Beurteilungen angestoßen würden. Die Messung ist dann also nur Mittel zum Zweck.

Da in der praktischen Organisationsberatung blindlings davon ausgegangen wird, dass ein derartiges Feedback erfolgsversprechend (zumindest nicht erfolgsmindernd) sei, ist es höchste Zeit, eine dritte Perspektive einzunehmen, in der beide Extreme vereint werden. Es bestehen verschiedene Gründe, weshalb Kenntnisse der Fremdperspektiven und in der Folge eine höhere Selbst-Fremd-Übereinstimmung wertvoll sind. Wie beschrieben, sind in der Regel Fremdwahrnehmungen valider. Was bringt es einer Führungskraft, wenn sie sich selbst als fair wahrnimmt, ihre Mitarbeiter diese aber anders wahrnehmen. Unabhängig davon, ob die Fremdbeurteilungen genau sind oder nicht, können nur die Wahrnehmungen der Fremdbeurteiler bei diesen selbst auch zu Veränderungen führen. Eine Führungskraft hat aber Vorteile, wenn sie weiß, aus welchen Gründen Veränderungen eintreten beziehungsweise nicht eintreten. Außerdem wurde in Kapitel 3.5.1 der Zusammenhang zwischen der Selbst-Fremd-Übereinstimmung und der Effektivität dargestellt, so dass es auch ökonomische Gründe für das Erkennen der Fremdwahrnehmungen gibt.

Aus verschiedenen Gründen muss die Wissenschaft also praxisrelevante Fragen erforschen, nämlich wie die Übereinstimmung erhöht werden kann und ob Feedback[117] immer ein geeignetes Mittel darstellt. Deshalb soll Feedback im Zentrum der weiteren Betrachung stehen. Erst in den letzten Jahren wurden vereinzelt empirische Daten publiziert, die die Wirksamkeit von Feedback in der Folge von multiperspektivischen Beurteilungen untersuchen. Damit hinkt die Feedbackforschung vielen anderen Bereichen hinterher, aus denen aber begründet Hoffnung auf die Wirksamkeit von Feedback gezogen werden kann (z.B.

[117] In der vorliegenden Arbeit wird unter Feedback immer so genanntes Bewertungsfeedback, mit dem unter anderem die Selbstwahrnehmung verbessert werden soll, und nicht Anforderungsfeedback, das Klarheit über die Arbeitsanforderung schaffen soll, verstanden (Einteilung vgl. Greller & Herold, 1975; Farr, 1991).

Biofeedback, Rief & Birbaumer, 2000). Eher skeptisch stimmen diejenigen Ergebnisse, die für Selbst-Fremd(Aufwärts)-Diskrepanzen eine hohe Retest-Reliabilität finden und diese folglich als relativ stabiles Merkmal interpretieren (z.B. Nilsen & Campbell, 1993).

Unter „Feedback“ wird die Rückmeldung an eine Person verstanden, wie sie von einem anderen wahrgenommen, verstanden und erlebt wird. Es stellt daher einen speziellen Aspekt der Kommunikation dar (Ilgen et al., 1979). Da Führung selbst ein interaktiver/kommunikativer Prozess ist, sollten Kommunikationsstörungen mit weit reichenden negativen Konsequenzen dadurch verhindert werden können, dass dem Sender (z.B. der Führungskraft) Rückmeldung darüber erteilt wird, wie dessen Kommunikation beim Empfänger (z.B. dem Mitarbeiter) angekommen ist.

Wie im vergangenen Kapitel dargelegt wurde, gehen mit diskrepanten Sichtweisen auch unerwünschte Begleiterscheinungen einher. Diese Begleiterscheinungen der auseinander klaffenden Sichtweisen des Führungsverhaltens legitimieren aus ökonomischer Perspektive den Feedbackprozess und daraufhin initiierte Veränderungsmaßnahmen, da die Effektivität der Organisation, nicht nur die Kompetenz einzelner Führungskräfte bedroht scheint (z.B. Bracken, 1997). Also geht der hier beschriebene Feedbackprozess immer von diesen verschiedenen, sich ergänzenden Sichtweisen aus. Auch wenn bisher nicht eindeutig geklärt ist, ob zur Vorhersage von Führungserfolg wirklich das Ausmaß der Diskrepanz die "kritischste" unter den Variablen ist, die in einem multiperspektivischen Beurteilungsprozess betrachtet werden (vgl. S. 187), wird ihr unisono eine Schlüsselrolle zugeschrieben. Auch Autoren, die für die Erfolgsprognose nicht die Diskrepanz, sondern das fremdeingeschätzte Leistungsniveau favorisieren (z.B. Fleenor et al., 1996), schreiben ersterer eine hohe motivationale Rolle im Entwicklungs-/Feedbackprozess zu. Ebenso kommen auch andere kritische Stimmen, die an sich eher auf methodologische Probleme hinweisen, zu einem optimistischen Fazit und verweisen auf das hohe Potential, das in 360°-*Feedbacks* steckt (Dunnette, 1993). In diesem Sinne ist die Selbstbewertung, die am wenigsten valide Beurteilungsperspektive darstellt, unabdingbar, da sie als Ausgangspunkt für den Weg hin zu weiteren Entwicklungszielen genutzt werden muss. Wie in jeder Beratung gilt, dass man die Fokusperson dort abholen sollte, wo sie derzeit steht. Daneben wird Feedback generell als Faktor betrachtet, der die Arbeitsmotivation beeinflussen kann (vgl. Zielsetzungstheorie von Locke & Latham [siehe Kapitel 3.6.3] oder „Job Characteristics Model“ von Hackman & Oldham, 1980).

In der Tat zeigt sich, dass die Karriereverläufe von Führungskräften erfolgreicher sind, wenn sie herausfordernde Situationen erleben und in der Lage sind, aus diesen Erfahrungen zu lernen (McCall, 1997; McCall, Lombardo & Morrison, 1988). Feedback wird dabei als Katalysator für Lerneinsichten konzipiert (McCall, 1997). Dadurch wird versucht zusammenzufassen, was sich beispielsweise bei Spreitzer, McCall und Mahoney (1997)

zeigt: International agierende Führungskräfte sind erfolgreicher und werden besser beurteilt, wenn sie offen für Kritik sind, nach Lerngelegenheiten suchen und aus Fehlern lernen, andere um Feedback bitten sowie die gewonnenen Einsichten in ihre Entwicklung integrieren.

Feedback wird also gegeben, um die Selbstreflexion der Führungskräfte (und damit verbunden die Selbstbewertung der eigenen Stärken und Schwächen) und das Vertrauen in die eigenen Kompetenzen zu erhöhen, um den Perspektivenwechsel zu trainieren (also auch um Gründe für die diskrepanten Wahrnehmungen zu erkennen)[118] und um letztlich die gewünschten Veränderungen im Führungsverhalten und folglich im Unternehmen zu erreichen (Scherm & Sarges, 2002; Waldman & Atwater, 1998). Dennoch gilt zu hinterfragen, ob Feedback uneingeschränkt nützlich ist.

3.6.1 Unter welchen Ausgangsbedingungen wirkt Feedback?

Trotz der lange währenden Überzeugung, dass Feedback niemals negative Konsequenzen nach sich ziehen könne, zeigt ein Review/eine Metaanalyse von mehr als 600 Feedback-Studien (Feedback zu den verschiedensten Leistungsbereichen, keineswegs auf Führung oder organisationale Leistung im Feld beschränkt), dass Feedback-Interventionen im Durchschnitt zwar eine Verbesserung im rückgemeldeten Leistungsbereich erbringen (um .41 Standardabweichungen); aber außerdem ergibt sich, dass ein Drittel der Effekte negativ ist, jede dritte Feedback-Intervention also zu verminderten Leistungen führt (Kluger & DeNisi, 1996). Deshalb erscheint es angebracht, einen Blick auf den Zusammenhang zwischen Feedback und Leistung zu werfen. In allen nachfolgend zitierten Primärstudien geht es um Feedback zu Führungsverhalten im weitesten Sinne (z.B. auch Führungskraft-Mitarbeiter-Beziehung), das vorher aus mehreren Perspektiven beurteilt wurde.

Eine wichtige Einflussgröße ist dabei der Ausgangswert: Die Führungskräfte, die ihr Führungsverhalten vorab deutlich besser bewerten als dies von anderen bewertet wird (*Überschätzer*), profitieren mit höchster Wahrscheinlichkeit von Feedback und zwar mit den höchsten Verbesserungsmargen. Sie verbessern ihre Leistung (gemessen durch Aufwärtsbeurteilung) und senken in der Folge das Niveau ihrer Selbstbeurteilung (Atwater, Roush & Fischthal, 1995;[119] Johnson & Ferstl, 1999;[120] Smither, London, Vasilopoulos,

[118] Wer gelernt hat, auch konträre Perspektiven anderer einzunehmen, der dürfte mit geringerer Wahrscheinlichkeit dem Gruppendenken (Janis, 1982; Schultz-Hardt, 1997) erliegen. So werden also Entscheidungsprozesse verbessert.

[119] Den Ergebnissen ist deshalb relativ geringe externe Validität zuzusprechen, weil als Führungskräfte Studenten der Marineakademie herangezogen wurden, die Anfänger (ein)führen.

[120] In dieser Studie wird zudem überprüft und bestätigt, dass sich die Leistungsverbesserung (wie auch in den anderen Studien gemessen durch Aufwärtsbeurteilung) unabhängig davon ergibt, wie die Mitarbeiter ihre Führungskraft zum ersten Messzeitpunkt bewerten.

Reilly, Millsap & Salvemini, 1995; Walker & Smither, 1999;[121] Übersichtsartikel von Yammarino & Atwater, 1997). In fast allen dieser Studien wurde geprüft, ob die festgestellten Effekte auch durch Regression zur Mitte erklärt werden können. Dies konnte mit einer Ausnahme (Veränderung der Selbstbewertung bei Smither, London et al., 1995) ausgeschlossen werden, so dass von tatsächlichen Veränderungen auszugehen ist.

Die Überschätzer sind zudem diejenigen Führungskräfte, die sich nicht nur nach individuell ausgerichtetem Feedback verbessern, sondern in gleichem Ausmaß auch generell nach dem Prozess der multiperspektivischen Beurteilung (Reilly, Smither & Vasilopoulos, 1996; Smither, London et al., 1995). Möglicherweise setzt bei diesen Führungskräften schon durch die Selbstbeschreibung (mit oder ohne anschließendem Feedback) einer der Prozesse ein, die Personen zur Gewinnung/Aktualisierung ihrer Selbstkonzepte nutzen (Kenrick, Neuberg & Cialdini, 1998, S. 51f.): Erstens könnten soziale Vergleichsprozesse im Sinne von Festinger (1954) ausgelöst werden; zweitens könnte durch die Antizipation (oder Rückmeldung) der Fremdbilder ein Reflexionsprozess angestoßen werden und drittens könnte erstmals eine bewusste Selbstwahrnehmung der eigenen Führung erfolgen. Alle drei Prozesse können letztlich über die (Re)Aktivierung von Zielen zu Leistungsverbesserungen führen. Außerdem nimmt ein verändertes Selbstkonzept zweifellos Einfluss auf das Kommunikations- und damit Führungsverhalten gegenüber anderen und führt damit unmittelbar zu einer Veränderung.

Da der Arbeitsalltag von Führenden durch viele verschiedene, kurze und unterbrochene Arbeitssequenzen charakterisiert ist (Mintzberg, 1973), bleibt wenig Zeit zur Selbstreflexion. Dadurch kann erklärt werden, weshalb bereits die pure Selbstbeschreibung (ein Prozess, der Selbstreflexion erfordert und die Aufmerksamkeit auf zu beurteilende Verhaltensweisen lenkt) ohne anschließendes Feedbackgespräch zu Verbesserungen führen kann (Reilly et al., 1996). Eine solche Selbstreflexion wäre zweifellos außerdem vorteilhaft für einen Feedbackprozess (vgl. Selbstregulationsmodell).

Im Gegensatz zu den Überschätzern nehmen Yammarino und Atwater (1997) für die sich unterschätzenden Führungskräfte (*Unterschätzer*) an, dass sie ihre Leistung halten und lediglich ihre Selbstbeschreibung erhöhen, nachdem sie Feedback von anderen bekommen haben. Einzelne Ergebnisse bestätigen dies (Atwater, Roush & Fischthal, 1995). Wo aber eine *Abnahme* der Leistung (Aufwärts-Rating) berichtet wird, kann sie gänzlich auf die Regression zur Mitte zurückgeführt werden, so dass von keinem wahren Leistungsabfall auszugehen ist (Reilly et al., 1996; Smither, London et al., 1995). Möglicherweise sehen diese Führungskräfte keine Notwendigkeit zur eigenen Verhaltensänderung, da sie von an-

[121] Da in dieser Studie nur die Leistungsveränderung, nicht die Selbstbeschreibung untersucht wird, kann nicht festgehalten werden, wie viele der Führungskräfte, die bei der ersten Messung eine schlechte Fremdbeschreibung erhalten, sich selbst besser einschätzen. Die Ergebnisse sind also allgemein für die Gruppe an Führungskräften zu verstehen, die am schlechtesten bewertet werden. Es ist – entsprechend der Forschungslage, die in vorliegendem Kapitel beschrieben wurde – aber davon auszugehen, dass sich darunter kaum Übereinstimmer befinden.

deren besser als erwartet beurteilt werden. Nur wenn die Führungskraft aufgrund einer zu hohen Diskrepanz ihr Anspruchsniveau senkt, könnte ein demotivierender Effekt eintreten, der in geminderter Anstrengung und Leistung resultiert. Als Folge dessen kann zumindest das Ergebnis von Johnson und Ferstl (1999) interpretiert werden, das eine negative Leistungsentwicklung für bestimmte Unterschätzer vorhersagt: Diejenigen, die zum ersten Messzeitpunkt von ihren Mitarbeitern noch besonders gut beurteilt werden, werden von diesen nach einem Jahr (nach Feedback) nämlich signifikant schlechter beurteilt. Dieses Ergebnis geht mit der meta-analytischen Schätzung einher, dass Lob (versus kein Lob) – welches wahrscheinlicher wird, je besser die Leistung bewertet wird – eine Feedback-Intervention nahezu ineffektiv macht (Kluger & DeNisi, 1996).

Auch das Führungsverhalten der *Übereinstimmer* wird von deren Geführten unverändert beurteilt, nachdem Feedback gegeben wurde (Atwater, Roush & Fischthal, 1995; Johnson & Ferstl, 1999). Dies gilt bei genauerer Betrachtung allerdings lediglich für diejenigen Übereinstimmer, die geringe Leistung zeigen. Ihnen gelingt es zwar, ihre Schwächen adäquat zu diagnostizieren (Übereinstimmung); in der Regel gelingt es ihnen jedoch nicht, diese signifikant zu verändern (Smither, Wohler & London, 1995; Yammarino & Atwater, 1997). Möglicherweise haben diese Führungskräfte für die notwendigen Veränderungen keine Selbsteffizienz oder sie akzeptieren die fehlenden Fertigkeiten, so dass sie keine Veränderungsmotivation aufbauen können. Taylor und Brown (1988, S. 203) sprechen in diesem Zusammenhang von "acknowledged pockets of incompetence". Umgekehrt stehen diejenigen Übereinstimmer, die hohe Leistung zeigen, vermutlich sowieso in häufiger Rückkopplung mit anderen, um ihr Verhalten an die sich verändernden Bedingungen anzupassen und können auch das organisierte Feedback für eine Leistungsverbesserung nutzen (Yammarino & Atwater, 1997).

Auch wenn erst begonnen wurde, Bedingungen zu identifizieren, unter denen Feedback mit hoher Wahrscheinlichkeit zu positiven Veränderungen führt, ermutigen die vorliegenden Befunde insofern, dass diejenigen Führungskräfte, die am dringendsten Verbesserungen brauchen (Überschätzer, vgl. Kapitel 3.5.1), diese auch zeigen, selbst wenn teilweise nur geringe Effektstärken festzustellen sind.[122] Im Vergleich zu einem meta-analytischen Ergebnis von Guzzo, Jette und Katzell (1985), das der Interventionsgruppe „Methoden zur Veränderung von Führung“ lediglich einen Effekt von d = .13 auf die Produktivität zuschreibt (die Operationalisierung von Produktivität ist unbekannt), sind die Effekte für Feedback sogar hoch. Dennoch müssen die Ergebnisse zum aktuellen Zeitpunkt noch vorsichtig interpretiert werden. Denn zum einen ist das einzig verwendete Leistungsmaß – eine Wiederholung der Aufwärtsbeurteilung – ein fehleranfälliges Maß. Zum anderen stimmen die kognitiven und affektiven Reaktionen auf negatives Feedback weniger opti-

[122] Beispielsweise d = .25-.40 bei Smither, London et al. (1995) beziehungsweise mittlere Effektstärke von d = .67 bei Reilly, Smither und Vasilopoulos (1996), für Führungskräfte mit niedriger bis mittlerer Ausgangsleistung

mistisch – also auf Feedback, in dem eine schlechtere Fremd- als Selbstbeurteilung zurückgemeldet wird: So wird es erstens von den Betroffenen als weniger akkurat und nützlich beurteilt; zweitens fördert es Ärger oder Enttäuschung (Brett & Atwater, 2001). Während bei Brett und Atwater die Fokuspersonen aber nach dem Feedback nicht mehr in ihrem Team tätig waren, fanden Atwater, Waldman, Atwater und Cartier (2000) auch in weiter existierenden Führungshierarchien (Polizei), dass Führungskräfte ihre Loyalität und ihr Commitment ihren Mitarbeitern gegenüber um so stärker reduzieren, je negativer das Aufwärtsfeedback ausfällt. Vor allem wenn ein besseres Informations- und Kommunikationsverhalten erzielt werden soll (siehe oben), scheint in der praktischen Umsetzung Coaching also besonders wichtig (vgl. auch DeNisi & Kluger, 2000).

Generell arbeitet multiperspektivisches Feedback mit den impliziten Annahmen, dass verschiedene Perspektiven verschiedene Informationen erbringen (z.B. Conway & Huffcutt, 1997) und dass eine Führungskraft Informationen von verschiedenen Perspektiven ernst nimmt. Letztere Annahme können Greguras, Ford und Brutus (2003) grundsätzlich bejahen. Allerdings stellen sie fest, dass Führungskräfte in ihre Selbstbewertung das Urteil ihres jeweiligen Vorgesetzten stärker einbeziehen als die Urteile von Mitarbeitern oder Peers. Für eine Beurteilung ihrer Führungskompetenzen gewichten Führungskräfte die Bewertungen ihrer Mitarbeiter stärker als die von Peers. In einem Coachingprozess könnten verschiedene Perspektiven gezielt bearbeitet werden.

3.6.2 Wie lange halten Feedbackveränderungen vor?

Angestrebt werden selbstverständlich anhaltende statt ephemere Veränderungen. Erste Belege sind dafür bereits zu verzeichnen. So ist eine höhere Korrelation zwischen dem Selbstbild der Führungskräfte und dem Fremdbild von Mitarbeitern beziehungsweise Vorgesetzten, Peers und Mitarbeitern nicht nur kurze Zeit nach dem Feedbackprozess, sondern auch 1 Jahr (London & Wohlers, 1991) beziehungsweise 2 Jahre später messbar (Hazucha, Hezlett & Schneider, 1993).[123] In einer Studie, die nicht auf die Erhöhung der Korrelation, sondern auf die Reduktion der Selbst-Fremd-Differenz (Niveau) fokussierte, zeigte sich jedoch keine signifikante Annäherung (Reilly et al., 1996). Der weiteren Forschung sind Studien zu empfehlen, deren Power groß genug ist, auch kleine Effekte erkennen zu können.

[123] In beiden Studien werden keine Ergebnisse für einzelne Führungskräfte-Gruppen – je nach ursprünglicher Übereinstimmung mit den Fremdbeurteilern – berichtet. Zu kritisieren ist, dass im Intervall zwischen erster und zweiter Messung, in dem die Rückmeldung liegt, keine Kontrollmessungen stattfinden, um potentielle kurzfristige Feedbackeffekte extrahieren zu können. Auch hier sind die Effektstärken gering.

Andere Studien finden stabile Leistungsverbesserungen derjenigen Führungskräfte auch noch nach einem Zeitraum von 2,5 (Reilly et al., 1996)[124] beziehungsweise 5 Jahren (Walker & Smither, 1999), die beim ersten Feedback geringe Leistung von ihren Mitarbeitern zurückgemeldet bekamen. In der Studie von Reilly et al. (1996) findet sich zusätzlich, dass das Ausmaß der Verbesserung wenig mit der Anzahl der Feedbacks zusammenhängt. Obwohl im betrachteten Zeitraum bis zu drei individuelle Feedbacks gegeben wurden, erzielten die meisten Manager bereits nach der ersten Rückmeldung ein Maß an Veränderung, das dann stabil blieb. Wegen fehlenden Kontrollgruppen ist jedoch die Rolle des Feedbacks nur schwer zu bestimmen. Insgesamt widerspricht der Befund der Annahme von DeNisi und Kluger (2000), dass wiederholtes Feedback, das auch Informationen über die jeweils erreichten Verbesserungen enthält, effektiver sei als einmaliges Feedback.

Zusammengefasst zeigen diese Studien, dass Manager ihr beurteiltes Führungsverhalten im Allgemeinen nach Feedback verbessern (Fremdbeurteilung). Insbesondere trifft das für diejenigen zu, die zunächst eine Rückmeldung über geringe Leistung erhielten oder sich überschätzten (diese beiden Gruppen überschneiden sich stark). Wichtig ist, dass die meisten dieser Studien nachwiesen, dass die Leistungsverbesserungen nicht vollständig durch Regression zur Mitte erklärt werden können (Atwater, Roush & Fischthal, 1995; Johnson & Ferstl, 1999; Smither, London et al., 1995; Walker & Smither, 1999). Erst noch untersucht werden muss, ob durch das besser bewertete Führungsverhalten auch bessere Führungsresultate erzielt werden können. Mit einem entsprechenden Längsschnittdesign könnte dann auch die Kausalität der oben dargestellten Befunde bestimmt werden, dass bessere Führung zu höherem Erfolg führt und nicht höherer Erfolg zu besserer Führungsbeurteilung.

3.6.3 Wie muss Feedback aussehen, damit es wirkt?

Wer sein Selbstbild mit Einschätzungen anderer vergleicht, beeinflusst dadurch unweigerlich auch seine Wahrnehmung der Diskrepanz zwischen erwarteter (im Idealfall auch angestrebter) und aktueller Leistung. Die Effektivität von Feedback basiert auf der Annahme, dass es diese Diskrepanz verringert und das Zielverhalten wahrscheinlicher macht. Diesen Punkt greifen sowohl die Kontrolltheorie (Carver & Scheier, 1982, 1998) als auch die Zielsetzungstheorie (Locke, Cartledge & Koeppel, 1968; Locke & Latham, 1990) auf.

[124] Die Studie von Reilly et al. (1996) ist eine Erweiterung der oben zitierten Untersuchung von Smither, London et al. (1995) um eine dritte Messung mit $N = 171$ Managern (von ursprünglich 238) und eine vierte Messung ($N = 92$).

Beide Theorien behandeln den Zusammenhang zwischen Zielen, der Motivation (genauer: Richtung, Anstrengung, Ausdauer und Aufmerksamkeit/aufgabenspezifische Strategien), diese zu erreichen, und der entsprechenden Leistung. Während die Kontrolltheorie die Veränderungsmotivation und -leistung eher auf die Aktivierung fokussiert, die durch die wahrgenommene Diskrepanz zwischen aktueller und erwarteter Leistung entsteht, beschreibt die Zielsetzungstheorie Charakteristika "guter" Ziele (z.B. Schwierigkeit und Spezifität) und Einflussfaktoren (z.B. Feedback und Zielcommitment), die über verschiedene Wirkmechanismen (Anstrengung, Ausdauer und Aufmerksamkeit) die Leistung beeinflussen. Beiden Theorien ist die Annahme gemein, dass schwierige Ziele (im Sinne der Kontrolltheorie eine hohe Diskrepanz zwischen aktueller und erwarteter Leistung) und spezifische Ziele sowie akzeptierte Ziele und spezifisches Feedback (bzw. [Selbst-]Verstärkung) zu höherer Leistung führen.

Optimale Ergebnisse sind also nicht zu erwarten, wenn man Manager lediglich mit purer Rückmeldung "abspeist". Die Leistungsverbesserung einer Person hängt davon ab, inwieweit diese das Feedback nutzt, um erstens ihre Leistung im Bezug auf ihre Ziele zu bewerten, zweitens sich im Sinne der Kontroll- und Zielsetzungstheorie "gute" Ziele zu setzen und drittens kontinuierlich den Weg zum Ziel zu beobachten (im Sinne der Kontrolltheorie "regulieren"). Um auch entsprechende Korrekturen vornehmen zu können, ist es wichtig, dass der Betroffene Kontrolle über das zu verändernde Verhalten hat. Da es im hier vorliegenden Fall um das Führungsverhalten von Führungskräften geht – also um Verhalten, das allein von der Führungskraft bestimmt und generiert wird – kann dies grundsätzlich als gegeben betrachtet werden. Das trifft allenfalls dann nicht zu, wenn Führungsziele angestrebt werden, die nicht mit den Unternehmensleitlinien vereinbar sind, oder die außerhalb des Handlungsspielraums der entsprechenden Führungskraft liegen. Für die Führungsdimensionen, die in der vorliegenden Arbeit beurteilt wurden, können diese Ausnahmen jedoch nicht eintreten.

Man darf dennoch nicht dem Irrglauben verfallen, dass sich Führungskräfte, die Feedback erhalten, automatisch adäquate Ziele zur Leistungsverbesserung setzen oder besser als Kontrollgruppen ohne Feedback abschneiden (Atwater, Waldman, Atwater & Cartier, 2000). Davon kann nicht einmal ausgegangen werden, wenn die Führungskräfte das Feedback als nützlich betrachten. Denn entgegen der intuitiven Annahme mündet individuelles Feedback im Vergleich zu normativem Feedback[125] zwar in einer höheren Nützlichkeitsbewertung, nicht jedoch in einer stärkeren Absicht, das Verhalten ändern zu wollen (Smither, Wohlers & London, 1995). Von einer Führungskraft akzeptiert wird Feedback vor allem, wenn folgende Faktoren gegeben sind: positive Bewertungen, Übereinstimmung zwischen verschiedenen Beurteilern, extreme Ratings, Konsistenz der Ratings über ver-

[125] Während die Führungskraft bei individuellem Feedback die Ratings ihrer Teammitglieder rückgemeldet bekommt, erhält sie bei normativem Feedback lediglich die Information, wie die durchschnittliche Führungskraft eingeschätzt wurde.

schiedene Dimensionen hinweg (London, 1995a). Möglicherweise ist es auch die Komplexität der Führungsaufgaben, die die Wirkmechanismen der Zielsetzungstheorie negativ beeinflusst. Jedenfalls zeigt eine Meta-Analyse, dass Feedback vor allem dann zu einer Veränderung führt, wenn es explizit mit einer Zielsetzung einhergeht (Kluger & DeNisi, 1996: $d = .51$ bei Feedback mit Zielsetzung versus $d = .30$ bei Feedback ohne Zielsetzung).

Wie aber können die empirisch gezeigten Leistungsveränderungen nach Feedback durch die Zielsetzungs- und Kontrolltheorie erklärt werden? Überschätzer nehmen die Diskrepanz wahr und gewinnen das Ziel, ihrem Leistungsanspruch in Zukunft gerecht zu werden. Dieser kann in der Nähe der Selbstbewertung angesiedelt werden. Legt man die Kriterien für Ziele aus der Zielsetzungs- und Kontrolltheorie an, dann stellt man fest, dass hoch motivierende Ziele resultieren. Denn als Ausgangspunkt steht erstens spezifisches Feedback; zweitens ist das Ziel schwierig, da die Diskrepanz hoch ist; drittens geht es um spezifische Ziele, da sie sich aus konkreten Beurteilungsdimensionen ableiten und außerdem kann viertens eine hohe Zielakzeptanz aufgebaut werden, da es um den Kernbereich der Führungstätigkeit geht. Die daraus entstehende Veränderungsmotivation führt zu einer Leistungsverbesserung (Verbesserung der Fremdeinschätzung).

Unterschätzern und Übereinstimmern wird bestätigt, dass sie ihr Ziel erreicht haben. Zumindest setzt die Rückmeldung allein kein neues Ziel, da das von anderen wahrgenommene Leistungsniveau bereits erreicht ist. Sie halten bestenfalls also ihre eigenen, schon vorher bestehenden Ansprüche und damit ihr Leistungsniveau (keine Veränderung der Fremdeinschätzung). Im Gegensatz zum hohen Anteil an negativen Feedbackeffekten, den die Meta-Analyse von Kluger und DeNisi (1996) erbrachte, finden sich unter den publizierten Feedbackstudien, die speziell auf Führungsleistung ausgerichtet sind, keine Leistungsverschlechterungen. Bei ursprünglichen Unterschätzern könnte die *Selbstbeschreibung,* die sich bei einer Messwiederholung an die Fremdbeschreibung angenähert hat, dann das wahre Leistungsniveau reflektieren, das aufgrund der Rückmeldung besser eingeschätzt werden kann.

Für die Praxis bedeutet dies nicht nur, Zielsetzungen zu integrieren,[126] sondern diese auch an die Bedingungen anzupassen. Einerseits sollten also mit den Überschätzern auf der Basis der Diskrepanzen adäquate Ziele formuliert werden. Dabei können verschiedene "Analyse-Einheiten" herangezogen werden: Ein Vergleich der individuellen Selbsteinschätzung mit Fremdeinschätzungen auf Itemebene oder dimensionaler Ebene, eine intraindividuelle Aggregation von Stärken versus Entwicklungsbereichen, ein interindividueller Vergleich auf Peer-Ebene oder gesamtorganisationaler Vergleich (internes Bench-

[126] „Kein Beurteilungsgespräch ohne Zielvereinbarung“ ist in Praxisleitfäden bereits etabliertes Wissen (z.B. Brenner & Brenner, 2002). Für die Umsetzung unabdingbar ist allerdings, dass gelernt wird, wie Feedback gegeben und angenommen wird (z.B. Praxisanleitung von Fengler, 1998).

marking) oder ein Vergleich der Organisation mit anderen Organisationen (externes Benchmarking). Je nach Ziel kann die Ergebnisrückmeldung über betriebliche Medien, Betriebsveranstaltungen oder individuelle Feedbackgespräche/-workshops erfolgen. Andererseits wird für die Unterschätzer und Übereinstimmer vorgeschlagen, im Feedbackprozess nicht auf die Diskrepanz zu fokussieren (da diese für Unterschätzer und Übereinstimmer keinen motivierenden Charakter hat), sondern auf die absolute individuelle Leistung. Diese muss in einen anderen Vergleichsrahmen als zur Fremdbewertung gestellt werden (zum Beispiel Vergleich zu erwarteter Leistung/Soll), um daraus individuelle Ziele mit möglichst konkreten Veränderungsschritten zu erarbeiten (vgl. auch DeNisi & Kluger, 2000). Dies ist auf jeden Fall solange anzuraten, bis geklärt ist, ob eine Zielsetzung tatsächlich von der Diskrepanz zwischen Selbst- und Fremdbeurteilung geleitet wird und nicht "nur" von einer schlechten Fremd- oder Selbstbeurteilung. Letzteres finden nämlich Brutus, London und Martineau (1999), die für einen Selbst-Fremd-Prädiktor kein signifikantes Beta-Regressions-Gewicht erhalten (zusätzlich zu den direkten Prädiktoren der einzelnen Perspektiven): Am stärksten wird die Zielsetzung der Führungskräfte von der Perspektive der Mitarbeiter beeinflusst (im Vergleich zu Selbst-, Peer- und Abwärtsperspektive).

Andere praktische Rahmenbedingungen, die Feedbackprozesse erfolgreicher machen, resultieren aus empirischen Arbeiten: Walker und Smither (1999) gewinnen in ihrer Langzeitstudie mit fünf Erhebungszeitpunkten beispielsweise Belege dafür, dass Teamgespräche zwischen der Führungskraft und ihren Aufwärtsbeurteilern (Mitarbeitern), in denen das vorangegangene Feedback diskutiert wird, höhere Leistungsveränderungen nach sich ziehen als keine Teamgespräche (Kontrollgruppe). Dieses Ergebnis erscheint plausibel, da durch derartige Gespräche die Zielsetzung konkretisiert, das Zielcommitment gestärkt und weitere Veränderungsmöglichkeiten (Fähigkeiten) erarbeitet werden können. Zudem kann aus einem derart offenen Umgang eine Feedbackkultur entstehen (vgl. auch Waldman & Atwater, 1998), die organisierte Feedbackprozesse überflüssig macht. Smither, London, Flautt, Vargas und Kucine (2003) zeigen in einer Feldstudie, dass diejenigen Führungskräfte, die nach dem multiperspektivischen Feedback bis zu sieben Coachingsitzungen erhalten, im Vergleich zur Kontrollgruppe (kein Coaching) ein Jahr nach der ersten Erhebung bessere Ratings von ihren Mitarbeitern und Vorgesetzten erhalten, eher Verbesserungsideen von ihren Vorgesetzten erbitten und in der Lage sind, spezifischere Ziele zu setzen. Leider blieben die Effektgrößen jeweils im niedrigen Bereich.

London, Smither und Adsit (1997) erarbeiten ein Modell, das multiperspektivisches Feedback effektiver gestaltet, indem die Verantwortung, die alle Beteiligten (Beurteilte für das Nutzen der Feedback-Information, Rater für das Geben von nützlichem Feedback und Management für die Bereitstellung von Ressourcen zur Veränderung) für den Prozess tragen, erhöht wird. Wer Verantwortung für etwas übernimmt, setzt sich die entsprechenden Ziele und kontrolliert seinen Weg dorthin, so dass die Wahrscheinlichkeit, das Ziel zu

erreichen, dadurch erhöht wird. Belohnungen für diejenigen in Aussicht zu stellen, die sich *genau* beurteilen, wird außerdem als eine Möglichkeit diskutiert, die Übereinstimmung von Ratings zu erhöhen (Murphy & Cleveland, 1995). Dies geht im Übrigen auch mit der Zielsetzungstheorie einher, weil eine hohe Übereinstimmung instrumentellen Charakter erwerben würde.

Wenn Feedback durch die vorgeschlagenen Möglichkeiten effektiver gestaltet wird, dann wäre anzunehmen, dass es auch stärkere Effekte als andere Interventionsstrategien zeigt. Denn in einer Meta-Analyse von Guzzo et al. (1985), die die Wirkung von 11 verschiedenen organisationspsychologischen Interventionsstrategien auf die Arbeitsproduktivität überprüft, liegt Leistungsbeurteilung mit Feedback zwar in einem ähnlichen Effektbereich (d = .35) wie bei Kluger und DeNisi (1996), jedoch unter dem mittleren Effekt aller Interventionen (d = .44). Nach der aufwändigen Intervention des Trainings finden Guzzo et al. (1985) für Zielsetzung indes den größten Effekt, was sich – wie oben beschrieben – gut mit Leistungsfeedback verbinden lässt.

Letztlich geht es also um Führungserfolg – für das Unternehmen, die Mitarbeiter und nicht zuletzt die Führungskräfte. Dass zentrale Organisationskennwerte auf dem Spiel stehen, macht die multiperspektivische Erfassung von Führung so wichtig. Die Feedback-Gespräche haben eine Unterbrechung des skizzierten Zusammenhangs zwischen schlechterer Aufwärtsbeschreibung (relativ zur Selbstbeschreibung) und negativen Unternehmens-Resultaten zum Ziel: Dadurch, dass – wo nötig – "blinde Flecken" aufgedeckt und konkrete Veränderungsschritte auf den Weg gebracht werden, sollen auch die Vorzeichen der Erfolgskriterien ins Positive gedreht werden. Und zwar mit Konsens aller Beteiligten. Hier sollte die Möglichkeit genutzt werden, die verschiedenen „Sprachen" der Beteiligten und damit die verschiedenen Bedeutungszuschreibungen einzelner Beurteilungsdimensionen zu berücksichtigen. Da Führungsverhalten keinen objektiv messbaren Gegenstand darstellt, gehen in jede Fragebogenbearbeitung auch idiosynkratische Verzerrungen des Bearbeiters ein, sein Erfahrungshintergrund und seine individuellen Ziele. Feedbackgespräche bieten aber die Chance, Führung als Interaktion zwischen Führenden und Geführten zu verstehen und damit den vielfältigen Formen der neuen Arbeitsmodelle nicht mit verkrusteten Führungsinstrumenten gegenüber zu stehen. Die „Wissensgesellschaft" braucht Austausch, keine Unantastbarkeit.

3.7 Zusammenfassung und Fazit

Multiperspektivische Führungsbeschreibung bedeutet, dass Führungskräfte aus verschiedenen Perspektiven hinsichtlich ihrer Leistung, dem Führungsverhalten, beurteilt werden. Zu diesen Perspektiven gehört die traditionelle *Abwärtsbeschreibung* durch die Vorgesetzten, der durchaus hohe Validität zuzuschreiben ist. Die *Aufwärtsperspektive* birgt aber fundamentale Vorteile, die sie mindestens ebenso stark für eine Führungsbeurteilung prädestinieren. Zum Beispiel sind Mitarbeiter unmittelbarer am zu beurteilenden Führungsverhalten beteiligt als andere Fremdbeurteiler, was sie zu exklusiven Beurteilern bestimmter Dimensionen macht. Weiter kann ihre Raterzahl leichter erhöht werden, so dass ihre Reliabilität und Validität sogar über die der Abwärtsbeurteilung steigen.

Insbesondere dann, wenn aus der multiperspektivischen Beurteilung, die gemeinsam mit den Abwärts- und Aufwärtsratings ein 270°-Design ergeben, eine Führungskräfte-Entwicklung resultieren soll, sind *Selbstratings* unabdingbar. Auch wenn diese dazu tendieren, erstens überhöht zu sein, zweitens nur gering mit Fremdratings zu korrelieren und drittens deutlich geringere Zusammenhänge zu Erfolgsmaßen aufzuweisen als Fremdratings, wird ihnen die zentrale Ausgangsposition für den Aufbau von Änderungsmotivation im *Feedbackprozess* zugesprochen. Insgesamt basiert multiperspektivisches Feedback auf der Annahme, dass aus unterschiedlichen Perspektiven unterschiedliche (perspektivenspezifische) Information gewonnen werden kann. Das geht mit dem methodischen Befund Hand in Hand, dass die verschiedenen Perspektiven inkrementelle Validität besitzen. Überschätzer verbessern nach einer Feedback-Intervention in der Regel ihre Leistung, während Unterschätzer und Übereinstimmer dazu offensichtlich keine Veranlassung sehen. Dies ist mit der Kontrolltheorie sowie der Zielsetzungstheorie erklärbar. Ziel des Feedbackprozesses sollte sein, die Diskrepanz zwischen Selbst- und Fremdratings zu minimieren (bei Überschätzern) beziehungsweise neue Ziele zu setzen (bei Unterschätzern und Übereinstimmern).

Es ist bislang nicht endgültig geklärt, welche Rolle die Selbst-Fremd-Diskrepanz im Entstehungsgefüge von Führung und Erfolg spielt. Die Fremd-Fremd-Diskrepanz blieb bis dato ganz außen vor. Uneindeutig sind die Ergebnisse dahingehend, dass zwar die interperspektivische Diskrepanz mit Erfolgsvariablen negativ kovariiert, dass aber kaum erforscht wurde, ob dies tatsächlich an der Diskrepanz liegt. Einzelne Hinweise sprechen für eine alleinige oder zumindest überwiegende Wirkung der Führungsleistung, wie sie aus der Fremdperspektive beurteilt wird. Jedenfalls scheint die Aufwärtsbeurteilung – also quasi die Umkehrung der Hierarchie – Erfolgsindikatoren anzuzeigen, die mit denen von Assessment Centern konkurrieren können (McEvoy & Beatty, 1989). Vor dem Hintergrund dieser dünnen Befundlage scheint eine Reduktion der mangelnden Übereinstimmung zwischen Selbst- und Fremdbeurteilern zweitrangig hinter einer Verbesserung der Führungs-

qualität an sich – auch wenn in Forschung und Praxis häufig ein anderes Fazit gezogen wird, weil die negativen Korrelate einer geringen Selbst-Fremd-Übereinstimmung so imposant sind. Sie reichen von Leistung über Gesundheit bis zur Zufriedenheit und dem Karrierestopp der Führungskräfte selbst.

Die Anforderung, sich mit anderen über seine Führung auseinander zu setzen, seine Führung selbst zu reflektieren und zu regulieren, reiht sich bereits jetzt ein in die Selbst-Kompetenzen, die im Katalog der Schlüsselqualifikationen eine immer wichtigere Rolle spielen. Auf eindeutigere Hinweise zum Zusammenhang zwischen Diskrepanz und Erfolgsmaßen sowie auf effektivere Interventionsstrategien zur Reduktion der Diskrepanzen ist zu hoffen, sobald mehr Erkenntnisse über deren Entstehungszusammenhang gewonnen sein werden. Aufgrund der aktuellen Erkenntnislage können lediglich plausible Erklärungen erarbeitet werden, die entweder eher perspektivenbezogenen oder idiosynkratischen Bias fokussieren. Von diesen erscheinen Selbstdarstellungstheorien besonders tauglich, empirisch überprüft wurde kein potentielles Erklärungsmodell in ausreichender Weise. Bekannt sind indes viele Moderatorvariablen, die zumindest für berufliche Leistung im Allgemeinen die Übereinstimmung zwischen Selbst- und Fremdbeurteilern beeinflussen können. Zu den wichtigsten gehören: Tätigkeit (Manager vs. Arbeiter), Anonymitätsinstruktion und Validierungserwartung. Sie stellen einen potentiellen Ausgangspunkt dar, wenn es um die Suche nach Erklärungen für die mangelnde Übereinstimmung geht. Im Bereich der Führungsbeurteilung müssen viele Moderatorvariablen allerdings erst noch untersucht werden.

In den Ausführungen zu multiperspektivischem Führungsfeedback zeigt sich so deutlich wie kaum anderswo, dass Führung eine wechselseitige Beziehung zwischen Führungskraft und Mitarbeiter bedeutet, wenn auch erstere aufgrund ihrer Positionsmacht mehr Gestaltungsmöglichkeiten hat. Wäre Führung allein durch die Perspektive der Führungskraft bestimmt, dann wäre die prognostische Validität der Fremdbeurteilungen und der Selbst-Fremd-Diskrepanz für Führungserfolg unplausibel.

4 Fragestellung

„In these papers [dealing with 360° feedback], I see a hodgepodge of techniques, testimonials, cautions, methodological difficulties, some axes being ground, and a rather confusing lack of cohesion“ (Dunnette, 1993, p. 374)

In den vorausgehenden Kapiteln wurde das Forschungsfeld der Personalführung und insbesondere das der multiperspektivischen Führungsbeurteilungen dargestellt, das derzeit noch durch eine Vielzahl nebeneinander stehender Befunde charakterisiert ist. Es konnte gezeigt werden, dass die transformationale Führung von Bass (1985) ein vielversprechendes Führungsmodell darstellt, in das viele herkömmliche Konzepte integriert werden können, das allerdings auch darüber hinaus geht.

Ferner wurden alle grundlegenden Untersuchungsgegenstände aus dem Forschungsfeld der multiperspektivischen Führungsbeurteilung behandelt. Dieses lässt sich in folgende Schwerpunkte gliedern: a) Selbst-Fremd-Übereinstimmung (z.B. London & Wohlers, 1991), b) Korrelate dieser Übereinstimmung (z.B. Atwater & Yammarino, 1992), c) Reaktionen auf Feedback (z.B. Bernardin et al., 1993), d) Nutzen von Feedback (z.B. London & Smither, 1995) und e) praktiker-orientierte Belange der Skalenkonstruktion und Befragungsdurchführung (z.B. Tornow, 1993). Als zentrale Erkenntnisse wurden aus den entsprechenden Forschungsarbeiten festgehalten, dass in allen bislang untersuchten Branchen deutliche interperspektivische Diskrepanzen auftreten, dass diese mit verschiedenen Erfolgsvariablen korrelieren und dass sie sich durch Feedback reduzieren lassen. Variablen, die die interperspektivische Übereinstimmung moderieren oder gar in kausalem Sinne hervorbringen, wurden im Bereich der *Führungs*beurteilung bislang vernachlässigbar selten untersucht. Die Ziele der vorliegenden Arbeit betreffen hauptsächlich die Bereiche a) und b), wobei zusätzlich die *Fremd-Fremd*-Übereinstimmung betrachtet wird, die bisher kaum Beachtung fand. Als Korrelate von Führungsbeurteilungen und interperspektivischer Übereinstimmung werden fast ausschließlich Erfolgskriterien betrachtet. Insofern ist die Studie in demjenigen Bereich der Führungsforschung anzusiedeln, der den Zusammenhang zwischen Führung und Führungserfolg im engeren Sinne (nicht Unternehmenserfolg) untersucht.[127] Auch für den Bereich e) können verschiedene Hinweise aus den gewonnenen Erkenntnissen extrahiert werden.

Die im Folgenden vorgestellte Untersuchung knüpft an den beschriebenen Erkenntnisstand an und fokussiert zum einen auf die interperspektivische Übereinstimmung zwischen Selbst-, Aufwärts- und Abwärtsbeurteilungen von Führungsbeurteilungen sowie auf darauf

[127] Die Betrachtung des Zusammenhangs zwischen Führungs- und Unternehmenserfolg beschreibt eher eine betriebswirtschaftliche Perspektive auf Personalführung. Weibler (2001) arbeitet die Unplausibilität der Annahme heraus, die Wirkung von Führungserfolg auf Unternehmenserfolg zu leugnen. Dennoch muss sie künftig gründlicher untersucht und zwischen den Wissenschaften kommuniziert werden. Denkbar ist, dass sich brachenabhängige Unterschiede ergeben. Andere situative Moderatorvariablen (z.B. Größe des Unternehmens) sind bereits untersucht (Gebert et al., 1987).

wirkende Moderatorvariablen und fokussiert zum anderen auf den Zusammenhang zwischen Führungsbeurteilungen beziehungsweise interperspektivischen Beurteilungsdiskrepanzen und Erfolgsvariablen. Alle Daten wurden mit Fragebögen erhoben, die überwiegend für die vorliegende Studie konstruiert wurden. Im Zentrum vieler Analysen stehen die Daten aus den Fragebögen, anhand derer Führungskräfte ihr eigenes Führungsverhalten beschrieben, dieses aber auch von ihren Vorgesetzten und Mitarbeitern beschrieben wurde. Diese 270°-Beurteilungsstudie macht – im Gegensatz zu vielen anderen Untersuchungen – den Versuch, einen Forschungsbeitrag zu verschiedenen Fragestellungen aus dem Gebiet der (multiperspektivischen) Führungsbeurteilung zu leisten, welche nun im Einzelnen mit Bezug zu entsprechender Literatur aufgeführt werden. Jedem bearbeiteten Themenkomplex (Fragestellung) sind mehrere Hypothesen zugeordnet.

Fragestellung 1: Unterschiede zwischen beobachtetem und erwartetem Führungsverhalten

> „So wie schriftlich fixierte Gesetze einer Gesellschaft keine Garantie für gesetzeskonformes Verhalten sämtlicher Mitglieder bieten, kann auch die tatsächlich gelebte Führungskultur eines Unternehmens erheblich von den schriftlich fixierten Führungsgrundsätzen abweichen, zumal diese nicht mit den individuellen Werthaltungen und -präferenzen der Organisationsmitglieder übereinstimmen müssen.“
> (Wunderer, 1997, S. 290)

Ausgangspunkt der vorliegenden Arbeit war die Ermittlung der Diskrepanzen zwischen erwartetem und beobachtetem Führungsverhalten. In der Regel wird von Führungskräften ein Führungsverhalten erwartet, das die Führungsgrundsätze des Unternehmens realisiert. Da die Führungsgrundsätze meistens sehr idealtypisch formuliert sind, ist zunächst von einem beobachteten Führungsverhalten ("**Ist**") auszugehen, das durchschnittlich unter dem erwarteten – in den Führungsgrundsätzen fixierten – Führungsverhalten (Unternehmens-"**Soll**") liegt. Diese Differenz ist in der Natur der Führungsgrundsätze begründet, die nach Gabele, Liebel und Oechsler (1992, S. 51) sogar „bewußt eine reale Utopie propagieren“ sollen.

Während Führungsgrundsätze und generell Unternehmensleitlinien eine Vision der gesamten Organisation verkörpern, spielt für den Alltag jeder einzelnen Führungskraft eine andere Art von Erwartungen an ihr Führungsverhalten eine unmittelbarere Rolle: Sowohl die Führungskräfte selbst als auch ihre direkten Mitarbeiter haben eine ideale Vorstellung von erwünschtem Führungsverhalten. In der vorliegenden Arbeit wird diese Art von „erwartetem“ Verhalten als „**Ideal**“ bezeichnet. Auch hierfür ist davon auszugehen, dass das beobachtete Führungsverhalten durchschnittlich unter dem idealen Führungsverhalten der Führungskräfte sowie der Mitarbeiter des Unternehmens liegt. Es ist also folgende Hypothese zu prüfen:

1) **Die Bewertung des aktuellen Führungsverhaltens (Ist) fällt negativer aus als die Einschätzung des erwarteten Führungsverhaltens.**

Vor dem Hintergrund, dass in der vorliegenden Studie Führungsbeurteilungen (Ist) aus drei Perspektiven (Selbst-, Aufwärts-, Abwärtsbeurteilung) und Führungsideale aus zwei Perspektiven (Selbst- und Aufwärtsperspektive) erhoben wurden, gilt für Hypothese 1):
Die Ist-Beurteilungen aus allen drei berücksichtigten Perspektiven fallen erstens negativer aus als die Soll-Kennwerte des Unternehmens sowie zweitens negativer als die Ideal-Beschreibungen der Aufwärts- und Selbstbeurteiler. Insbesondere liegt Aufwärts-Ist unter Aufwärts-Ideal und Selbst-Ist unter Selbst-Ideal.

Für Mitarbeiter kann eine höhere Diskrepanz zwischen beobachtetem Ist- und erwartetem Ideal-Führungsverhalten angenommen werden, da sie in der „Beoabachter-Rolle“ verschiedene situationale Faktoren, die Führung erschweren, unterschätzen. Unter der Bedingung von Fragestellung 3 ist für Aufwärtsbeurteiler prinzipiell von einer höheren Diskrepanz zwischen beobachtetem und erwartetem Führungsverhalten auszugehen (Hypothese 2).

2) **Mitarbeiter nehmen eine größere Ist-Ideal-Diskrepanz wahr als Führungskräfte.**

Da Führungsgrundsätze ein Instrument darstellen, mit dessen Hilfe die Werte und Visionen eines Unternehmens verbreitet und eine bestimmte Kultur implementiert werden (Albach, 1983; Wunderer, 2001), wird für den Fall, dass den befragten Führungskräften die Führungsgrundsätze bekannt sind, kein Unterschied zwischen Soll und Ideal erwartet. Hingegen steigt die Wahrscheinlichkeit, vom Soll (Ableitung aus Führungsgrundsätzen) abweichende Ideal-Vorstellungen zu konstruieren, wenn die Führungsgrundsätze unbekannt sind. Da diese in vielen Unternehmen (auch hier) in einem top-down-Prozess eingeführt werden/wurden, ist mit steigender Hierarchieebene eine höhere Übereinstimmung zwischen Soll und Ideal zu erwarten. Auf höheren Ebenen sollten sich die Soll-Vorstellungen bei einem solchen Prozess nämlich schneller verbreiten. Deshalb werden Hypothesen 3) und 4) aufgestellt.

3) **Die Ideal-Einschätzung der Führungskräfte (Selbst-Ideal) stimmt stärker mit dem Unternehmens-Soll überein, wenn diese die Führungsgrundsätze kennen (vs. nicht/kaum kennen).**

4) **Je höher die Hierarchieebene der fokussierten Führungskraft ist, umso besser stimmen Selbst-Ideal und Soll überein.**

Fragestellung 2: Korrelative Zusammenhänge zwischen Selbst-, Aufwärts- und Abwärtsbeurteilungen

Wie in Kapitel 3.2 dargestellt wurde, handelt es sich bei der geringen Selbst-Fremd-Übereinstimmung (Niveau-Unterschiede und Korrelationen) um ein vielfach bestätigtes Phänomen, das bereits für Beurteilungsgegenstände aus verschiedenen Inhaltsbereichen gefunden wurde. Allerdings liegen äußerst wenige Untersuchungen aus dem deutschsprachigen Raum vor, geschweige denn mit hier verorteten *Management*-Stichproben oder gar Management-Stichproben der *Automobilindustrie*. Erste Hinweise, dass sich deutsche Manager weniger stark überschätzen[128] als US-amerikanische, erbrachte Ludorf (1998). Die überwiegende Mehrheit der in Kapitel 3.2 berichteten Ergebnisse stammt indes aus amerikanischen Stichproben. Dass sich die Selbst-Fremd-Übereinstimmung für Fokus-Personen aus dem Management von der für Fokus-Personen aus dem Nicht-Management unterscheidet, wurde bereits in Kapitel 3.3.2 dargestellt (Meta-Analysen von Conway & Huffcutt, 1997; Heidemeier & Moser, 2002). Die korrelative Übereinstimmung verschiedener Perspektiven steht im Zentrum von Fragestellung 2, die Niveau-Unterschiede zwischen den Perspektiven stehen im Zentrum von Fragestellung 3.

In diesen beiden Fragestellungen wird also überprüft, ob die an anderen Stichproben ermittelte (mangelnde) Übereinstimmung zwischen Beurteilungen aus verschiedenen Perspektiven auf die vorliegende Stichprobe generalisierbar ist. Dafür spricht beispielsweise, dass sich die Ergebnisse derjenigen Studien, die den Zusammenhang zwischen der Aufwärtsbeurteilung von Führung und Erfolgskriterien untersuchten, über die Stichproben hinweg sehr ähnlich sind (vgl. Tabelle 9). Darunter finden sich auch Stichproben aus der deutschen Automobilbranche (Nickel & Krems, 1998; Schmidt, 1996). Geht man analog zu dieser stichprobenübergreifenden Ähnlichkeit der Kriteriumsvalidität der Aufwärtsbeurteilung von Führung auch von einer stichprobenübergreifenden Ähnlichkeit der Übereinstimmung zwischen verschiedenen Beurteilungsperspektiven aus, dann gelten Hypothesen 6) und 7).

Weiter wird angenommen, dass die Enge des Zusammenhangs nicht allein durch die Perspektive, sondern auch durch die zu beurteilende Dimension bestimmt wird (Hypothese 9). Durchschnittlich wird kein Unterschied zwischen der Selbst-Aufwärts- und der Selbst-Abwärts-Korrelation erwartet (Hypothese 8). Hypothese 9) ergibt sich aus der Tatsache,

[128] Überschätzung wird dabei zunächst „neutral“ gebraucht. Der Begriff soll lediglich die Richtung des Niveau-Unterschieds zwischen Selbst- und Fremdbeurteilern ausdrücken und nicht, ob dieser durch eine zu milde Selbst- oder eine zu strenge Fremdbeurteilung zustande kommt.

dass unterschiedliche Führungsdimensionen von unterschiedlichen Beurteilern unterschiedlich gut beobachtet werden können. Da es für verschiedene Analysen angezeigt ist, diejenigen Mitarbeiterratings, die sich auf dieselbe Führungskraft beziehen, zu aggregieren, wird gemäß Hypothese 5) überprüft, ob deren Übereinstimmung mit anderen Perspektiven dadurch steigt. Dies ist deshalb anzunehmen, weil die Reliabilität durch die Erhöhung der Rater-Anzahl gesteigert werden kann (vgl. Kapitel 3.1.1).

5) **Wenn man die Aufwärts-Beurteilungen, die die gleiche Führungskraft betreffen, aggregiert (versus nicht aggregiert), korrelieren sie stärker mit Beurteilungen aus anderen Perspektiven (Selbst, Abwärts).**
6) **Die Beurteilungen verschiedener Perspektiven (Selbst-Ist, Aufwärts-Ist, Abwärts-Ist) korrelieren gering ($.10 \leq r < .30$).**
7) **Die Korrelationen zwischen Selbst- und Fremdbeurteilungen sind geringer als die Korrelationen zwischen den Fremdbeurteilungen.**
8) **Die Korrelationen zwischen Selbst- und Aufwärtsbeurteilungen unterscheiden sich nicht von den Korrelationen zwischen Selbst- und Abwärtsbeurteilungen.**
9) **Die Korrelationen zwischen den Perspektiven variieren in ihrer Stärke über die Dimensionen hinweg.**

Fragestellung 3: Niveau-Unterschiede zwischen Selbst-, Aufwärts- und Abwärtsbeurteilungen

Crampton und Wagner (1994) überraschten mit dem meta-analytischen Befund, dass Selbstbeurteilungen nur in 38 von 143 untersuchten Fällen statistisch bedeutsam positiver ausfallen als die jeweils entsprechenden Fremdbeurteilungen. Um diejenigen Beurteilungsdimensionen besser identifizieren zu können, auf denen die Selbstbeurteilung besonders anfällig für eine Mildetendenz ist, plädieren die Autoren dafür, bereichsspezifische Untersuchungen durchzuführen. Der Führungsbereich scheint vor dem Hintergrund der empirischen Befunde, die in Kapitel 3.2.1 dargestellt wurden und unter der Annahme, dass Niveau-Unterschiede branchen- und kulturunabhängig auftreten, besonders anfällig zu sein. Folglich werden verschiedene Hypothesen zu Niveau-Unterschieden zwischen zwei verschiedenen Beurteilungsperspektiven formuliert.

10) **Zwischen allen erfassten Perspektiven (Selbst-, Abwärts-, Aufwärtsbeurteilung) bestehen Niveau-Unterschiede.**

Aus den Abhandlungen zum Above-Average-Effekt sowie zu den Verzerrungen bei Selbstbeurteilungen (Kapitel 3.2.3) ist zu schließen, dass die Selbstbeschreibung positiver ausfällt als die Fremdbeschreibungen. Da in der vorliegenden Studie aber alle Selbst- und Abwärtsbeurteiler Führungskräfte sind, kann gemäß der Sozialen Vergleichstheorie (Tajfel & Turner, 1986; Turner, 2000) angenommen werden, dass sich beide Beurteilergruppen unabhängig von ihrer Beurteilungsperspektive als Führungskräfte kategorisieren. Wenn sie sich zur selben Gruppe zählen, besteht keine Motivation zur Distinktheit. Die Kategorie der Führungskräfte (vs. Mitarbeiter) ist in streng strukturierten Unternehmen im Übrigen sehr salient. Also wird angenommen:

11) **Die Niveau-Unterschiede zwischen der Selbst- und Abwärtsbeurteilung sind geringer als die zwischen der Selbst- und Aufwärtsbeurteilung.**
12) **Die Selbstbeurteilungen fallen am positivsten, die Aufwärtsbeurteilungen am negativsten aus. Die Abwärtsbeurteilungen nehmen eine Zwischenstellung ein und unterscheiden sich von den Selbstbeurteilungen weniger als von den Aufwärtsbeurteilungen.**

Vor allem auch deshalb, weil sich Abwärts- und Aufwärtsbeurteiler in derart unterschiedlichen Rollen wiederfinden, ist nicht davon auszugehen, dass den sich selbst beurteilenden Führungskräften per se Selbstüberschätzung oder keine Selbstüberschätzung zugeschrieben werden kann. Viel plausibler erscheint es, die Selbstüberschätzung immer in Abhängigkeit von der Vergleichsperspektive zu berichten.

13) **Die Selbstüberschätzung der Führungskräfte im Vergleich zur Abwärtsbeurteilung ist unabhängig von deren Selbstüberschätzung im Vergleich zur Aufwärtsbeurteilung.**

Analog zu Hypothese 9) wird angenommen, dass die interperspektivischen Differenzen für die einzelnen Führungsdimensionen unterschiedlich hoch ausfallen. Da unterschiedliche Führungsanforderungen auf eine gemeinsame Basis an Kompetenzen zurück greifen, ist nicht davon auszugehen, dass die Diskrepanzen auf den einzelnen Dimensionen unabhängig voneinander sind. Doch eine deutliche Unterscheidung würde auf die Notwendigkeit hinweisen, zukünftig verstärkt mit mehrdimensionalen Instrumenten zu arbeiten und entsprechende Analysen auch häufiger als bislang faktorbezogen auszuwerten.

14) **Das Ausmaß der interperspektivischen Differenzen variiert über die Führungsdimensionen hinweg.**

Fragestellung 4: Moderatorvariablen für die Beurteilungs-Übereinstimmung zwischen verschiedenen Perspektiven

Moderatoranalysen wurden bislang überwiegend für die *Selbst*-Fremd-Übereinstimmung durchgeführt, weil das primäre Forschungsinteresse in der Bestimmung der Güte von Selbst-Ratings lag. Die Ratinggüte wurde in der Regel als *korrelative* Übereinstimmung operationalisiert. Da in der Beratungspraxis – bei der Rückmeldung multiperspektivisch erhobener Beurteilungen – hingegen vorwiegend auf Niveau-Unterschiede geachtet wird, sind auch hier wirksame Moderatorvariablen von weit reichendem (praktischen) Interesse. Heidemeier und Moser (2002) stellen eine Ausnahme dar und untersuchen auch potentielle Moderatoren für die Selbst-Abwärts-*Differenz* (vgl. Kapitel 3.3). Wenn sich zudem herausstellt, dass auch Fremd-Fremd-Übereinstimmung mit Erfolgsvariablen zusammenhängt (vgl. Fragestellung 10), dann sind moderierende Einflussfaktoren nicht nur für die Selbst-Fremd-, sondern eben auch für die Fremd-Fremd-Übereinstimmung wichtig zu erforschen.

Generell stellen Studien, die die moderierende Wirkung von Drittvariablen auf die Übereinstimmung von *Führungsbeurteilungen* untersuchen, eine Rarität dar.

In der vorliegenden Studie werden verschiedene Moderatorvariablen auf ihre Wirkung bezüglich der Übereinstimmung zwischen der Selbst- und Aufwärtsbeurteilung, der Selbst- und Abwärtsbeurteilung sowie der Abwärts- und Aufwärtsbeurteilung von Führung überprüft. Unter hoher „Übereinstimmung“ werden hier sowohl hohe Selbst-Fremd-Korrelationen als auch geringe Niveau-Unterschiede zwischen den Perspektiven bezeichnet.

15) Die in Tabelle 14 aufgeführten Variablen moderieren in jeweils angegebener Richtung die Stärke des korrelativen Zusammenhangs und das Ausmaß der Niveau-Unterschiede zwischen den angegebenen Beurteilungsperspektiven. Die Übereinstimmung zwischen den Perspektiven steigt oder fällt also mit der Ausprägung der Moderatorvariablen.

Tabelle 14 gibt einen Überblick über alle untersuchten Moderatorvariablen. In Klammern ist jeweils die Richtung der Hypothese notiert. ↑ (↓) bedeutet, dass die Selbst-Fremd-Übereinstimmung steigt (sinkt), je stärker die Moderatorvariable ausgeprägt ist.

Tabelle 14. Überblick über die untersuchten Moderatorvariablen

Untersuchte Moderatorvariable	S-Auf	S-Ab	Ab-Auf
Personale Faktoren (Selbstrater/Ratee):			
a) Führungsverhalten			
Transformationale Führung (↑)	x	x	x
Transaktionale Führung (CR↑, MBA↓, MBP↓)	x	x	x
Laissez Faire (↓)	x	x	x
b) Alter (↑)	x	x	
c) Wichtigkeit Ratee (o.H.)	x	x	
d) Ideal-Ist Ratee (o.H.)	x	x	
Personale Faktoren (Fremdrater)			
e) Wichtigkeit Aufwärts (↓)	x		x
f) Alter der Fremdrater (↑)	x	x	x
Organisations- und tätigkeitsbezogene Faktoren:			
g) Führungsspanne (Ratee) (↓)	x	x	x
h) Führungsspanne (Abwärtsrater) (↓)		x	x
i) Hierarchie Ratee (↓, Niveau)(↓, Korrelation)	x	x	
j) Dauer der Organisationszugehörigkeit (↑)	x	x	x
k) Dauer der Abteilungszugehörigkeit (↑)	x	x	x
l) Dauer der Zusammenarbeit zwischen Rater und Ratee (↑)	x	x	x
Durchführungsbedingungen (Format des Befragungsinstruments):			
m) Beschreibungsebene: ergebnisbezogene Itemformulierung	x	x	x
Durchführungsbedingungen (allgemeine Rahmenbedingungen):			
n) Erfahrung als Beurteilter (Ratee) (↑)	x	x	
o) Erfahrung als Beurteiler (Ratee) (↑)	x	x	
p) Beteiligungsquote (↑)	x	x	
q) Beteiligung des Ratees (↑)			x
r) Ideal Selbst-Aufwärts-Diskrepanz (↓)	x		
s) Kenntnis der Führungsgrundsätze (↑)	x		x
t) Führungsverantwortung (vs. keine Führungsverantwortung) der Aufwärtsrater (↑)	x		
u) Aufwärtsbeurteiler als Stellvertreter (↑)	x		x

Anmerkungen. S: Selbstbeurteilung, Auf: Aufwärtsbeurteilung, Ab: Abwärtsbeurteilung, o.H.: ohne Hypothese über die Richtung der moderierenden Wirkung.

Im Folgenden werden die Annahmen zu den einzelnen potentiellen Moderatorvariablen dargestellt.

Personale Faktoren (Selbstrater/Ratee):

a) Führungsverhalten: Transformationale und transaktionale Führung, Laissez Faire

Plausibel wäre, dass der Führungsstil der Ratee-Person selbst das Ausmaß der Übereinstimmung zwischen Selbst- und Fremdbeurteilung moderiert, da unterschiedliche Führungsstile unterschiedlich starke Kenntnisse über die Führenden mit sich bringen. Betrachtet man den „Full Range of Leadership“ (Bass, 1985; vgl. Kapitel 2.5.3.2), dann ist davon auszugehen, dass Führungskräfte, die die transformationalen Prinzipien verwirklichen oder im Sinne eines individuellen sozialen Austauschs kontingente Belohnung einsetzen (CR), eine stärkere Führungsbeziehung zu ihren Mitarbeitern haben als Führungskräfte, die nur in Ausnahmefällen intervenieren (Management by exception [MBA]) oder die passiv führen (Management by exception [MBP], Laissez Faire [LF]). Erwartungen sind klarer formuliert und es findet auch individuelle Führung statt. Insgesamt besteht für die Fremdbeurteiler eine größere Beobachtungsstichprobe hinsichtlich des zu beurteilenden Führungsverhaltens, so dass letztlich eine höhere Selbst-Aufwärts-, Selbst-Abwärts- und Abwärts-Aufwärts-Übereinstimmung zu erwarten ist.

b) Alter des Selbstraters

Je älter eine Person ist, umso häufiger war sie in der Vergangenheit bereits mit Beurteilungen und Feedback konfrontiert. Die dadurch gegebenen Möglichkeiten, Selbst- und Fremdwahrnehmungen zu reflektieren und gegebenenfalls aufeinander abzustimmen, sollten dazu führen, dass mit zunehmendem Alter der Führungskraft ihre Selbstbeurteilung valider wird und demzufolge die Selbst-Fremd-Übereinstimmung steigt.

c) Wichtigkeit Ratee

Für die Wichtigkeit von Beurteilungsinhalten sind Moderatorwirkungen in beide Richtungen plausibel. Auf der einen Seite kann angenommen werden, dass eine Führungskraft sich stärker mit den Dimensionen auseinandersetzt, die als wichtig anerkannt sind (Fox, Caspy & Reisler, 1994) oder die sie selbst für wichtig (erfolgsrelevant) erachtet. Aus einer solchen, intensiveren Beschäftigung (Zeit, Fortbildung, Selbstmanagement etc.) sollte letztlich eine Leistungsverbesserung auf diesen Dimensionen resultieren. In diesem Fall gäbe es keinen Anlass für eine Minderung der Beurteiler-Übereinstimmung. Im Gegenteil: Wenn das Verhalten dadurch häufiger beobachtet werden kann oder zum Gegenstand von Gesprächen wird, dann ist sogar eine besonders hohe Übereinstimmung zu erwarten. Auf

der anderen Seite kann angenommen werden, dass die Selbstbewertung in den als wichtig erachteten Dimensionen besonders gut ausfällt, zum Beispiel aufgrund des Motives nach Selbstwertschutz und –erhöhung. Dann wäre einseitig die Selbstbewertung verzerrt und die Selbst-Fremd-Übereinstimmung gemindert.

d) Ideal-Ist Ratee

Betrachtet man die Abweichung von einer idealen Führung als groben Indikator für die Güte der Führung, dann gilt: Je höher die Ideal-Ist-Diskrepanz von der Führungskraft selbst wahrgenommen wird, desto negativer kann ihre Führung bewertet werden. Da eine schlechte Führung mit höherer Wahrscheinlichkeit aber auch Facetten enthält, die eine Fremdbeurteilung erschweren (z.B. unzureichende Kommunikation), sollte mit steigender Ideal-Ist-Diskrepanz die Selbst-Fremd-Übereinstimmung sinken. Auf der anderen Seite ist aber auch denkbar, dass sich die Beurteiler aus verschiedenen Perspektiven über die schlechte Führung „einig sind“ und hohe Übereinstimmung auf niedrigem Niveau erzielen.

Personale Faktoren (Fremdrater):

e) Wichtigkeit Aufwärts

Aufwärtsbeurteilungen können von der Wichtigkeit beeinflusst werden, die die Beurteiler (Mitarbeiter) den zu beurteilenden Verhaltensweisen zusprechen. Mit steigender Wichtigkeit erhöhen sich die Erwartungen, das entsprechende Führungsverhalten auch erfahren zu wollen. Infolge dieser überhöhten Erwartungen können Fremdbeurteilungen auf wichtigen Items geringer als gerechtfertigt ausfallen. Bei hoher zugeschriebener Wichtigkeit ist also von einem höheren Niveau-Unterschied und einer geringeren Korrelation (Selbst-Aufwärts, Abwärts-Aufwärts) auszugehen als bei niedriger Wichtigkeit.

f) Alter der Fremdrater

Je älter Fremdrater sind, umso höher ist die Wahrscheinlichkeit, dass sie in der Vergangenheit bereits viel Erfahrung mit dem Beurteilen anderer gesammelt hat. Dass die psychometrischen Gütekriterien von Ratings generell durch Übung erhöht werden können, ist längst vielfach untersucht (z.B. Bernardin, 1978; Borman, 1975; Latham, Wexley & Pursell, 1975).

Organisations- und tätigkeitsbezogene Faktoren:

g) Führungsspanne (Ratee)

Je größer die Führungsspanne ist, umso weniger Zeit verbringt eine Führungskraft mit der Führung jedes einzelnen Mitarbeiter. Diese können ihre Beurteilung also auf vergleichsweise wenige Grundlagen stellen, so dass eine geringere Selbst-Aufwärts-Übereinstimmung resultieren sollte.

Wenn aufgrund mangelnder Beobachtungsmöglichkeiten die Aufwärtsbeschreibungen nicht akkurat sind, resultiert letztlich auch eine Minderung der Abwärts-Aufwärts-Übereinstimmung für steigende Führungsspannen.

Tragen die Mitarbeiter ihrerseits Führungsverantwortung, dann sollte bei großer Führungsspanne dadurch auch die Selbst-Abwärts-Übereinstimmung beeinträchtigt sein. Da in diesem Fall immer eine räumliche Trennung vorliegt, können die abwärts beurteilenden Führungskräfte das Führungsverhalten ihrer unterstellten Führungskräfte seltener beobachten (und folglich schlechter bewerten).

h) Führungsspanne (Abwärtsrater)

Auf ähnliche Weise kann angenommen werden, dass eine steigende Führungsspanne der Abwärtsbeurteiler ihre Beurteilungsgenauigkeit mindert und folglich die Selbst-Abwärts- sowie die Abwärts-Aufwärts-Übereinstimmung negativ beeinflusst. Abwärtsbeurteiler mit *vielen* zu führenden Führungskräften sind nicht nur räumlich von diesen getrennt, sondern haben auch mit steigender Anzahl weniger Zeit (also auch Beobachtungsmöglichkeiten) für den Einzelnen. Moser, Zempel und Schultz-Amling (2000, S. 222) sprechen von einer geringeren „Kommunikationsdichte“ und finden, dass Abwärtsbeurteiler mit zunehmender Führungsspanne angeben, ihre Ratings systematisch stärker zu verzerren. Dies muss letztlich in geringerer Übereinstimmung mit Ratings aus anderen Beurteilungsperspektiven resultieren.

i) Hierarchie Ratee

Steinmann und Schreyögg (1993, S. 674) beschreiben den so genannten „Hierarchie-Effekt“ folgendermaßen: „Die Beurteilung fällt um so besser aus, je höher der hierarchische Rang des Beurteilten ist. Dies ist ein Effekt, der in fast allen Organisationen anzutreffen ist und wohl in dem Bestreben seinen Ursprung hat, Konsistenz zwischen Status und Beurteilung herzustellen.“ Demnach sollten die Niveau-Unterschiede zwischen Selbst- und Fremdbeurteilern mit steigender Hierarchieebene abnehmen („alle beurteilen gut“). Da die positiven Bewertungen in diesem Fall jedoch nicht unbedingt auf positiven Leistungen beruhen, gibt es keinen Anhaltspunkt für die Übereinstimmung der Profile der Selbst- und Fremdbeurteiler. Es ist von einer Abnahme der korrelativen Zusammenhänge auszugehen.

Allerdings deuten die Ergebnisse von Harris, Smith und Champagne (1995) darauf hin, dass dieser Effekt lediglich für Ratings gilt, die für einen administrativen (versus entwicklungsorientierten) Zweck eingesetzt werden. Für einen moderierenden Einfluss in die andere Richtung könnte sprechen, dass Führungskräfte in höheren Hierarchieebenen in der Regel älter sind, mehr Erfahrung in der Beurteilung haben und bereits länger in der Organisation tätig sind (vgl. entsprechende Hypothesen), woraufhin die Selbst-Fremd-Übereinstimmung steigen sollte.

j) Dauer der Organisationszugehörigkeit

Je länger ein Beurteiler (Selbst- oder Fremdrater) bereits in der gleichen Organisation tätig ist, umso besser sind sein Verständnis der Organisationsziele und seine Kenntnis gern gesehener Verhaltensweisen in der Organisation (Organisationskultur). Zudem steigt die Wahrscheinlichkeit, bereits Erfahrung in der Beurteilung nach den Kriterien der Organisation gesammelt zu haben. Diese Faktoren beeinflussen die Interpretation der Items und sollten zu einer höheren Übereinstimmung von Selbst- und Fremd- sowie von Fremd- und Fremd-Ratern führen. Auch wenn Heidemeier und Moser (2002) nur einen tendenziell moderierenden Effekt der Organisationszugehörigkeit für die Selbst-Abwärts-Übereinstimmung feststellen konnten, ist anzunehmen, dass sie sich im Führungsbereich stärker auswirkt, da hier Ziele und Aufgaben insgesamt vager sind und stärker über Organisationen hinweg streuen. Eine bessere Kenntnis der spezifischen Organisationskultur muss sich hier also besonders klärend auswirken, allgemeine Berufserfahrung ist nicht ausreichend.

k) Dauer der Abteilungszugehörigkeit

Auch eine Abteilung entwickelt eigene Normen, Umgangsformeln und Prioritäten. Analog zur Dauer der Organisationszugehörigkeit sollten die Beurteiler-Übereinstimmungen also auch mit der Dauer der Abteilungszugehörigkeit steigen.

l) Dauer der Zusammenarbeit zwischen Rater und Ratee

Mitarbeiter (mit oder ohne Führungsverantwortung) oder Vorgesetzte von Führungskräften gewinnen im Laufe der Zeit, die sie derselben Führungskraft berichten bzw. delegieren, immer mehr Beobachtungsmöglichkeiten des Führungsverhaltens der fokussierten Führungskraft. Zudem steigt mit der Dauer der Zusammenarbeit die Wahrscheinlichkeit, dass Feedback (nach oben oder unten) gegeben wird oder explizite Beurteilungsgespräche geführt werden. Dies alles führt dazu, dass die gegenseitigen Schwerpunkte bekannt sind und gegebenenfalls eine Reflexion des eigenen Führungsverhaltens angeregt wurde. Die Beurteilungen sollten folglich genauer ausfallen und besser übereinstimmen. Für die Interrater-Reliabilität von Aufwärtsbeurteilungen wurde die Dauer der Beobachtungsmöglichkeit bereits als Moderatorvariable bestätigt (Rothstein, 1990), für die Übereinstimmung zwischen verschiedenen Perspektiven ist der Nachweis noch zu erbringen.

Durchführungsbedingungen (Format des Befragungsinstruments):

m) Beschreibungsebene: ergebnisbezogene Itemformulierung

Je konkreter der Führungsinhalt in den Items formuliert ist, umso weniger Interpretationsspielraum bleibt den Beurteilern, so dass die Übereinstimmung zwischen den Beurteilungsperspektiven steigen sollte. Eine Möglichkeit zur konkreten Formulierung von Items liegt darin, bereits die Konsequenzen zu nennen, die das Führungsverhalten hervorrufen soll. Die Moderatorwirkung der Itemformulierung kann deshalb überprüft werden, weil der zentrale Fragebogen der vorliegenden Studie sowohl ergebnisbezogene als auch nicht ergebnisbezogene Items enthält. Erstere sollten eine genauere Itembeantwortung ermöglichen und dadurch eine höhere Übereinstimmung zwischen den Beurteilungsperspektiven hervorrufen.

Konkreter bedeutet „ergebnisbezogen", dass das im Item formulierte Ergebnis des Führungsverhaltens mindestens *vom Mitarbeiter* selbst *zum Zeitpunkt* der Ausführung des zu beurteilenden Führungsverhaltens – also nicht erst zukünftig – beobachtet werden kann. Das zu beurteilende Führungsverhalten selbst ist jeweils im einleitenden Verb des Items ausgedrückt.[129]

Durchführungsbedingungen (allgemeine Rahmenbedingungen):

n) Erfahrung als Beurteilter (Ratee)

Je häufiger eine Person in ihrer beruflichen Leistung beurteilt wird, umso häufiger wird eine Selbstreflexion über die eigene Leistung und ein Abgleich zwischen der Fremdbeurteilung und der eigenen Einschätzung angestoßen. Daraus sollte dann eine validere und weniger stark überhöhte Selbstbeurteilung resultieren; die Selbst-Fremd-Korrelationen sollten also zunehmen, während die Selbst-Fremd-Niveau-Unterschiede abnehmen sollten. Mabe und West (1982) fanden, dass Beurteilungen der eigenen Leistung an Validität gewinnen, je häufiger eine Person sich bereits selbst auf dieser Leistungsdimension beurteilt hat. Befunde, die eine Annäherung der Selbstbeschreibung an vorausgegangenes Feedback zum Führungsverhalten erbringen, zeigen, dass auch erhaltene Rückmeldungen/Beurteilungen von anderen die Genauigkeit von Selbstbeurteilungen (Führung) erhöhen können (vgl. Kapitel 3.6).

[129] Manche Ergebnisse können neben dem Mitarbeiter auch von der Führungskraft beobachtet werden. Der Fragebogen enthält jedoch keine solchen ergebnisbezogenen Itemformulierungen, deren Ergebnis ausschließlich von der Führungskraft beobachtet werden kann.

o) Erfahrung als Beurteiler (Ratee)

Je häufiger eine Führungskraft ihre unterstellten Mitarbeiter beurteilt, umso mehr Erfahrung und -übung in der Beurteilung hat sie („Beurteilungserfahrung"). Dass Übung die Genauigkeit der Beurteilung erhöhen kann, wurde oben bereits erwähnt, so dass die Selbst-Fremd-Übereinstimmung steigen sollte.

p) Beteiligungsquote

Es ist anzunehmen, dass diejenigen Führungskräfte in ihrer Selbstbeurteilung besser mit den Aufwärtsbeurteilungen übereinstimmen, deren Arbeitsgruppe eine höhere Beteiligungsquote erreicht. Wenn nämlich mehr Mitarbeiter ihre Führungskraft beurteilen, bedeutet dies, dass das Klima in der Arbeitsgruppe dies zulässt. Bei einer ablehnenden Haltung der Führungskraft gegenüber Führungsbeurteilungen sowie bei starker Angst vor „Retourkutschen" ist weder eine hohe Beteiligung noch eine „ehrliche" Beurteilung zu erwarten. Die Beteiligungsquote kann also als Maß für ein offenes Feedback-Klima betrachtet werden. Dass weniger verzerrte, also weniger fehlerbehaftete Ratings besser übereinstimmen, ist müßig zu erwähnen.

q) Beteiligung des Ratees

Analog zur Beteiligungsquote kann die Beteiligung derjenigen Führungskraft (Selbstbeurteiler), die in der Sandwichposition zwischen Aufwärts- und Abwärtsbeurteilern liegt, als Moderatorvariable für den Abwärts-Aufwärts-Zusammenhang betrachtet werden.[130]

r) Ideal Selbst-Aufwärts-Diskrepanz

Je weniger die Idealvorstellungen von Führung zwischen den verschiedenen Beurteilern übereinstimmen, umso weniger sollten diese auch in der Beurteilung des Ist-Verhaltens übereinstimmen. Vor dem Hintergrund verschiedener Erwartungen können die Beurteilungsinhalte nämlich auch auf verschiedene Weise verstanden und kodiert werden. Von den Personen, die unter anderem als Abwärtsbeurteiler fungierten, liegen kaum Ideal-Beschreibungen vor, so dass diese Variable nur für den Selbst-Aufwärts-Zusammenhang untersucht werden konnte.

s) Kenntnis der Führungsgrundsätze

Da Items vor ihrer Beantwortung immer erst dekodiert werden müssen und unterschiedliche Dekodierungen die Beurteilerübereinstimmung mindern können, ist dann von einem ähnlicheren Verständnis auszugehen, wenn die Beurteiler die Führungsgrundsätze kennen, auf deren Basis die Items entwickelt wurden. Geringere Übereinstimmungen sind also zu

[130] Für die Berechnung der *Selbst*-Fremd-Zusammenhänge ist die Beteiligung der Führungskraft ohnehin vorausgesetzt.

erwarten, wenn mindestens einer von zwei Beurteilern angibt, mit den Führungsgrundsätzen nicht vertraut zu sein. Dies gilt nicht nur für Selbst-Fremd-Vergleiche, sondern auch für die Übereinstimmung zwischen Aufwärts- und Abwärtsbeurteilungen. Da fast allen Führungskräften die Führungsgrundsätze bekannt sind, kann die Moderatoranalyse für die Selbst-Abwärts-Übereinstimmung nicht durchgeführt werden.

t) Führungsverantwortung der Aufwärtsrater

Da aufwärtsbeurteilende Mitarbeiter *mit* eigener Führungsverantwortung (vs. ohne Führungsverantwortung) bereits Erfahrungen mit der Führungsrolle haben, können sie ihre Aufwärtsbeurteilungen vor einem ähnlicheren Hintergrund wie Selbst- und Abwärtsbeurteiler abgeben, die allesamt Führungsverantwortung tragen. Dies sollte die Selbst-Aufwärts- sowie die Abwärts-Aufwärts-Übereinstimmung steigern. Zudem spricht für diese Moderatorannahme, dass die aufwärtsbeurteilenden Mitarbeiter *mit* eigener Führungsverantwortung ebenso zur Gruppe der Führungskräfte gehören wie die Ratee-Führungskräfte selbst. Sie können also in einer positiven Aufwärtsbeurteilung ihre eigene positive soziale Identität (Tajfel & Turner, 1986; Turner, 2000) zum Ausdruck bringen bzw. diese halten oder stärken. Hingegen sind nach der Theorie der sozialen Identität aufwärtsbeurteilende Mitarbeiter *ohne* Führungsverantwortung eher dazu motiviert, Distinktheit zu schaffen. Sie können ihre eigene soziale Identität also sogar durch eine negative Aufwärtsbeurteilung erhöhen. Ob die Führungsverantwortung der Aufwärtsbeurteiler den Abwärts-Aufwärts-Zusammenhang moderiert, kann nicht überprüft werden, da nur für ein einziges Abwärts-Aufwärts-Paar die Aufwärtsbeurteiler Führungsverantwortung tragen.

u) Aufwärtsbeurteiler als Stellvertreter

Wenn die Aufwärtsbeurteiler aus eigener Erfahrung die Rolle der zu beurteilenden Führungskraft kennen, da sie diese bei deren Abwesenheit vertreten, ist von einer Beobachtungsgrundlage auszugehen, die eine genauere Aufwärtsbeurteilung erlaubt. Als Folge sollten sowohl die Selbst-Aufwärts-Korrelationen ansteigen als auch die Selbst-Aufwärts-Niveau-Unterschiede abnehmen. Gleiches gilt für den Abwärts-Aufwärts-Zusammenhang.

Fragestellung 5: Zusammenhang zwischen verschiedenen Übereinstimmungsindizes

Einzelne Studien, die leistungsbezogene, berufliche Selbstbeurteilung validieren, erbringen das Ergebnis, dass zwischen der Selbst- und Abwärtsbeurteilung kaum Niveau-Unterschiede bestehen, aber dennoch kein deutlicher korrelativer Zusammenhang besteht

(z.B. Moser et al., 1994; Somers & Birnbaum, 1991).[131] Dies wäre jedoch zu erwarten, wenn die beiden Übereinstimmungsindizes substitutiv wären („Substitutions-Annahme"). Diese Annahme liegt zumindest implizit der Mehrzahl der Studien zugrunde, da sie lediglich *ein* Übereinstimmungsmaß verwenden. Auch die Befunde der Meta-Analyse von Heidemeier und Moser (2002) demonstrieren, dass nur einer der sechs Moderatorvariablen, die die Selbst-Abwärts-*Niveau-Unterschiede* beeinflussen, auch die Selbst-Abwärts-*Korrelationen* beeinflusst. Während die angeführten Studien berufliche Leistung im Allgemeinen behandeln, soll in der vorliegenden Arbeit untersucht werden, ob diese Substitutions-Annahme erstens auch für den Führungsbereich und zweitens auch für andere Perspektiven-Vergleiche (Selbst-Aufwärts, Abwärts-Aufwärts, Selbst-Abwärts) abgelehnt werden kann. Dann müssten folgende Hypothesen *zurückgewiesen* werden.

16) **Je geringer die Selbst-Abwärts-Niveau-Unterschiede sind, umso höher sind die Selbst-Abwärts-Korrelationen.**
17) **Je geringer die Selbst-Aufwärts-Niveau-Unterschiede sind, umso höher sind die Selbst-Aufwärts-Korrelationen.**
18) **Je geringer die Abwärts-Aufwärts-Niveau-Unterschiede sind, umso höher sind die Abwärts-Aufwärts-Korrelationen.**

Fragestellung 6: Above-Average-Effekt

Bislang konnte der Above-Average-Effekt für verschiedenste Beurteilungsgegenstände bestätigt werden (vgl. Kapitel 3.2.3). In der Regel wurden dabei relationale Skalen verwendet, das heißt Skalen, auf denen sich der Selbstbeschreiber zum durchschnittlichen relevanten Anderen einordnen muss. So gingen auch diejenigen Studien vor, die den Above-Average-Effekt von Managern untersuchten (Larwood & Whittaker, 1977; Meyer, 1980). Eine relationale Einschätzung ist jedoch immer auch beeinflusst von sozialen Vergleichsprozessen. Weniger stark sollten solche Prozesse interferieren, wenn die Selbstbeurteilung nicht in einem direkten Vergleich mit anderen erfolgt, da dann dem Beurteiler die Vergleichsintention nicht bekannt ist. Einen solchen Rahmen bietet das Design der vorliegenden Untersuchung, das offensichtlich auf einen Vergleich von *verschiedenen* Beurteilungsperspektiven ausgerichtet war, aus denen die*selbe* Fokusperson beurteilt wurde (nicht umgekehrt).

[131] Es liegen keine Studien für den Vergleich der Übereinstimmungsindizes von Selbst-Aufwärts- oder Abwärts-Aufwärts-Perspektiven vor, da zur Validierung der Selbstbeurteilung (was häufig Ziel dieser Studien ist) in der Regel die traditionelle und akzeptierte Abwärtsbeurteilung durch den Vorgesetzten herangezogen wird.

Da zu diesem Zweck viele Führungskräfte jedoch nicht nur ihr eigenes Führungsverhalten, sondern auch das Führungsverhalten anderer Führungskräfte beschrieben, konnte die Selbstbewertung in Relation zur Bewertung anderer gesetzt werden. Es ist davon auszugehen, dass bisher erbrachte Befunde zum Above-Average-Effekt nicht allein auf soziale Vergleichsprozesse oder andere Störgrößen zurückzuführen sind, so dass angenommen wird:

19) Führungskräfte beurteilen ihr eigenes Führungsverhalten positiver als das anderer Führungskräfte.

Explorativ wird untersucht, inwiefern sich die Selbstüberschätzung auf den einzelnen Führungsdimensionen unterscheidet.

Erst wenn die *Mehrzahl* der Führungskräfte ihr eigenes Führungsverhalten besser beurteilt als das anderer Führungskräfte, kann von einer Selbstüberschätzung der Führungskräfte gesprochen werden. Erst dann nämlich entbehrt das Ergebnis – wie bei allen Above-Average-Effekten – der realen Grundlage, weil nicht die Mehrzahl der Führungskräfte besser sein kann als die anderen.

20) Die Mehrzahl der Führungskräfte beurteilt ihr eigenes Führungsverhalten positiver als das anderer Führungskräfte.

Als „andere Führungskräfte“ werden sowohl übergeordnete (Vorgesetzte) als auch nachgeordnete Führungskräfte (Mitarbeiter mit Führungsverantwortung) verstanden. Beurteilungen von Führungskräften auf gleicher Ebene (Peers) wurden nicht erhoben (vgl. z.B. Kapitel 5.5.1.4).

Unter den Annahmen, dass Führungskräfte erstens nach oben streben und dass zweitens die Dazugehörigkeit zur bereits erreichten Hierarchieebene durch Distinktheit nach unten gerechtfertigt werden kann, ist von einer besonders starken Selbstüberschätzung nach unten auszugehen:

21) Die Selbstüberschätzung nach oben fällt geringer aus als die Selbstüberschätzung nach unten.

Da das Führungsverhalten der zu beurteilenden Vorgesetzten vom Führungsverhalten der zu beurteilenden Mitarbeiter unabhängig ist und da für eine Überschätzung nach oben andere Motive aktiviert sind als für eine Überschätzung nach unten, wird angenommen,

dass die Überschätzung nach oben und unten voneinander unabhängig ist. Wäre dies nicht der Fall, könnte von stark verzerrten Ratings ausgegangen werden, die alle Möglichkeiten zur positiven Selbstdarstellung gleichförmig nutzen.

22) Die Selbstüberschätzung nach oben ist unabhängig von der Selbstüberschätzung nach unten.

Genau genommen wird also nicht überprüft, ob sich die befragten Führungskräfte positiver als die durchschnittliche Führungskraft beurteilen (above *average*), sondern ob sie sich durchschnittlich und mehrheitlich positiver beurteilen als bestimmte andere Führungskräfte. Da die Konkretisierung des Vergleichsgegenstands zu einer Reduzierung des Above-Average-Effekts führt (Alicke et al., 1995), können die Ergebnisse jedoch als konservative Schätzung dieses Effekts interpretiert werden.

Fragestellung 7: Kriteriumsvalidität der Führungsbeurteilungen

Personalführung in Unternehmen verfolgt das generelle Ziel, Prozesse so zu unterstützen, dass sie letztlich im Unternehmenserfolg münden (vgl. Kapitel 2.2). Deshalb ist es unabdingbar, Führungsbeurteilungen an Erfolgskriterien zu validieren. Meta-Analysen zeigen, dass sowohl Abwärts- (Bommer et al., 1995) als auch Aufwärts- und Peerbeurteilungen (Conway et al., 2001) mit objektiven Kriterien gering/mittel korrelieren (vgl. Kapitel 3). Die Beurteilungen, die in den Meta-Analysen berücksichtigt wurden, beziehen sich jedoch erstens auf Leistungsverhalten insgesamt. Führungsverhalten stellt dort also lediglich *eine* inhaltliche Komponente aus einem breiten Spektrum dar, die in den Meta-Analysen nicht isoliert berichtet wird. Außerdem ist zweitens nur für die dort berücksichtigten Aufwärtsbeurteilungen sichergestellt, dass sie sich auf Führungskräfte beziehen, also auf die für die vorliegende Studie relevante Ratee-Gruppe. Peer- und Abwärtsbeurteilungen können hingegen auch das berufliche Leistungsverhalten von Mitarbeitern *ohne* Führungsverantwortung betreffen.

Greift man auf Einzelstudien zurück, die sich auf die Beurteilung von *Führungsverhalten* beschränken, stellt man zwar einen Zusammenhang zwischen der Führungsbeurteilung und verschiedenen Außenkriterien fest (z.B. Hater & Bass, 1988; vgl. Tabelle 9). Man stellt aber gleichzeitig fest, dass die Studien häufig an Stichproben durchgeführt wurden (z.B. Marineoffiziere, Bass & Yammarino, 1991), die in Deutschland wenig Bedeutung haben. Auch das Werteverständnis, das sich hinter verschiedenen Führungsdimensionen verbirgt, ist nicht zweifellos universell. Die beiden Studien, die in der deutschen Automobilbranche verankert sind, reihen sich für einzelne Führungsfaktoren und Erfolgskriterien in die generelle Befundlage ein. Allerdings zogen sie zur Führungsbeurteilung den inzwi-

schen veralteten FVVB heran (Fittkau-Garthe & Fittkau, 1971; lediglich Aufwärtsbeurteilungen), befragten im Unterschied zur vorliegenden Studie die Ebene der Produktionsarbeiter und beschränkten sich auf ein beziehungsweise zwei Erfolgskriterien, die Ideeneinreichungsquote (Nickel & Krems, 1998) bzw. Fehlzeiten und Fluktuation (Schmidt, 1996). Ein solch enger Kriteriumsbereich erschwert es jedoch, bestimmte Führungsfaktoren als bedeutsame Prädiktoren extrahieren zu können. Außerdem finden Nickel und Krems (1998) lediglich in einem Extremgruppenvergleich unterschiedliche Ideeneinreichungsquoten.

Im Einklang mit den Ergebnissen aus anderen Branchen und für andere Prädiktor- und Erfolgskriterien wird für die vorliegende Studie erwartet, dass die Fremdbeurteilungen (Aufwärts- und Abwärtsbeurteilungen) von Führungsverhalten in der Automobilbranche mit den Erfolgskriterien positiv korrelieren.

23) Die Fremdbeurteilungen von Führung korrelieren positiv mit Erfolgskennwerten.

Als Erfolgskennwerte werden für diese wie für alle nachfolgend beschriebenen Fragestellungen folgende Indizes verstanden: Arbeitszufriedenheit (zwei Faktoren), Wohlbefinden, Leistung, Zufriedenheit mit Beurteilung, Zufriedenheit mit Engagement der Führungskraft, Extra Effort, Effektivität der Führung, Zufriedenheit mit der Führungskraft allgemein.

Für die Mehrzahl der Kriterien wurden Skalen konstruiert, die drei letztgenannten Kriterien wurden mit dem MLQ erfasst. Dadurch ergibt sich die Möglichkeit, MLQ-externe Führungsdimensionen zur Vorhersage von MLQ-internen Erfolgsvariablen heranzuziehen. Solche Studien stehen bislang aus, so dass die Ergebnisse der vorliegenden Studie nicht nur zur Konstruktvalidierung des konstruierten FzFü-Fragebogens, sondern auch zur Konstruktvalidierung der MLQ-Kriterien beitragen können.

Da die Erfolgskriterien das Erleben und Verhalten der Mitarbeiter (Aufwärtsbeurteiler) betreffen, wird angenommen, dass auch deren Führungsbeurteilungen eine bessere Erfolgsprognose abgeben:

24) Die Aufwärtsbeurteilungen von Führung korrelieren stärker mit den Erfolgskennwerten als die Abwärtsbeurteilungen.

Aufwärtsbeurteilungen von Führung wurden durch verschiedene Instrumente erfasst (Fragebogen zur Führungsbeschreibung [FzFü], Multifactor Leadership Questionnaire [MLQ], vgl. Kapitel 5.2).

Explorativ wird untersucht, durch welche Führungsdimensionen welche Erfolgsgrößen am besten prognostiziert werden können und ob mehrere Führungsdimensionen gemeinsam mehr Erfolg erklären können als einzelne Dimensionen. Eine solche inkrementelle Validität einzelner Führungsfacetten wäre ein Indiz für multidimensionale Führungsbeurteilungen. Seit langem werden auch in der Führungsliteratur *verschiedene* Führungsstile und -aufgaben beschrieben. Nicht umsonst gelten Führungstätigkeiten als sehr heterogen und komplex. Die spezifische Wirkung der verschiedenen Aufgaben muss aber erst noch erforscht werden.

Generell sollte eine Erhöhung des Zusammenhangs zwischen Führungsbeurteilung und Erfolg durch die Aggregation von Fremdbeurteilern erreicht werden (vgl. Mount & Scullen, 2001). Wie bereits ausgeführt, ist dies auf die Steigerung der Reliabilität zurückzuführen.

25) Aggregierte Aufwärtsbeurteilungen korrelieren mit den Erfolgsvariablen stärker als nicht aggregierte (individuelle) Aufwärtsbeurteilungen.

Es gibt einzelne Hinweise darauf, dass selbst die Aggregation von Ratings, die aus verschiedenen Perspektiven abgegeben wurden, eine höhere Kriteriumsvalidität ermöglichen (Wohlers & London, 1989). Da Abwärts- und Aufwärtsbeurteiler allerdings sehr unterschiedliche Ausschnitte des Führungsverhaltens beobachten können und da bereits in den Hypothesen 6) und 10) differierende Beurteilungen angenommen wurden, wird für die vorliegende Stichprobe kein Validitätsgewinn durch die Aggregation aller Fremdratings erwartet.

26) Aggregierte Fremdbeurteilungen (Aufwärts- und Abwärtsbeurteilungen) korrelieren mit den Erfolgsvariablen nicht stärker als aggregierte Aufwärtsbeurteilungen allein.

Für Selbstbeurteilungen, denen für allgemeine Leistungsdimensionen und für Management-Dimensionen im besonderen (Mabe & West, 1982) eine deutlich geringere Kriteriumsvalidität als Fremdbeurteilungen zugeschrieben wird (vgl. Kapitel 3.1.1), ist anzunehmen:

27) Selbstbeurteilungen korrelieren in geringerem Ausmaß mit Erfolg als Fremdbeurteilungen.

Dieser Unterschied wird sowohl für den Vergleich mit aggregierten als auch für den Vergleich mit nicht aggregierten Fremdbeurteilungen angenommen. Nur wenn er sich auch hinsichtlich Letzterer ergibt, kann die unterschiedliche Validität von Selbst- und Fremdbeurteilungen auf die Beurteilungsperspektive zurückgeführt werden und nicht auf die unterschiedlich gute Reliabilität, die durch Ratingaggregation erzeugt wird.

Fragestellung 8: Inkrementelle Validität der Führungsbeurteilungen

Bei dieser Fragestellung wird erstens die inkrementelle Validität verschiedener Führungsdimensionen *innerhalb* einer Beurteilungsperspektive und zweitens die inkrementelle Validität über die verschiedenen Beurteilungsperspektiven hinweg betrachtet.

Zunächst ist anzunehmen, dass sich verschiedene Führungsdimensionen in ihrer Wirkung auf die Erfolgskennwerte ergänzen. Insofern wird angenommen, dass die Ergebnisse bezüglich Fragestellung 7 durch die gleichzeitige Betrachtung mehrerer Führungsdimensionen verbessert werden können:

28) Verschiedene Führungsdimensionen klären gemeinsam mehr Erfolg auf als einzelne Führungsdimensionen.

Neben der Multidimensionalität ist die vorliegende Studie durch ihren Multiperspektivismus gekennzeichnet. Dabei verspricht sich jeder Anwender von multiperspektivischen Führungsbeurteilungen insgeheim einen Vorteil gegenüber einer uniperspektivischen Erhebung. Der Nutzen multiperspektivischer *Führungs*beurteilungen muss aber genauso belegt werden wie der von multiperspektivischen Leistungsbeurteilungen im Allgemeinen (dazu Meta-Analyse von Conway et al., 2001). Ein Nutzen mehrerer Perspektiven ergibt sich immer dann, wenn mehrere Perspektiven gemeinsam mehr Erfolgsvarianz aufklären als einzelne Perspektiven. Entsprechend der erwähnten Meta-Analyse für allgemeine Leistungsbeurteilungen wird auch für die vorliegende Stichprobe angenommen, dass die einzelnen Perspektiven inkrementelle Validität über die jeweils anderen hinaus erbringen können. Da Beobachter aus verschiedenen Perspektiven unterschiedliche Führungsverhaltensweisen unterschiedlich gut beobachten können, ist durchaus von spezifisch kriteriumsrelevanten, sich zu einer validen Beurteilung ergänzenden Beiträgen auszugehen. Der potentielle Beitrag, den die Selbstbeurteilung zu den Beiträgen der Abwärts- und Aufwärtsbeurteilungen hinzufügen kann, wurde bislang auch außerhalb des Führungskontextes nicht untersucht.

29) Die Selbst-, Aufwärts- und Abwärtsbeurteilung von Führung erklären gemeinsam mehr Erfolgsvarianz als die Beurteilung aus einzelnen Perspektiven.

Gemäß der Theorie zur transformationalen Führung (Bass, 1985) erbringt die transformationale Führung, die über die transaktionale Führung hinaus gezeigt wird, zusätzlichen Erfolg. Diese so genannte Augmentationshypothese (vgl. Kapitel 2.5.3.1) wurde in der Mehrzahl der bisherigen Studien für Erfolgsvariablen untersucht, die ebenso wie das Führungsverhalten mit dem MLQ gemessen wurden. In der vorliegenden Studie soll sie auch für MLQ-externe Erfolgskriterien geprüft werden und damit erste, dahingehend vorliegende Befunde stützen (z.B. Tartler, Goihl et al., 2003).

30) Die transformationale Führung klärt gemeinsam mit der transaktionalen Führung mehr Erfolgsvarianz auf als die transaktionale Führung allein.

Da klassische Führungsaufgaben, die der unternehmensspezifische Fragebogen (FzFü) erfasst, eher ins transaktionale Führungsspektrum einzuordnen sind, ist die transformationale Führung im Vergleich zu diesen höher im „Full Range of Leadership“ anzuordnen (vgl. Kapitel 2.6). Damit kann angenommen werden:

31) Die transformationale Führung klärt zusätzlich zu den klassischen Führungsdimensionen Erfolgsvarianz auf.

Da die transformationale Führung in der vorliegenden Studie aus der Aufwärtsperspektive erfasst wurde, die klassischen Führungsdimensionen, wie bereits erwähnt, zusätzlich aus der Selbst- und Abwärtsperspektive, kann eine Bestätigung von Hypothese 31) die grundlegende Annahme von Hypothese 29) – die inkrementelle Validität verschiedener Beurteilungsperspektiven – unterstützen.

Fragestellung 9: Moderatorvariablen für den Zusammenhang zwischen Führungsverhalten und Führungserfolg

Es ist anzunehmen, dass der Zusammenhang zwischen der Führungsbeurteilung und den Erfolgskriterien von der Wichtigkeitseinschätzung moderiert wird. Wird den beurteilten Dimensionen mehr Wichtigkeit (Erfolgsrelevanz) zugesprochen, dann sollte der Zusammenhang zwischen Führungsbeurteilung und Erfolg steigen. Dieser Annahme liegt eine kontingenztheoretische Sicht von Führungswirkung zugrunde. Danach sollte Führung umso erfolgreicher sein, je wichtiger (erfolgsrelevanter) die Mitarbeiter diejenigen Führungsverhaltensweisen einstufen, die ihre Führungskraft ausführt. Dann ist eine Passung zwischen der Führung der Führungskraft und den Bedürfnissen der Mitarbeiter nämlich am besten gegeben:

32) Je wichtiger die Aufwärtsbeurteiler die Führungsdimensionen bewerten, umso stärker ist der Zusammenhang zwischen diesen Führungsdimensionen und Erfolg.

Zur Überprüfung dieser Hypothese werden als Operationalisierung von Führung die Selbst-, Abwärts- *und* Aufwärtsbeurteilungen (jeweils getrennt voneinander) herangezogen (betrifft auch Hypothese 33).

Da die transformationale Führung gemäß der Annahmen von Bass (1985) darin resultiert, dass die Geführten nicht nur stärkeres Vertrauen in ihre Kompetenzen und damit höhere Erfolgserwartungen haben, sondern zusätzlich den angestrebten Ergebnisse auch noch erhöhte Bedeutsamkeit zusprechen, ist die Motivation der Zielerreichung insgesamt positiv beeinflusst. Dadurch kann die „gewöhnliche" Führungsleistung unterstützt werden. Es wird also folgender moderierender Effekt angenommen:

33) Je transformationaler die Aufwärtsbeurteiler ihre Führungskraft beschreiben, umso stärker ist der Zusammenhang zwischen Führung und Erfolg.

Auch die *Übereinstimmung zwischen verschiedenen Beurteilungsperspektiven* kann als Moderator für den Zusammenhang zwischen Führung und Erfolg konzeptualisiert werden. Diese Hypothese erfährt Unterstützung durch empirische Befunde, die besonders gute Leistungsträger durch besonders hohe Beurteilerübereinstimmung und besonders positives Führungsverhalten kennzeichnen (Atwater & Yammarino, 1992). Die Ergebnisse dieser Hypothese werden allerdings erst im Anschluss an die Darstellung der Validität der Perspektiven-Übereinstimmung (Fragestellung 10) berichtet. Es wird dann die moderierende Wirkung der Selbst-Aufwärts-, der Selbst-Abwärts- und der Abwärts-Aufwärts-Kongruenz überprüft (vgl. Hypothese 49).

Fragestellung 10: Kriteriumsvalidität der Selbst-Fremd- und Fremd-Fremd-Übereinstimmung

Lange Zeit wurde eine geringe interperspektivische Übereinstimmung lediglich unter dem Aspekt der Fehlerbehaftetheit von Ratings betrachtet. Neuere Arbeiten sprechen der Übereinstimmung aber eigene Validität zu (z.B. Atwater et al., 1998). So existieren inzwischen empirische Hinweise darauf, dass die Selbst-Fremd-Übereinstimmung in der Führungsbeurteilung mit Erfolgsvariablen der Führung (z.B. Zufriedenheit, Effektivitäts-

einschätzung) zusammenhängt (vgl. Kapitel 3.5). Doch nicht nur, weil deren Anzahl bislang sehr gering ist, sondern auch, weil die Untersuchungsstichproben ungewöhnlich sind (z.B. Marineoffiziere, vgl. Tabelle 13 in Kapitel 3.5.1), fordern beispielsweise Church (1997a) oder Greguras und Robie (1998) die Durchführung weiterer Studien, die die Selbst-Fremd-Kongruenz mit relevanten Einstellungs- und Verhaltenskonstrukten in Beziehung setzen. Alle existierenden theoretischen Annahmen über das Zusammenwirken von Selbst-Fremd-Übereinstimmung und Erfolg betrachten die Genauigkeit der Selbsteinschätzung als „kritische Variable“ (z.B. Ashford, 1989). Die dahinterstehende Logik besagt, dass bei einer genauen Selbsteinschätzung nicht nur Ziele adäquat gesetzt, sondern auch Ressourcen adäquat verwendet werden können. Eine Genauigkeit der Selbsteinschätzung ist oftmals über das Ausmaß der Abweichung von Fremdeinschätzungen operationalisiert worden.

Argumentiert man interaktionsorientiert anstatt allein auf die Selbstbewertung der Führungskraft abzuzielen, kann man eine geringe Selbst-Fremd-Abweichung als Indiz dafür bewerten, dass prinzipiell ausreichend Austausch über Führung stattfindet, so dass Selbst- und Fremdbild kontinuierlich adjustiert werden können. Aufgrund einer solchen Kommunikation sollten im Problemfall zügig Korrekturen vorgenommen werden können, so dass ein erfolgreiches Resultat wahrscheinlicher wird. Bei Einigkeit über Ziele und Vorgehensweisen ist außerdem davon auszugehen, dass weniger Reibungsverluste auftreten als bei Uneinigkeit.

Die beiden hier getrennt angerissenen Argumentationslinien gehen im Übrigen insofern miteinander einher, als kontinuierliche Kommunikationsprozesse auch die Selbsteinschätzung verbessern können. Möglicherweise handelt es sich also um zwei Prozesse, die aufeinander folgen. Beide Prozesse resultieren nicht nur in Hypothese 34), sondern auch in Hypothese 35).

34) Mit steigenden Niveau-Unterschieden zwischen der Selbst- und Fremdbeurteilung sinkt der Erfolg.

Als Erfolgskennwerte werden auch für die Beantwortung dieser Fragestellung alle erhobenen Erfolgsindizes herangezogen (siehe oben). Da es ausschließlich um Erfolgsmaße geht, die allein die Mitarbeiter der beurteilten Führungskräfte betreffen, ist davon auszugehen, dass die Übereinstimmung zwischen Selbst- und Aufwärtsbeurteilung stärker mit den Erfolgsmaßen zusammenhängt als die Übereinstimmung zwischen Selbst- und Abwärtsbeurteilung. Es wäre höchst unplausibel, wenn die Selbst-Abwärts-Abweichungen die Arbeitszufriedenheit, die Leistung und das Wohlbefinden der Aufwärtsbeurteiler besser vorhersagen könnten als die Selbst-Aufwärts-Abweichungen.

35) Hypothese 34) gilt stärker für die Diskrepanz zwischen der Selbst- und Aufwärts-Beurteilung als für die Diskrepanz zwischen der Selbst- und Abwärts-Beurteilung.

Da – wie erwähnt – in aller Regel die akkurate Selbsteinschätzung als Dreh- und Angelpunkt betrachtet wird, blieb die Abwärts-Aufwärts-Kongruenz als Erfolgsprädiktor bislang außen vor. Es kann allerdings durchaus angenommen werden, dass sehr inkonsistente Fremdeinschätzungen eine inkonsistente, schlecht greifbare Führung widerspiegeln. Ohne Klarheit über zentrale Führungsfragen sind Prozessverluste aber nicht auszuschließen (vgl. Kapitel 3). Daraus wird folgende Hypothese abgeleitet, wobei unter Erfolg wiederum alle erfassten Erfolgsvariablen zu verstehen sind.

36) Mit steigenden Niveau-Unterschieden zwischen der Abwärts- und Aufwärtsbeurteilung sinkt der Erfolg.

Explorativ wird untersucht, welche Kriterien sich durch welche Perspektiven-Diskrepanzen am besten prognostizieren lassen.

Wenn man der Übereinstimmung zwischen Selbst- und Fremdwahrnehmung mehr Gewicht zuschreibt als der inkonsistenten Fremdwahrnehmung, geht daraus Hypothese 37) hervor. Dies ist deshalb plausibel anzunehmen, weil Führungskräfte diejenigen Personen sind, die am unmittelbarsten Einfluss auf das Führen hin zu erfolgreicher Leistung haben. Eine Führungskraft gewinnt durch den Abgleich ihrer eigenen Wahrnehmung mit Fremdwahrnehmungen deutlich mehr Anhaltspunkte für eine regelmäßige Adjustierung der eigenen Führung als durch das Vorliegen inkonsistenter Fremdwahrnehmungen. Diese müssen der fokussierten Führungskraft gar nicht bekannt sein.

37) Die Selbst-Fremd-Übereinstimmung kann mehr Erfolgsvarianz aufklären als die Abwärts-Aufwärts-Übereinstimmung.

Aufgrund der in Kapitel 3.5 dargestellten Befunde ist davon auszugehen, dass auch die *Richtung* der Selbst-Fremd-Abweichung einen Beitrag zur Erfolgsaufklärung liefert. Es wird dementsprechend nicht nur angenommen, dass die Führungskräfte mit der stärksten Abweichung am wenigsten erfolgreich sind, sondern dass es unter den „Abweichlern“ die Überschätzer und nicht die Unterschätzer sind, die weniger Erfolg mit sich bringen. Den Unterschätzern wird vermehrt Motivation zur Verbesserung zugesprochen (vgl. Kapitel 3.6). Es wird also Hypothese 38) formuliert:

38) Die Mitarbeiter von Führungskräften, die sich im Vergleich zu Fremdbeurteilern überschätzen, sind am wenigsten erfolgreich, während der Erfolg bei Übereinstimmung am höchsten ausfällt. Die Mitarbeiter der Unterschätzer nehmen eine Zwischenstellung ein.

Unter der Annahme, dass sich die Kongruenzindizes[132] als wertvolle Kovariaten von Erfolg erweisen, muss weiter überprüft werden, ob wirklich die Kongruenz – und nicht die eingehenden einfachen Beurteilungsperspektiven – die kritische prognostische Leistung erbringt. Brutus, Fleenor und Tisak (1999) können für die Selbst-Peer-Diskrepanz in der Führungsbeurteilung keine prädiktive Validität hinsichtlich der Führungseffektivität feststellen. Sie plädieren deshalb in ihrem Fazit dafür, dies auch für andere Beurteilungsperspektiven zu überprüfen. So wird in der vorliegenden Studie die Selbst-Aufwärts-Kongruenz gegen die einfachen Selbst- und Aufwärtsbeurteilungen getestet. Bei Ablehnung von Hypothese 39) müssten die gängigen Interpretationen des Zusammenhangs zwischen Diskrepanz und Erfolg revidiert werden.

39) Es ist die Selbst-Aufwärts-Diskrepanz, die die kritische prognostische Leistung für Erfolg erbringt (vs. einfache Selbst- und Aufwärtsbeurteilungen).

In diesem Zusammenhang stellt sich weiter die Frage, ob die Vorhersagekraft von Kongruenzmodellen (= Diskrepanzmodellen) tatsächlich die der einfachen Modelle erreicht oder übersteigt. Andernfalls dürften sie allein aus Gründen der Sparsamkeit nicht gehalten werden. Auch wenn sie „nur“ gleich gut sind, wären aus methodischen Gründen die einfachen Modelle vorzuziehen. Allerdings sprechen praktische Gründe in der organisationspsychologischen Anwendung dafür, dann (auch) die Kongruenz/Diskrepanz zu betrachten (z.B. Akzeptanz). Entsprechend wird Hypothese 40) formuliert.

40) Die Selbst-Aufwärts-Diskrepanz eignet sich zur Erfolgsprognose mindestens ebenso gut wie die einfachen Beurteilungen aus der Selbst- und Aufwärtsperspektive.

Es ist nicht davon auszugehen, dass sich nur einfache Perspektiven gegenseitig in ihrer Validität ergänzen können (vgl. Hypothese 29). Da bei multiperspektivischen Beurteilungssystemen, in die mehr als zwei Perspektiven einbezogen sind, mehrere interperspektivische Diskrepanzen berechenbar sind, ist anzunehmen, dass diese gemeinsam eine höhere Erfolgsaufklärung leisten als einzelne Diskrepanzen (Hypothese 41).

[132] Entspricht Diskrepanzindizes, gemessen über Differenzmaße

41) Mehrere Diskrepanzen können mehr Erfolgsvarianz erklären als einzelne Diskrepanzen.

Durch die Analysen im Zusammenhang von Hypothese 41) wird für jedes Erfolgskriterium ein „Diskrepanzmodell“ entwickelt, das diejenige(n) Diskrepanz(en) umfasst, die im Vergleich zu anderen die höchste Varianzaufklärung erreichen. Dies wird im Folgenden als das „beste Diskrepanzmodell“ bezeichnet. Nur wenn dadurch mehr Erfolgsvarianz erklärt werden kann als durch das beste einfache Modell, ist die Verwendung von Diskrepanzen zur Erfolgsprognose gerechtfertigt (Hypothese 42). Das beste einfache Modell enthält als Prädiktoren diejenigen einfachen Führungsbeurteilungen aus den verschiedenen Perspektiven, die im Vergleich zu anderen die höchste Varianzaufklärung erzielen.

42) Das beste Diskrepanzmodell erklärt mehr Erfolgsvarianz als das beste einfache Modell.

Auch wenn nicht geklärt ist, was genau die interperspektivischen Diskrepanzen messen und worauf sich die Diskrepanz in den Wahrnehmungen von Führung auswirkt, kann davon ausgegangen werden, dass Diskrepanzen andere Informationen enthalten als einfache Führungsbeurteilungen. Vor diesem Hintergrund ist anzunehmen, dass interperspektivische Diskrepanzen einen inkrementellen Beitrag über die einfachen Führungsbeurteilungen hinaus leisten können (Hypothese 43).

43) Die Diskrepanz zwischen zwei Perspektiven leistet (über die einzelnen Beurteilungsperspektiven hinaus) einen inkrementellen Beitrag zur Erfolgsaufklärung.

Beispielhaft könnte angenommen werden, dass die einfachen Beurteilungen durch die entsprechende Führungsqualität und die Diskrepanzen durch Kommunikationshindernisse Einfluss auf die Erfolgsvariablen nehmen. Aufgrund solcher oder ähnlicher Annahmen, die selbstverständlich erst noch überprüft werden müssen, ist ein inkrementeller Beitrag der Diskrpanzen über die einfachen Perspektiven hinaus freilich plausibel. Das würde bedeuten, dass die interperspektivischen Diskrepanzen eine – zumindest partielle – Mediatorwirkung auf den Zusammenhang zwischen der Führungsbeurteilung und den Erfolgskriterien besitzt.

Es wäre unplausibel anzunehmen, dass Erfolg auch dann entsteht, wenn sich verschiedene Beurteiler über eine *negative* Führung einig sind. Deshalb darf die Beurteilungsübereinstimmung nicht losgelöst von der Führung an sich betrachtet werden. Insofern ist

Hypothese 38) folgendermaßen zu spezifizieren (Fleenor et al., 1996; vgl. auch Kapitel 3.5.1). Auch Hypothese 43) kann unter diesem Blickwinkel betrachtet werden.

44) Der Erfolg fällt für Übereinstimmer mit hoher Führungsqualität höher aus als für Unter- und Überschätzer mit hoher Führungsqualität.

45) Der Erfolg fällt für Übereinstimmer mit geringer Führungsqualität niedriger aus als für Unter- und Überschätzer mit hoher Führungsqualität.

Die Kategorisierung der Selbstbeurteiler in Unterschätzer, Übereinstimmer und Überschätzer wurde sowohl bezüglich der Selbst-Aufwärts- als auch bezüglich der Selbst-Abwärts-Abweichung vorgenommen.

Die Führungsqualität ist erstens durch die Beurteilung der transformationalen Führung operationalisiert sowie zweitens durch die Fremdbeurteilung (FzFü) derjenigen Perspektive, die jeweils nicht in die Kategorisierung als Überschätzer, Übereinstimmer oder Unterschätzer eingeht. Geht es also um die Selbst-*Aufwärts*-Übereinstimmung, wird die *Abwärts*beurteilung als Gradmesser der Führungsqualität herangezogen, und umgekehrt.

Für verschiedene wissenschaftliche und praxisorientierte Fragestellungen ist eine umgekehrte Analyserichtung häufig angezeigt. Es stellt sich dann also die Frage, wodurch erfolgreiche Mitarbeiter charakterisiert sind. Im Zentrum der vorliegenden Arbeit steht die Frage, ob sie durch eine höhere interperspektivische Beurteilerübereinstimmung charakterisiert sind. Church (1997a) berichtet entsprechende Ergebnisse bereits für die Selbst-Aufwärts-Übereinstimmung. Sie ist für starke Leistungsträger höher als für schwache Leistungsträger. Auf einer abstrakteren Analyseebene sprechen auch die Ergebnisse von Zimolong (2001) dafür, die in unterschiedlich erfolgreichen *Organisationen* unterschiedlich starke Selbst-Fremd-Übereinstimmungen ausmachen konnten.[133] Es ist im Einklang mit den bis dato aufgestellten Hypothesen anzunehmen:

46) Berichten Mitarbeiter überdurchschnittlichen (vs. unterdurchschnittlichen) Erfolg, sind die *Diskrepanzen* zwischen den Beurteilungsperspektiven geringer.

47) Berichten Mitarbeiter überdurchschnittlichen (vs. unterdurchschnittlichen) Erfolg, sind die *Korrelationen* zwischen den Beurteilungsperspektiven höher.

[133] Beide zitierten Studien wurden in Kapitel 2.5.4.1 bereits ausführlicher dokumentiert.

Sowohl Hypothese 46) als auch Hypothese 47) werden hinsichtlich der Selbst-Aufwärts-, der Selbst-Abwärts- und der Abwärts-Aufwärts-Übereinstimmung überprüft (vgl. Tabelle 15).

Tabelle 15. Untersuchte Variablen zur Charakterisierung erfolgreicher Mitarbeiter

Untersuchte Variablen	
Interperspektivische Diskrepanzen	▪ Selbst-Aufwärts-Diskrepanz ▪ Selbst-Abwärts-Diskrepanz ▪ Abwärts-Aufwärts-Diskrepanz
Interperspektivische Korrelationen	▪ Selbst-Aufwärts-Korrelation ▪ Selbst-Abwärts- Korrelation ▪ Abwärts-Aufwärts- Korrelation
Personale/organisationsbezogene Variablen	▪ Alter des Selbstbeurteilers ▪ Alter der Fremdrater ▪ Abteilungsgröße ▪ Beteiligungsquote ▪ Dauer der Zugehörigkeit zur Organisation ▪ Dauer der Zugehörigkeit zum Team

Daneben werden in der vorliegenden Studie aber auch verschiedene personale und organisationsbezogene Variablen betrachtet. Aufgrund von Führungserfahrung oder Berufserfahrung, aufgrund von vermehrter Zeit für die Führung Einzelner, aufgrund eines offenen Feedbackklimas und eines Klimas der Verbesserung beziehungsweise aufgrund guter Kenntnisse der organisations- und abteilungsinternen Abläufe und Ansprüche wird den Variablen, die in Tabelle 15 in der letzten Zeile aufgelistet sind, ein Einfluss auf den Führungserfolg zugesprochen. Es wird Hypothese 48) angenommen:

48) Berichten Mitarbeiter überdurchschnittlichen Erfolg, dann gilt im Vergleich zu Mitarbeitern mit unterdurchschnittlichem Erfolg:

a. Die Mitarbeiter sind älter.

b. Ihre Führungskraft ist älter.

c. Die Abteilung ist kleiner.

d. Die Beteiligungsquote an der Befragung ist höher.

e. Die Dauer der Zugehörigkeit zur Organisation und zum Team ist länger.

Vor dem Hintergrund, dass Personalführung an sich darauf abzielt Erfolge herzustellen, wäre eine Unterstützung dieses Wirkungszusammenhangs durch die interperspektivische Übereinstimmung von Führungsbeurteilungen folgendermaßen denkbar. Wenn Selbst- und

Fremd- oder Fremd- und Fremd-Beurteiler in ihrer Führungsbeschreibung übereinstimmen, ist dies ein Hinweis für konsistente und selbstreflektierte Führung. Das Ausmaß einer solchen Übereinstimmung kann Einfluss nehmen auf die Wirkungskraft, die von gezeigtem Führungsverhalten ausgeht, da Führung dann konsistent wirken kann, Sicherheit bietet und nicht durch andere Führungsverhaltensweisen gegenteilige Effekte wieder „reparieren" muss. Die Übereinstimmung zwischen den Beurteilungsperspektiven moderiert dann den Zusammenhang zwischen Führung und Erfolg (vgl. Atwater & Yammarino, 1992). Es wird folgende Hypothese geprüft, um die Rolle der interperspektivischen Übereinstimmung weiter aufklären zu können.

49) Liegt Übereinstimmung zwischen den Perspektiven vor (vs. Unter- oder Überschätzung), fällt der Zusammenhang zwischen Führung und Erfolg höher aus.

Übereinstimmung ist einmal als Kategorie im Unterschied zu Über- und Unterschätzung zu verstehen, einmal als geringes Ausmaß an Abweichung (im Unterschied zu stärkerer Abweichung, unabhängig von der Richtung der Abweichung). Ersteres betrifft die Selbst-Fremd-Gegenüberstellung, Letzteres zusätzlich die Abwärts-Aufwärts-Gegenüberstellung.

In Tabelle 82 sind am Ende der Ergebnisdarstellung alle Hypothesen der vorliegenden Studie überblicksartig zusammengefasst. Abschließend ist festzuhalten, dass die Hypothesen aufgrund des für viele Analysen geringen Stichprobenumfangs anhand der Effektstärken (vgl. Tabelle 27) abgelehnt oder bestätigt werden. Nur in Ausnahmefällen (z.B. Binomialtest) werden Nullhypothesen auf der Basis des 5%-Signifikanzniveaus verworfen. Für Hypothese 39) werden unmittelbar vor den entsprechenden Analysen spezifische Kriterien aufgestellt.

Im Vergleich zu den bislang dargestellten Fragestellungen sind Fragestellung 11 und Fragestellung 12 explorativ ausgerichtet, weshalb für sie keine prüfbaren Hypothesen aufgestellt werden.

Fragestellung 11: Hinweise auf die Entstehung von interperspektivischen Diskrepanzen

Die potentiellen Erklärungsmechanismen für interperspektivische Diskrepanzen bei Führungsbeschreibungen sind bislang sehr unzureichend erforscht. Unbestreitbar stehen experimentelle Studien aus, deren Analysen auf die kausalen Mechanismen schließen lassen. In der vorliegenden Studien generieren die Befragten am Ende des Fragebogens Ideen, wie sie sich die Entstehung von Selbst-Aufwärts-Diskrepanzen in ihrem Arbeitsalltag

erklären können. Dass neue Forschungsarbeiten solche Ideen der „Betroffenen“ explorieren sollen, wird zunehmend gefordert (Greguras & Robie, 1998). Mehr Wissen über das Zustandekommen von Selbst-Fremd-Diskrepanzen birgt einen Nutzen, der über den reinen Erklärungswert hinausgeht. Daraus könnten langfristig Ansatzpunkte erarbeitet werden, wie eine höhere Übereinstimmung erzielt werden kann. Zudem birgt das Wissen über subjektive Erklärungen der Betroffenen den Vorteil, diese in organisationspsychologischer Anwendung „dort abholen zu können, wo sie stehen“.

Fragestellung 12: Hinweise auf Ansatzpunkte zur Reduzierung der Beurteilungs-Diskrepanz

Unabhängig davon, ob die interperspektivischen Diskrepanzen wirklich den Dreh- und Angelpunkt in der Vorhersage von Erfolgsmaßen darstellen, kann das Ziel, eine höhere Übereinstimmung herzustellen, in der Praxis für sinnvoll erachtet werden. Von den beteiligten Führungskräften und Mitarbeitern werden Ideen erfragt, wie potentielle Diskrepanzen zwischen Selbst- und Fremdbeurteilern reduziert werden könnten. Eine solche Brücke in den Anwendungsbereich ist unabdingbar und muss zukünftig weiter ausgebaut werden, weil der Erfolg von Personalentwicklungsmaßnahmen immer auch davon abhängt, ob sie von den fokussierten Führungskräften oder Mitarbeitern akzeptiert werden.

Abschließend werden die einzelnen Fragestellungen so gruppiert, wie sie auch in der Ergebnisdarstellung zusammengefasst werden. Nach Neuberger (2000, S. 56f.) lassen multiperspektivische Erhebungen prinzipiell vier verschiedene Vergleichsmöglichkeiten zu: erstens den interpersonalen oder Intergruppenvergleich von Ist-Werten, zweitens den personalen oder interpersonalen Ist-Soll-Vergleich, drittens den Vergleich verschiedener Beurteiler hinsichtlich des selben Beurteilungsgegenstands und viertens den intertemporalen Vergleich verschiedener Messungen. Die Fragestellungen des ersten Schwerpunkts der vorliegenden Studie lassen sich diesen Vergleichskategorien zuordnen (Tabelle 16).

Tabelle 16. Klassifikation der bearbeiteten Fragestellungen (modifiziert und erweitert nach Neuberger, 2000)

Themenklasse	Fragestellung	Ergebniskapitel
Personaler/interpersonaler Ist-Soll-Vergleich	1	6.2
(interperspektivische) Beurteiler-Übereinstimmung	2, 3, 4, 5, 6	6.3
Interpersonaler/-gruppaler Vergleich von Ist-Werten	4	6.3.3
Intertemporaler Vergleich	-	5.5.1.9
Prädiktive Validität von Führungsbeurteilungen	7, 8, 9	6.4
Prädiktive Validität von interperspektivischer Übereinstimmung in der Führungsbeurteilung	10	6.5
Entstehung und Reduktion von Beurteilungs-Diskrepanzen	11, 12	6.6

Anmerkung. Wo die Bearbeitung mehrerer Fragestellungen in einem Ergebniskapitel zusammengefasst wird, wird dieses entsprechend der einzelnen Fragestellungen in Unterkapitel geteilt.

So beschäftigt sich Fragestellung 1 mit verschiedenen Ist-Soll-Vergleichen. Die Fragestellungen 2 bis 5 stellen die Übereinstimmung verschiedener Beurteilungen des Führungsverhaltens der selben Führungskraft (also des gleichen Beurteilungsgegenstands) in den Mittelpunkt. Im Zusammenhang mit der Beantwortung von Fragestellung 4 werden zusätzlich auch Intergruppen-Vergleiche durchgeführt. Die Stichprobe wird dazu anhand verschiedener potentiell moderierenden Kriterien (personale Kriterien, organisationsbezogene Kriterien, Durchführungsbedingungen) in Gruppen eingeteilt. Fragestellung 6 kehrt die Vergleichsrichtung um und fokussiert auf die Übereinstimmungen verschiedener Beurteilungen, die *eine* Person für verschiedene Führungskräfte (Beurteilungsgegenstände) abgibt. Für eine kleine Substichprobe kann ein intertemporaler Vergleich zur Bestimmung der Retest-Reliabilität der Aufwärtsbeurteilungen durchgeführt werden. Diese wird bereits vor der Ergebnisdarstellung in Kapitel 5.5.1.9 im Zusammenhang mit der Skalenanalyse dokumentiert. Über diese Vergleichskategorien hinaus gehen Fragestellung 7 bis 9, die sich mit dem Zusammenhang der Führungsbeurteilungen und verschiedenen Erfolgskriterien beschäftigen sowie Fragestellung 10, die den Zusammenhang zwischen der interperspektivischen Übereinstimmung und den Erfolgsvariablen betrachtet. Als weiterführend sind auch Fragestellung 11 und Fragestellung 12 zu verstehen, die eher explorativ angelegt sind. Es geht dabei um die Generierung von potentiellen Ansatzpunkten zur Erklärung beziehungsweise Reduzierung der Beurteilungs-Diskrepanz.

5 Methode

5.1 Stichprobe

Die Stichprobe umfasst Personen der vier höchsten Führungsebenen eines Automobil-Unternehmens sowie deren direkt unterstellten Mitarbeiter (mit oder ohne Führungsverantwortung). Insgesamt beteiligten sich an der Befragung 210 Personen, wovon 50 den vier höchsten Führungsebenen angehören. Sie werden im Folgenden als „*Führungskräfte*“ oder „*Mitarbeiter mit Führungsverantwortung*“ bezeichnet. Letztere Bezeichnung wird vor allem dann gewählt, wenn es um die Aufwärtsbeurteilung ihrer direkten Führungskraft geht. Beteiligte ab der fünften Hierarchieebene wurden ausschließlich zur Aufwärtsbeurteilung ihrer direkten Führungskräfte einbezogen. Sie werden durchgängig als „*Mitarbeiter“ (ohne Führungsverantwortung)* bezeichnet.[134]

Das Alter der Beteiligten schwankt zwischen 26 und 62 Jahren und liegt im Mittel bei 43 Jahren (*SD* = 8.9). Die überwiegende Mehrheit der Teilnehmer ist männlich sowie im Geschäftsbereich Technik tätig (je 85%). Die Befragung erstreckte sich prinzipiell über alle Geschäftsbereiche (Technik, Finanz/Controlling, Personalwesen) und über zwei Standorte. Die Führungskräfte tragen im Durchschnitt für 10 Personen Führungsverantwortung bei einer Spannbreite von 3 bis 27 Personen (*SD* = 6). Die mittlere Dauer, die die Befragten bereits in der gleichen Abteilung tätig sind, beträgt 6 Jahre und 7 Monate (*SD* = 4 Jahre und 2 Monate). Der Organisation gehören die Befragten im Durchschnitt 8 Jahre und 9 Monate an (*SD* = 4 Jahre und 6 Monate). Weitere Charakteristika zur Gesamtstichprobe sind Tabelle 17 zu entnehmen. Dort werden entsprechende Daten auch für die Selbst-, Aufwärts- und Abwärtsbeurteiler getrennt angegeben.

Von den Führungskräften wurden 48 Selbstbeschreibungen abgegeben sowie 48 Abwärtsbeschreibungen, die das Führungsverhalten ihrer direkt unterstellten Mitarbeiter mit Führungsverantwortung betrafen. Zusätzlich beurteilten 207 Personen jeweils das Führungsverhalten ihrer direkten Führungskraft (Aufwärtsbeschreibung). Daraus ergaben sich 40 Selbst-Aufwärts-Passungen, 32 Selbst-Abwärts-Passungen und 32 Aufwärts-Abwärts-Passungen. "Passung" bedeutet, dass es sich um Fragebögen aus verschiedenen Beurteilungsperspektiven handelt, die das Führungsverhalten derselben Fokusperson beschreiben. Der geringe Stichprobenumfang, was Passungen betrifft, ist nicht allein dem Rücklauf geschuldet, sondern liegt auch im Forschungsfeld begründet. Flache Hierarchien bringen die Reduktion von Führungspositionen mit sich.

[134] In dieser Gruppe der Mitarbeiter ohne Führungsverantwortung befinden sich unter anderem Gruppenleiter aus der Produktion. Diese tragen streng genommen Führungsverantwortung. Allerdings sind ihre Führungsaufgaben spezifischer und stärker an situationale Gegebenheiten gebunden (z.B. Zielsetzung ist nicht losgelöst von Taktzeiten möglich) als Personalführung in höheren Hierarchieebenen. Die Fragebogeninhalte sind deshalb für die Führungsarbeit der Gruppenleiter nicht ausreichend relevant, um sie als Selbstbeurteiler sinnvoll in die Befragung integrieren zu können.

Tabelle 17. Stichprobencharakteristik

Merkmal		Anteil an der Gesamtstichprobe	Anteilige Verteilung der Ist-Versionen aus den verschiedenen Perspektiven auf verschiedene Merkmale		
			Selbst	Aufwärts	Abwärts
Hierarchieebene	GF & 2. Ebene	4%	19%	3%	58%
	3. Ebene	8%	31%	8%	42%
	4. Ebene	12%	50%	12%	0%
	ab 5. Ebene	76%	0%	77%	0%
Alter	bis 35	22%	13%	23%	0%
	36 bis 45	31%	31%	32%	35%
	46 bis 55	31%	41%	31%	30%
	ab 56	11%	15%	11%	35%
Geschlecht	männlich	85%	90%	83%	100%
	weiblich	15%	10%	16%	0%
Geschäftsbereich	Technik	85%	81%	86%	79%
		davon: Fahrzeugfertigung: 36% Motorenfertigung: 18% Serienplanung/ Werktechnik: 12% Logistik/ Beschaffung: 26% Rest: 7%	aufgeteilt: Fahrzeugfertigung: 21% Motorenfertigung: 8% Serienplanung/ Werktechnik: 17% Logistik/ Beschaffung: 27% Rest: 8%	aufgeteilt: Fahrzeugfertigung: 31% Motorenfertigung: 16% Serienplanung/ Werktechnik: 10% Logistik/ Beschaffung: 22% Rest: 7%	aufgeteilt: Fahrzeugfertigung: 23% Motorenfertigung: 15% Serienplanung/ Werktechnik: 21% Logistik/ Beschaffung: 19% Rest: 2%
	Finanz/ Controlling	10%	15%	9%	15%
	Personalwesen	5%	4%	5%	6%
Standort[135]	Standort A	83%	92%	83%	86%
	Standort B	15%	8%	16%	14%

Anmerkungen. Es werden Prozentwerte berichtet, obwohl für die Abwärts- und Selbstbeurteilungen der prozentuale Anteil an lediglich 48 entsprechenden Beurteilungen angegeben werden kann. Fehlende auf 100%: Angaben (z.B. zu Geschlecht) fehlen im Fragebogen. GF = Geschäftsführung.

[135] Fehlende auf 100% sind die Personen (z.B. GF), die beiden Standorten gleichermaßen zuzuordnen sind oder aus anderen Gründen nicht eindeutig zuordenbar sind.

Die Anzahl der Aufwärtsbeschreibungen, die einem Selbstbild gegenüber gestellt werden können, variiert zwischen 1 (für 15 % der Führungskräfte) und 9 (für 5% der Führungskräfte). Die Anzahl der Aufwärtsbeschreibungen, die einem Abwärtsbild gegenüber gestellt werden können, variiert ebenso zwischen 1 (für 13% der Abwärtsbilder) und 9 (für 6% der Abwärtsbilder). Durchschnittlich können sowohl jedem Selbstbild als auch jedem Abwärtsbild vier Aufwärtsbeschreibungen gegenüber gestellt werden – vorausgesetzt es liegt überhaupt eine Passung vor. Selbst- und Abwärtsbeschreibungen liegen naturgemäß in einer Linienorganisation für das Führungsverhalten jeder Führungskraft niemals mehrfach vor. Entsprechend den Ergebnissen von Church, Rogelberg und Waclawski (2000) ist auch für die vorliegende Befragung nicht zu erwarten, dass die Anzahl der Rücksender pro Führungskraft mit deren Leistungsniveau zusammenhängt. Außerdem unterscheiden sich weder die Aufwärtsbeurteilungen noch die Abwärtsbeurteilungen in Abhängigkeit davon, ob die fokussierte Führungskraft selbst an der Befragung teilnahm oder nicht. Es sind allenfalls geringe Effekte festzuhalten, die über die verschiedenen Führungsfaktoren hinweg jedoch nicht in eine einheitliche Richtung weisen. Deshalb ist auszuschließen, dass durch potentielle Klimavariablen in bestimmten Abteilungen eine unerwünschte Beurteilerselektion stattgefunden hat.

Die Anzahl bearbeiteter Fragebögen zur Führungsbeschreibung pro beteiligter Person variiert zwischen einem und neun, wobei im Mittel drei Bögen von einer Person ausgefüllt wurden. Die Anzahl der bearbeiteten Fragebögen nimmt mit steigender Hierarchieebene zu. Zusätzlich zur Ist-Version wurden 118 Ideal- und 109 Wichtigkeitseinschätzungen des Führungsverhaltens abgegeben. Die Sollversion wurde gezielt von von 5 leitenden Führungskräften und der Geschäftsführung eingeholt. Neben diesen 536 Fragebögen zur Führungsbeschreibung (FzFü) wurden 91 Aufwärtsbeurteilungen von Führung auf einem anderen Instrument abgegeben, so dass insgesamt 627 Fragebögen vorliegen, die sich auf das Führungsverhalten beziehen. Hinzu kommen die von nahezu jedem Befragten eingeschätzten Erfolgsratings. In Abbildung 20 ist die Anzahl der Fragebögen getrennt für die verwendeten Versionen und Instrumente aufgelistet.

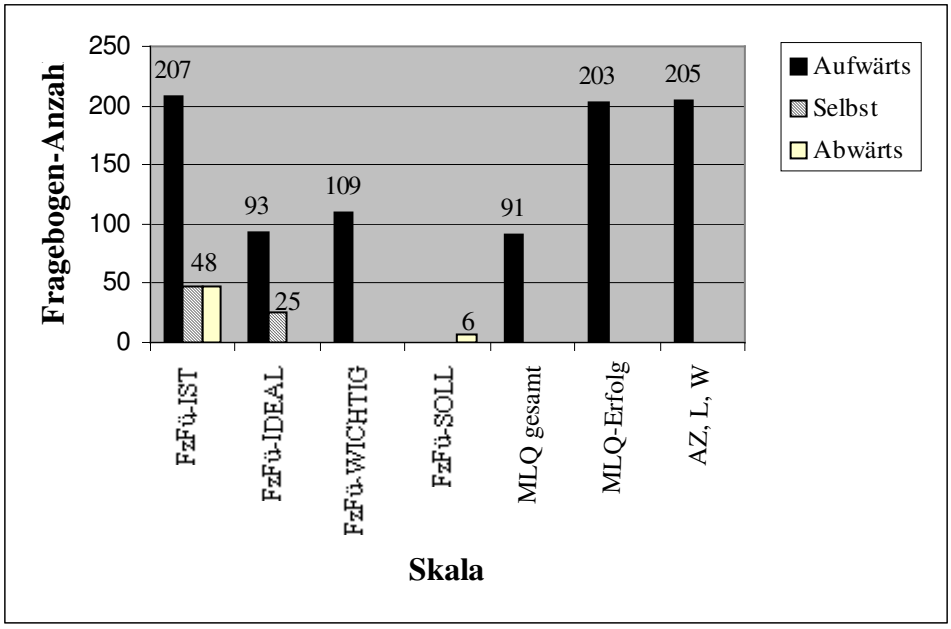


Abbildung 20. Rücklauf: Anzahl eingegangener Fragebogen-Versionen

Anmerkungen. MLQ-Erfolg: Erfolgsskalen des MLQ, AZ/L/W: Skalen zu Arbeitszufriedenheit, Leistung und Wohlbefinden. Die Instrumente sind im folgenden Kapitel näher beschrieben.

Als minimale Dauer, die die Fremdbeurteiler bereits in der aktuellen Funktion mit der beurteilten Führungskraft zusammen arbeiten mussten, um eine Führungsbeurteilung abgeben zu können, wurden 4 Monate festgesetzt. Im Mittel beträgt die Dauer der Zusammenarbeit von Führungskraft und Aufwärtsbeurteiler 3 Jahre und 2 Monate bei einer ebenso großen Standardabweichung (Range zwischen 4 Monaten und 12 Jahren) beziehungsweise von Führungskraft und Abwärtsbeurteiler 3 Jahre und 1 Monat bei einer Standardabweichung von 3 Jahren und 3 Monaten (Range zwischen 4 Monaten und 11 Jahren).

Von den angeschriebenen Führungskräften beteiligten sich 64%, von den Mitarbeitern 31%. Die Rücklaufquote beträgt damit 43%, wenn man alle angeschriebenen Personen berücksichtigt. Diese Zahlen reflektieren die generellen Erfahrungen, dass die Beteiligungsquote tendenziell ansteigt, je höher die hierarchische Position der Befragten ist (Jöns & Mataja, 1998). Für Mitarbeiterbefragungen im Allgemeinen liegt die Rücklaufquote damit im unteren Normbereich. Hinsichtlich der Führungskräftebeteiligung ist sie als gut zu bewerten (vgl. z.B. Becker et al., 2002). Desweiteren beantworteten vier Personen zur Bestimmung der Stabilität einen Auszug des Fragebogens circa 1,5 Jahre nach der Ersterhebung (August 2003).

5.2 Untersuchungsverfahren

Zur Datenerhebung wurden folgende Fragebögen eingesetzt:

- Fragebogen zum Führungsverhalten (FzFü)
- Multifactor Leadership Questionnaire (MLQ) in der deutschsprachigen Version von Felfe und Goihl (2003), enthält Skalen zu Führung sowie zu Erfolg
- Erfolgsskalen zu den Bereichen Arbeitszufriedenheit, Wohlbefinden und Leistung

Desweiteren wurden biografische sowie organisationsbezogene Daten strukturiert erfasst und offene Fragen zur multiperspektivischen Führungsbeurteilung in den Fragebogen integriert. Die einzelnen Skalen und Fragebogenabschnitte werden im Folgenden dargestellt. Ihre Konstruktion ist in Kapitel 5.5 ausführlich dokumentiert. Dort werden auch die Ergebnisse der Skalenanalysen berichtet, welche für alle, nicht nur für die neu konstruierten Verfahren durchgeführt wurden.

5.2.1 Fragebogen zur Führungsbeschreibung (FzFü)

Führungsbeschreibung. Der Fragebogen zur Führungsbeschreibung (FzFü) wurde eigens für die vorliegende Studie entwickelt. Sein Hauptteil setzt sich aus fünf Subskalen zusammen, die klassische Führungsaufgaben und zentrale Führungsverhaltensweisen erfassen. In Tabelle 18 sind sie jeweils mit einem Beispielitem aufgelistet.

Tabelle 18. Subskalen des Fragebogens zur Führungsbeschreibung (Hauptteil)

Entwickeln & Fördern, E/F (12 Items): „Meine direkte Führungskraft plant mit mir meine beruflichen Möglichkeiten und Entwicklungsschritte.“
Beurteilen & Steuern, B/S (10 Items): „Meine direkte Führungskraft informiert mich über den Anlass einer Beurteilung.”
Individuelle Berücksichtigung, IB (8 Items) „Meine direkte Führungskraft berücksichtigt meine Kenntnisse und Erfahrungen bei der Erarbeitung von Zielen.“
Klarheit, K (7 Items) “Meine direkte Führungskraft bespricht Zielsetzungen mit mir so lange, bis alle Unklarheiten beseitigt sind.”

Partizipation, P (5 Items) “Meine direkte Führungskraft beteiligt mich möglichst stark bei der Festlegung der Ziele, die ich im folgenden Zeitraum erreichen soll.“

Dieser Hauptteil des Fragebogens kann um zwei eindimensionale Skalen ergänzt werden, die flexibel eingesetzt werden können. Sie betreffen das Fördern von Führungsverhalten beziehungsweise die Stellvertretung (vgl. Tabelle 19). Erstere ist dann von Belang, wenn die Aufwärtsbeurteiler (direkte Mitarbeiter der Führungskraft) eigene Führungsverantwortung haben oder in Kürze haben werden. Letzere ist aus der Aufwärtsperspektive lediglich von denjenigen Mitarbeitern zu beantworten, die ihre Führungskraft bei deren Abwesenheit vertreten. In Anhang D und Anhang E sind alle Items nachzulesen.

Tabelle 19. Eindimensionale Skalen des Fragebogens zur Führungsbeschreibung

Fördern von Führungsverhalten, FvF (4 Items): „Meine direkte Führungskraft sieht es als ihre Aufgabe, auch mein Führungsverhalten zu entwickeln.“ Stellvertretung, S (4 Items): „Meine direkte Führungskraft bespricht mit mir alle wichtigen Aufgaben, bevor sie in Urlaub geht.”

Der FzFü liegt in drei „Perspektiv-“ und in drei „Modusversionen“ vor. Zwischen den „Perspektiv-Versionen“ variieren die Itemformulierungen insofern, dass die Führungsbeurteilung aus der Selbst-, der Abwärts- und der Aufwärtsperspektive erfolgen kann. Als „Modus-Versionen“ sind die Versionen zur Ist-, Ideal-, Wichtigkeits- und Sollbeurteilungen zu verstehen. Im Folgenden werden die einzelnen Versionen kurz beschrieben, wobei zusätzlich dargestellt wird, für welche Perspektiv-Versionen die einzelnen Modus-Versionen eingesetzt wurden.

Mittels der **Ist-Versionen** wird die Wahrnehmung des aktuellen Führungsverhaltens erhoben („beobachtetes Führungsverhalten“):

Aufwärtsperspektive: Mitarbeiter (mit oder ohne Führungsverantwortung) beschreiben das Führungsverhalten ihrer direkten Führungskraft.

Selbstperspektive: Führungskräfte beschreiben ihr eigenes Führungsverhalten.

Abwärtsperspektive: Führungskräfte beschreiben das Führungsverhalten ihrer direkt unterstellten Mitarbeiter *mit* Führungsverantwortung. Gemäß des Führungsmodells sind sie aufgefordert, deren Führungsverhalten zu fördern.

Die **Ideal-Versionen** fragen nach dersubjektiv idealen Vorstellung von Führungsverhalten:

Aufwärtsperspektive: Mitarbeiter beschreiben optimales Führungsverhalten eines Vorgesetzten.

Selbstperspektive: Führungskräfte beschreiben ein Führungsverhalten, wie sie es idealiter gerne zeigen würden.

Mithilfe der **Wichtigkeits-Version** werden Angaben darüber gewonnen, wie wichtig die einzelnen Führungsverhaltensweisen für den Erfolg gehalten werden.

Mitarbeiter mit und ohne eigene Führungsverantwortung schätzen ein, wie wichtig einzelne Führungsverhaltensweisen für ihre Leistung und Arbeitszufriedenheit sind. Da es sich nur im formalen Sinne um eine *Aufwärtsperspektive* handelt, inhaltlich jedoch per se die Item-Wichtigkeit für den Führungserfolg geratet werden sollte, wurde auf den Einsatz weiterer Wichtigkeitsversionen verzichtet. Es soll explizit darauf verwiesen werden, dass sich dadurch die Anzahl der potentiell zu Befragenden lediglich um $N = 3$ (oberste Führungsebene) verringert. Aus diesem Grund bot sich die Kombination der Wichtigkeits-Version mit den Aufwärtsbeurteilungen an, um von möglichst vielen Befragten eine Wichtigkeitseinschätzung erbitten zu können.

In der **Soll-Version** werden die einzelnen Aussagen danach bewertet, in welcher Ausprägung sie gemäß des Führungsmodells gefordert werden. Da es um eine Einschätzung der Unternehmensleitlinien geht, wurden nur die Geschäftsführung und leitende Führungskräfte, die entsprechende Schlüsselpositionen einnehmen, befragt, um die Soll-Werte zu bestimmen. Sowohl das Soll als auch das Ideal stellen „*erwartetes Führungsverhalten*“ dar.

Durch den Einsatz der genannten Fragebogen-Versionen ergeben sich also Angaben zum *beobachteten Führungsverhalten* der Führungskräfte selbst (Selbst-Ist), ihrer direkten Mitarbeiter (Aufwärts-Ist) und ihrer direkten Vorgesetzten (Abwärts-Ist), zum *idealen Führungsverhalten* der Führungskräfte selbst (Selbst-Ideal) und ihrer direkten Mitarbeiter (Aufwärts-Ideal), zur *Wichtigkeitseinschätzung* sowie zu dem *nach dem Unternehmensleitbild erwarteten Führungsverhalten* (Soll). Für alle Versionen wird als Antwortformat eine Skala von 1 („trifft überhaupt nicht zu“) bis 7 („trifft voll und ganz zu“) verwendet (vgl. Kapitel 5.5.1). In die Auswertung gehen die Skalenmittelwerte ein. In begründeten Ausnahmefällen wird der Mittelwert der ersten fünf Subskalen verwendet.

Übereinstimmung verschiedener FzFü-Beurteilungen. Aus den Angaben zu den einzelnen FzFü-Versionen wurden verschiedene Übereinstimmungsmaße errechnet. Within-subject wurden die *Übereinstimmung von Ist- und Ideal*-Beurteilungen sowie von *Ist- und Soll*-Beurteilung berechnet. Between-subject wurde die *Übereinstimmung von Ist-Aufwärts und Ist-Selbst*, von *Ist-Abwärts und Ist-Selbst,* von *Ist-Abwärts und Ist-Aufwärts* sowie von *Ideal-Aufwärts und Ideal-Selbst* berechnet.

Probleme in der methodischen Operationalisierung von „Übereinstimmung" betreffen die Aggregation von Ratings sowie die Wahl eines geeigneten Diskrepanz-Maßes. Was die Aggregation von Ratings betrifft, wurde je nach Fragestellung mit den individuellen Ratings einzelner Mitarbeiter beziehungsweise mit aggregierten Aufwärtsratings (Median der Aufwärtsratings für dieselbe Führungskraft) gerechnet. Schwieriger gestaltet sich das zweite Problem, da alle Indizes für die Diskrepanz zwischen zwei Urteilen verschiedene Nachteile in sich bergen und sich gegenseitig nicht uneingeschränkt ersetzen können. Deshalb wurden – entsprechend der Forderung von Atwater et al. (1998) – mehrere Indizes verwendet: erstens das *euklidische Distanzmaß* und zweitens die *Produkt-Moment-Korrelation* nach Pearson. Während Letztere eher die Ähnlichkeit der Profilverläufe angibt, entspricht die (euklidische) Distanz zwischen zwei Beurteilungen deren Abstand in einem zweidimensionalen Koordinatensystem. Das Abstandsmaß fokussiert also auf die Höhe der Niveau-Unterschiede zwischen verschiedenen Beurteilungen, ohne die Richtung der Abweichung zu berücksichtigen und wird folgendermaßen berechnet:

$$D_{1j} = \sqrt{\frac{\sum_{j=1}^{n}(x_1 - x_j)^2}{n}} \quad \text{(nach Cronbach \& Gleser, 1953)}$$

Anmerkung. *n*: Anzahl der Mitarbeiter, die derselben Führungskraft berichten.

Gegenüber dem Mittelwert einfacher Differenzen birgt das Distanzmaß den Vorteil, dass sich Abweichungen in verschiedene Richtungen nicht aufheben. Zudem wird durch die Quadrierung erreicht, dass größere Abweichungen stärker berücksichtigt werden als kleinere, welche mit höherer Wahrscheinlichkeit nur zufällige Messungenauigkeiten widerspiegeln. Dennoch wurde für verschiedene Vergleiche (z.B. Selbst-Abwärts-Vergleich) und verschiedene Unterfragestellungen auch auf *einfache Differenzen* zurückgegriffen. Dies wird an entsprechender Stelle berichtet.

Alle Differenzscores bergen das Problem, dass ihre Reliabilität von den Reliabilitäten der eingehenden Komponenten abhängt und sich daher eine Komponente stärker auswirkt als die andere. Auch wenn Differenzscores deswegen mehrfach kritisiert wurden (z.B. Edwards, 1994b), können sie unter verschiedenen Voraussetzungen als Maß für die Niveau-Unterschiede eingesetzt werden. Dazu gehört, dass die einzelnen Komponenten theoretisch

gerechtfertigt sind, von unterschiedlichen Beurteilungsquellen stammen und mit Skalen gemessen werden, die eine hohe interne Konsistenz besitzen (Johns, 1981). Alle Kriterien können für die vorliegende Studie als gegeben betrachtet werden.

Für Analysen, denen explizit die *Richtung* der Abweichung zugrunde gelegt werden sollte, wurden die Führungskräfte in Unterschätzer, Übereinstimmer und Überschätzer kategorisiert. Die Kategorisierung wurde in Anlehnung an die Arbeitsgruppe von Atwater und Yammarino vorgenommen. Dabei gelten diejenigen Führungskräfte als Übereinstimmer, die maximal eine halbe Standardabweichung von der Fremdbeurteilung abweichen. Wer weiter nach oben abweicht, gilt als Überschätzer, wer weiter nach unten abweicht, gilt als Unterschätzer. Eine solche Kategorisierung ist naturgemäß nur für die *Selbst*-Fremd-Gegenüberstellungen möglich. Diese drei Führungskräfte-Gruppen, die eine unterschiedliche Passung zu den Fremdbeurteilungen aufweisen, können anschließend bezüglich ihrer Erfolge miteinander verglichen werden (z.B. Atwater & Yammarino, 1992; Fleenor et al., 1996). Ausdrücklich sei hier noch einmal erwähnt, dass „Überschätzung“ als reine Deskription des Niveau-Unterschieds zu verstehen ist.

5.2.2 Multifactor Leadership Questionnaire (MLQ)

Der Multifactor Leadership Questionnaire (MLQ) basiert unmittelbar auf dem Führungskonzept von Bass zur transaktionalen und transformationalen Führung (vgl. Kapitel 2.5.3) und wurde in der vorliegenden Studie in der deutschen Version von Felfe und Goihl (2003) eingesetzt. Gemäß der Annahme von Bass, den „Full Range of Leadership“ abzubilden, ist das Führungskontinuum von Laissez-Faire-Führung (1 Subskala) über transaktionale Führung (3 Subskalen) bis zur transformationalen Führung (5 bzw. in der deutschen Version erweitert auf 6 Subskalen) abgebildet. Damit sind alle theoretisch angenommenen Führungsbausteine in einzelnen Subskalen des Fragebogens repräsentiert. Hinzu kommen drei Erfolgsskalen (Extra Effort, Effectiveness und Satisfaction).

Alle Subskalen sind in Tabelle 20 bis Tabelle 23 mit der jeweiligen Item-Anzahl und je einem Beispiel-Item aufgelistet. Der Fragebogen wurde zur Aufwärtsbeurteilung von Führung eingesetzt und ist in Anhang C abgedruckt.

Transformationale Führung umfasst Subskalen zu Charisma (zwei Faktoren zu Idealized Influence), Inspirierender Motivierung, Intellektueller Stimulation und Individueller Berücksichtigung. Im Vergleich zu frühen Fassungen des MLQ wird nicht nur Idealized Influence in einen Zuschreibungs- und einen Verhaltensfaktor getrennt, sondern zudem

Inspirierende Motivierung unabhängig davon erfasst. Die Subskala „Ausstrahlung“[136] wurde von Felfe und Goihl (2003) konstruiert und als sechste Skala der transformationalen Führung hinzugefügt, nachdem explorative Interviews und Vorstudien auf die Wichtigkeit des Konstrukts hingewiesen hatten. Sie ist also in den englischsprachigen Original-Versionen nicht enthalten.

Tabelle 20. Subskalen der Transformationalen Führung und Beispielitems

Idealized Influence (attributed), IIA (4 Items): „Meine direkte Führungskraft macht mich stolz darauf, mit ihr zu tun zu haben.“
Idealized Influence (behavior), IIB (4 Items): „Meine direkte Führungskraft spricht mit anderen über ihre wichtigsten Überzeugungen und Werte.”
Inspirational Motivation, IM (4 Items) „Meine direkte Führungskraft spricht mit Begeisterung über das, was erreicht werden soll.“
Intellectual Stimulation, IS (4 Items) “Meine direkte Führungskraft bringt mich dazu, Probleme aus verschiedenen Blickwinkeln zu betrachten.”
Individual Consideration, IC (4 Items) “Meine direkte Führungskraft erkennt meine individuellen Bedürfnisse, Fähigkeiten und Ziele.“
Ausstrahlung, AUS (4 Items) „Meine direkte Führungskraft vermag mich durch ihre Persönlichkeit zu beeindrucken und zu faszinieren.“

Neben den transformationalen Skalen sind im MLQ drei transaktionale Skalen enthalten. Im Gegensatz zu vorhergehenden Versionen des MLQ wurde in der Version 5x (Bass & Avolio, 1995), die der deutschen Fassung zugrunde lag, erstmals die Dimension „Management-by-exception“ in einen aktiven und einen passiven Faktor geteilt.

[136] Sie wird von Felfe und Goihl (2003) auch als „Charisma“ bezeichnet, während „Idealized Influence“ nicht übersetzt wird.

Tabelle 21. Subskalen der transaktionalen Führung und Beispielitems

Contingent Reward, CR (4 Items) “Meine direkte Führungskraft bietet im Gegenzug für meine Anstrengungen ihre Hilfe an.“ Management-by-exception (active), MBA (4 Items) “Meine direkte Führungskraft konzentriert sich überwiegend auf Unregelmäßigkeiten, Fehler, Ausnahmen und Abweichungen von Vorschriften.” Management-by-exception (passive), MBP (4 Items) “Meine direkte Führungskraft vertritt die Ansicht, dass Probleme erst wiederholt auftreten müssen, bevor man handeln sollte.“

Auf einer Subskala wird der untere Pol des Führungskontinuums, die Nicht-Führung, beschrieben.

Tabelle 22. Subskala Laissez Faire und Beispielitem

Laissez Faire, LF (4 Items) „Meine direkte Führungskraft klärt wichtige Fragen sofort.“ (umgepolt)

Neben dem Führungsverhalten werden durch den MLQ drei Erfolgsmaße erfasst. Diese messen erstens die Anregung durch die Führungskraft, sich über die Erwartungen hinaus anzustrengen (Extra Effort), zweitens die wahrgenommene Effektivität der Führungskraft (Effectiveness) sowie drittens die Zufriedenheit mit ihrer Führung (Satisfaction).

Tabelle 23. Subskalen zu Erfolgsmaßen (MLQ) und Beispielitems

Extra Effort, EEF (3 Items) „Meine direkte Führungskraft erhöht meine Bereitschaft, mich stärker anzustrengen.“ Effectiveness, EFF (4 Items) „Meine direkte Führungskraft kann eine Gruppe effektiv führen.“ Satisfaction, SAT (2 Items) „Meine direkte Führungskraft gestaltet die Zusammenarbeit so, dass ich wirklich zufrieden bin.“

Der Beantwortung aller Subskalen liegt ein Antwortformat von 1 bis 5 (nie, selten, hin und wieder, oft, regelmäßig/fast immer) zugrunde, das auf die Erfassung der Häufigkeit der einzelnen Iteminhalte ausgerichtet ist. In die Auswertung gehen die Skalenmittelwerte ein. Für eine ausführliche Darstellung der Vor- und Nachteile des MLQ vgl. Kapitel 2.7.4.

5.2.3 Kurzskalen zur Erfassung der Erfolgskriterien

„Viele Firmen haben eingesehen, dass es zu wenig ist, nur die Kosten zu senken und die Einhaltung der Arbeitszeit zu kontrollieren. Entscheidend ist, dass eine Firma über den engen Horizont der Gewinnmaximierung hinausschaut und einen Sinn in ihrem Wirtschaften erkennt.“ (Grün, 2003, S. 8)

Um den verschiedenen Ebenen der Führungswirkung Rechnung zu tragen, wurde auf multiple Erfolgskriterien zurückgegriffen. Als Kriterien erfolgreicher Führung wurden neben den im MLQ enthaltenen Erfolgsmaßen drei klassische Kriterien des Führungserfolgs erfasst, nämlich Arbeitszufriedenheit, Leistung und Wohlbefinden der geführten Mitarbeiter. Die Konstruktion der diesbezüglichen Kurzskalen ist gemeinsam mit den Skalenkennwerten in Kapitel 5.5.3 beschrieben. Auf die Verwendung bereits bestehender Instrumente wurde wegen ihrer Länge und ihres für die vorliegende Studie zu hohen Spezifitätsgrads verzichtet. Die drei Kurzskalen wurden mit einer gemeinsamen Instruktion eingeführt, so dass sie auch in Anhang F gemeinsam abgedruckt sind. Dort ist der gesamte Abschlussteil des Fragebogens nachzulesen, in dem auch die offenen Fragen enthalten sind. Alle Erfolgsskalen wurden aus der Aufwärtsperspektive bearbeitet, Beispielitems finden sich in Tabelle 24.

Tabelle 24. Erfolgsskalen und Beispielitems

Arbeitszufriedenheit – Interaktionsfaktor, AZ1 (6 Items) “Im Allgemeinen bin ich mit dem Verhältnis zu meinen Kollegen in unserem Team / unserer Abteilung (Skala von „gar nicht zufrieden“ [1] bis „voll und ganz zufrieden“ [7]).“
Arbeitszufriedenheit – Kontextfaktor, AZ2 (4 Items) “Im Allgemeinen bin ich mit den Maßnahmen zu meinem Gesundheitsschutz [Skala von „gar nicht zufrieden“ [1] bis „voll und ganz zufrieden“ [7]).“
Wohlbefinden, WOBE (2 Items) „Ich fühle mich an meinem Arbeitsplatz gestresst.“ (umgepolt)
Leistung, L (2 Items) „Ich bringe an diesem Arbeitsplatz meine volle Leistung.“

Arbeitszufriedenheit. Die Arbeitszufriedenheit wurde anhand eines zweifaktoriellen Fragebogens eingeschätzt, der zum einen die Zufriedenheit mit Interaktionen im weitesten Sinne erfasst und zum andern die Zufriedenheit mit Kontextbedingungen. „Interaktionen im weitesten Sinne“ ist so zu verstehen, dass neben der Zufriedenheit in der Zusammenarbeit mit anderen Organisationsmitgliedern auch die Zufriedenheit mit solchen Faktoren betrachtet wird, die entweder in Beziehungen stattfinden (z.B. Leistungsanerkennung) oder die individuell „verhandelt“ werden (z.B. Entwicklungsmöglichkeiten).

Wohlbefinden und Leistung. Wohlbefinden und Leistung gehen jeweils mit zwei allgemeinen Items in die Befragung ein. Im Zusammenhang mit der Leistung wurde zudem erfragt, wie häufig Beurteilungsgespräche stattfinden und wie häufig sich die Mitarbeiter Beurteilungsgespräche wünschen. Durch die Differenz zwischen tatsächlicher und gewünschter Häufigkeit von Beurteilungsgesprächen wird ein Aspekt von *Zufriedenheit mit Beurteilung* operationalisert.

Sowohl für Arbeitszufriedenheit als auch für Wohlbefinden wurde zusätzlich das Ausmaß erfasst, zu welchem die jeweilige Führungskraft aus der Perspektive ihrer Mitarbeiter die Arbeitszufriedenheit beziehungsweise das Wohlbefinden erstens prinzipiell beeinflussen könnte und zweitens tatsächlich beeinflusst. Die Differenz zwischen den beiden Werten kann als „*engagierte Führung*“ bezeichnet werden: Dies wird als Zeichen der *Zufriedenheit* mit dem Engagement der Führungskraft interpretiert. In einer offenen Frage wurde jeweils nach ungenutzten Einflussmöglichkeiten gefragt.

Für die drei Kriterien, die als Differenzmaße operationalisiert sind – Zufriedenheit mit Beurteilung, Engagierte Führung (Arbeitszufriedenheit), Engagierte Führung (Wohlbefinden) – gilt, dass die Zufriedenheit höher ausfällt, je geringer die Differenz ausgeprägt ist. Insofern sind für diese Kriterien im Vergleich zu den anderen Kriterien jeweils inverse Zusammenhänge zu erwarten.

Mit Ausnahme der Zufriedenheit mit Beurteilung wurde für die Erfassung aller Erfolgsmaße in Anlehnung an den FzFü als Antwortformat eine 7er-Skala verwendet. Für Wohlbefinden und Leistung wurde auch der Wortlaut der Verankerung von „trifft überhaupt nicht zu“ (1) bis „trifft voll und ganz zu“ (7) aus dem FzFü übernommen. Die Zufriedenheit mit Beurteilung ergibt sich aus Ratings auf einer 5er-Skala. Für die empirischen Analysen wurden jeweils die Skalenmittelwerte herangezogen. Diese wurden nicht zu einem Gesamtscore für Erfolg aggregiert, da weder eine theoretische noch eine empirische Rechtfertigung dafür besteht (vgl. frühe Diskussion bei Neuberger, 1976).

Betrachtet man alle MLQ-internen und –externen Erfolgsskalen im Überblick, so kann festgehalten werden, dass sowohl Einstellungen (z.B. Arbeitszufriedenheit) als auch Verhaltensintentionen (Extra Effort) und Verhaltensergebnisse (Effectiveness) als Erfolgskri-

terien erfasst werden. Alle Erfolgsmaße stellen subjektive Einschätzungen derjenigen Mitarbeiter dar, die gleichzeitig auch das Führungsverhalten ihrer Vorgesetzten beurteilten. Diese Methode wird häufig kritisiert, da sie einen Common-Method-Bias hervorrufen und damit die Korrelation zwischen Führungsverhalten und Effektivität ungerechtfertigterweise erhöhen kann. Deshalb war es wichtig, dass in der vorliegenden Studie erstens auch Abwärts-Führungsbeurteilungen eingeholt wurden und zweitens individuelle Beurteilungen aggregiert werden. Dies bleibt bei der Interpretation der Ergebnisse nicht unbeachtet (siehe dort). An zentralen Stellen wird deshalb auf ein Criss-Cross-Design zurückgegriffen (vgl. S. 262).

5.2.4 Erfassung biografischer und organisationsbezogener Daten

Insbesondere um moderierende Effekte für den Zusammenhang zwischen Beurteilungen aus verschiedenen Perspektiven ermitteln zu können, wurden biografische sowie organisationsbezogene Variablen erhoben. Erfasst wurden das *Geschlecht*, das *Alter*, die *Hierarchieebene*, die *Abteilungsgröße/Führungsspanne*, die *Kenntnis der Führungsgrundsätze,*[137] die *Dauer der Organisationszugehörigkeit*, die *Dauer der Abteilungszugehörigkeit und* die *Dauer, die der Mitarbeiter bereits derselben direkten Führungskraft* berichtet.

5.2.5 Offene Fragen

In Anlehnung an Borg (2000, S. 92f.) werden von Mess-Items, die ein Konstrukt möglichst reliabel und valide messen sollen (vgl. Operationalisierung aller bisher angeführten Variablen), so genannte Transport- und Aktions-Items unterschieden. Transport-Items sollen den Prozess unterstützen, bestimmte Themen im Unternehmen zu platzieren, Aktions-Items sollen eine spätere Einbindung der Befragten in einen Veränderungsprozess vorbereiten. Die offenen Fragen beziehen sich auf prinzipielle Einsatzmöglichkeiten des Fragebogens, auf Führungsgrundsätze und –workshops des Unternehmens sowie auf Führungsbeurteilung im Allgemeinen und potentielle Diskrepanzen in der Führungsbeurteilung im Speziellen (Anhang F). Die Fragen zu potentiellen Diskrepanzen enthalten eindeutig auch aktionsvorbereitende Komponenten, ebenso die auf Seite 252 angesprochenen Fragen zu Möglichkeiten der Führungskräfte, die Arbeitszufriedenheit und das Wohlbefinden ihrer Mitarbeiter zu verbessern. Dieser Befragungsteil soll die Akzeptanz, die Nützlichkeit und das Gefühl, fair behandelt zu werden, erhöhen (vgl. Runge, 1996) sowie die

[137] Ob tatsächlich die Kenntnis der „richtigen“ Führungsgrundsätze gegeben war, wurde anhand einer offenen Frage überprüft, in der der Kerngedanke der Führungsgrundsätze notiert werden sollte.

Wertschätzung von subjektiven Erklärungsmodellen der „Betroffenen“. In diesem Sinne ist auch die Partizipation von Führungskräften und Mitarbeitern bei der Implementierung der Maßnahme und bei einem Folgeprojekt zu betrachten (Senger, in Vorb.).

Zusammenfassend werden in Tabelle 25 alle verwendeten Instrumente dargestellt.

Tabelle 25. Übersicht der eingesetzten Instrumente und der erfassten Konstrukte unter Angabe der jeweils einbezogenen Beurteilungsperspektiven

Instrument	Konstrukte	Quelle
FzFü-IST	▪ Wahrgenommenes Führungsverhalten (sieben Dimensionen; Aufwärts-, Abwärts- und Selbst-Perspektive) ▪ Übereinstimmung zwischen Perspektiven	Eigene Skala
FzFü-IDEAL	▪ Ideales Führungsverhalten (sieben Dimensionen, Aufwärts- und Selbst-Perspektive), ▪ Übereinstimmung zwischen Perspektiven ▪ Übereinstimmung zwischen IDEAL und IST	Eigene Skala
FzFü-WICHTIG	▪ Als wichtig erachtetes Führungsverhalten (sieben Dimensionen, Aufwärtsbeurteiler)	Eigene Skala
FzFü-SOLL	▪ Nach dem Unternehmensleitbild erwartetes Führungsverhalten (sieben Dimensionen) ▪ Übereinstimmung zwischen SOLL und IST	Eigene Skala
MLQ	▪ Transformationale, transaktionale Führung und Laissez Faire (Aufwärtsperspektive), ▪ Erfolgsvariablen (Aufwärtsperspektive)	Felfe & Goihl, 2003
MLQ-externe Erfolgsskalen	▪ Subjektive Einschätzung der Arbeitszufriedenheit, des Wohlbefinden und der Leistung (Aufwärtsperspektive)	Eigene Skalen

Anmerkungen. FzFü = Fragebogen zur Führungsbeschreibun, MLQ = Multifactor Leadership Questionnaire, AZ = Arbeitszufriedenheit, WOBE = Wohlbefinden, L = Leistung.

Alle Fragebögen, also auch alle Perspektiv- und Modus-Versionen, erfassen verschiedene Dimensionen von Personalführung oder von Erfolg. Diese wurden im vorliegenden Kapitel beschrieben und sind in Tabelle 26 überblicksartig mit den entsprechenden Abkürzungen aufgelistet. Da diese im Ergebnisteil verwendet werden, sind sie auf dem beigefügten Lesezeichen noch einmal abgedruckt.

Tabelle 26. Abkürzungen der eingesetzten Skalen zur Messung von Führung und Erfolg

Bereich	Gesamter Fragebogen	Dimensionen	Abkürzung
Führung	▪ Fragebogen zur Führungsbeschreibung (FzFü)	Entwickeln/Fördern	E/F
		Beurteilen/Steuern	B/S
		Individuelle Berücksichtigung	IB
		Klarheit	K
		Partizipation	P
		Fördern von Führungsverhalten	FvF
		Stellvertretung	S
	▪ Multifactor Leadership Questionnaire (MLQ)	▪ Transformationale Führung:	
		Idealized Influence (attributed)	IIa
		Idealized Influence (behavior)	IIb
		Inspirational Motivation	IM
		Intellectual Stimulation	IS
		Individual Consideration	IC
		Ausstrahlung	AUS
		▪ Transaktionale Führung:	
		Contingent Reward	CR
		Management by exception (active)	MBA
		Management by exception (passive)	MBP
		▪ Laissez Faire	LF
Erfolg	▪ MLQ-intern	Extra Effort	EFF
		Effectiveness	EFF
		Satisfaction	SAT
	▪ MLQ-extern	Arbeitszufriedenheit (Interaktionsfaktor)	AZ1
		Arbeitszufriedenheit (Kontextfaktor)	AZ2
		Zufriedenheit mit engagierter Führung (Arbeitszufriedenheit)	EF-AZ
		Wohlbefinden	WOBE
		Zufriedenheit mit engagierter Führung (Wohlbefinden)	EF-WOBE
		Leistung	L
		Zufriedenheit mit Beurteilung	ZB

5.3 Durchführung

Das vorliegende Kapitel beschreibt den konkreten Ablauf der Befragung.

5.3.1 Grundgerüst

Das Grundgerüst der vorliegenden Studie ist als 270°-Beurteilungssystem im Sinne eines Sandwich-Modells zu bezeichnen (vgl. Kapitel 3.1.2). Dabei wurden die Führungskräfte gebeten, anhand des oben vorgestellten Fragebogens zur Führungsbeschreibung ein Selbstbild ihres Führungsverhaltens zu entwerfen. Im gleichen Zeitraum sollten ihre direkt unterstellten Mitarbeiter (mit oder auch ohne Führungsverantwortung) ein Aufwärtsbild sowie ihre direkten Vorgesetzten ein Abwärtsbild beschreiben[138]. Aus dem Sandwich-Modell ergibt sich generell eine personelle Überschneidung insofern, als die gleichen Personen unterschiedliche Fragebogen-Versionen für unterschiedliche Fokuspersonen bearbeiten (Abbildung 21; genauer Abbildung 22).

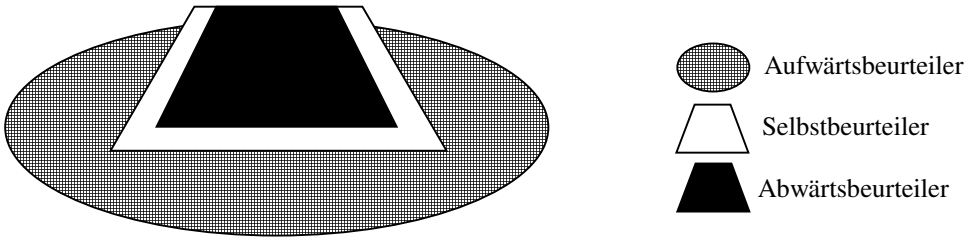


Abbildung 21. Schematische Darstellung der personellen Überschneidung bei Beurteilungen nach dem Sandwich-Modell bei einer 270°-Beurteilung

Auf Peer-Ratings wurde aus mehreren Gründen verzichtet, obwohl diese das unternehmensinterne 360°-Feedback hätten komplettieren können. Aus methodischer Sicht ist davon auszugehen, dass Peer-Ratings nicht deutlich mehr Information als

[138] Bezüglich der Abwärtsbeurteilung wird im Sinne der gültigen Führungsgrundsätze an den formalen Führungslinien festgehalten. Das impliziert erstens, dass Führungsbeziehungen, die sich nicht in einem Vorgesetztenverhältnis widerspiegeln, unberücksichtigt bleiben. Zweitens werden Vorgesetztenverhältnisse, die keine Führungs- sondern im Sinne von Weibler (2001, S. 34f.) nur eine Leitungsbeziehung darstellen, nicht ausgeschlossen (zu Begriffsklärung vgl. auch Kapitel 2.1). Während die Führungsgrundsätze zwar von jeder Führungskraft auch die Personalführung ihr unterstellter Mitarbeiter fordert, finden sich bei Betrachtung der informellen Wege auch andere oder zusätzliche Führungsbeziehungen. Astin und Astin (1996, S. 16) beschreiben dementsprechend alle Menschen als potentielle Führungskräfte und sagen eindeutig: „A leader ist not necessarily a person who holds some formal position of leadership [...].“

Abwärtsbeschreibungen enthalten. So weist zumindest das Ergebnis der Meta-Analyse von Harris und Schaubroeck (1988) auf hohe Korrelationen und geringe Niveau-Unterschiede der beiden Beurteilungsrichtungen hin. Außerdem gewinnt man durch Peer-Ratings kaum perspektivenbezogene Informationen, also Informationen, die nicht auch Beurteiler aus anderen Perspektiven geben könnten (Scullen et al., 2000; vgl. Kapitel 3.4.3). Das praktische Argument liefert der Zeitaufwand. Denn gerade diejenigen Personen, die für Peer-Beurteilungen in Frage gekommen wären, also wiederum ausschließlich Führungskräfte, sind durch Selbst-, Aufwärts- und teilweise Abwärtsbeurteilungen ohnehin bereits mehrfach belastet. Summa summarum ergibt sich aus den beiden Sachverhalten ein Kosten-Nutzen-Ergebnis, das gegen die zusätzliche Verwendung von Peer-Befragungen spricht.

Von den Moderatorvariablen, die an den Durchführungsbedingungen ansetzen und eine höhere Übereinstimmung von Selbst- und Fremd-Ratings erwarten lassen (vgl. Kapitel 3.3.3), wurden folgende umgesetzt:

- Validierungserwartung
- Anonymitätsinstruktion
- Verhaltensbezogene Items
- Ratingzweck: keine administrativen, sondern organisationsdiagnostische und forschungsbezogene Ratings

5.3.2 Fragebogenverteilung

Den Anfang bildeten schriftliche Ankündigungen der geplanten Fragebogenerhebung: Per Brief wurden alle später Befragten über die Vorgehensweise und die Ziele des Projekts informiert. Zusätzlich wurden in verschiedenen Managementkreisen Informationsveranstaltungen gehalten, bevor die entsprechenden Fragebögen (und je ein frankierter Rückumschlag) an alle Befragten verschickt wurden. Ein Begleitbrief enthielt weitere Informationen zum Projekt und zum Fragebogen. Auf folgende Weise wurden die verschiedenen Fragebogenversionen (Modus- und Perspektiv-Versionen) auf die einzelnen Befragten verteilt (vgl. Abbildung 22). Vorab sei angemerkt, dass sich die Fragebogenverteilung für die zweite und dritte Führungsebene nicht unterscheidet. Die Soll-Version ist in Abbildung 22 nicht aufgelistet, da sie gezielt von wenigen Führungskräften eingeholt wurde. Die Fragebögen konnten während der Arbeitszeit bearbeitet werden.

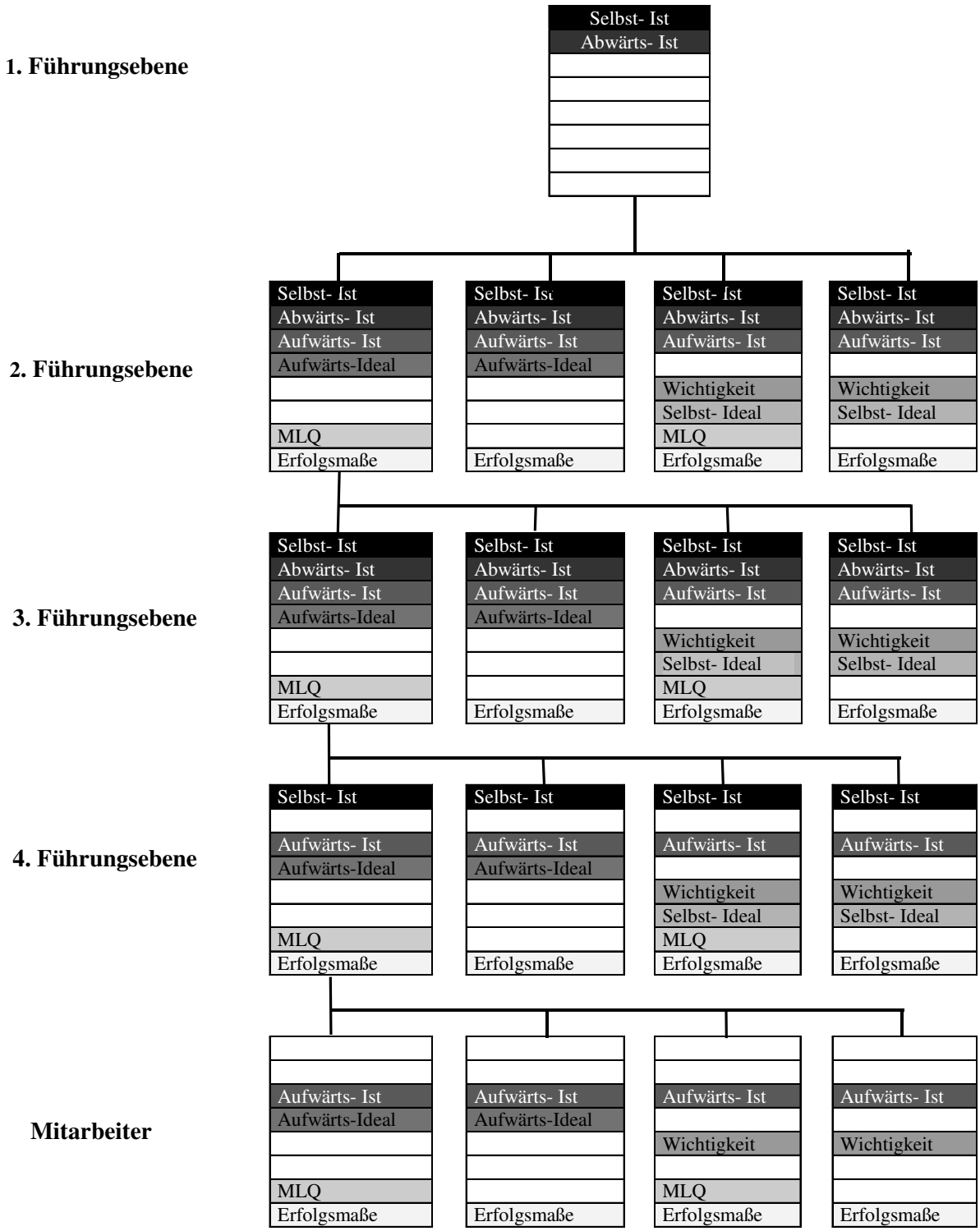


Abbildung 22. Schematische Darstellung des Verteilungsprinzips der Fragebogen-Versionen (ohne Soll-Version)

Alle Personen der obersten vier Führungsebenen erhielten die Version Selbst-Ist zur Beschreibung ihres eigenen Führungsverhaltens. Ihre jeweils unmittelbaren Vorgesetzten erhielten zusätzlich Abwärts-Ist, während ihre direkten Mitarbeiter Aufwärts-Ist bearbeiteten (bei eigener Führungsverantwortung zusätzlich Selbst-Ist). Alle Beteiligten beurteilten damit das Verhalten der Fokuspersonen. Ausbalanciert wurde die Verteilung der Ideal- und Wichtigkeits-Versionen insofern, dass die Hälfte der Aufwärtsbeurteiler gebeten wurde, ein Ideal-Aufwärts abzugeben, die andere Hälfte eine Wichtigkeitsschätzung. Diese Relation wurde nicht nur über die gesamte Personenstichprobe, sondern auch für jede Mitarbeitergruppe einer Führungskraft eingehalten. Die Version Selbst-Ideal erhielt jede zweite Führungskraft; genauer sind das jene Führungskräfte, die nicht Aufwärts-Ideal ausfüllten.

Für den MLQ gilt, dass die Gesamtversion (Aufwärts) der Hälfte der Befragten zugesandt wurde. Diese wiederum wurden ausbalanciert auf die Hälfte derjenigen, die Ideal-Aufwärts bearbeiteten sowie die Hälfte der Wichtigkeitsschätzer. Der MLQ war dem FzFü im Fragebogenheft nachgeordnet, so dass differentielle Einflussfaktoren auf die Beantwortung des FzFü ausgeschlossen waren.

Was die Erfolgsmaße angeht, so wurden diese von allen Aufwärtsbeurteilern und damit von allen Befragten mit Ausnahme der drei Führungskräfte, die die oberste Führungsebene bilden, eingeholt. Dies betrifft sowohl die Erfolgsmaße des MLQ als auch die Skalen zu Arbeitszufriedenheit, Wohlbefinden und Leistung.

5.3.3 Rücklaufprozess und Rückmeldung

Der Rücklauf erfolgte auf dem Postweg vom Befragten direkt zur TU Chemnitz. Nachdem der ursprünglich angekündigte Zeitraum von 5 Wochen (15. November bis 20. Dezember 2001) verstrichen war, wurden verschiedene Maßnahmen ergriffen, um den Rücklauf von 476 bis dahin eingegangenen Führungsfragebögen zu erhöhen. So wurden eine Rundmail mit Bitte um Teilnahme versendet, Telefonate mit Führungskräften sowie Teamgespräche vor Ort geführt. Dadurch konnte der Rücklauf um 32% auf 627 Fragebögen gesteigert werden (bis 15. März 2002). Für die Aufwärts- und Selbst-Versionen ergab sich zwischen den „Pünktlichen“ und den „Nachreichern“ kein Unterschied in der Beantwortung. Die Abwärtsversionen wurden aufgrund ihrer starken Bündelung auf wenige Personen (N = 16) von dieser Unterschiedsanalyse ausgeschlossen. Es gibt keinen Grund zur Annahme, dass sich die Rücksender von den Nicht-Rücksendern unterscheiden.

Die Ergebnisrückmeldung verlief über verschiedene Maßnahmen: Zum einen wurde ein Artikel in der Betriebszeitung veröffentlich, zum anderen fanden verschiedene Präsentationsveranstaltungen für das obere Management sowie auf Abteilungsebene statt. Damit wurden die in Deutschland gebräuchlichsten Methoden eingesetzt (Bungard et al., 1997).

5.4 Auswertungsmethoden

Alle Analysen werden zunächst für die Gesamtstichprobe durchgeführt. Der Selbst-Fremd-Zusammenhang unterscheidet sich auf keiner Führungsdimension statistisch signifikant zwischen dem technischen und den nicht-technischen Geschäftsbereichen, so dass die Analyse der Gesamtstichprobe gerechtfertigt ist. Das gleiche gilt für den Standortvergleich. Zudem brächte eine nach Standorten getrennte Analyse aufgrund des geringen Stichprobenumfangs von Standort B methodische Probleme mit sich. Unterschieden wird allerdings nach den Beurteilungsrichtungen. Einzelne andere Variablen werden außerdem im Rahmen von Moderatoranalysen als Trennkriterien verwendet. Dies ist an entsprechender Stelle in der Ergebnisdarstellung vermerkt.

Statistische Auswertung. Anhand verschiedener Methoden zu Mittelwertsvergleichen (v.a. t-Tests bzw. Mann-Whitney-U-Test bei sehr kleinen Stichproben, Einfaktorielle A-NOVA mit anschließenden Post-Hoc-Tests [Tukey-HSD-Test[139] bei Varianzhomogenität, Games-Howell-Test bei Varianzinhomogenität] und Test auf Linearität) wurden insbesondere Führungs- und Erfolgsbeurteilungen zwischen den Perspektiven sowie zwischen den verschiedenen Führungsdimensionen verglichen. Für viele Fragestellungen wurden Korrelationsanalysen durchgeführt. Neben der Berechnung von bivariaten (Pearson und Kendall-tau-b) und partiellen Korrelationen wurden zur Überprüfung, ob sich zwei Korrelationen unterscheiden, Hotellings *t*-Test[140] (bei abhängigen Stichproben) und Steigers *z*-Test (bei unabhängigen Stichproben) herangezogen. Während die Teststatistik des erstgenannten die Students t-Verteilung darstellt, ist letztere aufgrund der Fisher-Z-Transformation der Korrelationsindizes normalverteilt. Für Häufigkeitsvergleiche bei abhängigen Stichproben und dreifach gestuften Merkmalen wurde der Bowker-Test[141] herangezogen (vgl. Bortz, Lienert & Boehnke, 1990), seine Prüfgröße ist χ^2.

Zur Berechnung der Beurteilerübereinstimmung innerhalb einer Beurteilergruppe wurde die Intraklassenkorrelation nach Shrout und Fleiss (1979) bestimmt. Für die Moderatoranalysen wurden moderierte Regressionsanalysen und/oder Mediansplits mit anschließendem Mittelwertsvergleich vorgenommen. Zu moderierten Regressionen soll an dieser Stelle allgemein angemerkt werden, dass sie sich als konservative Methode herausgestellt haben, um Moderatoreffekte aufzudecken (Bobko, 1995, McClelland & Judd, 1993). Insofern sind auch kleine Interaktionseffekte bereits theoretisch und praktisch als relevant zu interpretieren. Mediansplits sind moderierten Regressionen vor allem dann vorzuziehen, wenn man

[139] „Tukey's honest significant difference"
[140] Auch bekannt als „Hotelling-Williams-Test" (vgl. Bobko, 1995, S. 56). Bobko (1995) resümiert, dass von den verschiedenen vorgeschlagenen Tests auf Korrelationsunterschiede für abhängige Stichproben kein anderer eindeutig präferiert werden könnte.
[141] Auch bekannt als „k x k-Felder-Symmetrietest" (vgl. Bortz et al., 1990, S. 166).

keine plausible Erklärung dafür hat, wie der Moderator die Wirkung vom Prädiktor auf das Kriterium beeinflusst. Da bei Selbst-Fremd-Übereinstimmung nicht begründetermaßen von einer Richtung des Zusammenhangs ausgegangen werden kann, wird dort auf moderierte Regressionen verzichtet. Generell bringen Mediansplits aber eine Erhöhung der Fehlervarianz und eine Verschwendung von Freiheitsgraden mit sich. Zudem ist das Ergebnis stark von der Verteilung abhängig.

Die Fragestellungen, die sich mit der prädiktiven Validität von Führungsbeurteilungen und Beurteilungsdiskrepanzen beschäftigen, wurden ebenfalls überwiegend regressionsanalytisch bearbeitet (schrittweise, hierarchische bzw. polynomiale Regressionen). Dafür gilt generell, dass die Führungsbeurteilungen (FzFü und MLQ) als Prädiktoren in die Studie eingehen, während die verschiedenen Erfolgsmaße als Kriterien fungieren.

Durch die verschiedenen Regressionsanalysen können Zusammenhänge zwischen verschiedenen Variablen differenzierter untersucht werden als durch einfache Korrelationsanalysen, beispielsweise können auch Interaktionseffekte festgestellt werden (Bobko, 1995). Mit dem Durbin-Watson-Test (Backhaus, Erichson, Plinke & Weiber, 2000) wurde das Vorliegen von Autokorrelationen[142] für alle Regressionen überprüft. Als Faustregel gilt: Liegt der d-Koeffizient des Durbin-Watson-Tests nahe bei 2, so liegt keine Autokorrelation vor, Werte nahe bei 0 weisen auf eine positive, Werte nahe bei 4 auf eine negative Autokorrelation hin.[143] Für alle berichteten Regressionsanalysen konnte eine Autokorrelation ausgeschlossen werden. Weiter wurden für alle Regressoren Toleranzwerte berechnet, um die Multikollinearität abschätzen zu können. Multikollinearität entsteht durch Korrelationen zwischen den Regressoren und mindert die Zuverlässigkeit der Schätzung der Regressionsgewichte. Als Schwellenwert wird .19 festgesetzt, was bedeutet, dass Interkorrelationen von maximal .90 vorliegen.[144] Bei Toleranzproblemen wird dies an entsprechender Stelle vermerkt. Die regressionsanalytischen Ergebnisse werden einer Analyse, die auf der Kategorisierung der Führungskräfte in Unterschätzer, Übereinstimmer und Überschätzer basiert (vgl. S. 248), gegenüber gestellt.

Um eine alleinige Rückführung der prädiktiven Validität auf den Single-Source-Bias auszuschließen, wurden in einem Criss-Cross-Design die Führungsbeurteilungen eines Mitarbeiters an den Erfolgsangaben eines anderen Mitarbeiters der gleichen Führungskraft validiert. An den entsprechenden Stellen finden sich im Ergebnisteil dazu genauere Erläuterungen.

---

[142] Autokorrelation liegt dann vor, wenn die Abweichungen von der Regressionsgerade nicht zufällig sind, sondern zum Beispiel von der Abweichung eines vorangegangenen Beobachtungswertes abhängig sind.

[143] Es exisitieren keine exakten Statistiken, die als Ober- bzw. Untergrenze des Durbin-Watson-Koeffizienten herangezogen werden können.

[144] Häufig werden Regressoren ab einem Toleranzwert von .10 bereits in das Regressionsmodell aufgenommen, was einer Interkorrelation von .95 entspricht. SPSS selbst legt die Toleranzschwelle bei .0001 an, was bedeutet, dass Variablen automatisch aufgenommen werden, wenn sie durch andere Prädiktoren nicht zu über 99.99% erklärt werden kann.

Generell werden Maße der Effektgrößen angegeben, um die praktische Relevanz der Ergebnisse abschätzen zu können. Im Unterschied zu Signifikanztests sind die Effektgrößen unabhängig von der Stichprobengröße und daher für die Interpretation der Ergebnisse vor allem wegen der teilweise geringen und unterschiedlich großen Stichprobenumfänge unumgänglich. Für die verschiedenen statistischen Analysen (Signifikanztests) werden verschiedene Maße herangezogen (Tabelle 27).

Tabelle 27: Übersicht über die verwendeten Maße der Effektstärke (nach Bortz & Döring, 2003; Dunlap et al., 1996)

Test	Effektgröße	kleiner	mittlerer Effekt ab	großer
t-Test für unabhängige Stichproben (t_i: t-Wert für unabhängige Stichproben)[145]	$d = \frac{\mu_A - \mu_B}{\sigma}$ oder $d = \frac{2t_i}{\sqrt{df}}$	.20	.50	.80
t-Test für abhängige Stichproben (t_c: t-Wert für gepaarte Stichproben, r: Korrelation zwischen zu vergleichenden Messwertreihen)	$d = \frac{\mu_A - \mu_B}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$ oder $d = t_c\sqrt{\frac{2(1-r)}{n}}$	.20	.50	.80
Korrelationstest (r)	r	.10	.30	.50
Test für Korrelationsdifferenzen ($r_A - r_B$)	$q = Z_A - Z_B$	.10	.30	.50
Multiple Korrelation (R), z.B. in Regressionsanalysen	$f^2 = \frac{R^2}{1-R^2}$	.02	.15	.35
Einfaktorielle Varianzanalyse	$f = \sqrt{\frac{\eta^2}{1-\eta^2}}$	.10	.25	.40

Anmerkung. Bei q: $Z_A > Z_B$

[145] Bei ungleich großen Stichproben: $d = \frac{\mu_A - \mu_B}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$ (Rosnow & Rosenthal, 1996) bzw. $d = \frac{2t_i}{\sqrt{df}}\sqrt{\frac{\bar{n}}{n_h}}$, wobei $\bar{n} = (n_1 + n_2)/2$ und das harmonische Mittel $n_h = \frac{1}{\frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ (Rosenthal, Rosnow & Rubin, 2000, S. 31).

Prinzipiell lassen sich die Maße der Effektgrößen in zwei Gruppen teilen: Zum einen können sie als standardisierte Differenz von Mittelwerten berechnet werden, zum anderen als Korrelation (Rosnow & Rosenthal, 1996). Da die *t*-Statistik bei gepaarten Stichproben aufgrund der Korrelation zwischen den zu vergleichenden Messwertreihen höher ist als die *t*-Statistik bei unabhängigen Stichproben, kann Cohens *d* bei gepaarten Stichproben nicht unmittelbar aus dem *t*-Wert errechnet werden. Während dieser Unterschied in der Literatur oft nicht gemacht wird, leiten Dunlap, Cortina, Vaslow und Burke (1996) her, dass bei Mittelwertsvergleichen abhängiger Stichproben entweder die ursprünglichen Standardabweichungen, der *t*-Wert für unabhängige Stichproben oder der um die Korrelation zwischen den Messwerten bereinigte *t*-Wert für abhängige Stichproben zur Berechnung von *d* herangezogen werden sollen.

Die Angaben zu Grenzwerten, ab denen ein Effekt als gering, mittel oder hoch interpretiert werden kann, müssen vor dem Hintergrund der jeweiligen Fragestellung auf ihre Gültigkeit überprüft werden. In begründeten Ausnahmen müssen die Grenzen in die eine oder andere Richtung aufgeweicht werden. Nach Cohen (1992) besagt ein mittlerer Effekt, dass er in der Praxis mit bloßem Auge sichtbar ist. Ein starker Effekt liegt deutlich darüber, ein schwacher deutlich darunter. Das bedeutet nicht, dass schwache Effekte keine Relevanz haben. Die Effektgröße lässt sich auch als (Nicht-)Überschneidungsbereich der Verteilungen definieren, die verglichen werden. Eine Effektgröße von 0 bedeutet dann, dass die beiden Verteilungen sich komplett decken. Bei einem starken Effekt von $d = .80$ überschneiden sich die beiden Verteilungen zu 69%.

Alle Analysen mit Ausnahme des Bowker-Tests und der Berechnungen der Effekgrößen wurden mit dem Programmpaket SPSS (Version 10.0) durchgeführt. Die genannten Ausnahmen wurden per Hand berechnet.

5.5 Exkurs: Konstruktion und Überprüfung der Fragebogeninstrumente

In diesem Kapitel werden die einzelnen Konstruktionsschritte des Fragebogens zur Führungsbeschreibung (Kapitel 5.5.1) und der Erfolgsskalen (Kapitel 5.5.3) sowie die Überprüfung der Skalenkennwerte und Gütekriterien aller verwendeten Fragebögen beschrieben (Kapitel 5.5.1.9, 5.5.2 und 5.5.3). Für alle eingesetzten Fragebögen wurden mit Hilfe des Statistik-Programmpakets SPSS für Windows 10.0 Item- und Reliabilitätsanalysen, für die neu konstruierten Fragebögen zusätzlich Faktorenanalysen durchgeführt.

Faktorenanalysen wurden durchgeführt, um die Dimensionalität der konstruierten Skalen zu überprüfen. Zur Extraktion der Faktoren wurde eine Hauptkomponentenanalyse mit anschließender Varimax-Rotation gerechnet. Gemäß Kaiser-Guttman-Kriterium besteht die endgültige Skala nur aus Faktoren mit einem Eigenwert größer 1.0. Bei der vorausgehenden Itemselektion war sogar das strengere Kriterium des Scree-Tests angelegt worden (vgl. Seite 272).

Für die *Itemanalysen* wurde auf den korrigierten Trennschärfekoeffizienten (TS) zurückgegriffen. Im Gegensatz zum Trennschärfekoeffizienten ergibt sich dieser aus der Korrelation des Items mit der Summe der *anderen* Items, nicht mit dem Gesamttestwert. Als Mindestkriterium für die Trennschärfe wurde ein Wert von .30 festgesetzt. Items, deren Trennschärfe geringer ausfiel, wurden für die weitere Auswertung ausgeschlossen. Letztlich liegen also alle verbleibenden Trennschärfekoeffizienten in einem mittleren bis hohen Bereich (Weise, 1975, S. 219). Auf die Berechnung von Schwierigkeitsindizes wurde verzichtet, weil die Iteminhalte normativ vorgegeben sind. In einem solchen Fall ist es im Unterschied zu Schnelligkeits- oder Niveau-Tests durchaus nicht untypisch, dass die Mehrzahl der Beurteiler einem Item zustimmen oder es ablehnen. Dies hat sogar große Aussagekraft für das entsprechende Unternehmen. Schiefe und Exzess werden für den neu konstruierten FzFü angegeben.

Im Rahmen der *Reliabilitätsanalysen* wurden für die Subskalen der verschiedenen Fragebogeninstrumente die internen Konsistenzen in Form des Alpha-Koeffizienten von Cronbach (1951) berechnet. An einer kleinen Stichprobe konnte für eine Auswahl der Instrumente die Retest-Reliabilität (Zwischenintervall 17 Monate) bestimmt werden. Befragt wurden diejenigen Personen, die an einer Interviewstudie beteiligt waren (Senger, in Vorb.) und die noch mit der gleichen Führungskraft zusammenarbeiteten wie zum ersten Befragungszeitpunkt. Für den FzFü wird außerdem die Konstruktvalidität betrachtet (Kapitel 5.5.1.10).

5.5.1 Fragebogen zur Führungsbeschreibung (FzFü)

Die Fragebogenkonstruktion verlief in mehreren Schritten, die in den folgenden Subkapiteln näher ausgeführt werden:

1. Festlegen relevanter Inhalte
2. Generieren des Itempools
3. Bestimmen des Antwortformats
4. Übertragen der Items auf verschiedene Versionen
5. Reduzieren des Itempools
6. Erstellen einer randomisierten Itemreihenfolge (und Umkehrung)
7. Formulieren der Instruktion
8. Datenerhebung und Itemselektion

5.5.1.1 Festlegen relevanter Inhalte

Da sich die inhaltliche Ausarbeitung des Fragebogens an den Führungsanforderungen des Unternehmens orientieren sollte, bot es sich an, als Grundlage die aktuell geltenden Führungsgrundsätze heranzuziehen. Sie sind allesamt in das globalere Führungsmodell des Unternehmens integriert. Dazu Wunderer (2001, S. 349): „Grundsätzlich ist es empfehlenswert, die in den Führungs- und Kooperationsgrundsätzen fixierten Inhalte abzufragen und dafür einen Fragebogen selbst (mit-)zuentwickeln.“ Zur Operationalisierung der Führungsgrundsätze wurden explorative Einzel-Interviews mit drei Führungskräften geführt, die bei der Einführung der Führungsgrundsätze mitgewirkt hatten. Die acht geltenden Führungsgrundsätze wurden bereits in Kapitel 2.3.6.2 vorgestellt (zur Erinnerung Anhang B)

Zusätzliche Materialien, die als Vertiefung einzelner Führungsgrundsätze im Unternehmen entwickelt wurden, wurden ebenfalls berücksichtigt. Da die inhaltliche Ausgestaltung der Führungsgrundsätze und angegliederter Instrumente durchaus mit klassischen Führungsaufgaben und auch mit führungstheoretischen Ansätzen vereinbar ist (vgl. Kapitel 2.6), wurde das insgesamt normative Vorgehen zugunsten der Anwendung einer theoriegeleiteten Fragebogenkonstruktion vorgezogen.

5.5.1.2 Generieren des Itempools

Auf der Basis der genannten Ausgangsmaterialien wurde aus der Aufwärtsperspektive eine hohe Anzahl an Items zu den genannten Führungsgrundsätzen generiert. Zu einem neunten Bereich wurden alle Verhaltensweisen zusammengefasst, die sich auf das Fördern

von Führungsverhalten beziehen. Sie sind in verschiedene Führungsgrundsätze integriert, beispielsweise in das Beurteilen und Steuern von Arbeits- und *Führungsverhalten*. Dieser inhaltliche Bereich musste deswegen vom „Korpus“ des Fragebogens getrennt werden, weil ihn nicht alle Befragten bearbeiten konnten, da die Items Aufgaben betreffen, die nur für Führungskräfte von Mitarbeitern *mit* Führungsverantwortung von Belang sind. Ebenso wurden die Items zur Stellvertretung, welche in einem eigenen Führungsgrundsatz geregelt ist, getrennt vom Hauptteil dargeboten, da zumindest die Inhalte zur aktiven Stellvertretung nicht alle Befragten betreffen.

Generell stellen die Iteminhalte aufgrund verschiedener Vorteile (vgl. Schuler, 1989) möglichst konkrete Führungs*verhaltensweisen* dar, niemals Persönlichkeitseigenschaften. Dies verspricht erstens genauere Ratingantworten. Denn auch wenn sich die entsprechende Moderatorvariable nicht immer zeigt (vgl. S. 143), ist davon auszugehen, dass die Güte von Ratingantworten steigt, je weniger Inferenzen der Beurteiler vornehmen muss. So finden Hoyt und Kerns (1999) in ihrer Meta-Analyse zu Fremdratings, dass explizite Ratings (z.B. Verhaltensratings) deutlich weniger verzerrt werden als Ratings, die vom Beurteiler mehr Inferenzen erfordern. Weiter berichten Binning, Zaba und Whattam (1986) in diesem Zusammenhang das Ergebnis, dass spezifische Items („ratings of specific observable behaviors“, S. 531) im Vergleich zu globalen Items einer bedeutend geringeren Verzerrung durch Leistungscues (z.B. Leistungsinformation zur zu beurteilenden Gruppe) unterliegen. Zweitens steigern Verhaltensratings die praktische Relevanz: Es zeigt sich eine unmittelbarere Passung zu den operationalisierten Anforderungen und es bieten sich konkretere Möglichkeiten, aus der Befragung auch Entwicklungsmaßnahmen abzuleiten. Beides steigert letztlich die Akzeptanz von Beurteilungen.

5.5.1.3 Bestimmen des Antwortformats

Als Antwortformat wurde eine siebenstufige Ratingskala im Likert-Format von 1 (stärkste Ablehnung) bis 7 (volle Zustimmung) festgelegt. Um Effekte der sozialen Erwünschtheit zu reduzieren, wurde dieses Rating einem Format von –3 bis +3 vorgezogen. Da bei Mitarbeiterbefragungen nicht davon auszugehen ist, dass alle Befragten zu allen Iteminhalten bereits abschließende Urteile gefasst oder eindeutige Einstellungen erworben haben (Borg, 2000), wurde ein ungerades Rating mit einer Mittelkategorie gewählt. Auch wenn beispielsweise Voltz (1998, S. 74) gegen eine Mittelkategorie argumentiert, indem er sagt, dass man „gegenüber einem Menschen, mit dem man täglich zu tun hat, nicht meinungslos“ ist, muss erwidert werden: Es reicht nicht aus, eine Meinungsäußerung aus dem Bauch heraus abzugeben; die Einschätzungen sollen auf Beobachtungen beruhen! Die Diskussion über die Wahl des Ratingformats wird ohnehin oft aufgebauscht; große Reichweite

auf praktische Ergebnisse hat sie nicht (London, 1995b, S. 217). Die Formulierungen der sieben Kategorien sind im nachfolgenden Punkt (Tabelle 29) für verschiedene Versionen aufgelistet.

5.5.1.4 Übertragen der Items auf verschiedene Versionen

Die Items der zuerst generierten Aufwärtsversion wurden lediglich in ihrer grammatikalischen Person verändert, um eine Selbst- bzw. Abwärtsversion zu konstruieren. Insgesamt wurden also drei sogenannte „*Perspektiv-Versionen*“ erstellt. Der Item-Beginn variierte innerhalb einer Perspektiv-Verion nicht. Beispielhaft sei auf Tabelle 28 verwiesen.

Tabelle 28. Beispielitem in den drei Perspektiv-Versionen der Ist-Beurteilung

Aufwärtsbeschreibung	Selbstbeschreibung	Abwärtsbeschreibung
Meine direkte Führungskraft ...	Ich ...	Die Führungskraft, die ich beschreibe, ...
bespricht Zielsetzungen mit mir so lange, bis alle Unklarheiten beseitigt sind.	bespreche Zielsetzungen mit meinen Mitarbeitern so lange, bis alle Unklarheiten beseitigt sind.	bespricht Zielsetzungen mit ihren Mitarbeitern so lange, bis alle Unklarheiten beseitigt sind.

Diesem Prinzip der minimalen Änderung wurde Folge geleistet, um die Vergleichbarkeit der Versionen zu gewährleisten.

Daneben wurden Versionen erstellt, mit denen nicht das beobachtete Führungsverhalten (Ist), sondern das ideale Führungsverhalten, die Wichtigkeit der Iteminhalte bzw. das Unternehmens-Soll erfasst werden können. Sie werden als „*Modus-Versionen*“ bezeichnet. Eine Ideal-Version wurde sowohl für die Selbst- als auch die Aufwärtsperspektive erarbeitet, während die Wichtigkeits-Version nur mit der Aufwärtsperspektive, die Soll-Version nur mit der Abwärtsperspektive kombiniert wurde. Um auch hier dem Prinzip der minimalen Änderung treu zu bleiben, wurden neben der Instruktion ausschließlich die Ankerbeschreibungen der Skalen an die entsprechende Version angepasst. Selbst dies war nur für die Wichtigkeitsversion umfassend notwendig (Tabelle 29).

Tabelle 29. Inhaltliche Verankerung der Skalierung von 1 bis 7 für die vier Modus-Versionen

	Ist	Ideal	Wichtigkeit	Soll	
	... so, wie es momentan ist	*... so, wie ich es ideal fände*	*... wie wichtig es ist*	*... was das Leitbild sagt*	
1	trifft überhaupt nicht zu	trifft überhaupt nicht zu	*absolut unwichtig*	trifft überhaupt nicht zu	**1**
2	trifft nicht zu	trifft nicht zu	*ziemlich unwichtig*	trifft nicht zu	**2**
3	trifft eher nicht zu	trifft eher nicht zu	*weniger wichtig*	trifft eher nicht zu	**3**
4	weder noch, unentschieden	weder noch, unentschieden	*weder noch, unentschieden*	weder noch, unentschieden	**4**
5	trifft eher zu	trifft eher zu	*von gewisser Wichtigkeit*	trifft eher zu	**5**
6	trifft zu	trifft zu	*wichtig*	trifft zu	**6**
7	trifft voll und ganz zu	trifft voll und ganz zu	*sehr wichtig*	trifft voll und ganz zu	**7**

Anmerkungen. Kursiv: Änderungen im Vergleich zur Ist-Version, Skalierung von 1 bis 7 linke und rechte Spalte.

Zusammenfassend gibt Tabelle 30 einen Überblick darüber, welche Versionen des FzFü konstruiert wurden.

Tabelle 30. Überblick über die Versionen des Fragebogens zur Führungsbeschreibung

	Ist	**Ideal**	**Wichtig**	**Soll**
Aufwärts	X	X	X	
Selbst	X	X		
Abwärts	X			X

5.5.1.5 Reduzieren des Itempools

Um den Itempool auf eine ökonomisch vertretbare Länge von maximal m = 100 Items zu reduzieren, wurde schrittweise folgendes Vorgehen gewählt:

- Zunächst beurteilten Studierende (N = 30)[146] die Items auf Verständlichkeit (Aufwärtsversion).[147]

[146] Studierende der Organisationspsychologie im Hauptstudium

- Im Unternehmen bewerteten anschließend diejenigen Führungskräfte, mit deren Hilfe die Führungsgrundsätze operationalisiert wurden, die Items ebenfalls nach Verständlichkeit und auch nach ihrer jeweiligen Passung zum Führungsmodell des Unternehmens (Selbstversion). Damit wurde ein erster Konsens über die Auswahl zentraler Inhalte erzielt (kommunikative Validierung).
- In einem Testlauf mit zehn Gruppenleitern aus verschiedenen Unternehmensbereichen konnten Hinweise zur Verständlichkeit der Items sowie zu deren Passung zu *unterschiedlichen* Arbeitsbereichen gewonnen werden (Aufwärtsversion der Ist-, Ideal- und Wichtigkeitsbeurteilung). Hinsichtlich der Verständlichkeit wurden die Teilnehmer nicht nur gefragt, bei welchen Items sie Schwierigkeiten hatten, sondern auszugsweise auch, was sie unter den einzelnen Items verstehen (vgl. Trost, 1997). So stellte sich heraus, dass negativ gepolte Items große Bearbeitungs- und Verständnisschwierigkeiten bei der Wichtigkeitsversion hervorriefen.
- Abschließend wurden die Items von Experten nach den Kriterien der Verständlichkeit und Dopplung von Inhalten bewertet (Expertenvalidierung). Um die Items verständlich zu halten, wurde versucht, auf Negationen zu verzichten, da sie in Verbindung mit der Antwortskala zu doppelter Verneinung führen können.

Als Selektionskriterien können hiermit festgehalten werden:

- Alle Items müssen gute Passung zum Führungsmodell aufweisen.
- Alle Items müssen sowohl von den Führungskräften als auch von den Gruppenleitern und Experten als eindeutig und verständlich bewertet werden.
- Alle Items müssen unternehmensweit für die verschiedenen Bereiche gültig sein.
- Kein Item ist umgepolt. So konnte gewährleistet werden, dass die Wichtigkeitsversion aus den gleichen Items bestehen konnte wie die anderen Versionen.
- Kein Iteminhalt wird prägnanter in einem anderen Item formuliert.

Das Ausschlussverfahren verlief im Sinne einer „Vereinigungsmenge“ der Selektionsannahmen. Keiner der ungünstigen Faktoren durfte weiterhin bestehen. Nach der Elimination der entsprechenden Items ergab sich ein Pool von $m = 98$ Items, wovon jeweils $m = 4$ Items auf Stellvertretung und Fördern von Führungsverhalten entfallen. Da diese Inhalten nicht für alle Befragten Relevanz haben, stellen die übrigen 90 Items den konstanten Hauptteil des Fragebogens dar.

An die erste Version des Fragebogens wurden also keine methodischen Testgütekriterien angelegt. Auf eine methodische Itemselektion musste zu diesem Zeitpunkt verzichtet werden, da die Zielstichprobe für die Erhebung begrenzt war und nicht in größerem

[147] Die Verständlichkeit von Items ist eine grundlegende Bedingung für die Fragebogenbeantwortung. Das Verstehen stellt also eine zentrale kognitive Aufgabe bei der Itembeantwortung dar (Sudman, Bradburn & Schwarz, 1996; Trost, 1997).

Umfang allein für Konstruktionszwecke herangezogen werden konnte. Aufgrund der unternehmensspezifischen Ausrichtung der Fragebogeninhalte konnte auch nicht auf andere Unternehmen ausgewichen werden. Die Führungskräfte und Gruppenleiter, die in die Voruntersuchung einbezogen waren, wurden nicht in die Erhebungsstichprobe aufgenommen. Da die Abwärtsbeurteiler in der Erhebungsstichprobe eine Teilmenge der Selbstbeurteiler darstellen, wurden für die Sichtung dieser Version im Vorfeld keine weiteren Personen engagiert.

5.5.1.6 Erstellen einer randomisierten Itemreihenfolge

Mittels eines Zufallsprogramms wurden die 90 Items des Hauptteils randomisiert. Die Skalen zu Stellvertretung und Fördern von Führungsverhalten wurden ans Ende des Fragebogens gestellt, um sie nur den Befragten vorlegen zu können, für die sie relevant sind.

Der Block aus m = 90 Items wurde zur Überprüfung von Reihenfolgeeffekten bei der Hälfte der Fragebögen in umgekehrter Reihenfolge dargeboten (für eine ausführliche Abhandlung von Reihenfolgeeffekten, siehe Sudman et al., 1996, Kapitel 6).

5.5.1.7 Formulieren der Instruktion

Die Instruktion überschreitet in keiner Fragebogen-Version die Länge einer Seite. Immer werden die Antwortskala erläutert sowie die Vorgehensweise beschrieben. Bei der Aufwärts- und Abwärtsversion war jeweils ein abtrennbarer Abschnitt mit dem Namen der Führungskraft angefügt, die beurteilt werden sollte (vgl. Anhang G). Dies war notwendig, nachdem im Testlauf mit den Gruppenleitern festgestellt wurde, dass die Instruktion, das Führungsverhalten der *direkten Führungskraft* zu beschreiben, aufgrund der teilweise komplexen Führungsstruktur uneindeutig war. So wurde sichergestellt, dass jeweils die formell hierarchisch übergeordnete Führungskraft beurteilt wurde.

In Anlehnung an die Ergebnisse der Meta-Analyse von Mabe und West (1982) wurde sowohl eine Anonymitätsinstruktion gegeben als auch die Validierungserwartung induziert, indem darauf verwiesen wurde, dass die Angaben mit denen anderer verglichen werden. Durch beides kann die Validität der Selbstbeurteilungen erhöht werden (Mabe & West, 1982; vgl. auch Kontrolle sozialer Erwünschtheit, Mummendey, 1995), da die Motivation, sich selbst richtig zu bewerten, zu Lasten der Motivation, die eigene Kompetenz zu erhöhen, gestärkt wird.

5.5.1.8 Itemselektion

Die Konstruktionsstichprobe entspricht der Erhebungsstichprobe, da keine äquivalente Stichprobe bestand, an der der Fragebogen vorab konstruiert werden konnte (siehe oben).

Die erste Hauptkomponentenanalyse mit Varimax-Rotation[148] erbrachte nach 56 Iterationen 14 Faktoren mit Eigenwerten größer 1 (Kaiser-Guttman-Kriterium; Guttman, 1954; Kaiser & Dickman, 1959). Die erklärte Gesamtvarianz betrug 75%. Da das Kaiser-Guttman-Kriterium die Anzahl der bedeutsamen Faktoren in der Regel überschätzt, wurden anhand des Scree-Tests (Cattell, 1966) weitere Items eliminiert. Hierbei werden diejenigen Faktoren aufgrund des Eigenwertediagramms als bedeutsam identifiziert, die über dem „Knick“ („Elbow“) liegen. Dementsprechend wurden alle Items der Faktoren 8 bis 14 ausgeschlossen. Auf den letzten drei Faktoren luden zudem weniger als drei Items. Dieses Procedere wurde einmal wiederholt, so dass bei m = 54 verbleibenden Items eine Fünf-Faktoren-Lösung mit einer Varianzaufklärung von 67% resultierte. Im Scree-Plot ist bei dieser Lösung kein „Elbow“ mehr zu erkennen. Im Anschluss daran wurden schrittweise weitere 12 Items eliminiert, da sie entweder auf einem fremden Faktor eine Ladung größer als a = .5 aufwiesen oder auf zwei fremden Faktoren je höher als a = .35 luden. Dass einzelne Items auf *einem* zusätzlichen Faktor über a = .4 laden wird aus inhaltlichen Gründen in Kauf genommen: Die Führungsverhaltensweisen sind nicht immer ausschließlich einer Führungsaufgabe zuzuordnen.

Bei der Itemselektion, die sich stark an methodischen Kriterien orientierte, wurden inhaltliche Gesichtspunkte dennoch nicht außer Acht gelassen. Der „Geist der Items“ sollte erhalten bleiben. Um nicht schwerpunktmäßig Items aus einzelnen inhaltlichen Bereichen auszuschließen, wurde je ein Item aus Faktor 4 (i42) und 5 (i40) trotz kritischer Ladungen auf fremden Faktoren nicht eliminiert. Die endgültige Version des Fragebogens umfasst m = 42 Items.

5.5.1.9 Skalenanalysen und inhaltliche Interpretation

Nachdem die Itemselektion des Fragebogens zur Führungsbeschreibung abgeschlossen war (vgl. 5.5.1.8), wurde die Endversion der Skala mit m = 42 Items[149] erneut einer Faktoren-, Item- und Reliabilitätsanalyse unterzogen. In Tabelle 31 sind die Skalenkennwerte dargestellt.

[148] Für den Korpus des Fragebogens (m = 90)

[149] Zusätzlich verblieben je vier Items zu den Bereichen „Stellvertretung“ und „Führen von Führungskräften“ in der Skala.

Faktorenanalyse. Die für die Version Ist-Aufwärts durchgeführte Hauptkomponentenanalyse mit anschließender Varimax-Rotation erbrachte nach neun Iterationen fünf Faktoren mit Eigenwerten größer 1.0 und eine Varianzaufklärung von insgesamt 67%. 17% leistet der erste Faktor, der jedoch kaum varianzstärker als der zweite ist. Die Faktorladungen der Items variieren zwischen .44 und .79; vereinzelt treten Ladungen auf fremden Faktoren auf. Die gesamte Faktorladungsmatrix der rotierten Lösung ist mit dem jeweiligen Item-Wortlaut in Anhang D abgebildet.

Tabelle 31. Faktorkennwerte der rotierten Lösung (Fragebogen zur Führungsbeschreibung)

Faktorbezeichnung	Item-Anzahl	*Eigenwert*	*Varianz-aufklärung (%)*	*Kumulierte Varianz-aufklärung*	*Cronbachs Alpha*	*∅TS*
Entwickeln & Fördern	12	6.93	16.50	16.50	.95	.76
Beurteilen & Steuern	10	6.67	15.87	32.37	.93	.73
Individuelle Berücksichtigung	8	5.32	12.67	45.04	.89	.67
Klarheit	7	5.11	12.16	57.20	.92	.76
Partizipation	5	4.00	9.53	66.73	.89	.73

Hervorzuheben ist, dass die drei varianzstärksten Faktoren dieser abschließenden Version bereits nach der ersten Faktorenanalyse mit den ursprünglich 90 Items als erster, zweiter und vierter Faktor extrahiert wurden. Dies deutet darauf hin, dass erstens eine insgesamt stabile Lösung vorliegt und zweitens durch die methodische Itemselektion keine einseitigen inhaltlichen Verzerrungen vorgenommen wurden.

Wenn im Folgenden von „Gesamtskala“ die Rede ist, dann sind darunter die fünf in Tabelle 31 aufgelisteten Faktoren zu verstehen, ohne die beiden einfaktoriellen Dimensionen (vgl. Tabelle 32) mit zu berücksichtigen. Diese werden getrennt betrachtet, um für die entsprechenden Analysen auf den jeweils maximalen Stichprobenumfang zurückgreifen zu können.

Die Hauptkomponentenanalysen der Skalen „Fördern von Führungsverhalten“ und „Stellvertretung“ bestätigten die Annahme einfaktorieller Strukturen bei jeweils hoher Varianzaufklärung (vgl. Tabelle 32).

Tabelle 32. Faktorkennwerte der einfaktoriellen Skalen Fördern von Führungsverhalten und Stellvertretung

Faktorbezeichnung	Item-Anzahl	*Eigenwert*	*Varianz-aufklärung (%)*	*Cronbachs Alpha*	*∅TS*
Fördern von Führungsverhalten	4	2.85	71.16	.86	.72
Stellvertretung	4	2.90	72.56	.86	.73

Im Folgenden wird von *sieben* FzFü-Dimensionen gesprochen und damit die Trennung in ein fünf-faktorielles und zwei einfaktorielle Instrumente, die sich in der statistischen Analyse aufgrund der unterschiedlichen Fallzahlen ergeben hat, als bekannt vorausgesetzt.

Item- und Reliabilitätsanalysen. Die Trennschärfen und die internen Konsistenzen, die durch Item- und Reliabilitätsanalysen ermittelt wurden, erweisen sich durchgängig als gut bis sehr gut – auch bei den Faktoren mit geringer Itemzahl (vgl. Tabelle 31 und Tabelle 32). Cronbachs Alphas bewegen sich in einem Range von $\alpha = .86$ (FvF, S) bis $\alpha = .95$ (E/F) und fallen damit durchgehend sehr hoch aus. Ebenso gut wie die interne Konsistenz sind die Trennschärfen zu bewerten, die auf Itemebene zwischen .59 und .84 (beide FvF) variieren und im Mittel bei .73 liegen. Dass die Skalenkennwerte für die Selbst- und Abwärtsbeurteilung schlechter ausfallen als für die Aufwärtsbeurteilung, kann daran liegen, dass die Selbst- und Abwärtsbeurteiler das Führungsverhalten beschreiben, das mehreren Mitarbeitern gegenüber gezeigt wird, während die Aufwärtsbeurteiler nur das Führungsverhalten beurteilen, das ihnen selbst gegenüber gezeigt wird. Die Kennwerte für die anderen Perspektiv- und Modus-Versionen sowie die jeweiligen *Item*kennwerte (Trennschärfe, Schiefe, Exzess) sind in Anhang E nachzulesen.

Die Retest-Reliabilität für die zweifach erhobenen Items (Items 21/2, 18/8, 15/3, 35/5, 41/5, Item-Nummern vgl. Anhang D) beträgt im Mittel $r = .90$ ($N = 4$). Dies deutet auf eine vergleichsweise hohe Stabilität hin, auch wenn sie aufgrund des geringen Stichprobenumfangs unter großen Vorbehalten zu betrachten ist.[150]

Inhaltliche Interpretation der Faktoren. In Tabelle 33 sind die einzelnen Führungsdimensionen inhaltlich beschrieben. In der rechten Spalte sind jeweils optimale Ausprägungen der entsprechenden Führungsdimension aufgelistet.

[150] Die *Interrater-Reliabilität* ist in den Kapiteln 6.1.2 und 6.3 ausführlich dargestellt.

Tabelle 33. Beschreibung der erfassten Führungsdimensionen

Führungsdimension	Konzeptualisierung (Bedeutung einer hohen Skalenausprägung) Die Führungskraft...
Entwickeln & Fördern (E/F)	▪ bietet ihren Mitarbeitern (MA) Möglichkeiten zu fachlicher Weiterentwicklung ▪ fördert die Selbständigkeit und Lernmotivation ihrer MA ▪ versucht, die individuellen Ziele der MA zu erkennen und dementsprechend zu fördern ▪ fördert die individuellen Stärken ihrer MA ▪ stellt das Fortkommen der MA über die Beibehaltung eines zufriedenstellenden Status quo ▪ bringt die Förderung der MA in Einklang mit den Zielen des Unternehmens
Beurteilen & Steuern (B/S)	▪ beurteilt das arbeitsbezogene Verhalten ihrer MA ▪ beurteilt nach transparenten Kriterien ▪ hat zum Ziel, dass Beurteilungen und Steuerungen nachvollzogen werden können ▪ vermittelt die aus einer Beurteilung folgenden Konsequenzen: neue Ziele, Verhaltensänderungen, Regeländerungen etc. ▪ erklärt notwendige Veränderungen
Individuelle Berücksichtigung (IB)	▪ erkennt individuelle Ressourcen der MA ▪ stellt den Erhalt und die Förderung dieser Ressourcen über ein „Aussaugen“ ▪ gestaltet Ziele und Arbeitsbedingungen ihrer MA unter der Berücksichtigung derer Kenntnisse/ Qualifikation ▪ versucht, für die Beurteilungen ihrer MA vergleichbare Standards heranzuziehen
Klarheit (K)	▪ informiert und korrigiert rechtzeitig ▪ informiert und korrigiert auf eindeutige Weise ▪ gibt das notwendige Maß an Information verständlich weiter ▪ erläutert übergeordnete und arbeitsrelevante Ziele

Partizipation (P)	▪ schafft Rahmenbedingungen, unter denen sich MA einbringen ▪ fordert ihre MA insofern, dass sie ihre Ziele und Leistungen auch selbst reflektieren ▪ bindet MA bei Entscheidungen mit weitreichenden Konsequenzen ein und stellt diese damit auf eine breite Basis ▪ versucht, Akzeptanz für ihre Interventionen zu gewinnen
Fördern von Führungsverhalten (FvF)	▪ führt Führungskräfte auf kooperative Art und Weise ▪ vermittelt, entwickelt und fordert von ihren MA mit Führungsverantwortung, dass sie kooperativ führen ▪ beurteilt MA nach deren Führungsverhalten ▪ fühlt sich für die Förderung des Führungsverhaltens ihrer MA mitverantwortlich
Stellvertretung (S)	▪ bereitet Stellvertreter auf dessen Rolle und Aufgaben vor ▪ überträgt dem Stellvertreter die Verantwortung ▪ delegiert auch Entscheidungsvollmachten

In Tabelle 34 wird ein Vergleich zwischen der Faktorstruktur des FzFü und den Führungsgrundsätzen (Seite 30) gezogen, deren Inhalte der Konstruktion des FzFü zugrunde lagen. Es zeigt sich, dass sich die ersten beiden Faktoren hauptsächlich aus je zwei Führungsgrundsätzen (FG) zusammen setzen (E/F: FG 1/2; B/S: FG 7/8). Dies ist wenig verwunderlich, da auch das Führungsmodell sehr enge Verbindungen zwischen den entsprechenden Führungsanforderungen heraus stellt. Die Items zu den Führungsgrundsätzen 3 (Delegation), 5 (Zusammenarbeit) und 6 (Ziele erarbeiten und setzen) rekategorisieren sich auf den Faktoren 3 bis 5. Während in den Führungsgrundsätzen klare Führungsaufgaben aufgelistet sind, verweisen die Antworten der Befragten darauf, dass sie die Führungsanforderungen eher nach Verhaltensweisen, die zur Erfüllung verschiedenster Aufgaben hilfreich sind, kategorisieren. Insofern legt die Faktorenstruktur auch eine Veränderung der Führungsgrundsätze nahe. Dies wäre insbesondere deshalb auch nützlich, weil sich mit der Arbeitswelt auch die Führungsaufgaben verändern, die Basisverhaltensweisen jedoch auch für neue Aufgaben nützlich erscheinen. Da Fördern von Führungsver-

halten und Stellvertretung einfaktoriell konzipiert wurden und dies bestätigt werden konnte, ergibt sich hierfür keine Änderung.

Tabelle 34. Passung zwischen Führungsgrundsätzen und Führungsfaktoren des FzFü

Führungsgrundsätze VWS	Faktoren des Fragebogens					
	F1	F2	F3	F4	F5	Itemanzahl gesamt
1 (Entwickeln)	3	1	1			5
2 (Fördern)	9					9
3 (Delegieren)		1	3	1		5
4 (Stellvertretung)						zus. Skala
5 (Zusammenarbeiten)				2	2	4
6 (Ziele Setzen)			1	3	2	6
7 (Beurteilen)		5	3		1	9
8 (Steuern)		3		1		4
Itemzahl gesamt	12	10	8	7	5	42

Weiter wurden die einzelnen Führungsdimensionen den klassischen Führungsaufgaben (vgl. Kapitel 2.3) gegenüber gestellt (Tabelle 35), die bereits in Kapitel 2.6 mit Konzepten aus der theoretischen Führungs- und Management-Forschung hinterlegt wurden. Eine solche Gegenüberstellung diente deshalb auch der Entscheidung, bei der Festlegung der Relevanzbereiche des Fragebogens die normativen Vorgaben durch die Führungsgrundsätze nicht um theoretische Konzepte zu erweitern.

Tabelle 35: Gegenüberstellung der FzFü-Dimensionen mit klassischen Führungsaufgaben

Skala	Führungsaufgabe
Entwickeln & Fördern	Personalentwicklung
Beurteilen & Steuern	Leistungsbeurteilung, Kontrollieren
Individuelle Berücksichtigung	Personalentwicklung
Klarheit	Organisieren, Entscheiden, Ziele setzen
Partizipation	Entscheiden, Ziele setzen
Fördern von Führungsverhalten	Personalentwicklung
Stellvertretung	Personalentwicklung, Organisieren

Da alle Führungsgrundsätze – also alle fixierten Führungsleitlinien – operationalisiert wurden und dies im Sinne einer kommunikativen Validierung (Gegenprüfverfahren von Führungskräften aus verschiedenen Unternehmensbereichen) bestätigt wurde, ist von einer hohen Inhaltsvalidität auszugehen.

5.5.1.10 Konstruktvalidität

Als erster Test der Operationalisierung der Führungsgrundsätze kann das allgemeine Rating, inwiefern die Führungskraft ihre Führung an den Führungsgrundsätzen ausrichtet, verstanden werden. Denn eine hohe Korrelation mit dem Fragebogenergebnis insgesamt spräche für einen engen Zusammenhang der beiden Konzepte und damit für eine hohe Konstruktvalidität. Alle Führungsbeurteilungen (Aufwärts) hängen hochsignifikant positiv – durchschnittlich zu $r = .72$ – mit dem allgemeinen Führungsverhalten im Sinne der Führungsgrundsätze zusammen. Am stärksten fällt der Zusammenhang für Klarheit aus ($r = .80$), der sich damit signifikant von den Korrelationen für andere Führungsfaktoren (E/F, B/S, IB) unterscheidet (Hotellings T-Test, einseitig, jeweils $df = 130$, E/F: $t = 2.29$, $p = .001$, B/S: $t = 3.81$, $p < .01$, IB: $t = 3.41$, $p < .001$, K: $t = 1.53$, $p = .06$). Klarheit scheint also ein Konzept zu sein, das stärker als andere mit den Führungsgrundsätzen im Allgemeinen assoziiert wird.

Weiter wurde erwartet, dass alle Faktoren des Fragebogens zur Führungsbeschreibung mit fast allen Faktoren des MLQ positiv korrelieren. Nur für MBP und LF werden negative Korrelationen angenommen. Dies konnte in dieser Form auch bestätigt werden. Genauer wurden für die einzelnen Faktoren vor allem folgende Zusammenhänge angenommen:

1) Entwickeln/Fördern (E/F). E/F sollte insbesondere mit den transformationalen Skalen positiv korrelieren, da diese explizit auf eine Weiterentwicklung der Mitarbeiter abzielen. Da E/F immer hinsichtlich des einzelnen Mitarbeiters zu verstehen ist, wird die höchste MLQ-Korrelation mit IC erwartet. Denn von den transformationalen Subskalen konzentriert sie sich explizit auf den einzelnen Mitarbeiter.

2) Beurteilen/Steuern (B/S). B/S sollte deshalb stark mit CR korrelieren, weil letzteres darauf ausgerichtet ist, gezeigte Verhaltensweisen kontingent zu verstärken. Dies ist jedoch nichts anderes als kontingente Beurteilung, die im Sinne der Lerntheorien eine Steuerung zukünftigen Verhaltens mit sich bringt. Daneben drückt sich MBA im Führungsfaktor B/S stärker aus als in anderen. Denn ein gezieltes Eingreifen in Ausnahmesituationen kommt eher einer steuernden als beispielsweise einer berücksichtigenden oder fördernden Intervention gleich.

3) Individuelle Berücksichtigung (IB). IB findet sein MLQ-Pendant in der transformationalen Subskala IC. Zu dieser sollte also ein besonders starker Zusammenhang bestehen. Da außerdem jede kontingente Verstärkung individuell sein sollte, ist auch eine hohe Korrelation zwischen IB und CR zu erwarten. Zu MBA sollte IB jedoch kaum Verbindung – zumindest geringere als die anderen Führungsfaktoren – zeigen, weil eine Intervention in Ausnahmefällen wenig Möglichkeiten zur individuellen Berücksichtigung lässt. Auch diejenigen Führungskräfte, die warten, bis etwas geschieht, verhalten sich umgekehrt zu IB, so dass eine hochnegative Korrelation zu MBP erwartet wird.

4) Klarheit (K). K sollte insbesondere negativ mit MBP und LF zusammenhängen, denn eine solch passive Führung hinterlässt alles andere als klare Regeln, Ziele oder Informationswege. Viel Klarheit schaffen hingegen kontingente Verhaltenskonsequenzen, so dass die höchste positive Korrelation zu CR erwartet wird. Zwar sollen auch die neuen Ziele klar sein, die durch transformationale Führung vermittelt werden, doch lassen sie mehr Spielraum für Unstimmigkeiten als konkrete Verstärkung von konkretem Verhalten.

5) Partizipation (P). P sollte sich hingegen insbesondere in der transformationalen Führung widerspiegeln, da diese zum Ziel hat, *gemeinsam* mit den Mitarbeitern neue Ebenen zu erreichen. Stark negativ sollte die Partizipation mit LF zusammen hängen, da man an einer Nicht-Führung auch nicht partizipieren kann.

6) Fördern von Führungsverhalten (FvF). Da durch eine Nicht-Führung auch nichts gefördert werden kann, geschweige denn gezielt Führungsverhalten gefördert wird, sollte insbesondere die negative Korrelation zwischen FvF und LF auffallen. Aufgrund der *inhaltlichen* Ausrichtung des Faktors, der Verhaltensweisen verschiedener anderer Führungsfaktoren hinsichtlich des Förderns von Führungsverhalten vereint, sollten keine anderen Korrelationen besonders hervortreten.

7) Stellvertretung (S). Ebenso wie mit FvF verhält es sich mit S.

In Tabelle 36 sind die angesprochenen Interkorrelationen dargestellt, die als Grundlage dafür dienen, anschließend die sieben beschriebenen Annahmen zu bearbeiten.

Tabelle 36: Korrelationen zwischen den Faktoren des FzFü und denen des MLQ

N=91	IIa	IIb	IM	IS	IC	TF gesamt	CR	MBA	MBP	TA gesamt	LF
E/F	.72**	.76**	.67**	.74**	.83**	.81**	.75**	.36**	-.50**	.41**	-.64**
B/S	.66**	.67**	.61**	.67**	.72**	.72**	.74**	.39**	-.43**	.47**	-.65**
IB	.73**	.73**	.67**	.76**	.78**	.80**	.78**	.25*	-.60**	.31**	-.69**
K	.77**	.77**	.75**	.80**	.77**	.84**	.81**	.37**	-.60**	.40**	-.73**
P	.75**	.79**	.67**	.77**	.80**	.82**	.76**	.37**	-.55**	.41**	-.70**
FvF, *N*=50	.64**	.63**	.56**	.68**	.61**	.70**	.68**	.29**	-.50**	.36**	-.64**
S, *N*=17	.39	.50*	.58*	.54*	.65**	.60*	.58*	.23	-.55*	.25	-.51*

Anmerkungen. Es gingen jeweils die individuellen Werte der Aufwärtsbeschreibungen in die Korrelationsanalysen ein. TF gesamt: Mittelwert aller transformationalen Skalen (außer „Ausstrahlung“), grau hinterlegt: besondere Erwartungen. *$p < .05$, **$p < .01$.

zu 1) Entwickeln/Fördern. E/F ist die Skala, die am drittstärksten mit dem Mittelwert der transformationalen Skalen korreliert. Wie erwartet ist der Zusammenhang zu IC signifikant höher als zu anderen transformationalen Subskalen (Hotellings T-Test im Vergleich zur nächsthöchsten transformationalen Skala [IIb]: $t = 2.073$, $p = 0.02$, $df = 88$; einseitig).

zu 2) Beurteilen/Steuern. Auch wenn CR mit vier Führungsfaktoren des FzFü tendenziell höher korreliert als mit B/S, ist es von allen MLQ-Skalen diejenige Dimension, die am höchsten mit B/S korreliert. Der Zusammenhang unterscheidet sich noch nicht signifikant vom nächsthöchsten (B/S mit IC). Der Unterschied zum dritt-höchsten Zusammenhang (B/S und IIb) erreicht fast Signifikanz (Hotellings T-Test: $t = 1.54$, $p = 0.06$, $df = 88$). Im Vergleich zu den anderen sechs Führungsfaktoren des FzFü korreliert B/S tendenziell stärker mit MBA.

zu 3) Individuelle Berücksichtigung. Auch wenn IB höher mit IC als mit anderen MLQ-Faktoren korreliert, unterscheidet sich der Zusammenhang nur von wenigen signifikant. Wie erwartet hängt IB kaum mit MBA zusammen. Vergleicht man IB mit den anderen Führungsfaktoren des Fragebogenkorpus des FzFü (F/E, B/S, K, P), dann ist der Zusammenhang zu MBA bereits deutlich niedriger als der zweitniedrigste Zusammenhang (F/E und MBA) (Hotellings T-Test: $t = 1.60$, $p = 0.06$, $df = 88$). Zu MBP weist IB gemeinsam mit K die höchste negative Korrelation auf und hängt dadurch stärker negativ mit MBP zusammen als F/E, B/S und FvF (jeweils $p < .05$).

zu 4) Klarheit. Für den Zusammenhang von K zu MBP gilt genau das gleiche wie für IB und MBP. Mit CR korreliert Klarheit sowohl im Vergleich zu den anderen Führungsfaktoren des Fragebogens zur Führungsbeschreibung als auch zu den anderen MLQ-Faktoren am höchsten, auch wenn sich die Korrelationen nicht durchgehend signifikant unterscheiden.

zu 5) Partizipation. Partizipation gehört zu den Faktoren, die am stärksten mit den transformationalen Subskalen korrelieren. Dabei scheinen insbesondere Individual Consideration sowie Idealized Influence (behavior) eine große Rolle zu spielen. Denn diese weisen erstens höhere Korrelationen zu P auf als die anderen transformationalen Skalen (nicht signifikant außer zu Korrelation zwischen IM und P, $p < .01$). Zweitens korrelieren IIb und IC zu keinem anderen Faktor aus dem Fragebogen zur Führungsbeschreibung höher als zu P (außer Korrelation zwischen E/F und IC, siehe Punkt 1). Allerdings unterscheiden sich diese Korrelationen in ihrer Höhe jeweils nur zur Korrelation zwischen B/S und IIb bzw. IC signifikant (Hotellings T-Test: B/S-IIb: $t = 1.92$, $p = 0.03$, $df = 88$; B/S-IC: $t = 2.62$, $p < .01$, $df = 88$). Stark negativ fällt der Zusammenhang zu Laissez Faire aus.

zu 6, 7) Fördern von Führungsverhalten, Stellvertretung. Beide Faktoren fallen nicht durch extrem negative Korrelationen zu LF auf. Fördern von Führungsverhalten scheint vielmehr mit Intellektueller Stimulation und Kontingenter Belohnung zusammenzuhängen. Ersteres verwundert nicht, da gerade für potentielle Führungskräfte kreatives Denken aus

verschiedenen Perspektiven wichtig ist. Letzteres überrascht dann nicht, wenn man davon ausgeht, dass Führungskräfte vor allem auf das Verhalten derjenigen achten (und entsprechend darauf reagieren), deren Führungsverhalten sie besonders fördern möchten, als auf das der Mitarbeiter *mit* Führungsverantwortung. Stellvertretung hängt am stärksten mit IC zusammen, was aus der Rolle des Stellvertreters, die die entsprechende Beurteilung ja auch abgaben, durchaus verständlich ist. Weiter sticht die hoch-negative Korrelation zu MBP ins Auge (r = -.55, N = 17). Dies ist damit erklärbar, dass man sich um eine Stellvertretung aktiv kümmern muss und man nicht warten kann, bis Probleme in der Zeit auftreten, die eigentlich klar vertreten werden müsste.

Festzuhalten ist also, dass im Großen und Ganzen die erwarteten Zusammenhänge gefunden wurden, so dass die vom FzFü erfassten Konstrukte insgesamt eher in die transformationale Führung einzuordnen sind. Erstaunlich sind zum einen die teilweise sehr hohen Zusammenhänge. Das deutet darauf hin, dass transformationale Führung von den Beurteilern nicht im originären Sinne verstanden wurde. Zum anderen verwundern die insgesamt sehr ähnlichen Korrelationen der FzFü- mit den MLQ-Faktoren, was für eine MLQ-Faktorisierung in aktive und passive Führung sprechen würde (vgl. Kapitel 2.7.4). Es muss für beide Fragebögen weiter daran gearbeitet werden, Subskalen mit spezifischerem Wert zu konstruieren. Die Kriteriumsvalidität wird in Kapitel 6.4 behandelt.

5.5.2 Multifactor Leadership Questionnaire (MLQ)

Da die Zuordnung der 49 Items zu den 13 Faktoren bereits festgelegt ist, geht es bei der Skalenanalyse des MLQ nicht um eine Konstruktion von Neuskalen. Stattdessen geht es um eine Überprüfung der Skalenkennwerte, auch um gegebenenfalls einzelne Modifikationen vornehmen zu können.

Item- und Reliabilitätsanalysen. In nachfolgender Tabelle 37 sind die internen Konsistenzen sowie die Trennschärfen für die MLQ-Skalen bzw. -Items zusammengestellt. Das Item f(5) aus der Skala Laissez Faire (LF) erreicht den Grenzwert für die Trennschärfe nicht und wird daher – wie übrigens bereits bei den Autoren der deutschsprachigen Version (Felfe & Goihl, 2003) – eliminiert. Nach dessen Elimination erhöht sich Cronbachs Alpha für LF sowie die Trennschärfen der selektierten Items entsprechend der ebenso in Tabelle 37 angegebenen Werte. In folgenden Punkten weicht die eingesetzte Version von der ursprünglichen deutschen Version ab: Die Items f(1) (Contingent Reward) und f(23) (Idealized Influence, behavior) wurden aufgrund ausreichender Kennwerte nicht eliminiert, während zwei weitere Items, die im Handbuch keinem Subfaktor zugeordnet sind, gar nicht aufgenommen wurden.

Tabelle 37. Skalen- und Itemkennwerte des MLQ

Skala	*Item-Nr.*	*TS*	*∅ TS*	*Cronbachs Alpha*	*N*
Idealized Influence, attributed (IIa)	f(10)	.76	.80	.91	85
	f(18)	.72			
	f(21)	.86			
	f(25)	.85			
Idealized Influence, behavior (IIb)	f(6)	.70	.66	.83	86
	f(14)	.69			
	f(23)	.54			
	f(34)	.71			
Inspirational Motivation (IM)	f(9)	.79	.79	.90	86
	f(13)	.79			
	f(26)	.83			
	f(36)	.72			
Intellectual Stimulation (IS)	f(2)	.74	.70	.85	87
	f(8)	.72			
	f(30)	.76			
	f(32)	.58			
Individual Consideration (IC)	f(15)	.71	.76	.89	87
	f(19)	.67			
	f(29)	.83			
	f(31)	.84			
Ausstrahlung (AUS)	f(46)	.67	.76	.89	198
	f(47)	.83			
	f(48)	.77			
	f(49)	.77			
Contingent Reward (CR)	f(1)	.61	.58	.77	87
	f(11)	.51			
	f(16)	.57			
	f(35)	.63			
Management-by-exception (MBA)	f(4)	.39	.43	.65	86
	f(22)	.50			
	f(24)	.38			
	f(27)	.45			
Management-by-exception (MBP)	f(3)	.58	.59	.78	87
	f(12)	.66			
	f(17)	.47			
	f(20)	.65			
Laissez Faire (LF)	f(5)	.18	.38	.57	88
	f(7)	.41			
	f(28)	.38			
	f(33)	.54			
nach Elimination von f(5)	*f(7)*	*.44*	*.50*	*.68*	*88*
	f(28)	*.50*			
	f(33)	*.57*			
Extra Effort (EEF)	f(39)	.71	.70	.84	194
	f(42)	.65			
	f(44)	.75			
Effectiveness (EEF)	f(37)	.70	.72	.87	187
	f(40)	.73			
	f(43)	.67			
	f(45)	.76			
Satisfaction (SAT)	f(38)	.80	.80	.89	197
	f(41)	.80			

Anmerkung. Item-Nr. vgl. Anhang D und E.

Die interne Konsistenz liegt für die MLQ-Skalen in einem Range von α = .65 (MBA) bis α = .91 (IIa), im Mittel bei α = .83. Dies ist mit dem bei Felfe und Goihl (2003) angegebenen Range von .63 bis .92 vergleichbar. Insbesondere unter Berücksichtigung der geringen Itemzahlen pro Subskala können die Reliabilitätsschätzungen als sehr zufriedenstellend bewertet werden. Entsprechende Werte wurden aufgrund der Angaben zur Originalversion auch erwartet (Bass & Avolio, 1997).

Die korrigierten Trennschärfekoeffizienten erreichen im Mittel .68 und variieren zwischen .38 und .86. Auch dies entspricht annähernd den Trennschärfeindizes bei Felfe und Goihl (2003), die einen Range zwischen .27 und .87 berichten, wobei hier wie da lediglich zwei Items der Skala MBA eine Trennschärfe unter .40 aufweisen.

5.5.3 Skalen zur Beurteilung der Erfolgskriterien

Arbeitszufriedenheit. Gemäß der Empfehlung von Rosenstiel (1999a, S. 198) wurden verschiedene Facetten der Arbeitszufriedenheit fokussiert anstatt lediglich die allgemeine Arbeitszufriedenheit zu betrachten. So setzt sich die Skala hauptsächlich aus den Dimensionen zusammen, denen Weinert (1998, S. 224f.) in einem zusammenfassenden Fazit über die Arbeitszufriedenheitsinstrumente zentrale Bedeutung zuspricht. Als „zentrale Dimensionen“ gelten: die Arbeit selbst, Führung, Organisation, Beförderungsmöglichkeiten, Kollegen, Arbeitsbedingungen, Entlohnung, Anerkennung. Aus organisationsspezifischen Gründen wurde zusätzlich die Zufriedenheit mit dem Gesundheitsschutz und mit dem Arbeitsplatz (hinsichtlich Ort und Status) aufgenommen. Abgeschlossen wird die Skala mit einem Item zur allgemeinen Arbeitszufriedenheit, so dass insgesamt m = 11 Items zur Arbeitszufriedenheit vorliegen.

Die entsprechenden Fragebogen-Items beginnen ausnahmslos folgendermaßen: „Im Allgemeinen bin ich mit ...“. An dieser Stelle folgt das Target der Zufriedenheit, bevor sich dann die Antwortskala von 1 (gar nicht zufrieden) bis 7 (voll und ganz zufrieden) unmittelbar anschließt und den Satz beendet. Ein Beispiel ist in Abbildung 23 dargestellt.

Im Allgemeinen bin ich mit ...	gar nicht zufrieden.	nicht zufrieden.	eher nicht zufrieden.	weder zufrieden noch unzufrieden.	eher zufrieden.	zufrieden.	voll und ganz zufrieden.
der finanziellen Entlohnung für meine Leistung ...	1	2	3	4	5	6	7

Abbildung 23. Beispielitem zu Arbeitszufriedenheit mit Skalenformat und -verankerung

Faktorenanalyse. Die Hauptkomponentenanalyse mit anschließender Varimax-Rotation (N = 205) erbrachte zwei Faktoren mit Eigenwerten größer 1. Die Rotation ist in drei Iterationen konvergiert. Bei einer gesamten Varianzaufklärung von 54% verteilen sich die Anteile der beiden Faktoren nahezu genau im Verhältnis 3 zu 2 (vgl. Tabelle 38). Die Faktorladungsmatrix sowie die Trennschärfen der einzelnen Items sind in Anhang H dargestellt.

Tabelle 38. Faktor- und Skalenkennwerte der rotierten Lösung (Arbeitszufriedenheit)

Faktor-bezeichnung	Itemanzahl	Eigenwert	Varianz-aufklärung (%)	Kumulierte Varianz-aufklärung	∅ *TS*	Cronbachs α
Interaktionsfaktor (AZ1)	6	3.17	31.68	31.68	.56	.80
Kontextfaktor (AZ2)	4	2.20	21.96	53.64	.48	.69

Die Bereiche, die in den ersten Faktor eingehen, umfassen die Zufriedenheit mit

- der Arbeitstätigkeit selbst (Inhalte, Variation etc.),
- der direkten Führungskraft,
- der Organisation,
- den Möglichkeiten der Fortbildung, Entwicklung und Förderung,
- dem Verhältnis zu Kollegen im Team/der Abteilung und mit
- der Anerkennung für Leistung.

Inhaltlich wird dieser Faktor als „Interaktionsfaktor“ interpretiert. Denn es sind insbesondere Beziehungen zu anderen Organisationsteilnehmern, die bewertet werden. Auch die Entwicklungsmöglichkeiten oder die Leistungsanerkennung werden zu großen Teilen mit solchen anderen „ausgehandelt“ und sind damit individuell ausgerichtet. Individuell – zumindest weniger allgemeingültig als die Items des zweiten Faktors – ist auch die Arbeitstätigkeit an sich, die auf den ersten Blick möglicherweise „aus der Reihe tanzt“. Doch auch die Arbeitstätigkeit geschieht im Gegensatz zu den Inhalten von AZ2 in Interaktion mit anderen. Im Sinne der Zwei-Faktoren-Theorie von Herzberg, Mausner und Snyderman (1967) wäre es zudem sehr erstaunlich, wenn zwischen Arbeitstätigkeit und dem zweiten Faktor nicht differenziert würde.

Der zweite Faktor ist charakterisiert durch die Zufriedenheit mit

- dem Arbeitsplatz (Ort, Status etc.),
- den Arbeitsbedingungen (Raum, Zeit, Material, Temperatur etc.),
- den Maßnahmen zum Gesundheitsschutz und mit
- der finanziellen Entlohnung für die Leistung.

Diese Zufriedenheitsdimensionen werden als „Kontextfaktor“ zusammenfassend interpretiert. Er kann durchaus in Anlehnung an Herzberg et al. (1967) verstanden werden, obwohl freilich einzuwenden ist, dass zu deren Hygienefaktoren auch interpersonale Beziehungen sowie Unternehmenspolitik zu zählen sind. Letztere ist von der Organisations-Dimension des ersten Faktors aber insofern abzugrenzen, dass unter eben diese auch Bausteine der Organisationskultur oder Commitment subsummiert werden müssen.

Item- und Reliabilitätsanalysen. Die Trennschärfen und internen Konsistenzen sind als befriedigend bis gut zu bezeichnen. Die Retest-Reliabilität, die für AZ1 bestimmt werden konnte, beträgt unter großem Vorbehalt aufgrund des sehr geringen Stichprobenumfangs $r = .87$ ($N = 4$). Für die Arbeitszufriedenheitsskalen kann ein „affektiver Halo-Effekt“, der alle Zufriedenheitsurteile überstrahlt, ausgeschlossen werden, da entgegen der Kriterien von Borg (2003) erstens die allgemeine Arbeitszufriedenheit nicht besser beurteilt wird als die einzelnen Arbeitszufriedenheiten und zweitens keine besonders hohen Item-Interkorrelationen bestehen (AZ1: $\bar{r} = .41$, AZ2: $\bar{r} = .37$).

Wohlbefinden und Leistung. Da die Skalen zu Wohlbefinden und Leistung, die jeweils aus nur zwei Items bestehen, unzufriedenstellende Skalenkennwerte aufweisen (Tabelle 39), sind diese bei der Interpretation der Ergebnisse mit Vorsicht zu behandeln.

Tabelle 39. Faktor- und Skalenwerte für die Skalen zu Wohlbefinden und Leistung

Faktor	Item-Anzahl	Eigen-wert	Varianz-aufklärung (%)	Faktor-ladung	Kommu-nalität	*TS*#	Cron-bachs *α*
Wohlbefinden	2	1.11	55.70	.75	.56	.11	.19
Leistung	2	1.26	63.02	.79	.63	.26	.35

Anmerkung. # Trennschärfe (entspricht der Interkorrelation der beiden Items).

Der Fragebogenteil zu den MLQ-externen Erfolgsskalen ist in Anhang F abgebildet, während die Items MLQ-internen Erfolgsskalen in Anhang C gemeinsam mit den MLQ-Führungsskalen abgedruckt sind.

6 Ergebnisse

Im vorliegenden Ergebnisteil werden die 12 beschriebenen Fragestellungen in je einem eigenen Kapitel dargestellt. Sie werden gemäß Tabelle 16 zu fünf „großen“ Kapiteln mit entsprechenden Subkapiteln zusammengefasst, an deren Beginn jeweils die entsprechenden Hypothesen erneut aufgelistet ist. Eine Zusammenschau am Ende des Ergebnisteils bietet Tabelle 82, die gleichzeitig einen Überblick über die Akzeptanz beziehungsweise Ablehnung der einzelnen Hypothesen gibt. Kapitel 6.5.6 schließt im Sinne einer Diskussion eine zentrale Fragestellung dieser Arbeit ab, nämlich die nach der Rolle der interperspektivischen Diskrepanz im Bedingungsgefüge von Führung und Erfolg. Dieser Abschnitt wird bereits in den Ergebnisteil aufgenommen, um die Fülle an dazu berichteten Befunden sofort integrieren und dadurch eine Antwort auf die prinzipiell zugrunde liegende Fragestellung festhalten zu können. Generell gilt, dass der Ergebnisteil bereits erste Interpretationen enthält, die im strengen Sinne auf die Ergebnisdiskussion vorgreifen. Die Vielzahl der Fragestellungen des bearbeiteten Forschungsgebiets legt es jedoch nahe, durch unmittelbar eingebaute Interpretationen der Ergebnisse die Verständlichkeit der Ergebnisse zu erhöhen und die Lesbarkeit der gesamten Arbeit zu verbessern. In Kapitel 7, der eigentlichen Diskussion, werden nur die zentralen Punkte erneut aufgegriffen und vertieft.

6.1 Deskriptive Statistik und Interrater-Übereinstimmung

In diesem einleitenden Kapitel des Ergebnisteils wird zunächst die deskriptive Statistik der Führungs- und Erfolgsvariablen dargestellt. Sie wird für den FzFü insofern fortgesetzt, dass die Beurteilungen der sieben Führungsdimensionen miteinander verglichen werden (Kapitel 6.1.1). Anschließend wird die Interrater-Übereinstimmung der aufwärtsbeurteilenden Mitarbeiter, die die gleiche Führungskraft beurteilten, dargestellt (Kapitel 6.1.2). Ab Kapitel 6.2 werden die Ergebnisse derjenigen Analysen, mit deren Hilfe die aufgeworfenen Fragestellungen beantwortet werden, dargestellt. Der eilige Leser kann den vorliegenden Abschnitt überspringen.

6.1.1 Deskriptive Statistik

In diesem Abschnitt wird die deskriptive Statistik aller erfassten Variablen dargestellt.

6.1.1.1 Fragebogen zur Führungsbeschreibung

In Tabelle 40 und Tabelle 41 sind die Mittelwerte und Standardabweichung der einzelnen FzFü-Dimensionen sowie deren Interkorrelationen für die drei Perspektiv-Versionen (Ist-Modus) beschrieben. Zur Berechnung der Skaleninterkorrelationen wurden Produkt-Moment-Korrelationen (Pearson) verwendet. Die Mittelwerte und Standardabweichungen der einzelnen *Items* sind in Anhang E aufgelistet. Das gleiche gilt für die drei anderen Modus-Versionen, also die Wichtigkeits-, Ideal- und Soll-Version.

Tabelle 40. Skaleninterkorrelationen, Mittelwerte und Standardabweichungen der FzFü-Subskalen (Aufwärtsbeurteilung)

	E/F	B/S	IB	K	P	FvF	S	*M*	*SD*
Entwickeln & Fördern (E/F)		.84**	.80**	.76**	.77**	.67**	.72**	4.21	1.28
Beurteilen & Steuern (B/S)	.85**		.76**	.75**	.76**	.75**	.66**	4.33	1.26
Individuelle Berücksichtigung (IB)	.79**	.73**		.77**	.75**	.60**	.74**	5.29	1.03
Klarheit (K)	.76**	.76**	.77**		.81**	.68**	.69**	5.03	1.22
Partizipation (P)	.83**	.80**	.74**	.77**		.72**	.75**	4.62	1.30
Fördern von Führungsverhalten (FvF)	.55**	.60**	58**	.46**	.54**		.66**	4.85	1.22
Stellvertretung (S)	.65**	.50**	.67**	.55**	.49**	.41*		5.47	1.17
M	4.30	4.42	5.38	5.05	4.80	4.88	5.56		
SD	1.01	.94	.75	1.01	.99	.86	1.13		

Anmerkungen. Rechts oberhalb der Diagonale: Interkorrelationen, Mittelwert und Standardabweichung für nicht aggregierte Aufwärtsratings (N = 207, FvF: N = 134, S: N = 54), links unterhalb der Diagonale: Interkorrelationen, Mittelwert und Standardabweichung für aggregierte Aufwärtsratings (Mediane der Gruppen von Mitarbeitern, die die gleiche Führungskraft beschreiben) (N = 56, FvF: N = 44, S: N = 36), zur Erinnerung: Skala von 1 bis 7 (vgl. Tabelle 29), ** $p < .01$, * $p < .05$.

Tabelle 41. Skaleninterkorrelationen, Mittelwerte und Standardabweichungen der FzFü-Subskalen (Abwärts- und Selbstbeurteilung)

	E/F	B/S	IB	K	P	FvF	S	*M*	*SD*
Entwickeln & Fördern (E/F)		.81**	.75**	.66**	.73**	.50**	.68**	5.09	.72
Beurteilen & Steuern (B/S)	.59**		.69**	.68**	.61**	.53**	.80**	5.13	.64
Individuelle Berücksichtigung (IB)	.52**	.50**		.83**	.83**	.49**	.57**	5.50	.66
Klarheit (K)	.75**	.52**	.56**		.73**	.37**	.56**	5.56	.73
Partizipation (P)	.47**	.70**	.34*	.53**		.50**	.57**	5.22	.67
Fördern von Führungsverhalten (FvF)	.53**	.56**	.21	.41**	.54**		.60**	4.84	.99
Stellvertretung (S)	.48**	.51**	.43**	.36*	.46**	.54**		5.60	.97
M	5.17	5.08	5.78	5.66	5.40	5.25	5.87		
SD	.67	.80	.45	.53	.57	.92	.79		

Anmerkungen. Rechts oberhalb der Diagonale: Interkorrelationen, Mittelwert und Standardabweichung für Abwärtsbeurteilung, links unterhalb der Diagonale: Interkorrelationen, Mittelwert und Standardabweichung für Selbstbeurteilung je (je N = 48).
** $p < .01$, * $p < .05$.

Es fällt auf, dass alle Mittelwerte über dem Skalenmittel liegen. Der Range ist in Anhang I angegeben. Dazu soll klar festgehalten werden, dass dies nicht bedeutet, dass es sich ausschließlich um „besonders gute Führungskräfte“ handelt. Als Referenzpunkt muss vielmehr der Stichprobenmittelwert herangezogen werden. Dass die Beurteilungen dennoch vergleichweise gut ausfallen, kann an der ostdeutschen Stichprobe liegen. Bungard (1997) berichtet Ergebnisse von einer Ost-West-vergleichenden Mitarbeiterbefragung im Jahre 1996 und stellt bei nahezu allen Items signifikant positivere Bewertungen im ostdeutschen Werk fest.

Die Standardabweichung der Selbstbeschreibung ist erwartungsgemäß geringer als die der Fremdbeschreibungen (Donat, 1991). Für die Varianzeinschränkung in den Selbstbeschreibungen wird deren stärkere Mildetendenz verantwortlich gemacht (Borman, 1993). Dass die Skaleninterkorrelationen für die Selbstbeschreibung geringer ausfällt als für die Fremdbeschreibungen geht mit bisherigen Befunden einher. Selbstbeurteilungen von beruflicher Leistung zeigen einen geringeren Halo-Effekt als Fremdbeurteilungen (Thornton, 1980).

Aufgrund der teilweise sehr hohen Interkorrelationen sind Korrelationstendenzen nicht vollständig auszuschließen. Es kann aber auch von einem "wahren Halo" ausgegangen werden, da für verschiedene Führungsdimensionen ähnliche Voraussetzungen gelten und sie entsprechend zusammenhängen sollten. Zum Beispiel wird eine Führungskraft, die

Klarheit vermittelt, auch auf der Dimension Beurteilung mit klar messbaren Kriterien arbeiten. Auf der ganz praktischen Führungsebene werden die Dimensionen wohl selten vollständig unabhängig voneinander umgesetzt werden können. Dass trotz der hohen Interkorrelationen nicht von einer Einfachstruktur, sondern von der Multidimensionalität auszugehen ist, wird durch Befunde gestützt, die beispielsweise facettenspezifische Zusammenhänge zwischen verschiedenen Beurteilungsrichtungen oder zwischen Führung und Erfolg erbringen.

Im Folgenden werden die Ist-Beurteilungen auf den sieben Führungsdimensionen miteinander verglichen (vgl. Abbildung 24). Ein multipler Mittelwertsvergleich mit Bonferroni-Anpassung zeigt für alle drei Beurteilungsperspektiven bedeutsame Unterschiede zwischen den Dimensionen (Abwärtsperspektive: η^2 = .72, Selbstperspektive: η^2 = .72, Aufwärtsperspektive: η^2 = .75, jeweils $p < .01$). Dass Unterschiede zwischen den Einschätzungen auf den verschiedenen Dimensionen gefunden werden, spricht im Übrigen zusätzlich dafür, die Multidimensionalität trotz hoher Interkorrelationen nicht aufzuheben.

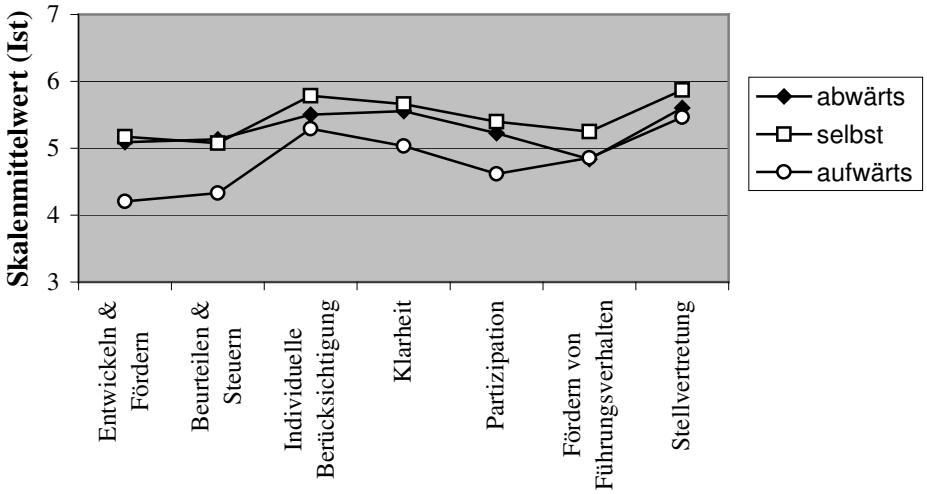


Abbildung 24. FzFü-Beurteilung aus der Abwärts-, Selbst- und Aufwärtsperspektive (jeweils N = 48)

Anmerkungen. Skala von 1 (negativer Pol) bis 7 (positiver Pol), von den 207 Aufwärtsbeurteilungen umfassen nur 48 alle sieben Dimensionen.

Die *Abwärtsbeurteilungen* fallen für die Dimensionen Individuelle Berücksichtigung (IB), Klarheit (K) und Stellvertretung (S) deutlich besser aus als für Entwickeln/Fördern (E/F), Beurteilen/Steuern (B/S), Partizipation (P) und Fördern von Führungsverhalten (FvF) ($p < .05$). Aus der *Selbstbeurteilungsperspektive* werden E/F und B/S ebenfalls bedeutend negativer bewertet als IB, K und S. P und FvF nehmen eine Zwischenstellung ein ($p < .05$). Das gleiche gilt für die *Aufwärtsperspektive* ($p < .05$). Insgesamt ist also für alle drei Beurteilungsperspektiven festzuhalten, dass E/F und B/S deutlich negativer beurteilt werden als IB, K und S.

Ein sehr ähnliches Bild ergibt sich für den interdimensionalen Vergleich der Ideal-Einschätzungen (vgl. Abbildung 25). Ein multipler Mittelwertsvergleich mit Bonferroni-Anpassung zeigt sowohl für die Selbst- als auch die Aufwärtsperspektive hochsignifikante Unterschiede zwischen den Führungsdimensionen (Selbstperspektive: $\eta^2 = .61$, Aufwärtsperspektive: $\eta^2 = .71$). Wie bei der Ist-Beurteilung liegen die Ratings für IB, K und S mindestens tendenziell höher als für die anderen Faktoren. Aus der Aufwärtsperspektive ist für P und FvF wiederum eine Mittelstellung auszumachen, während aus der Selbstperspektive E/F, B/S, P und FvF ähnlich eingestuft werden.

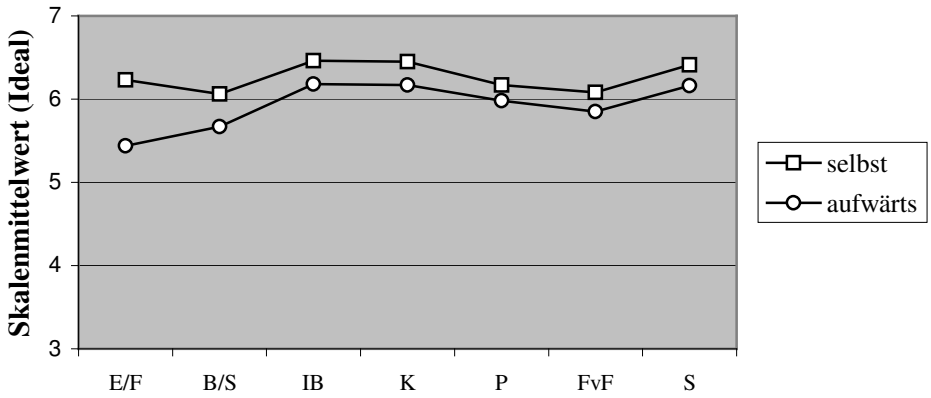


Abbildung 25. FzFü-Ideal-Beurteilung aus der Selbst- ($N = 25$) und Aufwärtsperspektive ($N = 17$)

Anmerkungen. Skala von 1 (negativer Pol) bis 7 (positiver Pol), von den 93 Aufwärtsbeurteilungen umfassen nur 17 alle sieben Dimensionen.

Betrachtet man die Wichtigkeitseinschätzungen, dann wird analog zur Ist- und Ideal-Beantwortung den Dimensionen E/F und B/S weniger Wichtigkeit zugeschrieben als den Dimensionen IB, K und S ($p < .05$), während P und FvF auch hier eine Mittelstellung erhalten (multipler Mittelwertsvergleich: $N = 26$, $p < .01$, $\eta^2 = .59$). Die zunächst als vergleichsweise verbesserungswürdig erscheinenden Dimensionen E/F und B/S (negativste Ist-Bewertung) stellen sich also nicht als solche heraus, da auch die Ideal- und Wichtigkeitseinschätzungen auf diesen Dimensionen geringer ausfallen als in den anderen Führungsbereichen.

6.1.1.2 Multifactor Leadership Questionnaire

Die nachfolgende Tabelle 42 zeigt die Korrelationsmatrix, in der die Interkorrelationen der transformationalen, der transaktionalen, der Laissez Faire- und der Erfolgssubskalen jeweils durch grauen Hintergrund voneinander abgegrenzt sind. Zudem sind die Skalenmittelwerte sowie die mittleren Standardabweichungen angegeben.

Tabelle 42. Skaleninterkorrelationen der MLQ-Subskalen, Mittelwerte, Standardabweichungen und Range

Skala	IIa	IIb	IM	IS	IC	AUS	CR	MBA	MBP	LF	EEF	EFF	SAT
IIa													
IIb	.85**												
IM	.83**	.81**											
IS	.83**	.84**	.81**										
IC	.82**	.83**	.71**	.82**									
AUS	.91**	.81**	.80**	.82**	.77**								
CR	.84**	.81**	.76**	.83**	.79**	.82**							
MBA	.30**	.43**	.34**	.39**	.30**	.33**	.39**						
MBP	-.61**	-.60**	-.63**	-.58**	-.53**	-.57**	-.53**	-.17**					
LF	-.81**	-.77**	-.75**	-.72**	-.69**	-.79**	-.76**	-.46**	.64**				
EEF	.84**	.83**	.78**	.83**	.82**	.78**	.83**	.39**	-.56**	-.75**			
EFF	.89**	.84**	.76**	.83**	.85**	.83**	.87**	.39**	-.55**	-.78**	.76**		
SAT	.86**	.85**	.79**	.80**	.86**	.82**	.83**	.29**	-.60**	-.75**	.68**	.88**	
N	91	91	91	91	91	201	91	91	91	91	201	201	201
M	3.43	3.45	3.55	3.38	3.15	3.32	3.43	3.31	1.93	2.27	3.30	3.43	3.48
SD	.98	.88	.96	.80	1.02	.94	.82	.70	.71	.81	.86	.86	.97
Min	1.00	1.00	1.00	1.50	1.00	1.00	1.25	1.50	1.00	1.00	1.00	1.00	1.00
Max	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	4.00	5.00	5.00	5.00	5.00

Wie auch in anderen Studien zeigt sich eine stärkere Ausprägung der aktiven als der passiven Führung (vgl. z.B. Lowe et al., 1996). Für die passiven Skalen (MBP und LF) liegen die Mittelwerte erwartungsgemäß unter dem Skalenmittel von 3.0, während sie für die aktiven Skalen jeweils leicht über der Skalenmitte liegen. Im Unterschied zu Felfe und Goihl (2003) liegt in der vorliegenden Studie der Mittelwert für Ausstrahlung höher als das Skalenmittel. Im Vergleich zu amerikanischen Stichproben liegt wie schon bei Felfe und Goihl (2003) der Skalenmittelwert für Management-by-exception (active) verhältnismäßig hoch. Das könnte in der vorliegenden Studie damit begründet werden, dass bei Großkonzernen eher auf den Ausnahmezufall gewartet werden kann als beispielsweise beim Flugverkehr oder im Militär. Dieses „Warten-Können“ gilt auch für die Untersuchungsstichprobe von Felfe und Goihl (2003), die öffentliche Verwaltung. Für alle Führungsskalen mit Ausnahme von MBP sind die Skalenmittelwerte der vorliegenden Stichprobe jedoch tendenziell höher als die der Manual-Stichprobe von Bass und Avolio (1997). Ob tatsächlich eine aktivere Führung besteht oder ob kulturabhängiges Antwortverhalten vorliegt, kann nicht beantwortet werden. Analog gilt dies für die Erfolgsskalen, deren Mittelwerte in der amerikanischen Stichprobe jeweils etwas höher ausfallen als in der vorliegenden Stichprobe.

Die Interkorrelationen zwischen den Führungsfaktoren sind fast ausschließlich sehr hoch (Range von $r = .17$ bis $r = .91$). Erwartungsgemäß zeigen sich hohe positive Interkorrelationen zwischen den transformationalen Subskalen. Der enge Zusammenhang, den die Subskala Ausstrahlung zu den transformationalen Skalen aufweist, unterstützt deren Konzeptualisierung als transformationale Führung (Felfe & Goihl, 2003). Aufgrund unspezifischer Zusammenhänge zu einzelnen Skalen wird in der Ergebnisdarstellung darauf verzichtet, sie getrennt zu betrachten. Weiter korreliert die transaktionale Subskala Contingent Reward mit den transformationalen Skalen durchgängig sehr hoch. Deutlich geringer, dennoch aber im mittleren Bereich, korreliert Management-by-exception (active) mit den transformationalen Skalen sowie mit Contingent Reward. Dieses Resultat bestätigt die theoretische Vorannahme, dass die aktiven Führungsweisen miteinander stärker als mit passiven (MBP und LF) korrelieren. Außerdem verwundert auch der deutliche Zusammenhang zwischen den aktiven transaktionalen und den transformationalen Skalen nicht, da sie auch im Führungsmodell nicht losgelöst voneinander konzeptualisiert sind.

Durchgängig negativ gestalten sich die Zusammenhänge zwischen Management-by-exception (passive) sowie Laissez Faire auf der einen Seite und allen anderen Subskalen auf der anderen Seite. Zueinander stehen die beiden passiven Dimensionen in einem positiven Zusammenhang. Während sie jedoch hochsignifikant negativ mit den drei Erfolgsskalen korrelieren, stehen alle andere transaktionalen und transformationalen Subskalen mit den Erfolgsskalen in hochsignifikant positivem Zusammenhang. Auch wenn der Zusammenhang zwischen Management-by-exception (active) und den Erfolgsskalen hochsignifikant ist, fällt er gleichzeitig deutlich geringer aus als für die anderen Skalen (Hotellings T-Test: $t_{[88]} = 5.21$, $p < .01$).[151] Insgesamt stimmen die Ergebnisse mit den Effektivitätsannahmen des „Full Range of Leadership"-Modells überein (vgl. Bass & Avolio, 1997).

Alle Befunde gehen konform mit bisherigen Untersuchungsergebnissen (Bass, 1985; Bass & Avolio, 1997, 2000; Den Hartog, Van Muijen & Koopman, 1997; Felfe & Goihl, 2003; Geyer & Steyrer, 1998; Lowe et al., 1996).

6.1.1.3 MLQ-externe Erfolgsvariablen

Für die MLQ-externen Erfolgsvariablen, die Arbeitszufriedenheit, das Wohlbefinden und die Leistung, sind die Interkorrelationen sowie die Mittelwerte, die Standardabweichungen und der Range in Tabelle 43 angegeben.

[151] Zum Zusammenhang zwischen Führung und Erfolg siehe Ergebnisdarstellung in Kapitel 6.4.

Tabelle 43. Skaleninterkorrelationen der MLQ-externen Erfolgsvariablen, Mittelwerte, Standardabweichungen und Range (***N*** = 205)

	AZ2	WOBE	L	*M*	*SD*	*Min*	*Max*
Arbeitszufriedenheit (Interaktionsfaktor, AZ1)	.45**	.41**	.46**	5.25	.90	2.33	7.00
Arbeitszufriedenheit (Kontextfaktor, AZ2)		.45**	.10	5.40	.87	2.75	7.00
Wohlbefinden (WOBE)			.25**	4.72	.87	2.00	7.00
Leistung (L)				5.33	1.02	2.50	7.00

Es fällt auf, dass die Skalenmittelwerte mittelmäßig miteinander korrelieren. Das deutet darauf hin, dass sie deutlich weniger Überlappungen aufweisen als beispielsweise die MLQ-internen Erfolgsvariablen untereinander. Ferner wird deutlich, dass die Skalenmittelwerte allesamt in der positiven Skalenhälfte liegen.

Die deskriptive Statistik der verschiedenen Differenzmaße ist in Anhang J dargestellt.

6.1.2 Zur Aggregation der Mitarbeiter-Ratings: Ihre Übereinstimmung

Zur Beantwortung verschiedener Fragestellungen ist es sinnvoll und anschaulich, auf die Analyse der individuellen Mitarbeiter-Ratings zu verzichten und stattdessen die Mitarbeiter-Ratings einer Arbeitsgruppe zusammenzufassen. Solche aggregierten Ratings können die Reliabilität und folglich die Validität der Urteile gegenüber Einzelratings erhöhen (z.B. Conway & Huffcutt, 1997; Hoyt & Kerns, 1999; Scullen, Mount & Goff, 2000; Wirtz & Caspar, 2002). Unter einer „Arbeitsgruppe“ werden diejenigen Mitarbeiter zusammengefasst, die die gleiche (ihre) Führungskraft beurteilen. Zur Rechtfertigung dieser Datenaggregation wurde die Interrater-Reliabilität innerhalb jeder Gruppe bestimmt.

Die Reliabilität der Beurteilungen wurde folglich für die Mitarbeitergruppen derjenigen 50 Führungskräfte berechnet, die von mehr als einem Mitarbeiter beurteilt wurden.[152] Dafür wurde das *zweifaktorielle* Modell der *justierten* Intraklassen-Korrelation (ICC) gewählt, in dem die Rater als *fixed-Faktor* zu betrachten sind (Shrout & Fleiss, 1979). Da es für Fragestellungen im vorliegenden Themenbereich immer angestrebt wird, Mitarbeiterbeurteilungen von mehr als einem Mitarbeiter zu erhalten, ist nicht die Reliabilität des Urteils *eines* beliebigen Aufwärtsbeurteilers (Mitarbeiters) von Interesse, sondern die Reliabilität der Mittelwerte der Mitarbeiter. Die Intraklassen-Korrelation entspricht in diesem Fall dem Alpha-Koeffizienten der internen Konsistenz und berücksichtigt lediglich die korrelative Übereinstimmung, nicht die Niveau-Unterschiede zwischen den Ratern. Sie kann nicht als

[152] Nur E/F, B/S, IB, K und P wurden von mehr als einem Beurteiler für 50 Führungskräfte beurteilt, FvF für 20 und S für zwei. Der unterschiedliche Stichprobenumfang liegt in der Beurteilungsdimension begründet. FvF wurde nur von den Mitarbeitern beurteilt, die selbst Führungsverantwortung tragen, S nur von denen, die ihre Vorgesetzten bei dessen Abwesenheit vertreten – und in aller Regel gibt es lediglich einen Stellvertreter.

Schätzung allgemeiner Rating-Reliabilitäten interpretiert werden, sondern als Schätzung der Reliabilitäten in der vorliegenden Stichprobe. Die Reliabilität der durchschnittlich vier Rater wurde für alle Führungsdimensionen berechnet ($ICC_{just,MW}[1,4]$). Sie liegt im Mittel bei .46.[153] Für Entwickeln/Fördern, Beurteilen/Steuern, Individuelle Berücksichtigung, Klarheit, Partizipation, Fördern von Führungsverhalten bzw. Stellvertretung liegt sie bei .50, .38, .44, .45, .55, .44 bzw. .30.

Legt man das Reliabilitätskriterium (r = .70) von Nunnally (1978) an, dann ist die Interrater-Reliabilität auf allen Führungsdimensionen relativ niedrig. Allerdings sollte entsprechend einer Empfehlung von Wirtz und Caspar (2002) das Reliabilitätskriterium von Nunnally nicht apodiktisch verwendet werden. Verschiedene Aspekte sprechen dafür, dass die ermittelten Koeffizienten als befriedigend akzeptiert werden können:

- *Erstens* kann die geringe Anzahl an Aufwärtsbeurteilern mitverantwortlich für die relativ geringen Koeffizienten sein. Entsprechend dieser Annahme hängt insbesondere die ICC der Dimension B/S, also die niedrigste ICC der vorliegenden Studie, positiv mit der Anzahl der Aufwärtsbeurteilungen zusammen (r = .37, p = .01, N = 48). Conway und Huffcutt (1997) berechneten beispielsweise, dass durchschnittlich sechs Aufwärtsbeurteiler aggregiert werden müssen, um eine Reliabilität von .70 zu erreichen. Die Autoren schätzten in ihrer Meta-Analyse die Interrater-Reliabilität von Aufwärtsbeurteilern im Übrigen lediglich auf .30 (berufliche Leistungsbeurteilung).
- *Zweitens* ist nicht davon auszugehen, dass eine Führungskraft alle Mitarbeiter auf die gleiche Art und Weise führt. Nach Konzepten der dyadischen Führung ist eine individuelle Ausrichtung der Führung sogar die erfolgsversprechende. So beobachten unterschiedliche Mitarbeiter verschiedene (Ausprägungen von) Führungsverhaltensweisen, was zwangsläufig die Beurteilerübereinstimmung gegenüber Beurteilungen eines stabilen Beurteilungsgegenstands mindert.
- *Drittens* haben die Aufwärtsbeurteiler der vorliegenden Stichprobe kaum Erfahrung mit Beurteilungen, da erstmalig Aufwärtsbeurteilungen vorgenommen wurden. Ohne Übung respektive Training kann demnach keine hohe Reliabilität erwartet werden.

Vor diesem Hintergrund werden die vorliegenden Interrater-Reliabilitäten als hoch genug interpretiert, um für entsprechende Fragestellungen die Aufwärtsratings, die die glei-

[153] Für einige Führungskräfte ergibt sich auf einzelnen Dimensionen ein negativer Koeffizient, welcher aufgrund von Korrekturfaktoren oft kleiner als -1.00 ausfällt. Da diese eine Mittelung stark verzerren würden, wurden sie nicht berücksichtigt.
Da die negativen Koeffizienten in einem *unjustierten* Modell weniger stark ausfallen, ist von einer besseren Übereinstimmung hinsichtlich des Niveaus auszugehen ist. Da jedoch auch dann keine hohen Koeffizienten erreicht werden, müssen neben der verhältnismäßig geringen Konsistenz auch Niveau-Unterschiede zwischen den Mitarbeitern einer Führungskraft für die relativ geringe Interrater-Reliabilität mitverantwortlich gemacht werden.

che Führungskraft betreffen, aggregieren zu können. Allerdings wird die Reliabilität methodisch nicht als hoch genug interpretiert, um ganz auf Analysen mit individiuellen Werten zu verzichten. Dies wird an den entsprechenden Stellen kenntlich gemacht. Manche Hypothesen zielen zudem explizit auf einen Vergleich zwischen individuellen und aggregierten Analysen ab.

Lediglich die Ratings der Aufwärtsperspektive – nicht die Ratings der Aufwärts- *und* Abwärtsperspektive – wurden aggregiert, auch wenn Hedge et al. (2001) resümieren, dass es psychometrisch keine Anhaltspunkte für eine Differenzierung der Perspektiven gäbe. Versteht sich dieses Forschungsgebiet jedoch als Bindeglied zur angewandten Organisationspsychologie, dann ist eine Trennung der verschiedenen Perspektiven allein aus Gründen der Akzeptanz von Fremdbeurteilungen zweifellos anzuraten (vgl. Kapitel 3.6). Die Übereinstimmung der Beurteilungen aus *verschiedenen Perspektiven* steht im Mittelpunkt von Kapitel 6.3.

6.2 Beobachtetes versus erwartetes Führungsverhalten

Die in diesem Abschnitt berichteten Ergebnisse nehmen Bezug auf Fragestellung 1, also auf folgende Hypothesen:

1) **Die Bewertung des aktuellen Führungsverhaltens (Ist) fällt negativer aus als die Einschätzung des erwarteten Führungsverhaltens.**
2) **Mitarbeiter nehmen eine größere Ist-Ideal-Diskrepanz wahr als Führungskräfte.**
3) **Die Ideal-Einschätzung der Führungskräfte (Selbst-Ideal) stimmt stärker mit dem Unternehmens-Soll überein, wenn diese die Führungsgrundsätze kennen (vs. nicht/kaum kennen).**
4) **Je höher die Hierarchieebene der fokussierten Führungskraft ist, umso besser stimmen Selbst-Ideal und Soll überein.**

Die Angaben von vier der fünf leitenden Führungskräfte, die die Soll-Version bearbeiteten, stimmen mit denen der Geschäftsführung überein, während die Antworten der fünften Führungskraft weit niedrigere Skalenausprägungen aufweisen (Abbildung 26). Diese werden im Folgenden nicht weiter beachtet. Die Angaben der vier konvergierenden Führungskräfte werden mit den Geschäftsführungs-Angaben zu einem Mittel zusammengefasst, welches im Folgenden als „Soll“ bezeichnet wird. Diese Vorgehensweise ist insbesondere inhaltlich begründet, denn jedes Unternehmen strebt *ein* Unternehmensleitbild an, das in der Regel top-down implementiert wird. Zumindest wird die Modellfunktion höherer Ebenen regelmäßig herausgestellt (Liebel & Oechsler, 1994; Schuhmacher, 1988; Kiessling & Spannagl, 1996; Weiss, 1977; Weiss, 1978). Wenn also einzelne Führungskräfte ein anderes Führungsleitbild vor Augen haben als die Geschäftsführung und die Mehrzahl ihrer Kollegen, dann spricht dies eher dafür, dass das Leitbild bis dato nicht deutlich genug kommuniziert wurde oder auf Misskredit stößt und nicht akzeptiert wird. Der Praxis sollten derartige Ergebnisse Anlass dazu geben, den Ursachen sowie potentiellen Konsequenzen nachzugehen und gegebenenfalls entsprechend zu intervenieren.

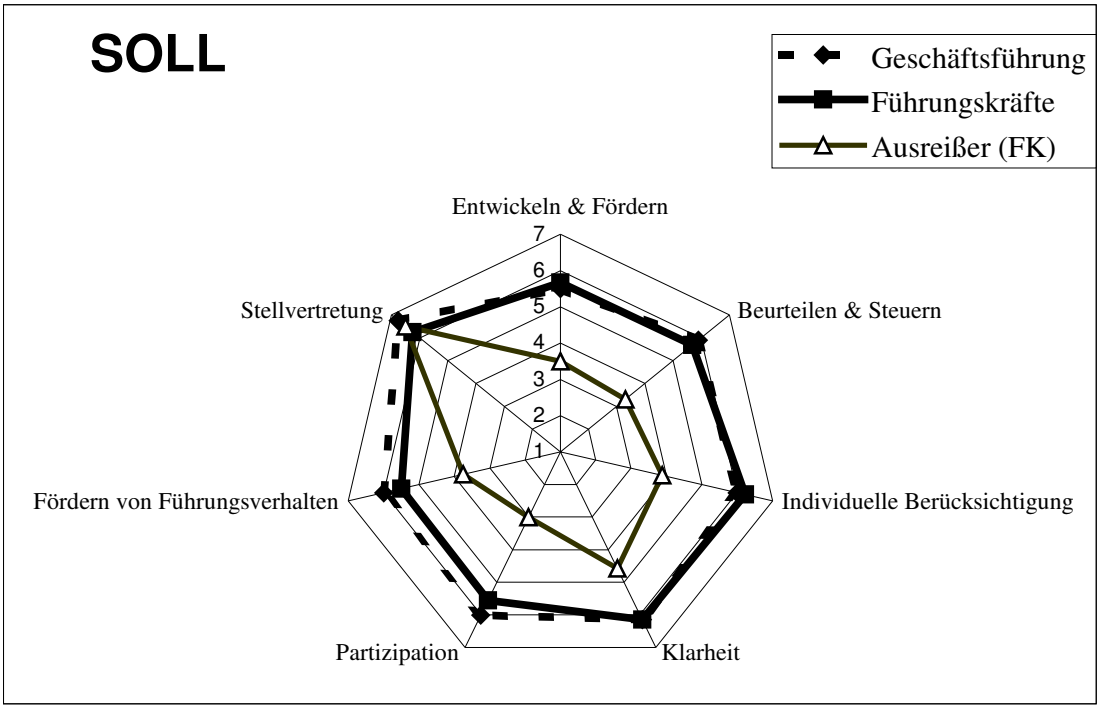


Abbildung 26. Soll-Beschreibungen auf den sieben FzFü-Dimensionen

Anmerkung. Skala von 1 bis 7.

Für alle erfassten Führungsdimensionen liegt der angestrebte Soll-Wert über den durchschnittlichen Selbst-, Aufwärts- und Abwärtsbeschreibungen (Ist-Version). Es ergeben sich durchgehend hochsignifikante Unterschiede (Tabelle 44).

Tabelle 44. Vergleich zwischen aktuellem und erwartetem Soll-Führungsverhalten: Mittelwerte, Standardabweichung und Signifikanzprüfung

FzFü-Faktor	Auf-Ist (1)	Ab-Ist (2)	S-Ist (3)	Soll (4)	*t* (1,4)	*df*	*t* (2,4)	*df*	*t* (3,4)	*p*[a]	*df*
E/F	4.21 (1.28)	5.09 (.72)	5.17 (.67)	5.63 (.36)	-15.98	206	-5.16	47	-4.78	<.001	47
B/S	4.33 (1.26)	5.13 (.64)	5.08 (.80)	5.72 (.28)	-15.86	206	-6.32	47	-5.60	<.001	47
IB	5.29 (1.03)	5.50 (.66)	5.78 (.45)	6.18 (.39)	-12.35	206	-7.01	47	-6.03	<.001	47
K	5.03 (1.22)	5.56 (.73)	5.66 (.53)	6.14 (.42)	-13.02	206	-5.55	47	-6.24	<.001	47
P	4.62 (1.30)	5.22 (.67)	5.40 (.57)	5.64 (.43)	-11.34	206	-4.30	47	-2.97	.005	47
FvF	4.85 (1.22)	4.84 (.99)	5.25 (.92)	5.60 (.65)	-7.06	133	-5.31	47	-2.66	.011	47
S	5.47 (1.17)	5.60 (.97)	5.87 (.79)	6.35 (.38)	-5.56	53	-5.30	47	-4.21	<.001	47

Anmerkungen. T-Test für gepaarte Stichproben, siebenstufige Skala, vor Klammer Faktor-Mittelwert, in Klammer Standardabweichung, [a]*p* nur für (3,4) angegeben, da für (1,4) und (2,4) durchgängig gilt: $p < .001$, für (1,4) ergeben sich auch hochsignifikante Ergebnisse für aggregierte Aufwärtsratings ($N = 40$).

Keine der drei Beurteilungsgruppen geht also davon aus, dass der Soll-Wert (das Unternehmensleitbild) bereits erreicht ist. Sogar die Selbstbeschreibung der Führungskräfte, die im Vergleich zur Aufwärtsbeschreibung ihrer Mitarbeiter überhöht ist (siehe unten, z.B. Tabelle 51), liegt unter dem Soll. Die Effektgrößen sind fast durchgehend hoch. Lediglich

auf den Dimensionen P und FvF liegen jeweils zwei Effekte „nur“ im mittleren Bereich. Grob kann jedoch festgehalten werden, dass Aufwärts- sowie Abwärtsbeurteilungen etwa eine Standardabweichung unter dem Soll zu verorten sind ($\bar{d}$ = 1.17 bzw. 1.01), während die Selbstbeurteilungen zwischen einer halben und einer Standardabweichung vom Soll abweichen ($\bar{d}$ = .80). Hypothese 1) ist für die Soll-Erwartungen also anzunehmen.

Das Ausmaß der Ist-Soll-Differenz unterscheidet sich in der Selbstbeschreibung nicht zwischen den sieben Dimensionen. Das bedeutet, dass Führungskräfte für keine Führungsdimension eine im Vergleich zu anderen Dimensionen stärkere Abweichung wahrnehmen. Vorausgesetzt, dass sie das Soll anstreben, entsteht aus der wahrgenommenen Ist-Soll-Diskrepanz also keine besondere Änderungsmotivation für bestimmte Dimensionen. Betrachtet man die Aufwärtsbeurteilung, so stellt man in einem multivariaten Vergleich (mit Bonferroni-Anpassung) der ersten fünf Faktoren (E/F, B/S, IB, K, P) fest, dass IB hochsignifikant geringer von Soll abweicht als E/F, B/S und K. Weiter liegen K und P hochsignifikant weniger unter dem Soll als E/F und B/S. Der Differenzwert zwischen Aufwärts-Ist und Soll unterscheidet sich nicht zwischen E/F und B/S bzw. zwischen K und P (N = 207). Bezieht man FvF und S in den multivariaten Vergleich ein (N = 48), dann ergibt sich außerdem, dass FvF den Soll-Wert hochsignifikant stärker erreicht als B/S, ebenso P hochsignifikant stärker als S. Vor dem Hintergrund, dass das Aufwärts-Ideal kaum vom Soll abweicht (vgl. S. 302), folgt daraus, dass die Dimensionen E/F und B/S aus Mitarbeitersicht nicht nur im Sinne des Leitbilds verbesserungswürdig sind, sondern auch im Sinne ihrer eigenen Führungsvorstellungen. Mit Blick auf die Abwärtsbeurteilung ergibt sich eine signifikant höhere Abweichung vom Soll-Wert auf IB im Vergleich zu P (multivariater Vergleich, p = .001, N = 48).

Zieht man als erwartetes Führungsverhalten nicht das Soll, sondern die Idealeinschätzungen heran, dann zeigt Tabelle 45, dass sowohl die Aufwärts-Ist-Beurteilung als auch die Selbst-Ist-Beurteilung auf allen sieben Führungsdimensionen hochsignifikant unter dem Aufwärts-Ideal bzw. dem Selbst-Ideal liegen. Die Effektgrößen variieren zwischen d = .72 und d = 1.45 (Aufwärts, $\bar{d}$ = 1.18) bzw. zwischen d = .71 und d = 1.92 (Selbst, $\bar{d}$ = 1.28).

Sowohl aus der Selbst- als auch aus der Aufwärtsperspektive ist die Ist-Ideal-Diskrepanz für E/F am größten und signifikant größer als für IB und K (multivariater Mittelwertsvergleich, p < .05). Für die Aufwärtsbeurteiler ist das Ideal für IB im Vergleich zu allen anderen Faktoren signifikant besser erreicht. Sowohl aus dem Ist-Soll- als auch dem Ist-Ideal-Vergleich leitet sich folglich für E/F ein besonderer Handlungsbedarf ab.

Unter der Annahme, dass Menschen ihre eigenen Idealvorstellungen anstreben, ist auf der *individuellen Ebene* die Differenz zwischen der Ist-Beurteilung und der *entsprechenden* Ideal-Einschätzung als Indiz dafür zu interpretieren, inwiefern die eigenen Ziele/Ansprüche bereits erreicht sind. Der Befund gewinnt zusätzlich dadurch an Relevanz, dass die Ist-Ideal-Diskrepanzen, die die *Aufwärtsbeurteiler* wahrnehmen, durchschnittlich

36% der Varianz der Erfolgskriterien aufklären (lineare Regression auf die Ist-Ideal-Diskrepanzen von E/F, B/S, IB, K und P, $N = 90$). Während sich für AZ2 und WOBE nur kleine Effekte ($f^2 > .02$) und für EF-AZ sowie EF-WOBE mittlere Effekte ($f^2 > .15$) ergeben, zeigen sich für alle anderen Erfolgskriterien sogar große Effekte ($f^2 > .35$).[154] Für die Ist-Soll-Diskrepanz gilt ähnliches.

Inwieweit sich die Führungskraft selbst von ihren Idealen entfernt sieht (*Selbstbeurteilung*), wirkt sich weniger deutlich auf die Erfolgskennwerte ihrer Mitarbeiter aus ($N = 21$), was aufgrund ihrer höheren Korrelation auch plausibel ist. Starken (erwartungsgemäß negativen) Einfluss nimmt diese Differenz auf das Wohlbefinden der Mitarbeiter ($f^2 > .63$) und ihre Bereitschaft zu zusätzlicher Anstrengung ($f^2 > .87$).

Tabelle 45. Vergleich von aktuellem und idealem Führungsverhalten (Selbst und Aufwärts): Mittelwerte, Standardabweichung und Signifikanzprüfung

FzFü-Faktor	Auf-Ist (1)	S-Ist (2)	Auf-Ideal (3)	S-Ideal (4)	*t* (1,3)	*p*	*df*	*d*	*t* (2,4)[a]	*p*	*d*
E/F	4.11 (1.36)	5.24 (.54)	5.70 (.74)	6.23 (.49)	-11.13	<.001	92	1.45	-12.22	<.001	1.92
B/S	4.26 (1.28)	5.14 (.76)	5.67 (.71)	6.06 (.70)	-11.35	<.001	92	1.36	-10.78	<.001	1.26
IB	5.18 (1.13)	5.80 (.47)	6.13 (.51)	6.46 (.41)	-8.14	<.001	92	1.08	-10.14	<.001	1.50
K	4.85 (1.32)	5.73 (.53)	6.15 (.44)	6.45 (.43)	-10.01	<.001	92	1.32	-9.89	<.001	1.49
P	4.52 (1.34)	5.38 (.61)	5.84 (.53)	6.17 (.55)	-10.26	<.001	92	1.30	-9.00	<.001	1.36
FvF	4.82 (1.00)	5.32 (.99)	5.78 (.88)	6.08 (1.08)	-7.14	<.001	51	1.02	-7.51	<.001	.73
S	5.42 (1.27)	5.85 (.80)	6.19 (.83)	6.41 (.78)	-3.00	.007	20	.72	-5.55	<.001	.71

Anmerkungen. T-Test für gepaarte Stichproben, siebenstufige Skala, [a]für alle Führungsdimensionen $df = 24$, vor Klammer Faktor-Mittelwert, in Klammer Standardabweichung.

Die Selbst-Ist-Beurteilung liegt aber nicht nur hochsignifikant unter dem Selbst-Ideal, sondern auch unter dem Aufwärts-Ideal. Der Effekt zeigt sich hochsignifikant mit starken Unterschieds-Effekten von $\bar{d} > .80$ für E/F ($t = -5.05$), B/S ($t = -3.84$), IB ($t = -4.27$), K ($t = -4.86$) und P ($t = -3.41$), jeweils $df = 35$. Für FvF und S unterscheidet sich die Selbst-Ist-Beurteilung nicht von dem Aufwärts-Ideal.

Ebenso liegt die Aufwärts-Ist-Beurteilung nicht nur hochsignifikant unter dem Aufwärts-Ideal, sondern auch in starkem Ausmaß ($\bar{d} > .80$) für E/F ($t = -8.56$), B/S ($t = -6.25$), IB ($t = -6.26$), K ($t = -8.67$), P ($t = -5.92$) bei $df = 20$ sowie für FvF ($t = -3.92$) bei $df = 15$ auch hochsignifikant unter dem Selbst-Ideal. S verfehlt das Signifikanzniveau ($t_{[14]} = -1.94$, $p = .07$).

Die Abwärtsbeurteilung (Ist) liegt in fünf der sieben Führungsdimensionen mindestens eine Standardabweichung unter dem Selbst-Ideal bzw. unter dem Aufwärts-Ideal (vgl. Tabelle 46). Auf einer *organisationsbezogenen Ebene* weisen die Ergebnisse somit darauf hin, dass keine der Beurteilergruppen – also keine der Gruppen, die eine Linienorganisa-

[154] Aus der Selbstperspektive erbringen die Regressionsanalysen lediglich für EF-AZ, WOBE, EF-WOBE und EEF mittlere Effekte, für die anderen Erfolgskriterien allenfalls geringe ($N = 21$).

tion charakterisieren – das Ideal annähernd erreicht sieht. Hypothese 1) kann folglich nicht nur für die Soll-Erwartungen, sondern auch für die Ideal-Erwartungen angenommen werden.

Tabelle 46. Vergleich von aktuellem (Abwärts) und idealem (Selbst und Aufwärts) Führungsverhalten: Mittelwerte, Standardabweichung und Unterschiedstestung

FzFü-Faktor	Ab-Ist (1)	Auf_{Med}-Ideal (2)	Ab-Ist (3)	S-Ideal (4)	*t* (1,2)	*p*	*df*	*d*	*t* (3,4)	*p*	*df*	*d*
E/F	5.16 (.68)	5.72 (.36)	5.20 (.72)	6.22 (.47)	-3.08	.005	22	1.03	-5.58	<.001	16	1.68
B/S	5.25 (.58)	5.57 (.64)	5.21 (.59)	6.08 (.62)	-2.00	.058	22	.52	-4.38	<.001	16	1.44
IB	5.64 (.57)	6.12 (.29)	5.65 (.53)	6.50 (.45)	-3.32	.003	22	1.06	-5.41	<.001	16	1.73
K	5.70 (.58)	6.19 (.31)	5.73 (.57)	6.50 (.43)	-3.80	.001	22	1.05	-4.97	<.001	16	1.53
P	5.29 (.57)	5.83 (.46)	5.36 (.52)	6.24 (.53)	-3.43	.002	22	1.04	-5.38	<.001	16	1.68
FvF	4.95 (.58)	5.69 (.79)	5.01 (.90)	6.01 (1.22)	-2.33	.042	10	1.07	-2.62	.019	15	.93
S	5.78 (.94)	6.23 (.49)	5.67 (1.07)	6.33 (.90)	-1.13	.296	7	.60	-2.05	.058	15	.67

Anmerkungen. T-Test für gepaarte Stichproben, siebenstufige Skala, vor Klammer Faktor-Mittelwert, in Klammer Standardabweichung, (1) unterscheidet sich deshalb minimal von (3), weil die Stichprobe für die Gegenüberstellung mit Auf- bzw. Selbst-Ideal leicht verändert ist (Passung der Perspektiven!).

Bevor die Selbst- und Aufwärtsperspektive hinsichtlich ihrer Ist-Ideal-Abweichungen verglichen werden, lohnt es, die Ideal-Einschätzungen der beiden Perspektiven einander gegenüberzustellen. Dabei zeigen sich für E/F ($t_{[116]} = 3.36$, $p < .01$), B/S ($t_{[116]} = 2.40$, $p = .02$), IB ($t_{[116]} = 3.03$, $p < .01$), K ($t_{[116]} = 3.00$, $p < .01$) und P ($t_{[116]} = 2.73$, $p < .01$) signifikante bis hochsignifikante Unterschiede. Das Selbst-Ideal übertrifft jeweils das Aufwärts-Ideal (vgl. auch Tabelle 45). Für FvF und S finden sich keine Unterschiede.

Zu beachten gilt, dass der Gruppe der Aufwärtsbeurteiler sowohl Mitarbeiter ohne als auch Mitarbeiter mit Führungsverantwortung (also Führungskräfte) angehören. Vernachlässigt man die Perspektive der Beurteilung und überprüft man, ob von Führungskräften (erste bis vierte Hierarchie-Ebene) und Mitarbeitern unterschiedlich hohe Ideal-Einschätzungen abgegeben werden, dann bestätigt sich die Tendenz des Aufwärts-Selbst-Vergleichs, auch wenn sich die Effekte von $\overline{d} = .46$ auf $\overline{d} = .33$ abschwächen. Insbesondere für E/F ($t_{[116]} = -2.58$, $p = .01$), IB ($t_{[116]} = -3.01$, $p < .01$) und P ($t_{[116]} = -2.00$, $p < .05$) ist das Ideal der Führungskräfte höher als das der Mitarbeiter. Die Ideal-Vorstellungen derjenigen Führungskräfte, die ihr Ideal aus der Aufwärtsperspektive abgeben, unterscheiden sich hingegen nicht von denen der Mitarbeiter.

Von praktischer Relevanz ist weniger, inwiefern sich Führungskräfte und Mitarbeiter in ihren Idealvorstellungen unterscheiden als vielmehr, ob und inwiefern sich Führungskräfte von *ihren* Mitarbeitern unterscheiden. In entsprechenden *t*-Tests für gepaarte Stichproben zeigen sich nur geringe (nicht signifikante) Unterschiede zwischen dem Selbst-Ideal und dem entsprechenden (aggregierten) Aufwärts-Ideal ($N = 21$). Die Effektgrößen variieren zwischen $d = .22$ (FvF, S) und $d = .56$ (E/F). Im Durchschnitt liegen die Ideale der Führungskräfte eine Drittel Standardabweichung höher als die ihrer Mitarbeiter ($\overline{d} = .32$).

Unter Kontrolle der Perspektive zeigt sich, dass die *Aufwärts*-Ideale der Führungskräfte den *Aufwärts*-Idealen *ihrer* Mitarbeiter tendenziell überlegen sind. Nur für IB ist der Unterschied statistisch bedeutsam (Wilcoxon-Vorzeichenrangtest: $Z = -2.04$, $p = .04$, $N = 9$). Unten wird zu prüfen sein, ob die Selbst-Aufwärts-Diskrepanz der Ideal-Einschätzungen moderierenden Einfluss auf die Selbst-Aufwärts-Übereinstimmung der Ist-Beurteilung nimmt (vgl. Kapitel 6.3.3).

Auf allen sieben Führungsdimensionen nehmen Aufwärtsbeurteiler prinzipiell eine höhere Ist-Ideal-Diskrepanz wahr als Selbstbeurteiler. Für E/F ($t_{[116]} = -.37$, $p < .01$), B/S ($t_{[116]} = -3.26$, $p < .01$), IB ($t_{[116]} = -2.08$, $p = .04$), K ($t_{[116]} = -3.89$, $p < .01$) und P ($t_{[116]} = -3.44$, $p < .01$) zeigen sich statistisch signifikante Unterschiede, für FvF ($t_{[75]} = -1.15$, $p = .26$) und S ($t_{[44]} = -.81$, $p < .42$) ergeben sich keine Unterschiede. Um auch hier ansatzweise zu kontrollieren, ob der Unterschied durch die Perspektive der Beurteilung verzerrt ist, wurde überprüft, ob die Ist-Ideal-Diskrepanz auch für aufwärtsbeurteilende Mitarbeiter größer ist als für aufwärtsbeurteilende Führungskräfte. Entsprechend der Hypothese 2) zeigen sich auch hier für die Mehrzahl der Führungsdimensionen insofern bedeutsame Unterschiede zwischen den Perspektiven, als die Mitarbeiter ohne Führungsverantwortung höhere Ist-Ideal-Diskrepanzen angeben. Sie erreichen für E/F ($t_{[91]} = 2.44$, $p = .02$), B/S ($t_{[91]} = 3.39$, $p < .01$), K ($t_{[91]} = 4.06$, $p < .01$) und P ($t_{[91]} = 2.25$, $p = .03$) Signifikanz.[155] Die Effektgrößen variieren zwischen $d = .06$ (FvF) und $d = .76$ (K) und erreichen durchschnittlich eine mittlere Stärke von $\bar{d} = .51$. Dieser Effekt gewinnt zudem an Gewicht, wenn man bedenkt, dass die Selbst-Ideale höher liegen als die Aufwärts-Ideale, letztere aber dennoch eine höhere Diskrepanz zum wahrgenommenen Führungsverhalten (Ist) aufweisen. Hypothese 2) kann somit für die vorliegende Stichprobe als bestätigt gelten. Mitarbeiter (ohne Führungsverantwortungen) geben stärkere Ist-Ideal-Diskrepanzen an als Führungskräfte (Selbstbeurteiler und Aufwärtsbeurteiler mit Führungsverantwortung).

Die größere Diskrepanz, die Mitarbeiter ohne Führungsverantwortung in ihrem Arbeitsalltag zwischen dem beobachteten und erwünschten Führungsverhalten im Vergleich zu Mitarbeitern mit Führungsverantwortung wahrnehmen, kann sich in geringerer Zufriedenheit ersterer auswirken (AZ1: $t_{[200]} = 2.15$, $p = .03$, $d = .30$, AZ2: $t_{[200]} = -.55$, $p = .58$, $d = -.08$, SAT: $t_{[113]} = 2.14$, $p = .03$, $d = .40$).

Denkbar ist, dass sich Idealvorstellungen stärker an die Soll-Erwartungen anpassen, je besser Führungsleitlinien kommuniziert werden. Erst dann besteht überhaupt die Möglichkeit, diese auch zu internalisieren. Will man dem entsprechend überprüfen, ob die Ideal-Soll-Differenz von der Vertrautheit mit den Führungsgrundsätzen moderiert wird, lohnt es, zunächst die Differenz selbst zu betrachten (vgl. Tabelle 47).

[155] Ergebnisse der anderen Führungsdimensionen: IB ($t_{[91]} = 1.23$, $p = .23$), FvF ($t_{[50]} = .19$, $p = .85$), S ($t_{[19]} = 1.02$, $p = .32$)

Tabelle 47. Mittelwertsvergleich zwischen Ideal (Aufwärts, Selbst) und Soll

FzFü-Faktor	S-Ideal (1)	Auf-Ideal (2)	Soll (3)	*t* (1,3)	*p*	*df*	*d*	*t* (2,3)	*p*	*df*	*d*
E/F	6.24 (.47)	5.70 (.74)	5.63 (.36)	-6.36	<.001	24	1.46	-.95	.344	92	.12
B/S	6.06 (.70)	5.67 (.71)	5.72 (.28)	-2.38	.026	24	.64	.66	.513	92	-.09
IB	6.46 (.41)	6.13 (.51)	6.18 (.39)	-3.42	.002	24	.70	.96	.342	92	-.11
K	6.45 (.43)	6.15 (.44)	6.14 (.42)	-3.56	.002	24	.74	-.26	.793	92	.02
P	6.17 (.55)	5.84 (.53)	5.64 (.43)	-4.78	<.001	24	1.07	-3.61	<.001	92	.41
FvF	6.08 (1.08)	5.80 (.88)	5.60 (.65)	-2.24	.035	24	.54	-1.66	.103	52	.26
S	6.41 (.78)	6.19 (.83)	6.35 (.38)	-.37	.718	24	.10	.88	.388	20	-.25

Anmerkung. T-Test für gepaarte Stichproben.

Für die Selbstbeurteiler ergibt sich für alle Führungsdimensionen mit Ausnahme von S, dass ihr Ideal deutlich höher ist als das entsprechende Soll ($\bar{d}$ = .86). Geht man davon aus, dass selbstgesetzte Ziele eine stärker motivierende Wirkung haben als fremde Ziele, zumal wenn die selbstgesetzten schwieriger sind (Locke & Latham, 1990), dann sollten die Führungskräfte eher durch ihre eigenen Ideal-Vorstellungen als durch extern vorgegebene Unternehmenserwartungen (Soll) motiviert sein. Es bestehen also durchaus hohe eigene Zielvorstellungen, die die Führungskräfte dazu anhalten sollten, ihr Führungsverhalten in die gewünschte Richtung (Soll) zu verändern.

Für die Aufwärtsbeurteiler zeigt sich lediglich auf der Dimension P ein deutlicher Unterschied zwischen dem Aufwärts-Ideal und dem Soll ($t_{[92]}$ = -3.61, p < .01, d = .41). Das Ideal geht auch hier über das Soll hinaus. Da das Selbst-Ideal über dem Aufwärts-Ideal liegt, ist die Ähnlichkeit von Aufwärts-Ideal und Soll nicht auf einen Deckeneffekt zurückzuführen. Der Vergleich zwischen dem Aufwärts-Ideal und dem Soll besitzt hohe Praxisrelevanz, denn wenn die Ideale der Mitarbeiter das Soll übersteigen, kommt es trotz Soll-Erfüllung durch die Führungskräfte zu Enttäuschungen auf Seiten der Mitarbeiter. Als Schlussfolgerung für die Praxis muss man an dieser Stelle Gespräche über die Ziele von Führung und über entsprechende Interventionsmöglichkeiten zur Anpassung der Erwartungshaltungen empfohlen werden.

Die Idealeinschätzung nähert sich dem Sollwert nicht in Abhängigkeit davon an, ob die Beurteiler mit den Führungsgrundsätzen vertraut sind. Die Differenz zwischen dem Selbst-Ideal und dem Sollwert ist auf der Dimension P sogar für diejenigen Führungskräfte stärker ausgeprägt, die mit den Führungsgrundsätzen vertraut sind (versus nicht/kaum vertraut) ($t_{[21]}$ = -2.72, p = .01, d = 1.88). Für alle anderen Faktoren ergibt sich kein Effekt. Es bleibt also festzuhalten, dass die Ideal-Einschätzung der Führungskräfte (Selbst-Ideal) nicht stärker mit dem Unternehmens-Soll übereinstimmt, wenn diese die Führungsgrundsätze kennen (vs. nicht/kaum kennen). Hypothese 3) muss also abgelehnt werden. Dies spricht dafür, dass andere Instrumente der Kultur-Implementierung (neben den Führungsgrundsätzen) die Führungswerte bereits transportiert haben. Jedoch ist festzuhalten, dass lediglich

drei Führungskräfte angeben, mit den Führungskräften nicht oder nur teilweise vertraut zu sein. Dies schränkt die Aussagekraft stark ein.

Mittels einer einfaktoriellen ANOVA mit anschließendem Tukey-HSD-Test bei homogenen Varianzen wurde überprüft, ob in Abhängigkeit der Hierarchie-Ebene Unterschiede in der Ideal-Soll-Übereinstimmung bestehen. Die Ergebnisse der ANOVA machen deutlich, dass sich bei keiner Führungsdimension ein bedeutsamer Unterschiedseffekt ergibt (E/F: $F_{[2,22]} = .28$, $p < .76$; B/S: $F_{[2,22]} = .65$, $p < .54$; IB: $F_{[2,22]} = .84$, $p < .44$; K: $F_{[2,22]} = .40$, $p < .67$; P: $F_{[2,22]} = .17$, $p < .84$; FvF: $F_{[2,22]} = .34$, $p < .72$; S: $F_{[2,22]} = .23$, $p < .80$). Hypothese 4) wird demzufolge abgelehnt. Das Ergebnis verweist darauf, dass die Führungswerte bereits über viele Ebenen hinweg implementiert sind. Alternativ kann interpretiert werden, dass das Führungsleitbild so gut zur Unternehmenskultur passt, dass sich entsprechende Ideale bereits hierarchieübergreifend gebildet haben, bevor die Führungsgrundsätze implementiert wurden.

Zusammenfassend kann aus der Gegenüberstellung zwischen beobachtetem und erwartetem Führungsverhalten festgehalten werden:

- Sowohl die Selbst- als auch die Fremdbeurteiler (Abwärts, Aufwärts) sehen das Soll, wie es in den Führungsgrundsätzen verankert ist, nicht erfüllt (circa 1 SD Abweichung).
- Sowohl die Selbst- als auch die Fremdbeurteiler (Abwärts, Aufwärts) sehen das Ideal (Aufwärts, Selbst) nicht erfüllt.
- Für Entwickeln/Fördern zeichnet sich sowohl im Vergleich zum Soll als auch im Vergleich zum Ideal die größte Diskrepanz und damit der größte Handlungsbedarf ab.
- Die Aufwärtsbeurteiler nehmen eine stärkere Diskrepanz zwischen beobachtetem und erwartetem (idealem) Führungsverhalten wahr als die Selbstbeurteiler. Unter den Aufwärtsbeurteilern sind es die Mitarbeiter ohne Führungsverantwortung (vs. mit Führungsverantwortung), die eine höhere Ist-Ideal-Diskrepanz angeben.
- Unter Vorbehalten (aufgrund des geringen Stichprobenumfangs) gilt, dass die Kenntnis der Führungsgrundsätze die Differenz zwischen Selbst-Ideal und Soll nicht moderiert.
- Die Hierarchieebene der fokussierten Führungskraft moderiert die Differenz zwischen Selbst-Ideal und Soll nicht.

6.3 Übereinstimmung zwischen verschiedenen Beurteilungsrichtungen

Gemeinhin werden unter „Übereinstimmung“ absolute Übereinstimmungen der Ratingantworten verstanden. Im Themenbereich der vorliegenden Studie geben korrelative Zusammenhänge zwischen verschiedenen Beurteilungsperspektiven auch Hinweise auf die Validität der Beurteilungen. In dieser Arbeit werden daher sowohl Niveau-Unterschiede als auch korrelative Zusammenhänge unter „Übereinstimmung“ subsumiert.
Da eine beträchtliche Meinungsverschiedenheit über die Güte unterschiedlicher Kongruenzmaße besteht (vgl. Cheung, 1999; Edwards, 1994a, 1994b; Johns, 1981; Tisak & Smith, 1994), wurden in der vorliegenden Studie verschiedene Maße verwendet. Nachdem in Kapitel 6.3.1 die Pearson-Korrelation zwischen Selbst- und Fremdratings und in Kapitel 6.3.2 die Niveau-Unterschiede (Differenzscore) betrachtet werden, werden sie in Kapitel 6.3.4 einander gegenübergestellt und in Kapitel 6.5 auf ihre Kriteriumsvalidität überprüft.

6.3.1 Korrelative Zusammenhänge zwischen Selbst-, Aufwärts- und Abwärtsbeurteilungen

Die Korrelation zwischen Selbstbeurteilungen und Fremdbeurteilungen wird häufig als Validierung für die Selbstbeurteilung herangezogen (z.B. Conway & Huffcutt, 1997; vgl. auch Moser, 1999). Da aber auch Fremdbeurteilungen verzerrt sein können, dürfen sie als Validierungskriterium nicht überinterpretiert werden. Daher werden in Kapitel 6.4 Erfolgskriterien auch noch zur Validierung der Beurteilungen herangezogen.

Die Hypothesen 5) bis 9) fokussieren auf den korrelativen Zusammenhang zwischen den verschiedenen Beurteilungsperspektiven der Ist-Version (Fragestellung 2). Dabei interessiert nicht nur die bloße Stärke des Zusammenhangs (Hypothese 6), sondern auch dessen Abhängigkeit von der Aggregation der Aufwärtsratings (Hypothese 5), von den einbezogenen Perspektiven (Hypothesen 7) und 8) und Führungsdimensionen (Hypothese 9). In Kapitel 7.2.3 wird die Höhe der Zusammenhänge mit denen verglichen, die aus anderen Studien an anderen Stichproben bekannt sind.

5) Wenn man die Aufwärts-Beurteilungen, die die gleiche Führungskraft betreffen, aggregiert (versus nicht aggregiert), korrelieren sie stärker mit Beurteilungen aus anderen Perspektiven (Selbst, Abwärts).

6) Die Beurteilungen verschiedener Perspektiven (Selbst-Ist, Aufwärts-Ist, Abwärts-Ist) korrelieren gering ($.10 \leq r < .30$).

7) Die Korrelationen zwischen Selbst- und Fremdbeurteilungen sind geringer als die Korrelationen zwischen den Fremdbeurteilungen.
8) Die Korrelationen zwischen Selbst- und Aufwärtsbeurteilungen unterscheiden sich nicht von den Korrelationen zwischen Selbst- und Abwärtsbeurteilungen.
9) Die Korrelationen zwischen den Perspektiven variieren in ihrer Stärke über die Dimensionen hinweg.

Vergleicht man die Selbst-Aufwärts- sowie die Abwärts-Aufwärts-Ratings für aggregierte vs. individuelle Aufwärtsratings,[156] stellt man fest, dass die Korrelationen, in die die aggregierten Aufwärtsratings eingehen, zwar fast durchgängig augenscheinlich höher sind, allerdings nur vereinzelt auch bedeutsame Effekte auftreten (vgl. Abbildung 27, Tabelle 48). Vor allem für die Abwärts-Aufwärts-Korrelation ergeben sich zumindest für drei Dimensionen geringe Effekte. Dass der Median für FvF und S den individuellen Ratings weniger stark überlegen ist, ist deshalb plausibel, weil er jeweils auf der Basis weniger Einzelratings gebildet wurde.[157] Die Reliabilität aggregierter Maße profitiert aber ja gerade von der Anzahl der Aufwärtsbeurteiler.

Tabelle 48. Gegenüberstellung der Selbst-Aufwärts- bzw. Abwärts-Aufwärts-Korrelationen für aggregierte vs. individuelle Aufwärtsratings

	Selbst-Aufwärts-Korrelation				Abwärts-Aufwärts-Korrelation			
	aggregiert (1)	individuell (2)	*N*	*q (1,2)*	aggregiert (3)	individuell (4)	*N*	*q (3,4)*
E/F	.15	(.20) **	40 (142)	-.05	.39*	(.25)**	32 (118)	**.16**
B/S	.06	(.04)	40 (142)	.02	.27	(.19)*	32 (118)	.09
IB	.08	(.11)	40 (142)	-.03	.43*	(.34)**	32 (118)	**.11**
K	.36*	(.22)**	40 (142)	**.15**	.11	(.14)	32 (118)	-.03
P	.27	(-.22)**	40 (142)	.05	.27	(.19)*	32 (118)	.09
FvF	.07	(-.01)	30 (79)	.08	.02	(.01)	24 (63)	.01
S	-.10	(-.01)	26 (36)	.09	.45’	(.36)	19 (26)	**.10**

Anmerkungen. $^{**}p < .01$, $^{*}p < .05$, $'p < .10$, fett: bedeutende Effektgrößen.

Dass eher die Fremd-Fremd- als die Selbst-Fremd-Korrelationen durch die Aggregation an Reliabilität gewinnt, geht stimmig mit der Annahme einher, dass die mangelnde korrelative Übereinstimmung eher auf eine ungenaue Selbst- als Fremdbeurteilung zurückgeht. Dazu passt auch, dass die Aggregation nur auf dem Faktor Klarheit, der offenbar auch von Selbstbeurteilern relativ gut einschätzbar ist, einen bedeutenden Effekt erzielt, also die

[156] Zur Erinnerung: „Aggregiert“ bedeutet, dass die Aufwärtsbeurteilungen, die dieselbe Führungskraft betreffen, zu ihrem Median zusammengefasst wurden.
[157] Zur Erinnerung: Die Items der beiden Faktoren wurden nur von denjenigen Mitarbeitern beantwortet, die entweder selbst Führungsverantwortung tragen (FvF) oder ihre Führungskraft bei Abwesenheit vertreten (S).

aggregierten Ratings deutlich stärker mit der Selbstbeurteilung korrelieren als die individuellen Ratings. Wenn die Selbstbeurteilung auf den anderen Faktoren aber ungenau ist, dann kann mit dieser auch keine reliablere (aggregierte) Aufwärtsbeurteilung korrelieren. Für die Abwärts-Aufwärts-Korrelation bringt die Aggregation der Aufwärtsratings vor allem für die Dimensionen Klarheit und Fördern von Führungsverhalten keinen Vorteil. Dies wird im Einklang mit oben genannten Hypothesen so erklärt, dass Klarheit von den Abwärtsbeurteilern, Fördern von Führungsverhalten von den Aufwärtsbeurteilern sehr ungenau beurteilt wird. Dafür spricht auch, dass die Interrater-Reliabilität der Aufwärtsbeurteiler für FvF sehr gering ausfällt. 8 der 20 Intraklassen-Koeffizienten fallen negativ aus (für 20 Führungskräfte liegen mindestens zwei Aufwärtsbeurteilungen auf FvF vor).

Für die Hypothese 5) lässt sich also sagen, dass die aggregierten Aufwärts-Beurteilungen, die die gleiche Führungskraft betreffen, nur auf wenigen Dimensionen einen höheren Zusammenhang mit Beurteilungen aus anderen Perspektiven (Selbst, Abwärts) aufweisen als die nicht-aggregierten Aufwärts-Beurteilungen.

Im Folgenden werden die korrelativen Zusammenhänge an sich genauer betrachtet: Die mittlere Selbst-Aufwärts-Korrelation (über die sieben Faktoren hinweg) beträgt $\bar{r}$ = .06, sie ist also unbedeutend (vgl.Abbildung 27, S. 310). Da jedoch zwei der sieben Korrelationen negativ sind, muss die mittlere Korrelation sehr kritisch betrachtet werden.[158] Aussagekräftiger sind Korrelationsindizes der einzelnen Führungsdimensionen (vgl. Tabelle 49).

Tabelle 49. Korrelative Zusammenhänge zwischen den Beurteilungsperspektiven der sieben FzFü-Dimensionen

	Selbst-Auf$_a$	*N*	Selbst-Ab	*N*	Auf$_a$-Ab		*N*
E/F	.15	40	.30$^+$	32	**.39***	(.44)*	32 (24)
B/S	.06	40	.22	32	**.27**	(.44)*	32 (24)
IB	.08	40	-.03	32	**.43***	(.55)**	32 (24)
K	**.36***	40	.10	32	.11	(.15)	32 (24)
P	-.27$^+$	40	.12	32	**.27**	(.29)	32 (24)
FvF	.07	30	**.19**	31	.02	(.08)	24 (17)
S	-.10	26	.28	31	**.45**$^+$	(.55)*	19 (14)

Anmerkungen. Auf$_a$: aggregierte Aufwärtsratings, fett: jeweils die höchste Korrelation für die entsprechende Führungsdimension, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .10$, in Klammern: Angaben nur für die Auf-Ab-„Paare“, für die auch eine Selbstbeschreibung vorliegt.

Für die Dimension Klarheit ergibt sich die höchste Korrelation (r = .36). Von der Selbst-Aufwärts-Korrelation auf allen anderen Dimensionen unterscheidet sie sich mit mittlerer Stärke. Die Effekte der Korrelationsdifferenzen schwanken in einem Bereich von q = .23 bis q = .66. Dass ausgerechnet Klarheit von Mitarbeitern und ihren Führungs-

[158] Positive und negative Korrelationen nivellieren sich bei einer Mittelung gegenseitig.

kräften ähnlich beurteilt wird, verwundert aus inhaltlichen Gründen nicht. Um Klarheit herzustellen, müssen die entsprechenden Inhalte zwischen den Beteiligten (Führungskraft und Mitarbeiter) „verhandelt“ werden.[159] Erst wenn der jeweils andere zustimmt, alles verstanden zu haben, ist dieser Prozess abgeschlossen. Insofern verlangt dieser per se viel Interaktion und gibt in der Regel unmittelbar in der Situation Rückmeldung über den Erfolg des entsprechenden Führungsverhaltens, das Klarheit schaffen soll. Dies begünstigt eine akkurate Beurteilung dieser Verhaltensweisen; andere Führungsdimensionen unterliegen stärker der individuellen Bewertung des Einzelnen.

Eine nicht unbedeutende, negative Selbst-Aufwärts-Korrelation findet sich für Partizipation (r = -.27). Dass dieser geringe negative Zusammenhang nicht durch einzelne Items zustande kommt, verdeutlicht die Analyse auf Item-Ebene. Eine negative Korrelation bedeutet jedoch, dass Mitarbeiter das Führungsverhalten ihrer direkten Führungskräfte umso weniger partizipativ beschreiben, je stärker diese es selbst als partizipativ beurteilen. Was von den Führungskräften als Gestaltungsfreiraum eingeführt werden soll, wird von den Mitarbeitern möglicherweise eher als undefinierter Raum wahrgenommen, in dem keine individuelle Förderung/Chance existiert, sondern eher Überforderung (vgl. auch Kapitel 6.4.1).

Die Selbst-Aufwärts-Korrelationen der anderen Führungsdimensionen liegen in unbedeutender bis sehr geringer Höhe zwischen r = -.10 bis r = .15. Zusammenfassend sind die Effekte also für alle Faktoren mit Ausnahme von K und P als unbedeutend bis äußerst gering zu interpretieren. Allein aufgrund der Tatsache, dass sowohl Mitarbeiter als auch Führungskräfte unmittelbar am Führungsgeschehen beteiligt sind, wäre zumindest auf allen Faktoren ein geringer Effekt zu erwarten. Eine Null-Korrelation ist sehr unerwartet, besagt sie doch, dass die beiden am Führungsprozess beteiligten Parteien – Führungskräfte und Mitarbeiter – voneinander unabhängige Bilder von Führung haben.[160]

Betrachtet man den Zusammenhang, den die **Selbstbeurteilung** zur **Abwärtsbeurteilung** aufweist, dann zeigen sich mit Ausnahme von IB, wofür eine Nullkorrelation besteht, geringe Effekte. Die Höhe der Korrelation liegt im Mittel bei $\bar{r}$ = .17 und schwankt zwischen r = -.03 (IB) und r = .30 (E/F). Was die Selbstbeurteiler als individuell berücksichtigend bewerten, scheint also nicht nur unabhängig davon zu sein, was die Aufwärtsbeurteiler, sondern auch davon, was die Abwärtsbeurteiler so charakterisieren. Dass der höchste Selbst-Abwärts-Zusammenhang für die Dimension E/F gefunden wird, erscheint zweifellos plausibel. Insbesondere wenn es um Karriereentwicklung geht, kann nicht ausschließlich abteilungsintern geplant werden, sondern ist die nächsthöhere Ebene einzubeziehen. Das bedeutet, dass über E/F Austausch zwischen den Selbst- und Abwärts-

[159] Beispielhaft sei an ein Item erinnert: „Meine direkte Führungskraft bespricht Zielsetzungen mit mir so lange, bis alle Unklarheiten beseitigt sind.“

[160] Nicht zu vergessen ist, dass die Prüfung auf Signifikanz bei Korrelationen in der Regel der Nullhypothese $\rho = 0$ folgt. Dies ist kein strenges Kriterium, da ohnehin immer von Zusammenhängen zwischen zwei Variablen auszugehen ist. So sagt Art Brief (zit. nach Webster & Starbuck, 1988, S. 113): „everything correlates .1 with everything else“.

beurteilern stattfindet, was letztlich die Beurteilungsgenauigkeit der Abwärtsbeurteiler erhöht. Ähnliches gilt für die Dimension S (r = .28), da dem Vorgesetzten (Abwärtsbeurteiler) zumindest gemeldet werden muss, wer in der Vertretungszeit als Ansprechpartner fungiert und für die Prozesse verantwortlich zeichnet. Ein solcher Austausch zwischen Selbst- und Abwärtsbeurteilern hinsichtlich derjenigen Führungsverhaltensweisen, die ja die Interaktion zwischen Selbst- und *Aufwärts*beurteilern betreffen, liegt für alle anderen Dimensionen weniger nahe.

Die **Abwärts-Aufwärts-Korrelationen**, also die Korrelationen zwischen zwei Fremdbeurteilern, liegen im Durchschnitt bei $\bar{r}$ = .31. Sie sind durchgehend positiv und erreichen für E/F, IB und S Signifikanz (vgl. Tabelle 49). Die gute Übereinstimmung auf den Dimensionen E/F und S ist deshalb sehr plausibel, weil sie in gewisser Weise abteilungsübergreifend ausgeübt werden. Eine gezielte Mitarbeiterförderung führt im Idealfall zu einer Karriereentwicklung, die mit übergeordneten Hierarchieebenen (also Abwärtsbeurteilern) abgestimmt sein muss. Die Stellvertretung der Führungskraft stellt per definitionem eine Schnittstelle zwischen der Hierarchieebene der Aufwärts- und Abwärtsbeurteiler dar. In ähnlicher Weise kann der deutliche Effekt für IB interpretiert werden. Denn wenn eine Führungskraft auf die individuellen Potentiale ihrer Mitarbeiter eingeht, können diese auch nach oben als Mitarbeiter mit individuellen Profilen kommuniziert werden. Nur für K und FvF liegen die Zusammenhänge in minimaler bzw. unbedeutender Höhe (r = .11 bzw. r = .02). Gerade Letzteres erstaunt, da die Förderung von Führungsverhalten (FvF) noch mehr als die generelle Entwicklung und Förderung (E/F) mit übergeordneten Ebenen abgestimmt sein sollte. Der Befund spricht deshalb eher dafür, dass Führungsverhalten kaum als spezifisch gefördert wahrgenommen wird.

Auffallend ist, dass auf allen sieben Dimensionen für diejenigen Abwärts-Aufwärts-Paare tendenziell höhere Korrelationen vorliegen, deren Ratee auch eine Selbstbeurteilung abgab (vgl. Tabelle 49). Der Befund wird in Kapitel 6.3.1 im Zusammenhang weiterer Moderatorvariablen für die interperspektivische Übereinstimmung genauer betrachtet.

Zusammenfassend fallen also die Selbst-Fremd-Korrelationen – je nach Führungsdimension – unbedeutend bis gering aus. Lediglich für die Dimension Klarheit ergibt sich ein mittlerer Zusammenhang zwischen der Selbst- und Aufwärtsbeurteilung sowie für die Dimension Entwickeln/Fördern zwischen der Selbst- und Abwärtsbeurteilung. Die Abwärts-Aufwärts-Korrelationen liegen für alle Dimensionen außer Fördern von Führungsverhalten im geringen oder mittleren Bereich. Betrachtet man die durchschnittlichen interperspektivischen Korrelationen (über die Dimensionen hinweg), dann kann Hypothese 6) für die Abwärts-Aufwärts- und die Selbst-Abwärts-Korrelationen angenommen werden, während sie für die Selbst-Aufwärts-Korrelationen abgelehnt werden muss (vgl. Abbildung 27).

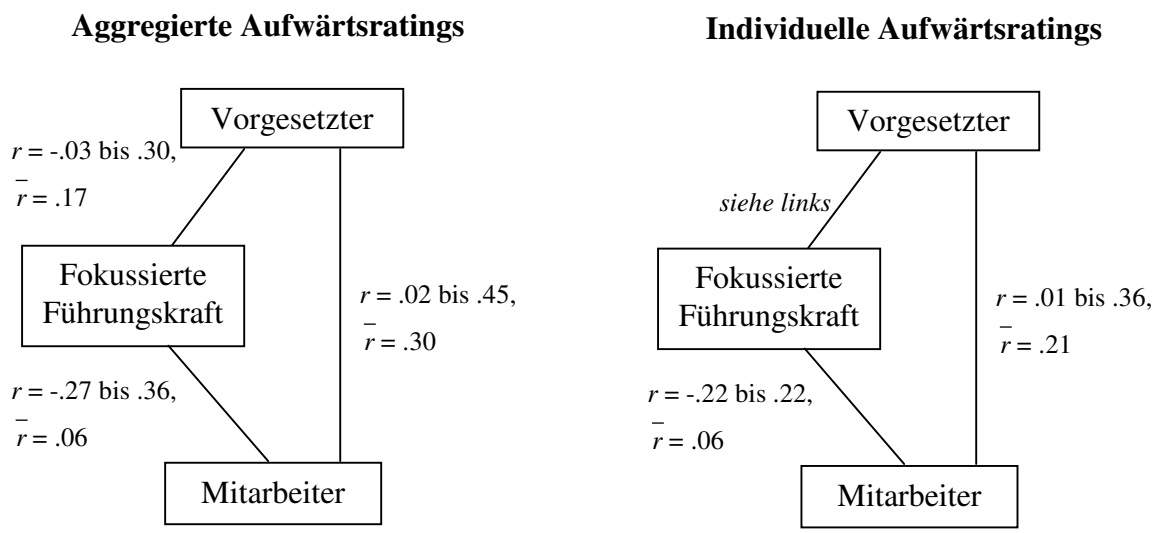


Abbildung 27. Korrelative Zusammenhänge zwischen Beurteilungsperspektiven (Ist), mittlere Korrelation der sieben FzFü-Faktoren

Anmerkungen. Vorgesetzter = Abwärtsbeurteilung, Fokussierte Führungskraft = Selbstbeurteilung, Aggregierte (vs. individuelle) Aufwärtsratings: Mitarbeiter = Median (individuelle Werte) der Aufwärtsbeurteiler, die der gleichen Führungskraft berichten). Es sind jeweils die niedrigste und höchste der Korrelationen auf den sieben Dimensionen angegeben sowie die mittlere Korrelation.

Stellt man die drei korrelativen Perspektiven-Vergleiche einander gegenüber (vgl. Tabelle 49, Tabelle 50), dann kann Folgendes festgehalten werden: Die Fremd-Fremd-Korrelationen sind für die Mehrzahl der Dimensionen höher als die Selbst-Fremd-Korrelationen, auch wenn dieser Unterschied nur für einzelne Dimensionen Signifikanz erreicht. Es zeigen sich mittlere Effekte für den Unterschied zwischen den Abwärts-Aufwärts-Korrelationen und den *Aufwärts*-Selbst-Korrelationen ($\bar{q}$ = .35). Die Effekte halbieren sich, wenn man die Abwärts-Aufwärts-Korrelationen mit den *Abwärts*-Selbst-Korrelationen vergleicht ($\bar{q}$ = .17); dies ist ein geringer Effekt, der erkennen lässt, dass die Fremd-Fremd-Korrelationen über den Selbst-Fremd-Korrelationen liegen (Hypothese 7).

Tabelle 50. Korrelationsunterschiede zwischen den Perspektiven

	Selbst-Auf$_a$ vs. Selbst-Ab				Ab-Auf$_a$ vs. Selbst-Auf$_a$				Selbst-Ab vs. Ab-Auf$_a$			
	q	*t*	*df*	*p*	*q*	*t*	*df*	*p*	*q*	*t*	*df*	*p*
E/F	-.16	.76	29	.23	.26	1.18	29	.12	-.10	-.41	29	.34
B/S	-.16	-.73	29	.24	.27	.94	29	.18	-.05	-.21	29	.42
IB	.11	.26	29	.40	.49	1.58	29	.06	-.50	-2.17	29	.02
K	.28	1.12	29	.14	-.27	-1.07	29	.15	-.01	-.05	29	.48
P	-.40	-1.84	29	.04	.54	2.40	29	.01	-.16	-.53	29	.30
FvF	-.12	-.48	27	.32	.04	.14	23	.44	.08	.27	23	.40
S	-.39	-1.98	23	.03	.56	2.32	15	.02	-.26	-.67	15	.26

Anmerkungen. Auf$_a$: aggregierte Aufwärtsratings, *q*: Effektgröße des Korrelationsunterschieds, *t*: Students's *t* als statistische Kenngröße des Hotellings-Williams-Tests auf Korrelationsunterschiede bei abhängigen Stichproben, positiver/negativer Student's *t*-Wert und *q*-Wert: Korrelation zwischen erstgenannten Perspektiven höher/geringer als Korrelation zwischen zweitgenannten Perspektiven (in erster Tabellenzeile).

Letztlich geben diese Befunde einen Hinweis darauf, dass die Selbstbeurteilungen weniger akkurat sind als die Fremdbeurteilungen, sie also stärkeren Verzerrungen unterliegen.

Auf den ersten Blick unerwartet zeigen sich nicht unbedeutende Effekte, was den Vergleich der beiden Selbst-Fremd-Korrelationen auf den einzelnen Dimensionen angeht. Mit Ausnahme von K und IB ist der Zusammenhang zwischen der Selbst- und Abwärtsbeurteilung höher als der zwischen der Selbst- und Aufwärtsbeurteilung, auch wenn die Vergleiche aufgrund des geringen Stichprobenumfangs nur auf wenigen Dimensionen Signifikanz erreichen.[161] Dass Selbst- und Abwärtsbeurteiler die Führung ähnlicher beschreiben als Selbst- und Aufwärtsbeurteiler, verweist darauf, dass *Führungskräfte* (also Selbst- und Abwärtsbeurteiler) ihre Beurteilungen vor einem ähnlichen Referenzrahmen vornehmen, der sich von dem der Mitarbeiter unterscheidet. Am stärksten fallen die Unterschiede auf dem Faktor Partizipation aus, obwohl dafür die Selbst-Abwärts-Korrelation gering ausfällt. Grund dafür ist die bereits oben diskutierte negative Selbst-Aufwärts-Korrelation auf dieser Dimension. Auch die durchschnittliche Korrelation über die Führungsdimensionen hinweg ist zwischen der Selbst- und Abwärtsperspektive höher (vgl. Abbildung 27). Hypothese 8) wird abgelehnt.

Vergleicht man die Zusammenhänge der Perspektiven auf der Ebene einzelner Führungsdimensionen, entwickelt sich ein inhaltlich genaueres Bild: Auf fünf der sieben Dimensionen ergibt sich der höchste Zusammenhang für die *Abwärts-Aufwärts*-Korrelation (vgl. Tabelle 49). Besonders stark sind die Effekte für den Unterschied auf der Dimension IB: Während sich für Abwärts-Aufwärts ein deutlicher Zusammenhang ergibt, sind die Selbst- mit den Fremdbeurteilungen quasi unkorreliert. Der Korrelationsunterschied, der sowohl zum Selbst-Aufwärts- als auch zum Selbst-Abwärts-Zusammenhang mittlere Stärke annimmt und mindestens annähernd statistisch bedeutsam ausfällt (vgl. Tabelle 50), verweist darauf, dass Führungskräfte selbst eine andere Vorstellung von individueller Berücksichtigung haben als Fremdbeurteiler.

Neben IB liegen die Abwärts-Aufwärts-Korrelationen auch für E/F und S im mittleren Effektbereich und damit deutlich höher als die Selbst-Fremd-Korrelationen (Tabelle 49 und Tabelle 50). Den Abwärtsbeurteilern gelingt eine Einschätzung von E/F und S also verhältnismäßig gut (vgl. S. 307f.). Für S kann fast anekdotisch resümiert werden, dass die Regelung der Abwesenheitszeiten von den Führungskräften und ihren Mitarbeitern invers beurteilt wird. Möglicherweise liegt das daran, dass die Konsequenzen einer Stellvertretungsregelung (z.B. unklare Verantwortungsverteilung) von der Führungskraft (Selbstbeurteiler) nicht unmittelbar erlebt werden, da sie in ihre Abwesenheitsperiode fallen.

[161] Es ist nebenbei anzumerken, dass es ungleich schwieriger ist, signifikante Korrelationsunterschiede zu entdecken als überhaupt signifikante Korrelationsindizes zu erreichen (Erläuterungen z.B. Bobko, 1995).

Auf den Dimensionen B/S und P ergeben sich generell keine besonders starken Korrelationen. Dennoch liegt der Zusammenhang zwischen der Abwärts- und Aufwärtsperspektive deutlich über dem der Selbst- und Aufwärtsperspektive (Tabelle 50).

Während für die Dimensionen E/F, B/S, IB, P, S der Abwärts-Aufwärts-Zusammenhang am stärksten ist, wird für die Dimension K die Selbst-Aufwärts-Korrelation und für FvF die Selbst-Abwärts-Korrelation am höchsten. Für FvF ist der Unterschied zur Abwärts-Aufwärts- sowie zur Selbst-Aufwärts-Korrelation jedoch sehr gering (Tabelle 50). Dennoch kann spekuliert werden, dass das Fördern von Führungsverhalten ein Metier von Führungskräften ist und Mitarbeiter (Aufwärtsbeurteiler) möglicherweise bis dato diese Dimension weniger reflektiert haben. Entsprechend stimmen die Beurteilungsperspektiven besser überein, an denen die Mitarbeiter nicht beteiligt sind. Für K wurde bereits ausgeführt, dass es sich um Inhalte handelt, die in der Interaktion zwischen Führungskraft und Mitarbeiter unmittelbare Konsequenzen erkennen lassen. Abwärtsbeurteiler, die nicht unmittelbar beteiligt sind, müssten demzufolge an Beurteilungsgenauigkeit einbüßen. Damit stimmig ist, dass sich sowohl die Selbst-*Abwärts*- als auch die *Abwärts*-Aufwärts-Korrelation mit fast mittlerem Effekt von der Selbst-Aufwärts-Korrelation unterscheiden. Weiter unten wird geprüft, ob eine ergebnisbezogene Itemformulierung (wie es für K überwiegend zutrifft) prinzipiell einen moderierenden Einfluss auf die Höhe der Übereinstimmung verschiedener Perspektiven hat (vgl. Kapitel 6.3.3). Nicht nur innerhalb einer der drei verschiedenen Perspektiven-Vergleiche unterscheiden sich die Korrelationen über die Führungsdimensionen hinweg, sondern für die unterschiedlichen Perspektiven-Vergleiche stellen sich auch unterschiedliche Dimensionen als besonders „gut“ heraus (Hypothese 9).

Stellt man die drei interperspektivischen Korrelationen der Führungsbeurteilungen aus der vorliegenden Automobil-Stichprobe denjenigen Korrelationsindizes gegenüber, die Conway und Huffcutt (1997) in ihrer Meta-Analyse für berufliche Leistungsbeurteilung angeben, dann zeigt sich die gleiche Reihenfolge.[162] So fallen die Abwärts-Aufwärts-Korrelationen höher aus als die Selbst-Fremd-Korrelationen, wobei die Selbst-Abwärts-Korrelationen höher sind als die Selbst-Aufwärts-Korrelationen. Während beide Selbst-Fremd-Übereinstimmungen der vorliegenden Stichprobe im Durchschnitt leicht unter den Koeffizienten aus Meta-Analysen liegen (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002, vgl. Kapitel 3.2), liegt die Abwärts-Aufwärts-Übereinstimmung leicht über den Angaben von Conway und Huffcutt (1997). Letztere blieb genauso wie die Selbst-Aufwärts-Übereinstimmung in den anderen Meta-Analysen unberücksichtigt. Die maximale Abweichung der Korrelationskoeffizienten der

[162] Conway und Huffcutt (1997) sind die einzigen, die alle drei Perspektiven (Selbst, Aufwärts, Abwärts) in einer
Meta-Analyse erfassen.

vorliegenden Stichprobe von den meta-analytisch geschätzten Koeffizienten beträgt jedoch lediglich r = .08 (betrifft die Selbst-Aufwärts- sowie die Abwärts-Aufwärts-Korrelation). Für die Selbst-Abwärts-Korrelation, die ohnehin nur um r = .05 von den meta-analytischen Angaben abweicht, verringert sich die Differenz auf r = .02, wenn man den Vergleich nur zur Substichprobe „Management" zieht (Conway & Huffcutt, 1997; Heidemeier & Moser, 2002). Diese Annäherung überrascht deshalb nicht, weil sich die Tätigkeit (Management vs. andere) bereits als Moderatorvariable für die Selbst-Abwärts-Übereinstimmung herausgestellt hat (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002; vgl. Kapitel 3.3.2). Insofern entsprechen die durchschnittlichen interperspektivischen Übereinstimmungen der vorliegenden Studie erwartungsgemäß den *durchschnittlichen* Übereinstimmungen anderer Branchen.

Abschließend soll angemerkt werden, dass die Korrelationen zwischen den Beurteilern verschiedener Perspektiven erwartungskonform niedriger sind als die in Kapitel 6.1 berichteten Korrelationen innerhalb der Mitarbeitergruppen. Dieses Ergebnis unterscheidet sich von Studien, deren Daten auf gleich starke Unterschiede zwischen Ratings derselben und Ratings unterschiedlicher Perspektiven verweisen (z.B. Abwärts- und Peerbeurteilungen u.a. von Führung: Meta-Analyse von Viswesvaran, Schmidt & Ones, 2002). Auch die meta-analytischen Schätzungen von Conway und Huffcutt (1997) zu beruflicher Leistung fallen für die Interrater-Reliabilität der Mitarbeiter (r = .30) nicht deutlich höher aus als für die Korrelation der Mitarbeiter- mit Abwärts- oder Peerratings (je r = .22). Wenn die durchschnittliche Interrater-Reliabilität der Mitarbeiter – wie in der vorliegenden Studie – aber deutlich über den interperspektivischen Korrelationen liegt, dann sind die geringen Übereinstimmungen zwischen den Perspektiven eher auf perspektivenbezogene statt auf idiosynkratische Ursachen zurückzuführen. Ähnliche Ergebnisse berichten auch Tsui und Ohlott (1988) für Mitarbeiter und Peers.

Zusammenfassend lässt sich Folgendes festhalten:

- Die Selbst-Aufwärts- sowie sie Abwärts-Aufwärts-Korrelationen der Führungsbeurteilungen unterscheiden sich nur gering in Abhängigkeit davon, ob die Aufwärtsratings in aggregierter oder nicht aggregierter Form in die Berechnung eingehen. Wo Unterschiede auftreten, fallen diese erwartungsgemäß zugunsten der aggregierten Ratings aus, so dass im weiteren Verlauf der Arbeit die nicht aggregierten Ratings nur noch in Ausnahmefällen herangezogen werden.
- Während die mittlere Selbst-Aufwärts-Korrelation unbedeutend ist, liegt die mittlere Selbst-Abwärts-Korrelation im geringen und die mittlere Abwärts-Aufwärts-Korrelation im mittleren Bereich (vgl. Abbildung 27). Insgesamt korrelieren die

Fremdbeurteilungen untereinander also deutlich stärker als sie jeweils mit der Selbstbeurteilung korrelieren.

- Auf einzelnen Führungsdimensionen können hingegen für alle interperspektivischen Vergleiche (Selbst-Aufwärts, Selbst-Abwärts, Abwärts-Aufwärts) bedeutende Korrelationen festgehalten werden. Als besonders hoch hervorzuheben sind die Selbst-Aufwärts-Korrelation für Klarheit, die Selbst-Abwärts-Korrelationen für Entwickeln/Fördern und Stellvertretung sowie die Abwärts-Aufwärts-Korrelationen für Entwickeln/Fördern, Individuelle Berücksichtigung und Stellvertretung. Letztlich ist also zu schließen, dass nicht ausschließlich die Beurteilungsperspektive, sondern auch die Führungsdimension den korrelativen Zusammenhang stark beeinflusst. Kritisch sind daher alle Studien zu betrachten, die eine Führungsbeurteilung global vornehmen, ohne zwischen unterschiedlichen Facetten zu unterscheiden. Langfristig muss auf diesem Wege empirisch erarbeitet werden, welche Beurteilungsperspektiven auszuwählen sind, wenn es um eine verlässliche Beurteilung bestimmter Dimensionen geht.
- Alle drei interperspektivischen Zusammenhänge liegen in ähnlicher Höhe wie die entsprechenden Zusammenhänge bei der allgemeinen beruflichen Leistungsbeurteilung. Einen spezifischeren Vergleich zu Befunden aus anderen Studien und Stichproben nimmt Kapitel 7.2.3 vor.

6.3.2 Niveau-Unterschiede zwischen Selbst-, Aufwärts- und Abwärtsbeurteilungen

In der Unternehmenspraxis werden die Ergebnisse von multiperspektivischen Führungsbeurteilungen in aller Regel anhand der Mittelwertsunterschiede zurückgemeldet. Deshalb ist es nützlich, auch aus wissenschaftlicher Perspektive Erkenntnisse über die Art und Höhe der Diskrepanzen (Kapitel 6.3.2), über das Zustandekommen und potentielle Reduktionsmöglichkeiten der Diskrepanzen (vgl. Kapitel 6.6) sowie über moderierende Einflüsse auf die Diskrepanzen (vgl. Kapitel 6.3.2) zu gewinnen. Die vorliegende Arbeit verwendet also – im Unterschied zur Mehrzahl der Studien dieses Themenbereichs – neben den korrelativen Zusammenhängen die Perspektiv-Diskrepanzen als weiteres Ähnlichkeits- bzw. Differenzmaß.

Im Zuge von Fragestellung 3 werden folgende Hypothesen untersucht:

10) Zwischen allen erfassten Perspektiven (Selbst-, Abwärts-, Aufwärtsbeurteilung) bestehen Niveau-Unterschiede.

11) Die Niveau-Unterschiede zwischen der Selbst- und Abwärtsbeurteilung sind geringer als die zwischen der Selbst- und Aufwärtsbeurteilung.

12) Die Selbstbeurteilungen fallen am positivsten, die Aufwärtsbeurteilungen am negativsten aus. Die Abwärtsbeurteilungen nehmen eine Zwischenstellung ein und unterscheiden sich von den Selbstbeurteilungen weniger als von den Aufwärtsbeurteilungen.

13) Die Selbstüberschätzung der Führungskräfte im Vergleich zur Abwärtsbeurteilung ist unabhängig von deren Selbstüberschätzung im Vergleich zur Aufwärtsbeurteilung.

14) Das Ausmaß der interperspektivischen Differenzen variiert über die Führungsdimensionen hinweg.

Im Einklang mit früheren Forschungsarbeiten kann festgehalten werden, dass sich Führungskräfte im Vergleich zu ihren Mitarbeitern generell überschätzen. Für alle Faktoren mit Ausnahme von S erreicht der Vergleich mindestens das 1%-Signifikanzniveau (vgl. Tabelle 51). Für S liegt möglicherweise ein Deckeneffekt vor, denn auch die Mitarbeiterbeurteilungen liegen nahe am sechsten (von sieben) Skalenpunkt. Dennoch darf der Unterschied nicht unterschätzt werden, da die Effektgröße auf eine Differenz von etwa einer Drittel Standardabweichung hinweist. Am größten ist die Differenz auf der Dimension E/F, auf der die Selbstbeurteilung fast eine ganze Standardabweichung über der Aufwärtsbeurteilung liegt. Hypothese 10) ist folglich für den Selbst-Aufwärts-Vergleich anzunehmen.

Tabelle 51. Niveau-Unterschied zwischen den Beurteilungsperspektiven (Ist-Beurteilung)

Faktor	Selbst – Aufwärts$_a$				Selbst – Abwärts				Abwärts – Aufwärts$_a$[1]			
	t	*df*	*p*	*d*	*t*	*df*	*p*	*d*	*t*	*df*	*p*	*d*
E/F	4.66	39	<.01	.97	-.19	31	.85	.04	6.18 (5.09)	31 (23)	<.01 (<.01)	1.24 (1.12)
B/S	3.53	39	<.01	.76	-.85	31	.40	.19	5.89 (5.70)	31 (23)	<.01 (<.01)	.85 (1.30)
IB	3.25	39	<.01	.70	.98	31	.34	.25	2.66 (2.32)	31 (23)	.01 (.03)	.75 (.45)
K	3.76	39	<.01	.70	-.13	31	.90	.03	3.48 (2.79)	31 (23)	<.01 (.01)	.83 (.75)
P	3.13	39	<.01	.78	.70	31	.49	.16	3.59 (2.73)	31 (23)	.01 (.01)	.80 (.69)
FvF	2.71	29	.01	.67	2.03	30	.05	.47	1.07 (-.37)	23 (16)	.29 (.72)	.31 (.14)
S	.80	25	.43	.36	.68	30	.50	.15	1.29 (.48)	18 (14)	.22 (.64)	.31 (.13)

Anmerkungen. T-Test für gepaarte Stichproben, für entsprechende *M* und *SD* vgl. Tabelle 50, Aufwärts$_a$: aggregierte Aufwärtsratings [1]in Klammern: Werte für die Auf-Ab-Paare, für die auch eine Selbstbeschreibung vorliegt. Negativer *t*-Wert: Abwärts > Selbst bzw. Aufwärts > Abwärts.

Hypothese 10) nimmt nicht nur Selbst-Aufwärts-Differenzen an, sondern auch Differenzen zwischen den anderen Perspektiven. Wie in Kapitel 6.3.2 bereits deutlich wurde, unterscheiden sich die Abwärts- und Selbstbeschreibungen kaum voneinander. Ein t-Test für gepaarte Stichproben erbringt entsprechend auch nur für FvF einen signifikanten Unterschied in der Form, dass die Selbstbeurteilung positiver ausfällt als die Abwärtsbeurteilung. Da FvF impliziert, nicht nur selbst gut zu führen, sondern zusätzlich andere zu guten Führungskräften auszubilden, kann diese Dimension von den Selbstbeurteilern auch besonders gut dafür herangezogen werden, sich als verantwortungsvolle Führungskraft zu kategorisieren. Diese Annahme wird dadurch unterstützt, dass die Führungskräfte der zweiten Hierarchieebene – entgegen der Führungskräfteanforderungen – FvF als weniger wichtig einstufen als die Führungskräfte der dritten und vierten Hierarchieebene.[163] Wenn man berücksichtigt, dass in einem Sandwichdesign Angehörige höherer Hierarchieebenen als Abwärtsbeurteiler fungieren, resultieren daraus nach oben verzerrte Selbstbeurteilungen.

Neben FvF zeigt sich für IB ein bedeutender, wenn auch geringer Effekt in die gleiche Richtung (d = .25). Dies ist damit erklärbar, dass individuell berücksichtigende Verhaltensweisen im Vergleich zu den anderen erfassten Führungsdimensionen nur sehr schwierig von Dritten überprüfbar sind. Eine solche Validierungserwartung stellte sich jedoch als robuste Moderatorvariable für die Güte von Selbstbeurteilungen heraus (Mabe & West, 1982). Ist sie nicht gegeben, fehlt quasi ein korrigierender Faktor, so dass die Selbstbeurteilung milder ausfällt. Für alle anderen Faktoren ergibt sich kein bedeutender Effekt.

Hypothese 10) ist bezüglich des Selbst-Abwärts-Vergleichs also für die überwiegende Mehrzahl der Führungsdimensionen abzulehnen. Weitere Studien müssen überprüfen, ob sich einzelne Führungsdimensionen, z.B. FvF, herauskristallisieren, auf denen eine überzufällige Selbst-Abwärts-Differenz bestätigt werden kann.

Betrachtet man die Mittelwertsunterschiede zwischen den beiden Fremdbeurteilungsperspektiven, dann zeigt sich, was auch aus den beiden bereits berichteten Vergleichen zu schließen ist: Die Abwärtsbeurteiler bewerten die fokussierte Führungskraft deutlich positiver als die Aufwärtsbeurteiler. Bei hochsignifikanten Mittelwertsunterschieden für E/F, B/S, IB, K und P sind die Effekte als mittel bis hoch zu interpretieren. Die Abwärtsbeurteilung liegt für diese Faktoren zwischen .75 und 1.24 Standardabweichungen über den dazugehörigen Aufwärtsbeurteilungen. Für FvF und S sind die Niveau-Unterschiede deutlich geringer. Sie erreichen keine Signifikanz, auch wenn sie sich um eine Drittel Standardabweichung unterscheiden. Hypothese 10) kann also für den Abwärts-Aufwärts-Vergleich bestätigt werden.

[163] Zweite Ebene (N = 4): Skalenmittelwert 5.25, dritte Ebene (N = 8): 6.03, vierte Ebene (N = 12): 6.08. Von der Geschäftsführung (erste Hierarchieebene) liegen keine Wichtigkeitseinschätzungen vor.

Nachdem im vorangehenden Kapitel eine höhere korrelative Abwärts-Aufwärts-Übereinstimmung für diejenigen „Paare“ berichtet wurde, für die auch die entsprechende Selbstbeurteilung abgegeben wurde (vgl. Tabelle 49), gilt hier analog Folgendes für die Niveau-Unterschiede: Die Unterschiede fallen mit Ausnahme von B/S[164] tendenziell geringer aus als für die Gesamtstichprobe. Die moderierende Wirkung der Beteiligung verschiedener Ratergruppen wird in Kapitel 6.3.3 im Zusammenhang anderer Moderatorvariablen weitergehend interpretiert.

Bereits in der bisherigen Darstellung wurde deutlich, dass Hypothese 11) angenommen werden kann. Während die Selbstbeurteilung durchschnittlich um .71 Standardabweichungen positiver als die Aufwärtsbeurteilung ausfällt, zeigen sich für den Selbst-Abwärts-Vergleich lediglich für zwei Faktoren überhaupt bedeutende Effekte ($d \geq .20$). Entsprechend kann auch Hypothese 12), die eine Rangreihe annimmt, in der die Selbstbeurteilung deutlich positiver als die Abwärtsbeurteilung ausfällt und diese wiederum – mit noch größerem Abstand – über der Aufwärtsbeurteilung liegt, nur für einzelne Führungsdimensionen bestätigt werden. Während die Selbst- und Abwärtsbeurteiler das Führungsverhalten ähnlich positiv bewerten, geben die Aufwärtsbeurteiler deutlich negativere Beurteilungen ab. Erklärungsansätze werden in Kapitel 7.3 diskutiert.

Es kann festgehalten werden, dass die Selbstbeurteilungen also nicht generell – gegenüber allen anderen Beurteilungsperspektiven – als inflationär zu bewerten sind (vgl. auch Crampton & Wagner, 1994). Im Vergleich zur meta-analytischen Schätzung der Selbst-Abwärts-Diskrepanz für Manager (Heidemeier & Moser, 2002: $d = .34$) fällt sie in der vorliegenden Studie unerwartet gering aus. Verschiedene Motive und implizite Annahmen können den milden Abwärtsbeurteilungen zugrunde liegen. So bewerten die Abwärtsbeurteiler nicht nur ihre unterstellten Führungskräfte, sondern werden auch von diesen in ihrer eigenen Führungskompetenz bewertet. Auch wenn die Befragung anonym durchgeführt wurde, könnte durch dieses Design eine Pseudo-Gerechtigkeitsnorm aktiviert worden sein, die zu vergleichsweise milden Abwärtsbeurteilungen führt.

Da in der Literatur selten Effektgrößen für Niveau-Unterschiede oder die entsprechenden Kennwerte zu ihrer Berechnung angegeben sind, ist ein Vergleich zu anderen Einzelstudien aus dem Führungskräftebereich nicht möglich.

Manche Studien legen aus Gründen der Reliabilität nahe, alle Fremdbeurteiler – unabhängig von ihrer Beurteilungsrichtung – gemeinsam zu betrachten und ihren mittleren Beurteilungswert der Selbstbeurteilung gegenüberzustellen (Wohlers & London, 1989). Deshalb soll abschließend in dieser Fragestellung geklärt werden, ob die beiden Fremd-

[164] Die Sonderstellung von B/S ist damit zu erklären, dass sechs der acht Abwärtsbeurteilungen, die in die größere Stichprobe zusätzlich eingehen, von nur *zwei* Abwärtsbeurteilern gegeben wurden. Sie beurteilen B/S tendenziell höher als der Durchschnitt. Insofern ist der Unterschied eher auf diese Bündelung der Rater zurückzuführen statt auf inhaltliche Gründe.

perspektiven bei den folgenden Auswertungen getrennt voneinander oder gemeinsam miteinander zu betrachten sind. Bereits die Niveau-Unterschiede deuten darauf hin, dass die beiden Fremdbeurteiler unterschiedliche Information einbringen – wofür im Übrigen auch die Schätzungen von Scullen et al. (2000) plädieren (vgl. Kapitel 3.4.3). Und auch wenn die Niveau-Differenz zwischen Selbst- und Abwärtsbeurteilern auf vier (E/F, B/S, IB, P) der sieben Faktoren mit der Diskrepanz zwischen Selbst- und Aufwärtsbeurteilern signifikant korreliert ($\bar{r}$ = .35), kann nicht davon ausgegangen werden, dass sich diejenigen Führungskräfte, die sich bezüglich der Abwärtsbeurteiler überschätzen (unterschätzen), auch bezüglich der Aufwärtsbeurteiler überschätzen (unterschätzen). Dies zeigt auch Tabelle 52, in der die Selbstbeurteiler in einem entsprechenden Kontingenzschema verzeichnet sind. Beispielsweise gilt für E/F (Dimension 1), dass sich von 24 Führungskräften drei hinsichtlich der Abwärts- *und* Aufwärtsbeurteilung unterschätzen, eine Führungskraft mit beiden Fremdperspektiven übereinstimmt und sich acht Führungskräfte hinsichtlich beider Fremdperspektiven überschätzen. Die restlichen 12 Führungskräfte fallen für den Vergleich mit der Aufwärts- und Abwärtsperspektive in unterschiedliche Kategorien. Bowkers Chi^2-Test (df = 3) erreicht für vier der sieben Führungsfaktoren (E/F, B/S, K, P) Signifikanz, so dass die Differenzen für die beiden Perspektiven-Vergleiche (Selbst-Aufwärts und Selbst-Abwärts) keine vollkommen redundante Information enthalten, die Annahme von Hypothese 13) unterstützt.

Dies ist Grund genug, um die Fremdbeurteiler getrennt voneinander zu betrachten. Durch eine Aggregation aller Fremdbeurteilungen soll keine wertvolle Information verloren werden. Neben methodischen sprechen auch praxisrelevante Gründe dafür, die Perspektiven getrennt zu betrachten: Erstens integrieren Ratees Beurteilungen aus unterschiedlichen Perspektiven unterschiedlich stark in ihr Selbstbild (Greguras, Ford & Brutus, 2003), so dass die Perspektiven in einem Feedbackprozess zu trennen sind. Zweitens greifen Ratees in unterschiedlichem Maße auf Beurteilungen aus unterschiedlichen Perspektiven zurück, wenn es um konkrete Zielsetzungen geht (Brutus, London & Martineau, 1999).

Tabelle 52. Klassifikation der Übereinstimmung: Selbst-Aufwärts und Selbst-Abwärts

		Selbst-Abwärts																			
		Unterschätzer (A)							Übereinstimmer (B)							Überschätzer (C)					
Führungsdimensionen	1 bis 7	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Selbst-Aufwärts	A	3	2	4	2	3	1	0	0	1	1	1	2	1	0	0	0	0	0	1	0
	B	4	3	1	1	0	0	2	1	4	2	5	6	2	2	0	0	4	3	0	3	1
	C	3	5	2	5	1	1	1	5	3	2	0	6	1	2	8	6	8	7	6	7	3

Anmerkungen. Führungsdimensionen: Kategorisierung in Unterschätzer, Übereinstimmer und Überschätzer vgl. zur Erinnerung S. 248, 1=E/F, 2=B/S, 3=IB, 4=K, 5=P, 6=FvF, 7=S, N = 24 (E/F, B/S, IB, K, P), N = 17 (FvF), N = 11 (S), grau hinterlegt: 3x3-Kontingenztafel für Faktor 1, fett: zur besseren Lesbarkeit.

Ein multipler Mittelwertsvergleich mit Bonferroni-Anpassung für Mehrfachvergleiche erbringt für die Selbst-Aufwärts-Differenz keine Unterschiede über die sieben Führungsdimensionen hinweg (N = 23).[165] Diese verhältnismäßig stabile Tendenz zeigt sich zudem in den hohen Korrelationen der Selbst-Aufwärts-Diskrepanzen über die Führungsdimensionen hinweg (mit Ausnahme von Stellvertretung jeweils $r > .50$). Dies ist ein Befund, der in anderen Bereichen bereits gezeigt wurde.[166] Auch die Differenz zwischen der Selbst- und Abwärtsbeurteilung schwankt nicht überzufällig zwischen den sieben Faktoren (N = 31). Für die Abwärts-Aufwärts-Differenz (N = 16) kann festgehalten werden, dass sie auf E/F und B/S höher ist als auf IB und P ($p < .05$).[167] Hypothese 14) wird deshalb für diese Stichprobe zurückgewiesen.

Da die interperspektivischen Differenzen über die Führungsdimensionen hinweg kaum variieren, wird in den folgenden Analysen häufig auf eine faktorspezifische Auswertung verzichtet und stattdessen lediglich die Differenz im Gesamtscore verwendet. Dies erleichtert auch den Überblick über die Ergebnisse.

Zusammenfassend gilt für Fragestellung 3:

- Führungskräfte beurteilen ihr eigenes Führungsverhalten besser als es von ihren Vorgesetzten und von ihren Mitarbeitern beurteilt wird.
- Dabei ist die Diskrepanz zwischen der Selbst- und Abwärtsbeurteilung deutlich geringer als die zwischen der Selbst- und Aufwärtsbeurteilung. Die Vorgesetzten beurteilen das Führungsverhalten ihrer unterstellten Führungskräfte deutlich positiver als die Mitarbeiter dieser Führungskräfte.
- Eine Selbstüberschätzung im Vergleich zum Fremdurteil der Vorgesetzten bedeutet nicht zwangsläufig, dass auch eine Selbstüberschätzung im Vergleich zum Fremdurteil der Mitarbeiter vorliegt. Den Führungskräften ist also kein generelles Attribut „Selbstüberschätzer“ zuzuschreiben.
- Die Diskrepanz zwischen den Selbst- und Fremdbeurteilungen variiert über die verschiedenen Führungsdimensionen hinweg kaum.

[165] Vergleicht man nur die fünf Führungsfaktoren (N = 40), dann zeigt sich, dass die Differenz auf dem Faktor E/F deutlich größer ist als auf dem Faktor IB (p = .02). Wird der Mittelwertsvergleich nicht aufgrund der Differenzen, sondern der D-Statistik durchgeführt, dann ist die Selbst-Aufwärts-Distanz auf IB deutlich geringer als auf E/F ($p < .01$), B/S (p = .01), K (p = .03) und P ($p < .01$). Dies ist deshalb plausibel, weil die Antworten auf IB weniger stark streuen, große Abweichungen aber in die D-Statistik mit stärkerem Gewicht eingehen als geringe Abweichungen.

[166] Vgl. Selbst-Lehrer-Differenz in der Beurteilung von Schülerverhalten bei Cillessen und Bellmore (1999)

[167] Bezieht man in den Vergleich nur die fünf Führungsfaktoren ein (N = 32), bleiben die Ergebnisse stabil.

6.3.3 Moderatoreffekte auf die interperspektivische Übereinstimmung

Grundsätzlich können Annahmen über die Wirkung vieler verschiedener Moderatorvariablen hergeleitet werden (vgl. Fragestellung 4). Die Auswahl der in der vorliegenden Arbeit untersuchten Moderatorvariablen stellt eine deutliche Erweiterung zum bisherigen Forschungsstand dar, der sich ohnehin nicht dadurch auszeichnet, Moderatoreffekte auf den Selbst-Aufwärts- oder gar Abwärts-Aufwärts-Zusammenhang herausgearbeitet zu haben. Für die Untersuchung moderierender Effekte auf die Stärke des Selbst-Abwärts-Zusammenhangs von beruflicher Leistungsbeurteilung allgemein wurden bereits in einer Meta-Analyse eine Vielzahl an Moderatorvariablen untersucht (Heidemeier & Moser, 2002).

Moser et al. (1994) schlagen wegen Schwierigkeiten bei Moderatoranalysen in Einzelstudien (vgl. Kapitel 7.2.4) vor, die untersuchten Variablen im Aggregat zu betrachten, also zu untersuchen, ob überzufällig *viele* Moderatorvariablen Effekte in die erwartete Richtung aufweisen. Dabei ist es unerheblich, wie groß die Effekte ausfallen. Dieses Vorgehen wurde auch in der vorliegenden Arbeit gewählt.

Sowohl für den Selbst-Aufwärts-, den Selbst-Abwärts- sowie den Abwärts-Aufwärts-Zusammenhang (Korrelation und Niveau-Unterschied) wurde eine Vielzahl an potentiellen Moderatorvariablen untersucht (vgl. Tabelle 14).

Alle kontinuierlichen potentiellen Moderatorvariablen wurden anhand des Medians dichotomisiert. Aus inhaltlichen Gründen wurden für folgende fünf Variablen andere Fraktionierungskriterien gewählt: Die *„Hierarchie der Ratees“* variiert zwischen der ersten und vierten Führungsebene. Bei der Gruppenbildung wurden die Führungskräfte der ersten drei Ebenen denen der vierten Ebene gegenübergestellt. Letztere bekleiden Tätigkeitsfelder, die stärker im operativen Geschäft verankert sind. *„Kenntnis der Führungsgrundsätze“* teilt sich danach, ob beide Beurteiler angaben, die Führungsgrundsätze zu kennen oder nicht. Die *„Führungsverantwortung der Aufwärtsrater“* ist ohnehin ein dichotomes Merkmal, sie ist gegeben oder nicht. *„Stellvertretung*“ wurde mit ja (nein) kodiert, wenn der Aufwärtsbeurteiler seine Führungskraft bei deren Abwesenheit (nicht) vertritt. Letztlich wurde für die Variable *„Beschreibungsebene: ergebnisbezogene Itemformulierung“* nicht die Stichprobe nach einem Kriterium (z.B. Hierarchie) in zwei Kategorien geteilt. Stattdessen wurden die zu beurteilenden Verhaltensweisen (Items) kategorisiert. Diese Kategorisierung erfolgte danach, ob das Ergebnis der zu beurteilenden Verhaltensweise zum Zeitpunkt der Verhaltensausführung vom Mitarbeiter prinzipiell beobachtbar (vs. nicht beobachtbar) ist (Kodierung der Items vgl. Fragestellung und Anhang K). Der Vergleich der interperspektivischen Übereinstimmung zwischen ergebnisbezogenen Items (m = 14) und nicht ergebnisbezogenen Items (m = 28) kann also für die Gesamtstichprobe durchgeführt werden, da alle Befragten alle einbezogenen Items beantworteten. Um den Stichprobenumfang nicht

zu verringern, wurden lediglich die 42 Items der fünf Führungsfaktoren, nicht die acht Items der zwei zusätzlichen Dimensionen berücksichtigt.

Zur besseren Übersichtlichkeit sind für die Mehrzahl der Moderatorvariablen die Selbst-Fremd- bzw. Fremd-Fremd-Zusammenhänge für die *Gesamtskala*, bestehend aus den fünf FzFü-Faktoren, dargestellt. Bei einer dimensionalen Auswertung wäre nämlich eine unübersichtliche Menge an Ergebnissen entstanden, da dann über 500 Moderatoranalysen notwendig gewesen wären. Die Moderatoranalysen wurden lediglich für diejenigen Variablen, die selbst faktorbezogen operationalisiert sind, auf Faktorebene durchgeführt. Das betrifft „Wichtigkeit Ratee“, „Ideal-Ist Ratee“, „Wichtigkeit Aufwärts“, „Ideal Selbst-Aufwärts-Diskrepanz“. Auf den Seiten 326ff. sind auch für das Führungsverhalten (MLQ-Kontinuum) faktorbezogene Ergebnisse dargestellt.

Tabelle 53 bis Tabelle 58 berichten die Ergebnisse der Moderatoranalysen. Die ersten vier Tabellen geben die *Selbst*-Fremd-Korrelationen bzw. -Diskrepanzen für die jeweiligen Subgruppen an, die aus der Dichotomisierung der Moderatorvariablen entstanden sind, die letzten beiden Tabellen die entsprechenden Werte für die Abwärts-Aufwärts-Übereinstimmung. In der vorletzten Spalte sind jeweils die im Folgenden nicht weiter beachteten Ergebnisse der Signifikanzprüfung auf Unterschiede zwischen den beiden Indizes aufgeführt. Ausgangspunkt für die aggregierte Auswertung der Moderatoreffekte stellt die letzte Spalte dar, in der angegeben ist, ob die Richtung der Moderatorhypothese bestätigt werden konnte (unabhängig von der Stärke des Effekts).

Da es allein um die Richtung geht, wurde auf Angaben der Effektgrößen genauso verzichtet wie auf die Durchführung nonparametrischer Verfahren für die Analyse derjenigen Variablen, bei denen sich aufgrund der Dichotomisierung für eine Gruppe ein sehr geringer Stichprobenumfang ergab. Da einzelne Variablen nicht übergewichtet werden sollten, gingen alle faktorbezogen ausgewerteten Variablen sowie die transformationale Führung jeweils nur mit einem einzigen Wert in die aggregierte Analyse ein. Entsprechend blieben in der Auswertung die in Klammern gesetzten Angaben aus den Tabellen jeweils unberücksichtigt. Die Richtung, in die die Mehrzahl der Faktoren bzw. der Gesamtwert der transformationalen Skalen wies, wurde der Variable für die aggregierte Auswertung zugesprochen. *„Wichtigkeit Ratee“* und *„Ideal-Ist Ratee“* gingen gar nicht in die Gesamtauswertung ein, da für sie lediglich explorative Fragestellungen formuliert wurden (vgl. Kapitel 4).

Zunächst werden moderierende Effekte auf die **Selbst-Aufwärts**-Übereinstimmung betrachtet. Auch wenn im vorliegenden Kapitel nicht jede Variable gesondert interpretiert wird, sondern der Zugang überwiegend in aggregierter Form stattfindet, gehen viele Ergebnisse in das Kapitel 7.3 erneut ein.

Tabelle 53. Ergebnisse der Moderatoranalysen zur Selbst-Aufwärts-Korrelation

Moderatorvariable (MV)	Selbst-Aufwärts-Korrelation (*N*) unter Dichotomisierungskriterium der MV				*z*, *p*	Richtung der Hypothese bestätigt
			über Dichotomisierungskriterium der MV			
Personale Faktoren (Selbstrater/Ratee):						
Iia	.17	(15)	<.01	(16)	.43, $p = .67$	(nein)
Iib	.03	(15)	.12	(16)	-.23, $p = .41$	(ja)
IM	.12	(12)	-.31	(19)	1.06, $p = .86$	(nein)
IS	.16	(15)	-.05	(16)	.53, $p = .70$	(nein)
IC	.04	(16)	.03	(15)	.03, $p = .51$	(nein)
AUS	.06	(18)	.31	(22)	-.75, $p = .23$	(ja)
Transformationale Führung ges.	.08	(16)	.04	(15)	.10, $p = .54$	nein
CR	.20	(15)	-.33	(16)	1.36, $p = .91$	nein
MBA	-.02	(18)	.01	(13)	-.07, $p = .47$	nein
MBP	.27	(15)	.09	(16)	.47, $p = .68$	ja
Laissez Faire	-.16	(15)	.20	(16)	-.91, $p = .18$	ja
Alter Ratee	.25	(18)	.21	(20)	.12, $p = .55$	nein
Wichtigkeit Ratee (E/F)	-.57	(11)	.63	(11)	-2.78, $p < .01$	
Wichtigkeit Ratee (B/S)	.10	(12)	.05	(10)	.10, $p = .54$	
Wichtigkeit Ratee (IB)	.64	(11)	.33	(11)	.83, $p = .80$	
Wichtigkeit Ratee (K)	.11	(13)	.78	(9)	-1.81, $p = .04$	
Wichtigkeit Ratee (P)	-.29	(12)	-.48	(10)	.45, $p = .67$	
Wichtigkeit Ratee (FvF)	-.05	(6)	-.38	(10)	.51, $p = .69$	
Wichtigkeit Ratee (S)	-.53	(6)	-.14	(3)	zu geringes *N*	
Ideal-Ist Ratee (E/F)	.69	(10)	.03	(11)	1.58, $p = .94$	
Ideal-Ist Ratee (B/S)	-.03	(9)	.19	(12)	-.42, $p = .34$	
Ideal-Ist Ratee (IB)	-.15	(9)	.38	(12)	-1.05, $p = .15$	
Ideal-Ist Ratee (K)	.48	(9)	.22	(12)	.57, $p = .71$	
Ideal-Ist Ratee (P)	-.01	(9)	-.32	(12)	.61, $p = .73$	
Ideal-Ist Ratee (FvF)	-.40	(8)	.21	(8)	-1.01, $p = .16$	
Ideal-Ist Ratee (S)	.38	(6)	-.47	(10)	1.32, $p = .91$	
Personale Faktoren (Fremdrater):						
Wichtigkeit Aufwärts (E/F)	.23	(20)	.24	(13)	-.03, $p = .49$	(nein)
Wichtigkeit Aufwärts (B/S)	.06	(22)	-.28	(11)	.83, $p = .80$	(ja)
Wichtigkeit Aufwärts (IB)	.06	(18)	-.01	(15)	.18, $p = .57$	(ja)
Wichtigkeit Aufwärts (K)	.54	(16)	.22	(17)	.99, $p = .84$	(ja)
Wichtigkeit Aufwärts (P)	-.37	(21)	-.16	(12)	-.56, $p = .29$	(nein)
Wichtigkeit Aufwärts (FvF)	.09	(13)	.09	(13)	.00, $p = .50$	(=)
Wichtigkeit Aufwärts (S)	-.24	(10)	.23	(8)	-.82, $p = .21$	(nein)
Wichtigkeit Aufwärts						ja[#]
Alter Aufwärtsrater	.03	(22)	.34	(18)	-.94, $p = .17$	ja
Organisations- und tätigkeitsbezogene Faktoren:						
Führungsspanne (Ratee)	.35	(21)	-.03	(19)	1.15, $p = .88$	ja
Hierarchie (Ratee)[a]	<.01	(22)	.26	(18)	-.77, $p = .22$	ja
Dauer der Organisationszugehörigkeit (Aufwärtsrater)	.16	(16)	.05	(23)	.31, $p = .62$	nein
Dauer der Abteilungszugehörigkeit (Aufwärtsrater)	.44	(21)	-.09	(19)	1.64, $p = .95$	nein
Dauer der Zusammenarbeit FK-MA (Aufwärtsrater)	.27	(19)	.04	(21)	.69, $p = .75$	nein
Dauer der Organisationszugehörigkeit (Selbstrater)	.41	(19)	.03	(19)	1.15, $p = .87$	nein
Durchführungsbedingungen (Format des Befragungsinstruments):						
Beschreibungsebene: ergebnisbezogene Itemformulierung[b]	.18	(40)	.21	(40)	-.13, $p = .45$	ja

Durchführungsbedingungen (allgemeine Rahmenbedingungen):						
Erfahrung als Beurteilter (Ratee)	.11	(23)	.15	(13)	-.13, p = .45	ja
Erfahrung als Beurteiler (Ratee)	.16	(24)	.21	(16)	-.15, p = .44	ja
Beteiligungsquote	-.12	(19)	.41	(21)	-1.62, p = .05	ja
Ideal S-Auf-Diskrepanz (E/F)	.52	(10)	-.46	(11)	2.07, p = .98	(ja)
Ideal S-Auf-Diskrepanz (B/S)	.37	(10)	-.18	(11)	1.10, p = .86	(ja)
Ideal S-Auf-Diskrepanz (IB)	.49	(10)	-.02	(11)	1.07, p = .86	(ja)
Ideal S-Auf-Diskrepanz (K)	.43	(10)	.11	(11)	.68, p = .75	(ja)
Ideal S-Auf-Diskrepanz (P)	.29	(10)	-.67	(11)	2.14, p = .98	(ja)
Ideal S-Auf-Diskrepanz (FvF)	-.38	(5)	-.78	(5)	.65, p = .74	(ja)
Ideal S-Auf-Diskrepanz (S)	-.80	(4)	-.39	(5)	-.56, p = .71	(nein)
Ideal S-Auf-Diskrepanz						ja
Kenntnis der Führungsgrundsätze[c]	.06	(77)	.18	(80)	.75, p = .77	ja
Führungsverantwortung der Aufwärtsrater[d]	.12	(28)	.67	(7)	1.28, p = .90	ja
Aufwärtsrater als Stellvertreter[e]	.16	(106)	.12	(35)	.20, p = .58	nein

Anmerkungen. [a]links: Hierarchieebene 1-3, rechts: Hierarchieebene 4, [b]links: nicht ergebnisbezogene Itemformulierung, rechts: ergebnisbezogene Itemformulierung, [c]nicht-aggregierte Ratings, links: mindestens ein Beurteiler ohne Kenntnis der Führungsgrundsätze, rechts: beide Beurteiler kennen die Führungsgrundsätze, [d]links: höchstens 15% der Aufwärtsbeurteiler tragen Führungsverantwortung, rechts: mindestens 75% der Aufwärtsbeurteiler tragen Führungsverantwortung. (In der Regel haben alle oder keine Mitarbeiter Führungsverantwortung.), [e]nicht-aggregierte Ratings, links: Aufwärtsbeurteiler ohne Stellvertreterrolle, rechts: Aufwärtsbeurteiler mit Stellvertreterrolle, [#]Die Variable geht mit „ja“ in die aggregierte Auswertung ein, da das Skalenmittel in die erwartete Richtung tendiert (z = .36).

Tabelle 53 zeigt, dass sich bei 13 der 22 untersuchten Variablen die Selbst-Aufwärts-Korrelation in die erwartete Richtung verändert. Der Binomialtest stellt dies als nicht signifikant heraus (p = .26).[168] Allerdings zeigt sich erstens für diejenigen Moderatorvariablen, die den personalen Faktoren des Fremdraters zuzuordnen sind, dass beide untersuchten Moderatorvariablen in die richtige Richtung weisen (p = .25) und zweitens für diejenigen Moderatorvariablen, die den Durchführungsbedingungen zuzuordnen sind, dass sechs der sieben untersuchten Variablen in die richtige Richtung weisen (p = .06). Unter den organisations-/tätigkeitsbezogenen Variablen weisen beide strukturellen Bedingungen (Führungsspanne, Hierarchie) Effekte in die richtige Richtung auf, während die Variablen, die auf die Zugehörigkeit abzielen, allesamt unerwartete Vorzeichen tragen.

Von den Einzeleffekten erreichen lediglich vier der faktorenbezogenen Analysen Signifikanz. Letztlich sind überwiegend nur geringe Effekte festzuhalten. Hypothese 16 ist für die Selbst-Aufwärts-Korrelation vorläufig abzulehnen, wenn man alle Moderatorvariablen gemeinsam betrachtet. Vor allem die Variablen der Durchführungsbedingungen sollten weiter betrachtet werden.

[168] Wahrscheinlichkeit, dass mindestens 13 von 22 Effekten in die erwartete Richtung weisen.

Tabelle 54. Ergebnisse der Moderatoranalysen zur Selbst-Aufwärts-Diskrepanz

Moderatorvariable (MV)	Selbst-Aufwärts-Diskrepanz (*N*) unter Dichotomisierungskriterium der MV	Selbst-Aufwärts-Diskrepanz (*N*) über Dichotomisierungskriterium der MV	*t, p*	Richtung der Hypothese bestätigt
Personale Faktoren (Selbstrater/Ratee):				
IIa	1.71 (15)	.87 (16)	$t(29) = 4.44, p < .01$	(ja)
IIb	1.67 (15)	.91 (16)	$^{\circ}t(23) = 3.71, p < .01$	(ja)
IM	1.62 (12)	1.06 (19)	$t(29) = 2.46, p = .02$	(ja)
IS	1.67 (15)	.91 (16)	$^{\circ}t(23) = 3.78, p < .01$	(ja)
IC	1.55 (16)	.99 (15)	$^{\circ}t(25) = 2.58, p = .02$	(ja)
AUS	1.61 (18)	.73 (22)	$^{\circ}t(28) = 5.01, p < .01$	(ja)
Transformationale Führung ges.	1.60 (16)	.93 (15)	$^{\circ}t(26) = 3.23, p < .01$	ja
CR	1.47 (15)	1.10 (16)	$t(29) = 1.59, p = .12$	ja
MBA	1.32 (18)	1.22 (13)	$t(29) = .44, p = .66$	nein
MBP	.94 (15)	1.59 (16)	$^{\circ}t(23) = -3.10, p < .01$	ja
Laissez Faire	1.06 (15)	1.49 (16)	$t(29) = -1.86, p = .07$	ja
Alter Ratee	1.20 (18)	1.10 (20)	$t(36) = .43, p = .67$	ja
Wichtigkeit Ratee (E/F)	1.10 (11)	1.76 (11)	$t(20) = -2.79, p = .01$	
Wichtigkeit Ratee (B/S)	1.12 (12)	1.47 (10)	$t(20) = -1.35, p = .19$	
Wichtigkeit Ratee (IB)	.65 (11)	1.07 (11)	$^{\circ}t(15) = -2.02, p = .06$	
Wichtigkeit Ratee (K)	1.15 (13)	1.14 (9)	$t(20) = .04, p = .97$	
Wichtigkeit Ratee (P)	1.38 (12)	1.48 (10)	$t(20) = -.30, p = .77$	
Wichtigkeit Ratee (FvF)	.85 (6)	1.47 (10)	$t(14) = -1.87, p = .08$	
Wichtigkeit Ratee (S)	1.12 (6)	2.21 (3)	$t(7) = -1.78, p = .12$	
Ideal-Ist Ratee (E/F)	1.60 (10)	1.29 (11)	$t(19) = 1.11, p = .28$	
Ideal-Ist Ratee (B/S)	1.26 (9)	1.30 (12)	$t(19) = -.17, p = .87$	
Ideal-Ist Ratee (IB)	.94 (9)	.82 (12)	$t(19) = -.79, p = .44$	
Ideal-Ist Ratee (K)	1.14 (9)	1.20 (12)	$t(19) = .03, p = .98$	
Ideal-Ist Ratee (P)	1.46 (9)	1.46 (12)	$t(19) = -.82, p = .42$	
Ideal-Ist Ratee (FvF)	1.34 (8)	1.08 (8)	$t(19) = -.39, p = .70$	
Ideal-Ist Ratee (S)	1.15 (10)	1.18 (11)	$t(13) = -.97, p = .35$	
Personale Faktoren (Fremdrater):				
Wichtigkeit Aufwärts (E/F)	1.28 (20)	1.58 (13)	$t(31) = -1.09, p = .28$	(ja)
Wichtigkeit Aufwärts (B/S)	1.43 (22)	1.67 (11)	$t(31) = -.79, p = .44$	(ja)
Wichtigkeit Aufwärts (IB)	.89 (18)	1.11 (15)	$t(31) = -97, p = .34$	(ja)
Wichtigkeit Aufwärts (K)	1.41 (16)	.91 (17)	$t(31) = 2.23, p = .03$	(nein)
Wichtigkeit Aufwärts (P)	1.70 (21)	1.15 (12)	$t(31) = 1.80, p = .08$	(nein)
Wichtigkeit Aufwärts (FvF)	.87 (13)	1.43 (13)	$t(24) = -2.10, p = .05$	(ja)
Wichtigkeit Aufwärts (S)	1.11 (9)	.82 (8)	$t(15) = .59, p = .57$	(nein)
Wichtigkeit Aufwärts				ja
Alter Aufwärtsrater	1.23 (22)	.99 (18)	$t(36) = .43, p = .67$	ja
Organisations- und tätigkeitsbezogene Faktoren:				
Führungsspanne (Ratee)	1.00 (21)	1.26 (19)	$t(38) = -1.18, p = .24$	ja
Hierarchie (Ratee)[a]	.96 (22)	1.32 (18)	$t(38) = -1.68, p = .10$	ja
Dauer der Organisations-zugehörigkeit (Aufwärtsrater)	1.15 (16)	1.11 (23)	$t(37) = .14, p = .89$	ja
Dauer der Abteilungs-zugehörigkeit (Aufwärtsrater)	1.03 (21)	1.23 (19)	$^{\circ}t(29) = -.87, p = .38$	nein
Dauer der Zusammenarbeit FK-MA (Aufwärtsrater)	1.02 (19)	1.21 (21)	$t(38) = -.90, p = .38$	nein
Dauer der Organisations-zugehörigkeit (Selbstrater)	.99 (19)	1.30 (19)	$t(36) = -.37, p = .18$	nein
Durchführungsbedingungen (Format des Befragungsinstruments):				
Beschreibungsebene: ergebnis-bezogene Itemformulierung[b]	1.16 (40)	1.12 (40)	$t(39) = -.63, p = .53$	ja

Durchführungsbedingungen (allgemeine Rahmenbedingungen):						
Erfahrung als Beurteilter (Ratee)	1.04	(23)	1.35	(13)	$t(34) = -1.33$, $p = .19$	nein
Erfahrung als Beurteiler (Ratee)	1.21	(24)	1.00	(16)	$t(37) = 1.04$, $p = .31$	ja
Beteiligungsquote	1.21	(19)	1.04	(21)	$t(38) = .76$, $p = .45$	ja
Ideal S-Auf-Diskrepanz (E/F)	1.38	(10)	1.49	(11)	$^{\circ}t(13) = -.40$, $p = .70$	(ja)
Ideal S-Auf-Diskrepanz (B/S)	1.09	(10)	1.46	(11)	$t(19) = -1.41$, $p = .18$	(ja)
Ideal S-Auf-Diskrepanz (IB)	.67	(10)	1.05	(11)	$t(19) = -1.71$, $p = .10$	(ja)
Ideal S-Auf-Diskrepanz (K)	1.16	(10)	1.19	(11)	$t(19) = -.14$, $p = .89$	(ja)
Ideal S-Auf-Diskrepanz (P)	1.15	(10)	1.75	(11)	$t(19) = -1.92$, $p = .07$	(ja)
Ideal S-Auf-Diskrepanz (FvF)	1.07	(5)	1.81	(5)	$t(8) = -1.82$, $p = .11$	(ja)
Ideal S-Auf-Diskrepanz (S)	.93	(4)	.98	(5)	$t(7) = -.12$, $p = .91$	(ja)
Ideal S-Auf-Diskrepanz						ja
Kenntnis der Führungsgrundsätze[c]	.94	(77)	.61	(80)	$t(155) = -1.72$, $p = .09$	ja
Führungsverantwortung der Aufwärtsrater[d]	1.23	(28)	.43	(7)	$t(33) = -3.02$, $p < .01$	ja
Aufwärtsrater als Stellvertreter[e]	1.05	(106)	38	(35)	$t(139) = -2.86$, $p < .01$	ja

Anmerkungen. °keine Varianzgleichheit (also weniger *df*), [a]links: Hierarchieebene 1-3, rechts: Hierarchieebene 4, [b]links: nicht ergebnisbezogene Itemformulierung, rechts: ergebnisbezogene Itemformulierung, [c]nicht-aggregierte Ratings, links: mindestens ein Beurteiler ohne Kenntnis der Führungsgrundsätze, rechts: beide Beurteiler kennen die Führungsgrundsätze, [d]links: keine Führungsverantwortung der Aufwärtsbeurteiler, rechts: Führungsverantwortung der Aufwärtsbeurteiler, [e]nicht-aggregierte Ratings, links: Aufwärtsbeurteiler ohne Stellvertreterrolle, rechts: Aufwärtsbeurteiler mit Stellvertreterrolle.

Aus Tabelle 54 wird ersichtlich, dass auch bezüglich der Selbst-Aufwärts-Diskrepanz (Niveau) die Mehrheit der Einzelprüfungen von den Ergebnissen her in die Richtung der formulierten Hypothesen geht. Hypothese 16) kann für die Selbst-Aufwärts-Diskrepanz also aufrechterhalten werden, da insgesamt 17 der 22 Indizes erwartungsgetreu ausfallen, was im Binomialtest Signifikanz erreicht ($p = .01$).

Generell ist also festzuhalten, dass eher moderierende Effekte auf die Selbst-Aufwärts-*Niveau-Unterschiede* als auf die Selbst-Aufwärts-*Korrelation* anzunehmen sind. Ein solcher Unterschied ergab sich für die Selbst-*Abwärts*-Übereinstimmung auch in der Meta-Analyse von Heidemeier und Moser (2002). Das bedeutet, dass bislang eher Moderatorvariablen für die Genauigkeit und weniger für die Validität von Selbstratings aufgedeckt werden konnten.

Genauer betrachtet kristallisiert sich heraus, dass personale Faktoren (sieben von acht Effektrichtungen erwartungsgemäß, $p = .04$) sowie Durchführungsbedingungen (sechs von sieben Effektrichtungen erwartungsgemäß, $p = .06$) moderierende Wirkung auf die Selbst-Aufwärts-Niveau-Unterschiede nehmen. Hier ergeben sich teilweise sogar große Effekte, die deshalb besonders stark zu gewichten sind, weil – wie oben erklärt – in Einzelstudien, in denen bereits validitätserhöhende Bedingungen realisiert sind, kaum große Effekte zu erwarten sind. Hervorzuheben ist die Führungsverantwortung der Aufwärtsrater, die sowohl auf die Selbst-Aufwärts-Korrelation als auch auf die Selbst-Aufwärts-Diskrepanz einen starken moderierenden Effekt ausübt ($q = .69$ bzw. $d = 1.05$). Dieser Befund wird unten ausführlicher diskutiert (Kapitel 7.3).

Uneindeutig verhalten sich die organisations-/tätigkeitsbezogenen Moderatorvariablen, von denen jeweils drei in die erwartete bzw. unerwartete Richtung weisen. Das deckt sich mit den Moderatoranalysen für die Selbst-Aufwärts-Korrelation insofern, dass auch hier die Effekte der strukturbezogenen Variablen (Führungsspanne, Hierarchie) erwartungsgemäß ausfallen und am stärksten sind, nicht die der Zugehörigkeitsdauer. Bei höherer Führungsspanne und niedrigerer Hierarchieebene sind die Diskrepanzen größer.

Es kann also festgehalten werden, dass von den untersuchten Moderatorvariablen für den Selbst-Aufwärts-Zusammenhang (Korrelation, Niveau-Unterschiede) die Mehrheit der Einzelprüfungen in die erwartete Richtung weist. Insbesondere personale und strukturbezogene Variablen sowie Durchführungsbedingungen scheinen Einfluss auf den Selbst-Aufwärts-Zusammenhang zu nehmen. Auffallend ist, dass die personalen Faktoren der Selbstrater (Führungskraft) insbesondere die Niveau-Unterschiede moderieren. So zeigt sich im Unterschied zum Einfluss auf den korrelativen Zusammenhang ein mittlerer moderierender Effekt von CR (d = .59) und ein großer Effekt der transformationalen Führung (d = 1.27) auf die Diskrepanz.[169] Für die MLQ-Dimension AUS (Ausstrahlung) kann zusätzlich angemerkt werden, dass sie nicht nur für die Selbst-Aufwärts-Niveau-Unterschiede, sondern auch für die Selbst-Aufwärts-Korrelation unter den transformationalen Führungsdimensionen diejenige mit der stärksten moderierenden Wirkung darstellt. Das spricht für die Erweiterung des MLQ genau um diese Dimension, die erst von Felfe und Goihl (2003) vorgenommen wurde.

Auf den ersten Blick scheint es also so zu sein, dass die personalen Faktoren der Selbstrater, insbesondere ihr Führungsstil, starken Einfluss auf das Ausmaß der Selbstüberschätzung nehmen. Je aktiver die Führung, desto geringer scheint die Selbst-Aufwärts-Diskrepanz zu sein. Um genauer zu untersuchen, ob das Führungsverhalten selbst das Ausmaß der Übereinstimmung zwischen Selbst- und Fremdbild moderiert, wird der Full Range of Leadership im Folgenden weiter betrachtet.

Abbildung 28/links illustriert den Zusammenhang zwischen der Selbst-Aufwärts-Diskrepanz (FzFü) und verschiedenen Führungsfacetten (MLQ-Kontinuum von Laissez Faire bis Charisma). Mitarbeiter, deren Führungsbeurteilung gering (vs. stark, Mediansplit) von der Selbstbeurteilung ihrer Führungskraft abweicht, schätzen deren Führungsverhalten stärker transformational und weniger passiv ein. Mit Ausnahme von MBA erreichen die Unterschiede mindestens das 5%-Signifikanzniveau. Diese Ergebnisse entsprechen denen von Tartler, Goihl et al. (2003), gehen aber in einem zentralen Punkt über diese hinaus: Während bei Tartler, Goihl et al. (2003) auch die Führungsbeurteilungen, aus denen die Selbst-Aufwärts-Diskrepanz errechnet wurde, auf MLQ-Skalen beruhen, sind dies in der vorliegenden Untersuchung konkretere Führungsaufgaben (FzFü-Dimensionen). Das

[169] Im Unterschied zum korrelativen Zusammenhang weist für die Niveau-Unterschiede auch der Effekt des Alters (Ratee) in die richtige Richtung, auch wenn die Stärke des Effekts unbedeutend bleibt (d = .14).

unterstützt auf den ersten Blick sogar mit zusätzlichem Gewicht die Annahme, dass eine transformationale Führungsweise mit einer genaueren Führungsbeurteilung (auch Selbstbeurteilung) einhergeht.

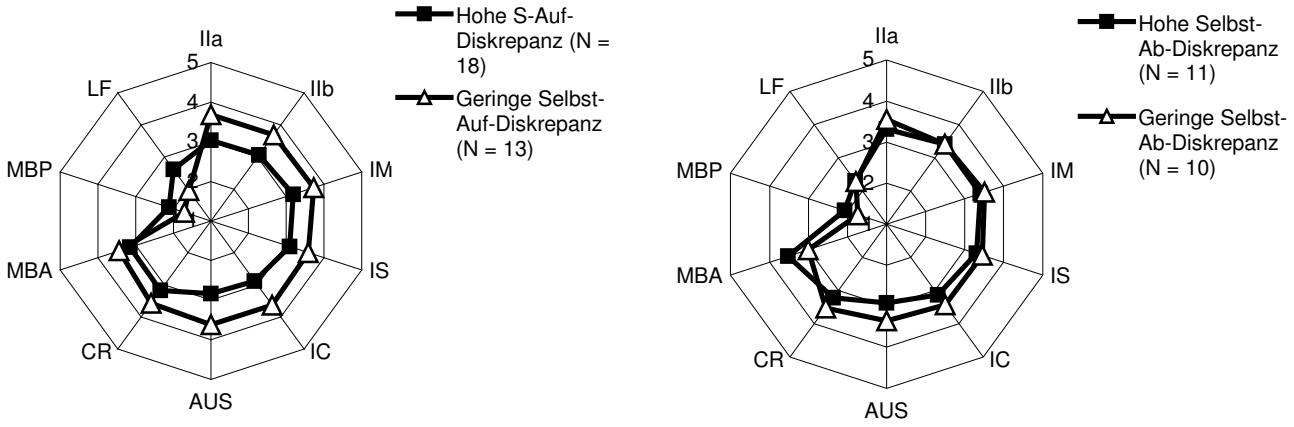


Abbildung 28. MLQ-Beurteilungen (Aufwärts) in Abhängigkeit von den Selbst-Fremd-Niveau-Unterschieden

Entgegen der soeben genannten Interpretation können die moderierenden Effekte der transformationalen Führung und von CR (siehe oben) aber auch als Artefakt erklärt werden. Es zeigt sich nämlich, dass sich nicht die Selbstbeurteilung, sondern lediglich die Aufwärtsbeurteilung zwischen den hoch vs. gering transformational führenden Ratees sowie zwischen den hoch vs. gering kontingent belohnenden Ratees unterscheidet. Hoch transformational und kontingent belohnende Führungskräfte bewerten sich selbst auf den FzFü-Faktoren nicht positiver, werden aber von ihren Mitarbeitern auf allen FzFü-Faktoren signifikant positiver bewertet als gering transformational und kontingent belohnende Führungskräfte. Nach den Befunden dieser Studie ist also eher anzuzweifeln, dass die geringeren Diskrepanzen tatsächlich der Führung an sich zuzuschreiben sind. Es ist vielmehr davon auszugehen, dass die Aufwärtsbeurteiler vor dem Hintergrund ihrer impliziten Führungstheorien ihre Führungskräfte generell positiv bewerten, so dass quasi automatisch eine positive Bewertung der transformationalen Führung mit einer positiven Bewertung der FzFü-Faktoren einhergeht und letztere ebenso automatisch mit geringeren Selbst-Aufwärts-Niveau-Unterschieden.

Die Diskrepanz zwischen der Selbst- und Aufwärtsbeurteilung fällt bei stark transformational und kontingent belohnenden Führungskräften also „nur“ deshalb geringer aus, weil die Aufwärtsbeurteiler ihre Führungskräfte auch auf den FzFü-Faktoren positiver bewerten. Vor diesem Hintergrund sind auch Interpretationen anderer Studien (z.B. Tartler, Goihl et al., 2003) zu überdenken, in denen die geringeren Selbst-Aufwärts-Diskrepanzen bei gleichzeitig stärkerer transformationalen Führung auf eben letztere zurückgeführt werden.

Ein besserer Beleg für den Einfluss des Führungsverhaltens auf eine akkuratere Selbstbeurteilung und demzufolge eine geringere Selbst-Fremd-Diskrepanz könnte allenfalls dann erbracht werden, wenn als Fremdperspektive die Abwärtsbeurteilung herangezogen wird. Dann sind nämlich beide Beurteilungsperspektiven, die zur Berechnung der Diskrepanz herangezogen werden (Selbst, Abwärts), unabhängig von der Beurteilung des transformationalen/transaktionalen Führungsverhaltens (Aufwärtsbeurteilung durch Mitarbeiter).

Für den Vergleich der Selbst- mit der Abwärtsperspektive, die sich im Niveau ohnehin nur gering unterscheiden (siehe oben), zeigt sich auf den einzelnen Dimensionen der erwartete Trend in der Ausprägung der Führungsdimensionen (vgl. Abbildung 28/rechts), jedoch jeweils mit sehr geringen bis unbedeutenden Effekten (maximal d = .18). Die Führungskräfte, deren Selbstbeurteilung kaum von der entsprechenden Abwärtsbeurteilung abweicht, erhalten von ihren Mitarbeitern minimal höhere Werte auf den Dimensionen aktiver Führung als Führungskräfte, deren Selbstbeurteilung stark von der entsprechenden Abwärtsbeurteilung abweicht. Führungskräfte mit hohen Werten auf den transformationalen Skalen scheinen sich selbst tendenziell eher negativer zu beurteilen als sie von ihren Vorgesetzten beurteilt werden (FzFü-Dimensionen) (vgl. Tabelle 56).

Aufgrund dieses Designs, das den Single-Source-Bias ausschließt, ist schlussfolgernd festzuhalten, dass die Art der Führung allenfalls einen geringen Einfluss auf die Selbst-Fremd-Diskrepanz nehmen kann. So geht eine geringere Selbst-Abwärts-Diskrepanz in geringem Ausmaß mit einer Führung einher, die von unten als transformational charakterisiert wird.

Im Folgenden werden potentielle Moderatorwirkungen auf den **Selbst-Abwärts**-Zusammenhang überprüft. Insbesondere für die Selbst-Abwärts-Niveau-Unterschiede sind im Vergleich zum Selbst-Aufwärts-Zusammenhang weniger Effekte zu erwarten, da sich die Mittelwerte ohnehin kaum unterscheiden. Wo sich aber Tendenzen ergeben, sind diese hervorzuheben, da sie im Vergleich zum Selbst-Aufwärts-Zusammenhang für die Variablen, die durch Mitarbeiterratings erhoben wurden, den Single-Source-Bias ausschalten.

Tabelle 55. Ergebnisse der Moderatoranalysen zur Selbst-Abwärts-Korrelation

Moderatorvariable (MV)	Selbst-Abwärts-Korrelation (*N*) unter Dichotomisierungskriterium der MV				*z, p*	Richtung der Hypothese bestätigt
			über Dichotomisierungskriterium der MV			
Personale Faktoren (Selbstrater/Ratee):						
IIa	.30	(10)	-.52	(11)	1.71, *p* = .96	(nein)
IIb	.08	(11)	-.10	(10)	.35, *p* = .64	(nein)
IM	.29	(8)	-.25	(13)	.82, *p* = .79	(nein)
IS	.17	(11)	-.20	(10)	.72, *p* = .77	(nein)
IC	.17	(11)	-.17	(10)	.66, *p* = .75	(nein)
AUS	.25	(12)	-.23	(12)	1.04, *p* = .85	(nein)
Transformationale Führung ges.	.10	(10)	-.20	(11)	.59, *p* = .72	nein
CR	.28	(9)	-.26	(12)	1.05, *p* = .85	nein
MBA	-.30	(11)	.19	(10)	-.97, *p* = .17	nein
MBP	-.22	(10)	.41	(11)	-1.27, *p* = .10	nein
Laissez Faire	-.20	(10)	.14	(11)	-.66, *p* = .25	nein
Alter Ratee	-.14	(13)	.41	(17)	-1.39, *p* = .08	ja
Wichtigkeit Ratee (E/F)	.11	(8)	.32	(9)	-.37, *p* = .36	
Wichtigkeit Ratee (B/S)	.34	(9)	.40	(8)	-.11, *p* = .45	
Wichtigkeit Ratee (IB)	.18	(8)	-.10	(9)	.47, *p* = .68	
Wichtigkeit Ratee (K)	.16	(7)	-.02	(10)	.29, *p* = .61	
Wichtigkeit Ratee (P)	-.12	(9)	-.17	(8)	.08, *p* = .53	
Wichtigkeit Ratee (FvF)	-.29	(6)	-.03	(11)	-.40, *p* = .35	
Wichtigkeit Ratee (S)	.21	(4)	.97	(4)	-1.33, *p* = .09	
Ideal-Ist Ratee (E/F)	.03	(6)	.45	(11)	-.82, *p* = .21	
Ideal-Ist Ratee (B/S)	.15	(6)	.54	(11)	-.67, *p* = .25	
Ideal-Ist Ratee (IB)	.03	(7)	-.10	(10)	.21, *p* = .58	
Ideal-Ist Ratee (K)	.46	(7)	-.27	(10)	1.20, *p* = .89	
Ideal-Ist Ratee (P)	-.14	(8)	-.07	(9)	-.12, *p* = .45	
Ideal-Ist Ratee (FvF)	-.18	(8)	.48	(9)	-1.16, *p* = .12	
Ideal-Ist Ratee (S)	.50	(6)	-.03	(11)	.86, *p* = .80	
Personale Faktoren (Fremdrater):						
Alter Abwärtsrater	-.05	(9)	.19	(15)	-.48, *p* = .31	ja
Organisations- und tätigkeitsbezogene Faktoren:						
Führungsspanne (Ratee)	.48	(16)	-.25	(16)	1.98, *p* = .98	ja
Führungsspanne (Abwärtsrater)	.05	(21)	-.31	(10)	.83, *p* = .80	ja
Hierarchie (Ratee)[a]	.18	(12)	.26	(20)	-.20, *p* = .42	ja
Dauer der Organisationszugehörigkeit (Selbstrater)	.28	(14)	.39	(16)	-.30, *p* = .38	ja
Dauer der Abteilungszugehörigkeit (Selbstrater)	.17	(13)	.33	(17)	-.41, *p* = .34	ja
Dauer der Zusammenarbeit von Selbst- und Abwärtsrater	-.22	(14)	.49	(15)	-1.82, *p* = .03	ja
Durchführungsbedingungen (Format des Befragungsinstruments):						
Beschreibungsebene: ergebnisbezogene Itemformulierung[b]	.08	(31)	.18	(31)	-.38, *p* = .35	ja
Durchführungsbedingungen (allgemeine Rahmenbedingungen):						
Erfahrung als Beurteilter (durch Abwärtsbeurteiler)	.17	(16)	.14	(15)	.08, *p* = .53	nein
Erfahrung als Beurteiler (Ratee)	.04	(18)	.32	(6)	-.46, *p* = .32	ja
Beteiligungsquote	-.02	(19)	.28	(12)	-.74, *p* = .23	ja

Anmerkungen. [a]links: Hierarchieebene 1-3, rechts: Hierarchieebene 4, [b]links: nicht ergebnisbezogene Itemformulierung, rechts: ergebnisbezogene Itemformulierung.

Für die Selbst-Abwärts-Korrelation weisen 11 der 17 untersuchten Moderatorvariablen mindestens tendenzielle Effekte in die erwartete Richtung auf (p = .17) (vgl. Tabelle 55). Deutliche Wirkungen zeigen die organisations-/tätigkeitsbezogenen Variablen. Alle sechs untersuchten Moderatoren – also sowohl die strukturellen als auch die Zugehörigkeitsbedingungen – weisen in die erwartete Richtung (p = .02). Zwei davon erreichen sogar große Effekte (q = .78 bzw. q = .76):

Die Selbst-Abwärts-Korrelation ist dann höher, wenn die Selbstbeurteiler bereits länger der Organisation (q = .13) oder der Abteilung angehören (q = .17). Ein starker Effekt ergibt sich für die Dauer der Zusammenarbeit von Selbst- und Abwärtsbeurteilern (q = .76). Es könnte angenommen werden, dass sich im Laufe der Zusammenarbeit die Leistungskriterien zwischen Abwärts- und Selbstbeurteilern aneinander angleichen, so dass die Beurteilungs-Items vor dem gleichen Hintergrund interpretiert werden. Dies sollte letztlich in einer übereinstimmenden Beurteilung münden. Dem widerspricht jedoch, dass sich diejenigen Führungskräfte, die bereits länger der Organisation bzw. der Abteilung angehören oder mit ihrem Vorgesetzten zusammen arbeiten, selbst positiver beurteilen als sie von ihren Vorgesetzten beurteilt werden, während sich die Führungskräfte mit jeweils kürzerer Zugehörigkeitsdauer selbst negativer beurteilen als ihre Abwärtsbeurteiler (vgl. Tabelle 56). Dies könnte zum einen auf einer Unsicherheit zu Beginn einer neuen Position beruhen. Dafür spricht, dass der Effekt für die Organisationszugehörigkeit am stärksten ist (d = 1.15), denn ein Organisationswechsel ist auch mit den meisten Unsicherheitsfaktoren verbunden. Zum anderen könnte sich darin eine Anpassung an die Beurteilungskultur der Vorgesetzten widerspiegeln, selbst wenn sie in diesem Fall eher als „Unkultur“ der Überschätzung bezeichnet werden müsste.

Der zweite große Effekt ist für die Führungsspanne des Selbstbeurteilers (Ratee) festzuhalten (q = .78). Dies ist plausibel dadurch zu begründen, dass die Abwärtsbeurteiler das Führungsverhalten ihrer nachgeordneten Führungskräfte (Ratee) zunehmend schlechter beobachten und folglich beurteilen können, je höher deren Führungsspanne ist.

Alle untersuchten Organisationsvariablen bringen letztlich mit sich, dass vermehrt Möglichkeiten des Austauschs zwischen Ratee und Abwärtsrater gegeben sind. Wenn man davon ausgeht, dass sich insbesondere die Führungsspanne auf die Beobachtungsmöglichkeiten auswirkt, wird plausibel, weshalb bei hohen Führungsspannen (Ratee bzw. Abwärtsrater) die Selbst-Abwärts-Korrelation sogar im mittleren negativen Bereich liegt. Dass im Unterschied zum korrelativen Zusammenhang die Selbst-Abwärts-Niveau-Unterschiede von der Zugehörigkeits-/Zusammenarbeitsdauer nicht oder in die unerwartete Richtung beeinflusst sind (vgl. Tabelle 56), mag an dem bereits angesprochenen Sicherheitsgefühl liegen, das sich mit der Dauer zunehmend einstellt und eine positivere Selbstbeurteilung begünstigt.

Deutlich unerwartet fallen die Effekte des Führungsstils auf die Selbst-Abwärts-Korrelation aus (personenbezogene Faktoren). Je mehr aktive Führung und je weniger MBA und passive Führung einer Führungskraft von ihren Mitarbeitern zugesprochen wird, umso „inverser" korrelieren die Selbst- und Abwärtsbeurteilung. Die Korrelationskoeffizienten sind in der Regel ähnlich hoch, kehren jedoch ihr jeweiliges Vorzeichen um. Dies erklärt sich aus den Ergebnissen bezüglich der Selbst-Abwärts-Niveau-Unterschiede (vgl. Tabelle 56). Bei einer stark transformationalen Führung fällt die Selbstbeurteilung unter die Abwärtsbeurteilung, so dass sich auch der korrelative Zusammenhang invertieren muss.

Bei 11 der 17 untersuchten Moderatorvariablen für die Selbst-Abwärts-*Differenz* verändert sich die Selbstüberschätzung in die erwartete Richtung (p = .17). Die erwarteten Einflussbedingungen umfassen insbesondere das Führungsverhalten (transformationale Führung), strukturbezogene Organisations- sowie alle Durchführungsbedingungen. Allerdings ist festzuhalten, dass die Selbstüberschätzung bei vielen Moderatorvariablen nicht nur erwartungsgemäß abnimmt, sondern in eine Selbstunterschätzung „kippt". Da dies nicht hypothesenkonform ist, wird in Tabelle 56 das Ausmaß der Selbst-Abwärts-Differenz (also deren Betrag) als Kriterium zugrunde gelegt. Dann finden sich generell nur sehr geringe Unterschiede und es weisen auch nur 6 der 17 untersuchten Variablen in die erwartete Richtung (p = .93).

Da sich die Niveau-Unterschiede aufgrund dessen, dass sich die Selbst- von der Abwärtsbeurteilung prinzipiell nicht unterscheidet (Kapitel 6.3.2), sehr leicht umkehren, sollen die entsprechenden Moderatoranalysen nicht weiter interpretiert werden. Vorläufig ist Hypothese 16) bezüglich der Selbst-Abwärts-Differenz abzulehnen.

Der meta-analytische Befund von Mabe und West (1982), dass die Validität von Selbstbeurteilungen von der Erfahrung mit Selbstbeurteilungen beeinflusst wird, kann insofern erweitert werden, dass sowohl die Selbst-Aufwärts- als auch die Selbst-Abwärts-Korrelation höher ist, wenn die Führungskraft ihre Mitarbeiter häufig (vs. selten) beurteilt. Unter Berücksichtigung der sehr geringen Effekte soll die Tendenz festgehalten werden, dass Beurteilungserfahrung der Führungskräfte auch in der Rolle als Fremdbeurteiler validitätsfördernd sein kann. Selbst mehr (vs. wenig) beurteilt zu werden, scheint hingegen keine Prozesse zu initiieren, die zu einer genaueren oder valideren Selbstbeurteilung führen.

Tabelle 56: Ergebnisse der Moderatoranalysen zur Selbst-Abwärts-Diskrepanz

Moderatorvariable (MV)	Selbst-Abwärts-Diskrepanz (*N*) unter Dichotomisierungskriterium der MV	über Dichotomisierungskriterium der MV	*t, p*	Richtung der Hypothese bestätigt [a]
Personale Faktoren (Selbstrater/Ratee):				
IIa	.07 (10)	-.15 (11)	$t(19) = .73$, $p = .48$	(nein)
IIb	.09 (11)	-.19 (10)	$t(19) = .87$, $p = .39$	(nein)
IM	-.06 (8)	-.04 (13)	$t(19) = -.07$, $p = .94$	(ja)
IS	.09 (11)	-.20 (10)	$t(19) = .93$, $p = .36$	(nein)
IC	.08 (11)	-.18 (10)	$t(19) = .82$, $p = .42$	(nein)
AUS	.06 (12)	-.16 (12)	$t(22) = .78$, $p = .44$	(nein)
Transformationale Führung ges.	.01 (10)	-.10 (11)	$t(19) = .35$, $p = .73$	nein
CR	-.05 (9)	-.04 (12)	$t(19) = -.05$, $p = .96$	ja
MBA	-.23 (11)	.16 (10)	$t(19) = -1.26$, $p = .23$	nein
MBP	-.41 (10)	.29 (11)	$t(19) = -2.51$, $p = .02$	nein
Laissez Faire	-.02 (10)	-.07 (11)	$t(19) = .16$, $p = .88$	ja
Alter Ratee	.12 (13)	-.03 (17)	$t(28) = .65$, $p = .52$	ja
Wichtigkeit Ratee (E/F)	-.27 (8)	-.11 (9)	$t(15) = -.44$, $p = .67$	
Wichtigkeit Ratee (B/S)	-.58 (9)	.19 (8)	$t(15) = -2.22$, $p = .04$	
Wichtigkeit Ratee (IB)	-.36 (8)	.40 (9)	$t(15) = -2.39$, $p = .03$	
Wichtigkeit Ratee (K)	-.04 (7)	-.04 (10)	$t(15) = .00$, $p = 1.00$	
Wichtigkeit Ratee (P)	-.07 (9)	.05 (8)	$t(15) = -.30$, $p = .77$	
Wichtigkeit Ratee (FvF)	-.03 (6)	.33 (10)	$t(14) = -.47$, $p = .65$	
Wichtigkeit Ratee (S)	-.75 (4)	1.13 (4)	$t(6) = -2.47$, $p < .05$	
Ideal-Ist Ratee (E/F)	.21 (6)	-.20 (11)	$t(15) = 1.09$, $p = .29$	
Ideal-Ist Ratee (B/S)	.42 (6)	-.37 (11)	$t(15) = .76$, $p = .46$	
Ideal-Ist Ratee (IB)	.41 (7)	-.05 (10)	$t(15) = .02$, $p = .98$	
Ideal-Ist Ratee (K)	.29 (7)	.19 (10)	$t(15) = .06$, $p = .96$	
Ideal-Ist Ratee (P)	.15 (8)	-.04 (9)	$t(15) = .19$, $p = .85$	
Ideal-Ist Ratee (FvF)	.07 (7)	.23 (9)	$t(14) = 1.51$, $p = .15$	
Ideal-Ist Ratee (S)	.15 (5)	-.09 (11)	$t(14) = 3.02$, $p < .01$	
Personale Faktoren (Fremdrater):				
Alter Abwärtsrater	.21 (9)	-.12 (15)	$t(22) = 1.21$, $p = .24$	ja
Organisations- und tätigkeitsbezogene Faktoren:				
Führungsspanne (Ratee)	-.02 (16)	.02 (16)	$t(30) = -.17$, $p = .87$	=
Führungsspanne (Abwärtsrater)	.04 (21)	-.12 (10)	$t(29) = .63$, $p = .53$	ja
Hierarchie (Ratee)[b]	-.36 (12)	.21 (20)	$t(30) = -2.61$, $p = .01$	nein
Dauer der Organisationszugehörigkeit (Selbstrater)	-.29 (14)	.32 (16)	$t(28) = -3.05$, $p < .01$	nein
Dauer der Abteilungszugehörigkeit (Selbstrater)	-.21 (13)	.22 (17)	$t(22) = -1.99$, $p = .06$	nein
Dauer der Zusammenarbeit von Selbst- und Abwärtsrater	-.03 (14)	.14 (15)	$t(27) = -.73$, $p = .47$	nein
Durchführungsbedingungen (Format des Befragungsinstruments):				
Beschreibungsebene: ergebnisbezogene Itemformulierung [c]	.03 (31)	.01 (31)	$t(30) = -.26$, $p = .79$	ja
Durchführungsbedingungen (allgemeine Rahmenbedingungen):				
Erfahrung als Beurteilter (durch Abwärtsbeurteiler)	.12 (16)	-.15 (15)	$t(29) = 1.15$, $p = .26$	nein
Erfahrung als Beurteiler (Ratee)	.06 (18)	-.39 (6)	$t(22) = 1.42$, $p = .17$	nein
Beteiligungsquote	-.02 (19)	-.08 (12)	$t(29) = .38$, $p = .71$	nein

Anmerkungen. Negative Diskrepanz = Unterschätzer, [a]bezogen auf das Ausmaß der Diskrepanz, [b]links: Hierarchieebene 1-3, rechts: Hierarchieebene 4, [c]links: nicht ergebnisbezogene Itemformulierung, rechts: ergebnisbezogene Itemformulierung.

Durchgehend (alle vier berichteten Selbst-Fremd-Koeffizienten) wird die Selbst-Fremd-Übereinstimmung vom Alter des Fremdraters und von der ergebnisbezogenen Itemformulierung in die erwartete Richtung beeinflusst. Ersteres kann durch eine mit dem Alter steigende Erfahrung mit Beurteilungen erklärt werden. Auch das Alter des Selbstraters moderiert 3 der 4 Koeffizienten. Der Einfluss der ergebnisbezogenen Itemformulierung gibt einen Hinweis darauf, dass der Interpretationsspielraum, den Itemformulierungen in aller Regel lassen, in Richtung Selbstüberschätzung „genutzt“ wird. Diese Moderatorwirkung kann erklären, weshalb die Selbst-Aufwärts-Korrelation auf der Dimension Klarheit besonders hoch ausfällt. Die Skala setzt sich überwiegend aus ergebnisbezogen formulierten Items zusammen, was den jeweiligen Ratinggegenstand für die Beurteiler fassbarer macht.

Die Moderatoranalysen zu den beiden Variablen, für die vorab keine eindeutige Hypothese formuliert wurde, erbringen unterschiedliche Hinweise. Inwieweit sich die Führungskräfte von ihrem eigenen Idealbild entfernt sehen (Ideal-Ist Ratee), scheint keinen Einfluss auf die Selbst-Fremd-Übereinstimmung zu nehmen. Zumindest ist in den vorliegenden Daten kein regelmäßiges Muster zu erkennen. Die Wichtigkeit, die die Führungskräfte den einzelnen FzFü-Dimensionen zuschreiben (Wichtigkeit Ratee), nimmt eher Einfluss auf die Selbst-Fremd-Übereinstimmung. Jedenfalls fällt die Selbst-Aufwärts-Diskrepanz bei hoher Wichtigkeitsbewertung für viele Führungsdimensionen deutlich höher aus als bei niedriger Wichtigkeitsbewertung. Dass man sich selbst auf wichtig erachteten Dimensionen positiver bewertet als auf weniger wichtigen, kann vor dem Hintergrund von Selbstwert-Theorien interpretiert werden.

Abschließend werden nun die Ergebnisse der Moderatoranalysen für die **Fremd-Fremd**-Übereinstimmung berichtet. Auf die tabellarische Darstellung der faktorbezogenen Auswertungen von „Wichtigkeit Aufwärts“ wurde der Übersichtlichkeit zuliebe verzichtet, da sie ausnahmslos mit der globalen Auswertung übereinstimmen.

Aus Tabelle 57 ist zu entnehmen, dass sich die Abwärts-Aufwärts-Korrelation insgesamt bei 12 der 17 untersuchten Variablen mindestens tendenziell in die erwartete Richtung verändert, was das Signifikanzniveau im Binomialtest nur knapp verfehlt (p = .07). Insbesondere die organisations- und tätigkeitsbezogenen Variablen nehmen moderierenden Einfluss auf die Abwärts-Aufwärts-Korrelation, da sie durchgehend die erwarteten Effekte zeigen, was im Binomialtest Signifikanz erreicht (p = .03). Zudem sind davon alle drei Effekte der Variablen zur Zugehörigkeits-/Zusammenarbeitsdauer als groß zu bezeichnen (q = .71 bis q = .74). Das verweist darauf, dass im Laufe der Zugehörigkeit Organisationsnormen übernommen werden, vor deren Hintergrund ähnliche Führungsinterpretationen und folglich -beurteilungen abgegeben werden. Für den Bereich der organisations- und tätigkeitsbezogenen Variablen ist Hypothese 16 also nicht zu verwerfen.

Tabelle 57. Ergebnisse der Moderatoranalysen zur Abwärts-Aufwärts-Korrelation

Moderatorvariable (MV)	Abwärts-Aufwärts-Korrelation (*N*) unter Dichotomisierungskriterium der MV	über Dichotomisierungskriterium der MV	*z*, *p*	Richtung der Hypothese bestätigt
Personale Faktoren (Selbstrater/Ratee):				
IIa	.62 (10)	.06 (11)	1.24, *p* = .90	(nein)
IIb	.68 (11)	.16 (10)	1.29, *p* = .90	(nein)
IM	.70 (8)	.56 (13)	.43, *p* = .67	(nein)
IS	.59 (11)	.45 (10)	.37, *p* = .65	(nein)
IC	.47 (11)	.62 (10)	-.42, *p* = .34	(ja)
AUS	.57 (12)	-.09 (12)	1.57, *p* = .94	(nein)
Transformationale Führung ges.	.71 (10)	.39 (11)	.92, *p* = .82	nein
CR	.69 (9)	.35 (12)	.92, *p* = .82	nein
MBA	-.08 (9)	.75 (10)	-2.03, *p* = .02	ja
MBP	-.20 (10)	.63 (11)	-1.82, *p* = .03	nein
Laissez Faire	.59 (10)	.54 (11)	.14, *p* = .56	ja
Personale Faktoren (Fremdrater):				
Wichtigkeit Aufwärts (ges.)	.49 (11)	.20 (14)	.72, *p* = .76	ja
Alter (Aufwärtsrater)	.35 (9)	.61 (13)	-.67, *p* = .25	ja
Alter (Abwärtsrater)	-.39 (12)	.42 (11)	-1.77, *p* = .04	ja
Organisations- und tätigkeitsbezogene Faktoren:				
Führungsspanne (Ratee)	.62 (11)	.36 (13)	.73, *p* = .77	ja
Führungsspanne (Abwärtsrater)	.45 (15)	.25 (8)	.43, *p* = .67	ja
Dauer der Organisations-zugehörigkeit (Aufwärtsrater)	-.08 (9)	.58 (21)	-1.57, *p* = .06	ja
Dauer der Abteilungs-zugehörigkeit (Aufwärtsrater)	.03 (14)	.64 (16)	-1.78, *p* = .04	ja
Dauer der Zusammenarbeit von Selbst- und Aufwärtsrater	.08 (14)	.66 (16)	-1.74, *p* = .04	ja
Durchführungsbedingungen (Format des Befragungsinstruments):				
Beschreibungsebene: ergebnis-bezogene Itemformulierung [a]	.46 (29)	.32 (29)	.60, *p* = .72	nein
Durchführungsbedingungen (allgemeine Rahmenbedingungen):				
Beteiligung des Ratees [b]	-.18 (8)	.44 (24)	-1.31, *p* = .09	ja
Kenntnis der Führungs-grundsätze [c]	.14 (62)	.41 (50)	-1.51, *p* = .07	ja
Aufwärtsrater als Stellvertreter [d]	.32 (91)	.28 (25)	.18, *p* = .57	nein

Anmerkungen. [a]links: nicht ergebnisbezogene Itemformulierung, rechts: ergebnisbezogene Itemformulierung, [b]links: nein, rechts: ja, faktorbezogene Korrelationskoeffizienten vgl. Tabelle 49, [c]nicht-aggregierte Ratings, links: mindestens ein Beurteiler ohne Kenntnis der Führungsgrundsätze, rechts: beide Beurteiler kennen die Führungsgrundsätze, [d]nicht-aggregierte Ratings, links: Aufwärtsbeurteiler ohne Stellvertreterrolle, rechts: Aufwärtsbeurteiler mit Stellvertreterrolle.

Außerdem zeigen einzelne Variablen, das Alter der Abwärtsrater (q = .86) sowie die Beteiligung des Ratees (q = .65), starke Effekte. Während Ersteres – analog zu den Befunden für die *Selbst*-Fremd-Zusammenhänge – durch erhöhte Beurteilungserfahrung zu erklären ist, kann Letzteres auf das generelle Beurteilungsklima zurückgeführt werden. Wenn nämlich die Beteiligung der sich selbst beurteilenden Führungskräfte als Indiz für deren Offenheit gegenüber multiperspektivischen Beurteilungen gewertet wird, dann kann ein Klima angenommen werden, in dem Beurteilungen und Feedback im Arbeitsalltag

existieren (dürfen). Das heißt, dass mit höherer Wahrscheinlichkeit Beurteilungserfahrung besteht und mit geringerer Wahrscheinlichkeit angstmotivierte Mildetendenzen auftreten.

Für die Abwärts-Aufwärts-Diskrepanz verändern sich die Koeffizienten für 15 der 17 Variablen in die erwartete Richtung (Tabelle 58). Der Binomialtest stellt dies im Sinne von Hypothese 16 als überzufällig heraus (p = .001). Fast durchgängig zeigen sich mittlere bis starke Effekte, insbesondere für das Führungsverhalten des Ratees und für die Durchführungsbedingungen.

Tabelle 58. Ergebnisse der Moderatoranalysen zur Abwärts-Aufwärts-Diskrepanz

Moderatorvariable (MV)	Abwärts-Aufwärts-Diskrepanz (*N*) unter Dichotomisierungskriterium der MV		Abwärts-Aufwärts-Diskrepanz (*N*) über Dichotomisierungskriterium der MV		*t, p*	Richtung der Hypothese bestätigt
Personale Faktoren (Selbstrater/Ratee):						
IIa	1.59	(10)	.89	(11)	$t(19) = 3.44, p < .01$	(ja)
IIb	1.57	(10)	.83	(11)	$t(19) = 3.79, p < .01$	(ja)
IM	1.73	(8)	.91	(13)	$t(19) = 4.40, p < .01$	(ja)
IS	1.64	(11)	.77	(10)	$t(19) = 5.28, p < .01$	(ja)
IC	1.57	(11)	.84	(10)	$t(19) = 3.62, p < .01$	(ja)
AUS	1.53	(12)	.85	(12)	$t(19) = 3.38, p < .01$	(ja)
Transformationale Führung ges.	1.65	(10)	.83	(11)	$t(19) = 4.62, p < .01$	ja
CR	1.66	(9)	.90	(12)	$t(19) = 3.89, p < .01$	ja
MBA	1.35	(11)	1.08	(10)	$t(19) = 1.07, p = .30$	ja
MBP	1.01	(10)	1.41	(11)	$t(19) = -1.67, p = .11$	ja
Laissez Faire	.89	(10)	1.52	(11)	$t(19) = -2.91, p < .01$	ja
Personale Faktoren (Fremdrater):						
Wichtigkeit Aufwärts (ges.)	1.45	(11)	1.19	(14)	$t(23) = .97, p = .34$	nein
Alter (Aufwärtsrater)	1.34	(9)	1.11	(13)	$t(20) = .85, p = .41$	ja
Alter (Abwärtsrater)	1.31	(12)	1.25	(11)	$t(21) = .23, p = .82$	ja
Organisations- und tätigkeitsbezogene Faktoren:						
Führungsspanne (Ratee)	1.07	(11)	1.29	(13)	$t(22) = -.87, p = .39$	ja
Führungsspanne (Abwärtsrater)	1.21	(15)	1.28	(8)	$t(21) = -.27, p = .79$	ja
Dauer der Organisationszugehörigkeit (Aufwärtsrater)	1.27	(9)	1.23	(21)	$t(28) = .17, p = .87$	ja
Dauer der Abteilungszugehörigkeit (Aufwärtsrater)	1.22	(14)	1.26	(16)	$t(28) = -.14, p = .89$	nein
Dauer der Zusammenarbeit von Selbst- und Aufwärtsrater	1.39	(14)	1.11	(16)	$t(28) = 1.23, p = .23$	ja
Durchführungsbedingungen (Format des Befragungsinstruments):						
Beschreibungsebene: ergebnisbezogene Itemformulierung[a]	.79	(29)	.69	(29)	$t(28) = -.96, p = .35$	ja
Durchführungsbedingungen (allgemeine Rahmenbedingungen):						
Beteiligung des Ratees[b]	1.53	(8)	1.19	(24)	$t(30) = -1.18, p = .25$	ja
Kenntnis der Führungsgrundsätze[c]	1.03	(62)	.58	(50)	$t(110) = -2.17, p = .03$	ja
Aufwärtsrater als Stellvertreter[d]	1.02	(91)	.23	(25)	$^{\circ}t(54) = -3.94, p < .01$	ja

Anmerkungen. [a]links: nicht ergebnisbezogene Itemformulierung, rechts: ergebnisbezogene Itemformulierung, [b]links: nein, rechts: ja, [c]nicht-aggregierte Ratings, links: mindestens ein Beurteiler ohne Kenntnis der Führungsgrundsätze, rechts: beide Beurteiler kennen die Führungsgrundsätze, [d]nicht-aggregierte Ratings, °keine Varianzgleichheit, links: Aufwärtsbeurteiler ohne Stellvertreterrolle, rechts: Aufwärtsbeurteiler mit Stellvertreterrolle.

Genauer: Je mehr aktive Führung dem Ratee zugeschrieben wird, umso geringer sind die Niveau-Unterschiede zwischen Abwärts- und Aufwärtsbeurteiler. Die Niveau-Unterschiede für aktive Führungskräfte liegen zum Teil sehr deutlich unter denen für nicht aktive Führungskräfte (d = .49 bis d = 2.42). Da die Effekte erstens für alle MLQ-Dimensionen erwartet und zweitens überwiegend stärker als für die *Selbst*-Fremd-Diskrepanz ausfallen, kann geschlossen werden, dass tatsächlich das Führungsverhalten und nicht ausschließlich die Wahrnehmung durch die geführten Mitarbeiter die Abwärts-Aufwärts-Übereinstimmung mitbegünstigt. Plausibel ist, dass aktive Führung auch für die Abwärtsbeurteiler, die am Führungsprozess selbst nicht beteiligt sind, besser beobachtbar und dadurch beurteilbar ist als passive Führung.

Alle vier untersuchten Durchführungsbedingungen beeinflussen die Abwärts-Aufwärts-Diskrepanz in erwarteter Weise, was im Binomialtest nahezu signifikant ist (p = .06). Für die ergebnisbezogene Itemformulierung (d = .36), die Beteiligung des Ratees als Selbstbeurteiler (d = .43) und die Kenntnis der Führungsgrundsätze (d = .41) sind die Effekte gering, für die Stellvertreter-Erfahrung der Aufwärtsbeurteiler sogar stark (d = 1.07).

Aus den Moderatoranalysen ist zusammenfassend festzuhalten:

- Die Abwärts-Aufwärts-Übereinstimmung wird von *mehr* (der untersuchten) Variablen moderiert als die Selbst-Fremd-Übereinstimmung.
- Die moderierenden Effekte für die Abwärts-Aufwärts-Übereinstimmung fallen deutlich stärker aus als für die Selbst-Fremd-Übereinstimmung. Besonders hervorzuheben ist, dass hierfür – trotz der auf Seite 450 dargestellten Problematik von Moderatoranalysen in Einzelstudien – deutliche Effekte festgehalten werden können. Das spricht zusätzlich zu anderen Ergebnissen der vorliegenden Studie (z.B. Kapitel 6.4) dafür, dass die Selbstbeurteilungen weniger genau und valide sind als Fremdbeurteilungen.
- Es kristallisieren sich einige Variablen heraus, die sowohl die Selbst-Fremd- als auch die Fremd-Fremd-Übereinstimmung in erwarteter Richtung beeinflussen: Alter der Fremdrater, Führungsspanne, ergebnisbezogene Itemformulierung, Beteiligung (Fremd- und Selbstrater) und Kenntnis der Führungsgrundsätze. Das Alter der Fremdrater kann seine moderierende Wirkung aus der Beurteilererfahrung ziehen, eine geringe Führungsspanne bietet eine bessere Beobachtungsgrundlage, ergebnisbezogene Itemformulierungen geben den Beurteilungsgegenstand klarer vor und verringern den Verzerrungsspielraum, eine hohe Beteiligung aller signalisiert ein offenes Beurteilungsklima, so dass auch insgesamt von besserer Kommunikation und regelmäßigen Anpassungen von Selbst- und Fremdwahrnehmungen auszugehen ist, und die Kenntnis der Führungsgrundsätze (von beiden beteiligten Beurteilerperspektiven) legt eine ähnliche Definition von Führung nahe. All diese Faktoren sollten die strategischen Beurteilungsverzerrungen reduzieren und in der Folge die interperspektivischen Beurteilungsübereinstimmungen erhöhen.

- Wie in der Meta-Analyse von Heidemeier und Moser (2002) zur Selbst-Abwärts-Übereinstimmung beruflicher Leistungsbeurteilung konnten auch in der vorliegenden Stichprobe mehr moderierende Effekte auf Niveau-Unterschiede als auf korrelative Zusammenhänge aufgedeckt werden. Das bestätigt einmal mehr das in der vorliegenden Untersuchung gewählte methodische Vorgehen, mehr als nur einen Index für die Perspektiven-Übereinstimmung zu verwenden. Ein detaillierterer Vergleich zu den Befunden der Meta-Analyse von Heidemeier und Moser (2002) muss sich auf eine Variable beschränken. Nur die Dauer der Organisationszugehörigkeit wurde sowohl hier als auch dort untersucht. Während sich bei Heidemeier und Moser (2002) weder für die Selbst-Abwärts-Niveau-Unterschiede noch -Korrelation ein bedeutsamer Effekt herausstellt, zeigt sich in der vorliegenden Studie für die Abwärts-Aufwärts-Korrelation ein starker Effekt (q = .74, p = .06). Das kann am Beurteilungsgegenstand liegen. Denn gerade *Führungs*definitionen und -kultur werden erst im Laufe der Zeit transportiert, während die Grundlagen zur allgemeinen beruflichen Leistungsbeurteilung weniger organisationsabhängig sind.

6.3.4 Zusammenhang zwischen verschiedenen Übereinstimmungs-Indizes

Während in der Literatur regelmäßig darüber diskutiert wird, welche Vor- und Nachteile verschiedene (Un-)Ähnlichkeitsmaße haben (z.B. Atwater et al., 1998; Edwards, 1994a, 2001), gibt es nur ansatzweise den Versuch, die Prognoseleistung verschiedener Methoden gegeneinander zu testen. Church (1997a) schließt aus seinen wenig kritischen Analysen, dass Korrelationen und Differenzmaße zwischen Selbst- und Aufwärtsbeurteilungen zu keinen unterschiedlichen Leistungsprognosen kommen. Brutus, Fleenor und Tisak (1999) testeten für den Selbst-Peer-Vergleich die Prognoseleistung von polynomialen Regressionen gegen ein typisches varianzanalytisches Vorgehen auf der Basis einer Kategorisierung nach Differenzmaßen. Dabei zeigte sich, dass aufgrund unterschiedlicher Methoden durchaus unterschiedliche Vorhersagen gemacht werden. Wie stark die Niveau-Unterschiede überhaupt mit dem korrelativen Zusammenhang einhergehen, blieb in der bisherigen Forschung unbeantwortet (vgl. Moser et al., 1994).

Mehr Kenntnis über die Ersetzbarkeit bzw. die unterschiedliche Aussagekraft verschiedener Indizes wäre aber eine wichtige Entscheidungshilfe dafür, sich gezielt für den Einsatz eines Maßes bzw. – wenn kein starker Zusammenhang zwischen den Maßen besteht – mehrerer sich ergänzender Maße zu entscheiden. Diese Entscheidung liegt im Übrigen nicht nur bei Forschungsarbeiten an, sondern bei allen multiperspektivischen Führungsbeurteilungen, auf deren Basis Entscheidungen (Entwicklung, Gehalt, Beförderung etc.) getroffen werden. Wegen der hohen Relevanz, die einem schmalen Erkenntnisstand gegenüber steht, soll dieser Fragestellung 5 in der vorliegenden Arbeit nachgegangen

werden. Da aus den unterschiedlichen Perspektiven möglicherweise unterschiedliche Beurteilungsinformationen einfließen, muss die Frage auch für die jeweiligen Perspektiven getrennt voneinander bearbeitet werden.

Folgende Hypothesen werden untersucht:

17) Je geringer die Selbst-Abwärts-Niveau-Unterschiede sind, umso höher sind die Selbst-Abwärts-Korrelationen.
18) Je geringer die Selbst-Aufwärts-Niveau-Unterschiede sind, umso höher sind die Selbst-Aufwärts-Korrelationen.
19) Je geringer die Abwärts-Aufwärts-Niveau-Unterschiede sind, umso höher sind die Abwärts-Aufwärts-Korrelationen.

Es ist anzunehmen, dass die Beurteiler, deren Ratings sich im Niveau weniger von denen anderer Beurteiler unterscheiden, akkuratere Beurteilungen abgeben und ihre Urteile in Folge dessen auch höhere Interkorrelationen aufweisen. Gemäß dieser Annahme wurde für alle drei Perspektiven-Vergleiche (Selbst-Aufwärts, Selbst-Abwärts, Abwärts-Aufwärts) und sieben Führungsdimensionen ein Mediansplit bezüglich der „positiven“ Niveau-Unterschiede vorgenommen. „Positiv“ bedeutet, dass bei negativen Differenzen der Betrag herangezogen wurde, um ausschließlich das Ausmaß, nicht aber die Richtung der Abweichung zu fokussieren. In Tabelle 59 sind die entsprechenden Interkorrelationen zwischen je zwei Perspektiven angegeben, die gemäß der Hypothesen 16) bis 18) für Spalte 1 (niedriger Niveau-Unterschied) höher sein müssten als für Spalte 2 (hoher Niveau-Unterschied). Dies bestätigt sich in der Signifikanzprüfung. Nahezu alle Korrelationsunterschiede erreichen mindestens das 5%-Signifikanzniveau. Die Effekte sind für alle drei Vergleiche (im Durchschnitt über die sieben Führungsdimensionen) stark (Selbst-Aufwärts: $\bar{q} = .78$, Selbst-Abwärts: $\bar{q} = 1.06$, Abwärts-Aufwärts: $\bar{q} = .76$). Bei einer starken (vs. schwachen) Abweichung (unabhängig davon, in welche Richtung die Abweichung vorliegt) ergeben sich engere, positive korrelative Zusammenhänge. Die Höhe der Abweichung moderiert also den korrelativen Zusammenhang zwischen den betrachteten Perspektiven. Dies entspricht der Hypothese und ist deshalb plausibel, weil eine Abweichung, die keine konstante Niveau-Verschiebung von der Referenzperspektive darstellt, immer eine ungenaue Einschätzung impliziert. Je ungenauer die Beurteilung aber ausfällt, umso schwächer muss eine Korrelation sein.

Tabelle 59. Zusammenhang zwischen Korrelationen und dem Betrag der Niveau-Unterschiede

	Korrelation zwischen Selbst und Aufwärts$_a$		Korrelation zwischen Selbst und Abwärts		Korrelation zwischen Abwärts und Aufwärts$_a$	
	NU gering	NU hoch	NU gering	NU hoch	NU gering	NU hoch
F/E	.67**	-.13	.75**	.07	.62*	.44
B/S	.50**	-.03	.75**	-.06	.19	.37
IB	.38	-.03	.50	-.15	.85**	.38
K	.39	.39	.76**	-.18	.63*	-.12
P	.39	-.33	.83**	-.09	.45	.08
FvF	.71**	-.20	.53*	.00	.76*	-.29
S	.87**	-.31	.89**	-.29	.97**	.24

Anmerkungen. NU gering/hoch = Niveau-Unterschied unter/über Median (nur Ausmaß der Abweichung berücksichtigt, Richtung unberücksichtigt). Für Selbst-Aufwärts: D-Statistik als Mediansplit-Grundlage, *N* vgl. Tabelle 49, **p < .01, *p < .05.

Häufig werden in Forschungsstudien jedoch keine Differenz-*Beträge* verwendet, sondern absolute Niveau-Unterschiede. Da sich Unter- und Überschätzer dann nivellieren können, kommen solche Studien eher zu dem Schluss, dass im Durchschnitt keine deutlichen Niveau-Unterschiede bestehen. Dennoch können beispielsweise geringe Korrelationen gefunden werden (vgl. gesamte Datenbreite von Abbildung 29). Niveau-Unterschiede und korrelative Zusammenhänge sind scheinbar unabhängig voneinander (vgl. auch Anhang L).

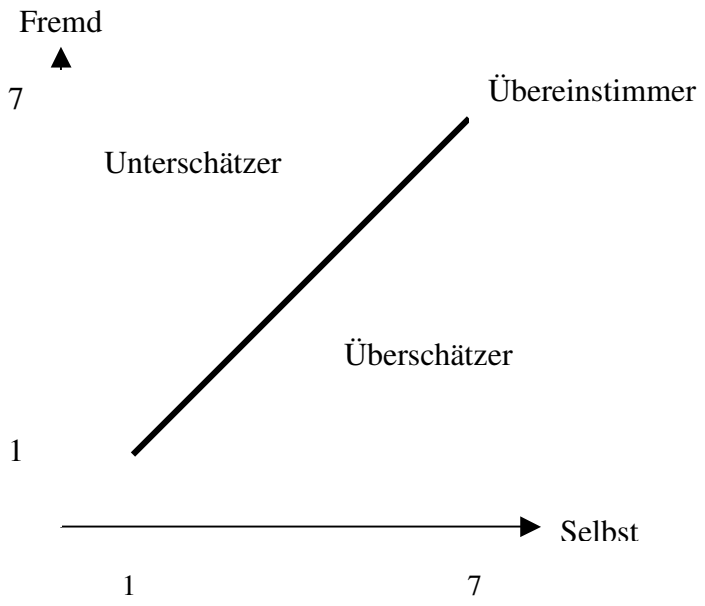


Abbildung 29. Skizzierung der Selbst-Fremd-Diskrepanz

Anmerkung: Skala von 1 (negativer Pol) bis 7 (positiver Pol).

Zieht man die Richtung stärker in Betracht und vergleicht man die Korrelationen, die die Übereinstimmer, Über- und Unterschätzer zum Urteil ihrer Mitarbeiter aufweisen (Kategorisierung vgl. S. 248), ergibt sich ein differenzierteres Bild (vgl. Tabelle 60). Sowohl die Selbstbeurteilungen der Unterschätzer als auch die der Übereinstimmer zeigen sehr hohe Korrelationen zu den entsprechenden Aufwärtsbeurteilungen. Die Überschätzer hingegen stimmen nur mittelmäßig mit diesen überein. Dies kann trivialerweise daran liegen, dass das Ausmaß der Überschätzung stärker variiert als das Ausmaß der Unterschätzung und ohnehin der Übereinstimmung. In der skizzenhaften Darstellung von Abbildung 29 müssten sich also die Selbst-Fremd-Punkte unterhalb der Diagonale inhomogener über das Dreieck verteilen als die Selbst-Fremd-Punkte oberhalb der Diagonale, die beispielsweise bei generell nur geringer Unterschätzung in einem Streifen über der Übereinstimmer-Diagonale liegen. Die geringere Selbst-Aufwärts-Korrelation für die Überschätzer zeigt aber auch, dass die Überschätzer als besonders schlechte Selbstbeurteiler bezeichnet werden können.

Tabelle 60. Korrelative Selbst-Aufwärts-Übereinstimmung der Unterschätzer, Übereinstimmer und Überschätzer

	Unterschätzer	Übereinstimmer	Überschätzer
$\bar{r}$	.86	.87	.49
N	5-8	8-15	12-25

Anmerkung. $\bar{r}$: mittlere Selbst-Aufwärts-Korrelation über die sieben FzFü-Dimensionen hinweg, *N* variiert über die Dimensionen hinweg.

Letztlich kann festgehalten werden, dass das korrelative Übereinstimmungsmaß für alle drei Perspektiven-Vergleiche stimmig mit dem Ausmaß der Niveau-Unterschiede einhergeht, dass jedoch auch die Richtung der Niveau-Abweichungen eine Rolle spielt, wenn es um die Vorhersage des korrelativen Zusammenhangs geht. Abbildung 29 macht diesen quasi-natürlichen Befund plausibel, denn in Abhängigkeit von der Richtung zieht man gezielt „Stichproben“ ober- bzw. unterhalb der Übereinstimmer-Linie, nicht jedoch aus dem ganzen potentiell möglichen Feld. Es ist fast trivial, dass die korrelativen Zusammenhänge geringer ausfallen, wenn man das gesamte potentielle Feld an Abweichungen betrachtet (absolute Differenzen), als wenn man lediglich eine Hälfte betrachtet (Beträge). Insofern ist plausibel, dass Niveau-Unterschiede bei starker Streuung der Abweichungen sowie bei der Betrachtung absoluter Differenzmaße die Korrelationsmaße nicht ersetzen können und zu anderen Vorhersagen kommen.

Eine weitere Überprüfung des Zusammenhangs der beiden Ähnlichkeitsmaße stellt der Vergleich über die sieben Faktoren hinweg dar. Für den Fall, dass sie sich substituieren könnten, müsste der korrelative Zusammenhang zwischen zwei Perspektiven auf denjenigen Faktoren hoch sein, auf denen die Diskrepanz gering ist – und umgekehrt. Ein mono-

toner Zusammenhang liegt also dann vor, wenn mit steigender Korrelation die Diskrepanz abnimmt. Geprüft wurde diese Hypothese mittels Rangkorrelationen (Kendalls τ).[170] Jeder Führungsdimension wurde also ein Rang für Korrelationen und ein Rang für Niveau-Unterschiede zugewiesen. Weder für den Selbst-Aufwärts- noch für den Selbst-Abwärts- oder Abwärts-Aufwärts-Vergleich ergab sich ein hoher, geschweige denn signifikanter Rangkorrelationskoeffizient zwischen den Niveau-Unterschieden und den Korrelationen der Perspektiven[171]. Dabei ist es unerheblich, ob die Rangplätze für die Niveau-Unterschiede auf der Basis der absoluten Diskrepanzen (τ = -.20 bis -.15, p = .66 bis .77) oder ihrer Beträge (τ = -.25 bis .29, p = .46 bis .77) zugeteilt wurden. Dass die Korrelationen mit einer Ausnahme (Selbst- vs. Abwärtsbeurteilung, Betrag) negativ sind, deutet zumindest die richtige Tendenz an: Je höher der Niveau-Unterschied, umso geringer die Korrelation.

In der organisationspsychologischen Praxis sind in der Regel führungskraft-bezogene Kennwerte von Bedeutung. Beispielsweise werden in einem Mitarbeiter-Feedback den Selbstbeurteilungen der Führungskraft die Beurteilungen ihrer Mitarbeiter gegenüber gestellt. Der Fokus liegt also eindeutig auf Niveau-Unterschieden (vgl. auch Kapitel 3.6). Es stellt sich generell die Frage, ob dies gerechtfertigt ist. Mehr Unterstützung würde dieses Vorgehen finden, wenn diejenigen Führungskräfte, die hohe Diskrepanzen zu ihren Mitarbeitern aufweisen, auf korrelativer Ebene nur gering mit diesen übereinstimmen. Für die vorliegende Stichprobe zeigt sich kein bedeutender monotoner Zusammenhang zwischen dem mittleren Selbst-Aufwärts-Niveau-Unterschied über die beurteilten Dimensionen hinweg und der entsprechenden Selbst-Aufwärts-Korrelation (τ = -.14, p = .26, N = 36).[172]

Wenn man nicht mit aggregierten, sondern mit individuellen Aufwärtsratings arbeitet, dann gehen die Niveau-Unterschiede zwischen Führungskraft und Mitarbeiter (Gesamtskala) mit dem korrelativen Zusammenhang zwischen der Selbst- und Aufwärtsbeurteilung über die sieben Führungsdimensionen deutlicher miteinander einher (τ = -.18, p < .01, N = 137). Dies gilt nicht für die individuelle Abwärts-Aufwärts-Übereinstimmung (τ = -.04, p = .52, N = 114) und gilt sogar in umgekehrter Richtung für die Selbst-Abwärts-Übereinstimmung (τ = .25, p < .05, N = 31).

Insgesamt muss das Fazit gezogen werden, dass sich kein ausreichender Zusammenhang zwischen den beiden Ähnlichkeitsmaßen zeigt, was den Annahmen entspricht (Hypothesen 16) bis 18). Für die folgenden Auswertungen werden deshalb jeweils beide Maße verwendet. In Kapitel 6.5, in dem die prognostische Validität der interperspektivischen Übereinstimmung untersucht wird, werden weitere Methoden hinzugenommen, die dann

[170] Kendalls τ wurde statt Spearmans ρ herangezogen, weil aufeinanderfolgende Ränge auf dem untersuchten Merkmal keine äquidistanten Positionen darstellen. Dies ist für eine sinnvolle Anwendung von Spearmans ρ jedoch Voraussetzung (Gegenüberstellung vgl. Bortz, Lienert & Boehnke, 1990).

[171] Die Signifikanzprüfung erfolgte als „exakte Signifikanzprüfung“, die auf der exakten Stichprobenverteilung basiert. Sie wird für kleine Stichproben grundsätzlich empfohlen (vgl. Diehl & Staufenbiel, 2001).

[172] Für die Rangkorrelation wurde jeder Führungskraft ein Rang für die Selbst-Aufwärts-Korrelation und ein Rang für die Selbst-Aufwärts-Abweichung zugewiesen.

mit der Vorhersageleistung der Korrelation sowie der Niveau-Unterschiede verglichen werden können. Möglicherweise sind die unterschiedlichen Maße und Methoden für unterschiedliche Fragestellungen unterschiedlich gut tauglich, so dass ihr Einsatz künftig von der Fragestellung abhängig gemacht werden sollte. Differenzen sind jedenfalls nicht als Substitute für korrelative Zusammenhänge zu betrachten

6.3.5 Above-Average-Effekt

In diesem Kapitel werden die erhobenen Daten im Hinblick auf folgende Hypothesen untersucht:

20) Führungskräfte beurteilen ihr eigenes Führungsverhalten positiver als das anderer Führungskräfte.
21) Die Mehrzahl der Führungskräfte beurteilt ihr eigenes Führungsverhalten positiver als das anderer Führungskräfte.
22) Die Selbstüberschätzung nach oben fällt geringer aus als die Selbstüberschätzung nach unten.
23) Die Selbstüberschätzung nach oben ist unabhängig von der Selbstüberschätzung nach unten.

In Fragestellung 6 interessiert also nicht die Übereinstimmung von Beurteilungen aus verschiedenen Perspektiven, die das Führungsverhalten derselben Führungskraft betreffen. Stattdessen folgt die Fragestellung einer umgekehrten Betrachtungsrichtung. Es werden die Beurteilungen, die eine Führungskraft hinsichtlich ihres eigenen Führungsverhaltens abgibt, mit denen verglichen, die *die gleiche* Führungskraft hinsichtlich des Führungsverhaltens ihres Vorgesetzten bzw. ihrer Mitarbeiter mit Führungsverantwortung abgibt (vgl. Abbildung 30).

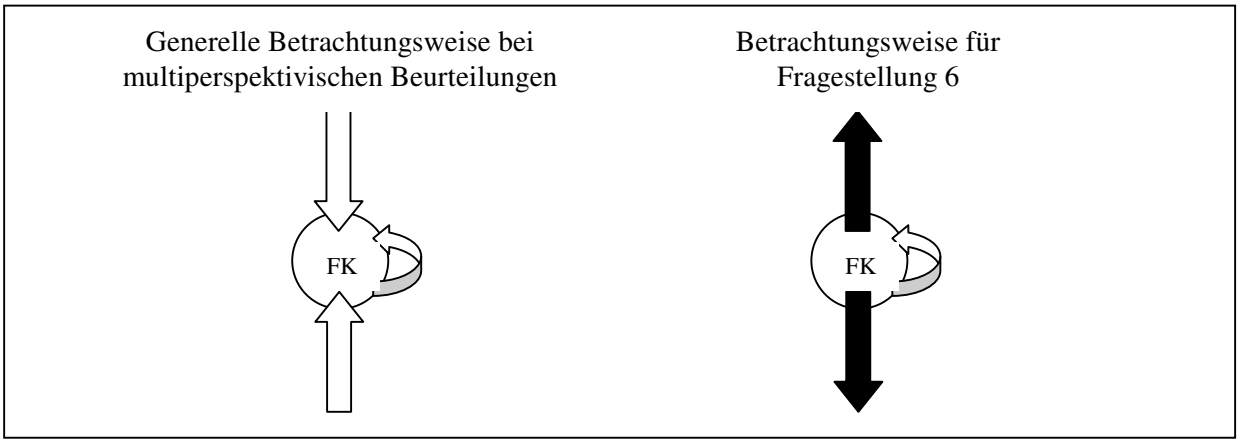


Abbildung 30. Schematische Darstellung der beiden Betrachtungsrichtungen
Anmerkung. FK: Führungskraft.

Es geht also um die Frage, ob Führungskräfte ihr eigenes Führungsverhalten besser beurteilen als das anderer Führungskräfte. Wenn dies die Mehrzahl der Führungskräfte angibt, dann kann dem de facto keine betriebliche Wirklichkeit zugrunde liegen, da – einfach gesagt – nicht alle besser sein können als die anderen. Diese Annahme gilt auch dann, wenn der Vergleich nicht mit Führungskräften der gleichen Ebene gezogen werden kann, sondern aufgrund der fehlenden Peer-Beurteilungen in einem Sandwich-Design zwischen Führungskräften verschiedener Hierarchieebenen gezogen werden muss (wie in der vorliegenden Studie). Die Alternativhypothese, dass die Führungskräfte einer Ebene tatsächlich besser führen als die anderer Ebenen, kann nämlich vorab ausgeschlossen werden, da auch nicht auf allen Ebenen besser geführt werden kann als auf den jeweils anderen. Weiter treffen die im Folgenden berichteten Ergebnisse nicht nur für einzelne Ebenen zu, sondern für die gesamte Führungskräfte-Stichprobe. Im Übrigen unterscheidet sich das Ausmaß der Selbstüberschätzung nicht signifikant in Abhängigkeit von der Hierarchie.[173] „Selbstüberschätzung“ bedeutet in diesem Kapitel, dass eine Führungskraft ihr eigenes Führungsverhalten positiver bewertet als das anderer Führungskräfte.

Vorweg sei angemerkt, dass die Selbstbeurteilung erstaunlich gering mit der Beurteilung der vorgesetzten (r = .30, N = 41) und nachgeordneten Führungkräfte (r = .34, N = 45) korreliert.

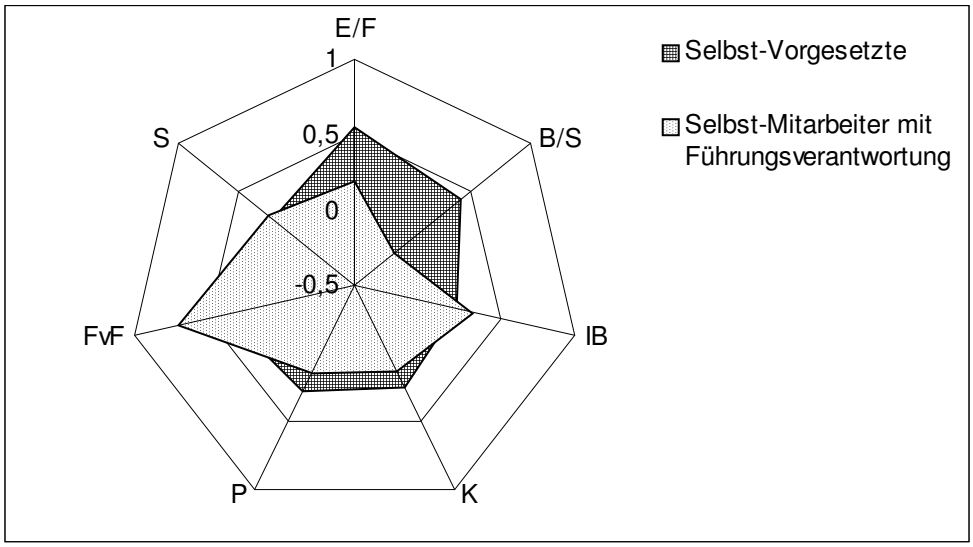


Abbildung 31. Differenz zwischen der Selbstbeurteilung der Führungskräfte und der Beurteilung des Führungsverhaltens anderer Führungskräfte (Vorgesetzte, Mitarbeiter)

Anmerkung. Skala von 1 bis 7.

[173] Die einfaktorielle ANOVA (Faktor Hierarchie) mit Trendtest erbringt weder für den Selbst-Vorgesetzten-Vergleich noch für den Selbst-Mitarbeiter-Vergleich eine lineare Beziehung derart, dass mit steigender Hierarchieebene das Ausmaß der Selbstüberschätzung zu- oder abnehme. Insofern müssen einzelne kleine Effekte als zufällig interpretiert werden.

Es wird zunächst überprüft, ob sich die Selbstbeurteilungen der Führungskräfte von denjenigen Beurteilungen bedeutsam unterscheiden, die die Führungskräfte für das Führungsverhalten ihrer Vorgesetzten und Mitarbeiter abgeben. Abbildung 31 illustriert das Ausmaß der relativ positiven Selbstbewertung in der vorliegenden Studie, in Tabelle 61 sind die genauen Ergebnisse dargestellt.

Mit einer Ausnahme zeigt sich für alle Führungsdimensionen sowie für beide Vergleichs-Ebenen (Selbstüberschätzung im Vergleich nach oben [Vorgesetzter] und unten [Mitarbeiter]), dass die befragten Führungskräfte ihr eigenes Führungsverhalten besser beurteilen als das anderer Führungskräfte ($d > .20$, vgl. Tabelle 61), was der Annahme von Hypothese 20) entspricht.[174] Ein starker Effekt ergibt sich für die Selbstüberschätzung relativ zu den unterstellten Führungskräften auf der Dimension FvF. Dies scheint zunächst besonders plausibel, da Führungskräfte von höheren Hierarchieebenen mit dem Fördern von Führungsverhalten bereits mehr Erfahrung gesammelt haben.[175] Doch es ist zusätzlich festzuhalten, dass Führungskräfte ihr FvF-Verhalten auch besser beurteilen als das ihrer *übergeordneten* Führungskräfte. Insofern ist von internen Prozessen auszugehen, die insbesondere zur Selbstüberschätzung auf FvF motivieren. Denkbar ist, dass sich diese Führungsdimension besser als andere dafür eignet, sich von niedrigeren Positionen positiv abzugrenzen, da sie ein spezifisches Charakteristikum wichtiger Leitungspositionen ausmacht.

Tabelle 61. Above-Average-Effekt: Selbstbeurteilung im Vergleich zur Beurteilung der eigenen Vorgesetzten bzw. der unterstellten Mitarbeiter mit Führungsverantwortung (FV)

FzFü-Faktor	Selbst (1)	Mitarbeiter mit FV (2)	Vor-gesetzter (3)	(1,2) *t*	*p*	*df*	*d*	(1,3) *t*	*p*	*df*	*d*
E/F	5.13 (.50)	4.94 (.67)	4.59 (1.12)	1.69	<.01	40	.32	3.07	<.01	44	.60
B/S	4.84 (.82)	4.99 (.57)	4.67 (.97)	-1.32	.19	40	.21	2.68	.01	44	.45
IB	5.70 (.48)	5.40 (.65)	5.57 (.85)	3.16	<.01	40	.53	1.44	.16	44	.28
K	5.58 (.36)	5.45 (.72)	5.39 (.82)	1.15	.26	40	.23	2.14	.04	44	.37
P	5.31 (.53)	5.16 (.70)	5.10 (.91)	1.09	.28	40	.25	2.22	.03	44	.38
FvF	5.38 (.61)	4.69 (.98)	4.85 (1.14)	4.16	<.01	40	.85	1.86	.07	44	.36
S	5.68 (.93)	5.45 (.95)	5.51 (.86)	1.21	.23	40	.24	.95	.36	19	.26

Anmerkungen. T-Test für eine Stichprobe, siebenstufige Skalen, vor Klammer Faktor-Mittelwert, in Klammer Standardabweichung, der Mittelwert der Selbstbeurteilung unterscheidet sich für die Stichproben, die dem Selbst-Vorgesetzten bzw. dem Selbst-Mitarbeiter-Vergleich zugrunde liegen, aufgrund der unterschiedlichen Anzahl an Vergleichspaaren minimal (berichtet werden die Werte, die in den Selbst-*Mitarbeiter*-Vergleich eingingen).

[174] Eine Ausnahme bildet lediglich die Dimension B/S, auf der die Führungskräfte ihren Mitarbeitern tendenziell bessere Bewertungen zukommen lassen als sich selbst. Dies soll aufgrund des äußerst geringen Effekts nicht überbewertet werden. Denkbar wäre, dass steuernde Verhaltensweisen eher zu den Anforderungen unterer Hierarchieebenen gezählt werden.

[175] Zwei Alternativerklärungen werden ausgeschlossen. Erstens trifft es nicht zu, dass die FvF-Fremdbeurteilungen durch die Mitarbeiter bzw. Vorgesetzten mit steigender Hierarchieebene der beurteilten Führungskraft positiver ausfallen ($F_{[3,129]} = .81$, $p = .49$ bzw. $F_{[2,45]} = 2.19$, $p = .12$). Zweitens trifft es nicht zu, dass FvF mit steigender Hierarchieebene von den Führungskräften als wichtiger betrachtet wird ($F_{[2,21]} = 2.00$, $p = .16$) und die Selbstüberschätzung aufgrund dieser Prioritätensetzung zustande kommt.

Es wurde weiter untersucht, ob die Selbstüberschätzung der Führungskräfte für die sieben Dimensionen gleich stark ausfällt. Nach Bonferroni-Anpassung für Mehrfachvergleiche zeigen sich für die Überschätzung relativ zum Führungsverhalten der *Vorgesetzten* keine Unterschiede zwischen den FzFü-Faktoren (N = 20 für den Vergleich aller sieben Faktoren, N = 45 für den Vergleich der ersten fünf Faktoren). Hingegen überschätzen sich Führungskräfte im Vergleich zu ihren nachgeordneten *Mitarbeitern* mit Führungsverantwortung insbesondere auf der Dimension FvF (N = 41). Die diesbezügliche Selbstüberschätzung ist deutlich stärker als für E/F und B/S. Auf IB ist sie stärker als auf B/S und auf E/F stärker als auf B/S (je $p < .05$). Für alle anderen paarweisen Vergleiche treten keine statistisch bedeutsamen Unterschiede auf.

Weiter stellt sich die Frage, ob die Überschätzung nach oben oder unten stärker ausfällt. Unter der Annahme, dass Führungskräfte karriereorientiert sind, streben sie eher nach der sozialen Identität hoher Führungsebenen. Durch eine besonders hohe Überschätzung im Vergleich zu den Führungskompetenzen ihrer unterstellten Mitarbeiter kann Distinktheit geschaffen werden, was die Zugehörigkeit zu einer höheren Ebene „rechtfertigt“. Diese Annahme konnte nur für IB bestätigt werden ($t_{[25]}$ = 2.84, p = .01, d = 1.14). Für B/S zeigt sich eine deutlich höhere Überschätzung hinsichtlich der Vorgesetzten ($t_{[25]}$ = -2.66, p = .01, d = -1.06; Erklärung vgl. Fußnote 174), in sehr geringem Ausmaß auch für K ($t_{[25]}$ = -1.34, p = . 91, d = -.27). Für die anderen Führungsdimensionen unterscheidet sich die Selbst-Mitarbeiter-Differenz nicht von der Selbst-Vorgesetzten-Differenz, so dass Hypothese 22) abgelehnt wird.

Eine hohe positive Korrelation zwischen den Differenzen nach oben und unten würde anzeigen, ob Führungskräfte die Beurteilung anderer gleichförmig dazu nutzen, sich selbst positiv darzustellen. Entsprechend dieser Annahme ergibt sich für B/S ein starker und statistisch signifikanter Zusammenhang (r = .56, $p < .01$, N = 26). Kleine bis mittlere Effekte ($r > .10$) ohne statistische Signifikanz ($p > .05$) zeigen sich für E/F, IB, P und FvF. Für Klarheit und Stellvertretung zeigt sich quasi eine Nullkorrelation. Fast beruhigend ist, dass die Korrelationskoeffizienten so gering sind, deutet dies doch darauf hin, dass sich neben den Rater-Verzerrungen auch das Ratee-Verhalten in den Ratings widerspiegelt (und keine gleichförmige Rater-Verzerrung).

Um festzustellen, ob es hauptsächlich die gleichen Führungskräfte sind, die sich gegenüber ihren unterstellten Mitarbeitern mit Führungsverantwortung sowie auch gegenüber ihren Vorgesetzten besser einschätzen, wurden Häufigkeitsvergleiche gerechnet. Das Schema in Tabelle 62 zeigt die Kontingenztafeln, die diesen zugrunde gelegt wurden. Die Einordnung ist lediglich als grobe Annäherung zu verstehen, weil nicht unterschieden wurde, ob sich eine Führungskraft gering oder stark unter- bzw. überschätzt. Es gingen also auch minimale Abweichler in die Kategorisierung ein. Genau genommen gibt es aber einen

Range, innerhalb dessen von einer übereinstimmenden Bewertung gesprochen werden müsste. Zur besseren Illustration wird darauf verzichtet, zumal jede Range-Festlegung ein willkürliches Kriterium darstellen würde.[176]

Durchschnittlich (über die sieben Führungsdimensionen hinweg) beurteilen 63% der Führungskräfte ihr eigenes Führungsverhalten besser als das ihrer Vorgesetzten (Range über die sieben Führungsdimensionen: 33% bis 95%) und 65% der Führungskräfte ihr eigenes Führungsverhalten besser als das ihrer Mitarbeiter (Range: 45% bis 80%). Das bedeutet, dass sich auf keiner Dimension überzufällig viele Führungskräfte *unterschätzen*, dass sich jedoch – bei einem Signifikanzniveau von 6% – auf fünf Dimensionen überzufällig viele Führungskräfte *überschätzen*. Dabei unterscheidet sich die Anzahl der Selbstüberschätzer relativ zur vorgesetzten Führungskraft nicht von der Anzahl der Selbstüberschätzer relativ zur nachgeordneten Führungskraft.[177]

Zwischen 28% (IB) und 75% (FvF) der Führungskräfte bewerten sich relativ nach oben *und* nach unten positiver, im Durchschnitt sind das 45% der Führungskräfte. Dies erweist sich für E/F, P und FvF als überzufällig ($p < .05$). Durchschnittlich sind es nur 17% der Führungskräfte (Range: 0% bis 30%), die sich selbst sowohl im Vergleich zum Führungsverhalten ihrer Vorgesetzten *als auch* im Vergleich zum Führungsverhalten ihrer nachgeordneten Führungskräfte (Mitarbeiter) *negativer* bewerten. Lediglich für FvF (keine Unterschätzer) stellt sich dies als überzufällig wenig heraus, für keine Dimension als überzufällig viele (vgl. Tabelle 62).

Tabelle 62. Vierfelder-Schema: Vergleich des Above-Average-Effekts nach oben vs. unten

	Selbst > bzw. < Vorgesetzter													
Faktoren	E/F		B/S		IB		K		P		FvF		S	
	>	<	>	<	>	<	>	<	>	<	>	<	>	<
Selbst > Mitarbeiter	13	3	7	2	5	7	8	5	9	3	15	1	2	2
Selbst < Mitarbeiter	4	3	5	6	1	5	5	2	4	4	4	0	1	1
Kontinuitätskorr. χ^2	.04		.89		3.78		.03		.04		1.25		(.08)	
p für >/>	<.001		.21		.48		.10		.04		<.001		.47	
p für </<	.16		.79		.72		.09		.41		<.01		.53	

Anmerkungen. Die kontinuitätskorrigierte χ^2-Formel des McNemar-Tests empfehlen Bortz et al. (1990, S. 161) für (b + c) < 30, $df = 1$, für S ist der χ^2-Wert wegen der geringen Stichprobe nicht interpretierbar, „*p* für >/>“: Binomialtest für die Wahrscheinlichkeit, dass sich *mindestens* soviele Führungskräfte sowohl nach oben als auch nach unten überschätzen, wie es für die einzelnen Faktoren angegeben ist (z.B. 13 für E/F), „*p* für </<“: Binomialtest für die Wahrscheinlichkeit, dass sich *höchstens* soviele Führungskräfte sowohl nach oben als auch nach unten unterschätzen, wie es für die einzelnen Faktoren angegeben ist (z.B. 3 für E/F), die Stichprobe schwankt für die einzelnen Faktoren, da die Werte für Gleich-Beurteilung nicht kodiert wurden.

[176] Die Schätzungen sind daher als konservativ zu bezeichnen, denn die Überschätzungen sind deutlich höher als die Unterschätzungen und lägen daher mit höherer Wahrscheinlichkeit über einem solchen Range.

[177] Die McNemar-χ^2-Tests bleiben für alle sieben Führungsdimensionen unter dem Signifikanzniveau von 5% (vgl. Tabelle 62).

Die Mehrzahl der Führungskräfte beurteilt das eigene Führungsverhalten positiver als das anderer Führungskräfte, was die Annahme von Hypothese 21) bestätigt. Dabei gilt auch, dass die Selbstüberschätzung nach oben unabhängig von der Selbstüberschätzung nach unten ist (Hypothese 23). Ein direkter Vergleich mit den Befunden traditioneller A-bove-Average-Studien (z.B. Meyer, 1980) ist nicht möglich, da in der vorliegenden Arbeit das eigene Führungsverhalten nicht *in Relation* zum Führungsverhalten der *durchschnittlichen* Führungskraft eingeordnet wurde.

Umso mehr untermauern die vorliegenden Ergebnisse, dass auch im Führungsbereich von einer Selbstüberschätzung auszugehen ist. Wenn auch die obigen prozentualen Angaben nicht so imposant wirken wie die von Meyer (1980, vgl. Kapitel 3.2.3), so wurde in der vorliegenden Studie jedoch gewährleistet, dass sie nicht durch ein relationales Ratingformat verfälscht wurden. Eine negative Bewertung anderer führt hier also nicht unmittelbar zu einer vergleichsweise positiven Selbstdarstellung, während relationale Items buchstäblich zu einem solchen Vergleich einladen. Daneben ist anzunehmen, dass die Selbstüberschätzung auch deshalb nicht so gravierend ausfällt, weil eine *konkrete* Person beurteilt wurde, zu der außerdem *Kontakt* besteht. Beide Bedingungen (konkrete Person, Kontakt) konnten in Laborstudien als Faktoren identifiziert werden, die die Selbstüberschätzung reduzieren (Alicke et al., 1995, für Beurteilung von Traits; vgl. Kapitel 3.2.3).

Die vorliegenden Ergebnisse erweitern die aktuelle Befundlage folglich insofern, als letztgenannte Bedingungen auch im Feld bei alltagsrelevanten Beurteilungsdimensionen wirksam sind und dennoch eine Selbstüberschätzung auftrat. Das spricht entweder dafür, dass Vergleichsprozesse auch dann zu positiver Selbstdarstellung genutzt werden, wenn sie nicht offensichtlich provoziert werden, oder dass Vergleichsprozesse keine notwendige Voraussetzung für derartige Selbstüberschätzung darstellen. Dann wäre das beobachtete Ausmaß an Selbstüberschätzung darauf zurückzuführen, dass für die Beurteilung anderer eine weniger konkrete und umfassende Informationsgrundlage besteht als für die Beurteilung eigener Verhaltensweisen und dass diese Unsicherheit zum eigenen Vorteil genutzt wird (vgl. Alicke et al., 1995; Perloff & Fetzer, 1986). Alicke et al. (1995, p. 806) beenden die Erklärungsnot, indem sie kurzerhand eine „I am better than others“-Heuristik annehmen.

Man kann sich dieser Problematik durch die Suche nach charakterisierenden Momenten der Überschätzer nähern. Eine erste Frage, die sich stellt, bezieht sich auf den Zusammenhang zwischen der Selbstüberschätzung und der Fremdbeurteilung *durch* andere (Mitarbeiter, Vorgesetzte). Fest steht, dass ein Above-Average-Effekt per definitionem nicht die tatsächlichen Leistungen widerspiegeln kann und daher auch nicht von einem hohen positiven Zusammenhang zu den Fremdbeurteilungen auszugehen ist. Mehr Erkenntnisse über den Entstehungsmechanismus der Selbstüberschätzung könnten sich aus der Gegenüberstellung mit den entsprechenden Fremdbeurteilungen dennoch ergeben. Eine negative Korrelation könnte einen Hinweis auf mangelnde Führungskompetenz liefern.

Die Ergebnisse sind freilich ernüchternder Natur: Es bestehen quasi keine korrelativen Zusammenhänge zwischen der Selbstüberschätzung nach unten und der mittleren Aufwärtsbeurteilung durch die Mitarbeiter. Für alle Führungsdimensionen mit Ausnahme von Partizipation ergibt sich zwar ein positiver Korrelationskoeffizient, im Mittel beträgt er jedoch nur $\bar{r} = .13$. Für Partizipation ergibt sich ein hochsignifikant negativer Zusammenhang ($r = -.49$, $N = 37$). Je stärker die Selbstüberschätzung nach unten also ausfällt, umso negativer wird die Führungskraft von unten beurteilt. Möglicherweise ist die Selbstüberschätzung folglich auch dadurch genährt, dass die sich stark überschätzenden Führungskräfte nicht in der Lage sind, das Führungsverhalten ihrer unterstellten Führungskräfte adäquat wahrzunehmen und sie deshalb von diesen negativ beurteilt werden. Man kann schließen, dass die überdurchschnittlich gute Beurteilung der eigenen Führungsfertigkeiten auf keiner von außen beobachtbaren Leistung beruht, sondern auf kognitiven und motivationalen Verzerrungen der Führungskräfte.

Eine zweite Frage betrifft die Erfolgsvariablen. Sind Überschätzer in den Augen ihrer Mitarbeiter weniger erfolgreich als Unterschätzer? Eine entsprechende Testung auf Mittelwertsunterschiede (U-Test) erbrachte für keine der Führungsdimensionen auch nur annähernd bedeutende Effekte. Ob sich eine Führungskraft im Vergleich zum Führungsverhalten ihrer vorgesetzten oder nachgeordneten Führungskraft selbst positiver oder negativer beurteilt, ist also unabhängig davon, wieviel Erfolg ihr von ihren Mitarbeitern zugesprochen wird.

Das gleiche gilt drittens für die Zuschreibung eines bestimmten Führungsverhaltens aus dem MLQ-internen Range an Führungsverhalten und für verschiedene demografische/strukturelle Variablen (Alter des Raters und des Ratees, Hierarchie, Dauer der Zugehörigkeit zu Organisation und Abteilung). Es sind also keine zusätzlichen Hinweise oder moderierende Einflussfaktoren auf den Entstehungsmechanismus der Selbstüberschätzung auszumachen.

Zusammenfassend wird festgehalten:

- Die Führungskräfte beurteilen ihr eigenes Führungsverhalten durchschnittlich positiver als das ihrer Vorgesetzten sowie positiver als das ihrer Mitarbeiter mit Führungsverantwortung.
- Die Selbstüberschätzung variiert in unbedeutendem Ausmaß über die Führungsdimensionen hinweg.
- Die Selbstüberschätzung fällt im Vergleich zu vorgesetzten und nachgeordneten Führungskräften ähnlich stark aus.
- Überdurchschnittlich viele Führungskräfte beurteilen ihr eigenes Führungsverhalten positiver als das anderer – übergeordneter und untergeordneter – Führungskräfte.
- Die Selbstüberschätzung ist unabhängig von der Fremdbeurteilung durch andere.
- Die Selbstüberschätzung ist unabhängig vom zugesprochenen Führungserfolg, vom zugesprochenen Führungsstils sowie von verschiedenen demografischen und strukturellen Variablen.

6.4 Validität von Führungsbeurteilungen

6.4.1 Zusammenhang zwischen Führungsverhalten und Erfolg

In diesem Kapitel wird überprüft, ob die Führungsbeurteilungen aus den verschiedenen Perspektiven (Selbst-, Abwärts- und Aufwärtsperspektive) geeignete Prädiktoren für die erhobenen Erfolgskriterien darstellen. Tabelle 63 gibt zur Erinnerung einen Überblick über die Erfolgskriterien und die verwendeten Abkürzungen, die zur Validierung herangezogen wurden.

Tabelle 63. Überblick über die Erfolgskriterien

Kriterium	Abkürzung
Arbeitszufriedenheit (Interaktionsfaktor)	AZ1
Arbeitszufriedenheit (Hygiene-Faktor)	AZ2
Engagierte Führung (Arbeitszufriedenheit)	EF-AZ
Wohlbefinden	WOBE
Engagierte Führung (Wohlbefinden)	EF-WOBE
Leistung	L
Zufriedenheit mit Beurteilung	ZB
Extra Effort (MLQ)	EEF
Effectiveness (MLQ)	EFF
Satisfaction (MLQ)	SAT

Anmerkung. EF-AZ, EF-WOBE und ZB sind Differenzmaße und folglich negativ gepolt. Je höher die jeweilige Differenz, umso negativer.

Insbesondere geht es um den Vergleich der Kriteriumsvalidität der unterschiedlichen Perspektiven sowie um den Vergleich der Kriteriumsvalidität zwischen aggregierten und individuellen Fremdratings. Genauer geht es um die Überprüfung folgender Hypothesen:

24) Die Fremdbeurteilungen von Führung korrelieren positiv mit Erfolgskennwerten.

25) Die Aufwärtsbeurteilungen von Führung korrelieren stärker mit den Erfolgskennwerten als die Abwärtsbeurteilungen.

26) Aggregierte Aufwärtsbeurteilungen korrelieren mit den Erfolgsvariablen stärker als nicht aggregierte (individuelle) Aufwärtsbeurteilungen.

27) Aggregierte Fremdbeurteilungen (Aufwärts- und Abwärtsbeurteilungen) korrelieren mit den Erfolgsvariablen nicht stärker als aggregierte Aufwärtsbeurteilungen alleine.

28) Selbstbeurteilungen korrelieren in geringerem Ausmaß mit Erfolg als Fremdbeurteilungen.

Um diese Fragestellung 7 zu beantworten, wurden Korrelationsanalysen mit anschließender Unterschieds-Prüfung der Höhe der Korrelationskoeffizienten durchgeführt. Dadurch können die Ergebnisse, die in Kapitel 3.1.1 berichtet wurden, um Ergebnisse aus einem klassischen Industrieunternehmen ergänzt werden. Außerdem kann zur dort bereits angerissenen Diskussion über die Validität von Selbstratings sowie die Validität von aggregierten Ratings beigetragen werden. Möglicherweise stellen sich einzelne Führungsdimensionen oder Erfolgsvariablen als besonders brauchbar heraus. Untersucht werden die verschiedenen Führungsdimensionen (sieben FzFü-Dimensionen aus allen drei Perspektiven sowie die zehn MLQ-Faktoren aus der Aufwärtsperspektive).

Ob beziehungsweise welche Führungsdimensionen und Beurteilungsperspektiven *gemeinsam* mehr Varianz erklären als einzeln, ob also inkrementelle Validität vorliegt, wird im sich anschließenden Kapitel 6.4.2 untersucht. Dadurch kann dann auch der Masse an Korrelationskoeffizienten besser begegnet werden. Zunächst ist zu einer korrelationsanalytischen Auswertung in diesem Rahmen nämlich anzumerken, dass eine unübersichtliche Anzahl an Korrelationskoeffizienten entsteht, von denen aufgrund des diesbezüglich unzureichenden Forschungsstandes keine Auswahl a priori stattfinden konnte. So werden für jede einzelne Perspektive 70 Koeffizienten für die FzFü-Dimensionen (sieben Dimensionen, zehn Erfolgsvariablen) berechnet. Das ergibt für drei Richtungen 210 Koeffizienten. Hinzu kommen 100 Korrelationskoeffizienten für die MLQ-Skalen (zehn Führungsdimensionen, zehn Erfolgsvariablen). Bei einer solchen Masse an Koeffizienten ist es wenig sinnvoll, die einzelnen signifikanten Koeffizienten „herauszupicken“, denn von 210 Koeffizienten werden elf allein zufällig auf dem 5%-Niveau signifikant (vgl. Bobko, 1995). Eine Reduktion und aggregierte Betrachtung ist also schon deshalb dringend angesagt, weil man nicht weiß, welche der signifikanten Korrelationskoeffizienten sich zufällig ergeben haben.

Eine Aggregation stellt deshalb auch Tabelle 64 dar, in der jeweils die geringste und höchste Korrelation sowie die mittlere Korrelation[178] zwischen der Führungsbeurteilung und jedem Erfolgskriterien angegeben ist. (faktorenspezifische Darstellung vgl. Anhang M). Für die Aufwärtsratings werden dort neben den Koeffizienten, die auf individuellen (Auf_i) und aggregierten Werten (Auf_a) beruhen, auch die Ergebnisse derjenigen Korrelationsanalysen berichtet, die auf einem gekreuzten Design basieren (Auf_c, Erklärung siehe Ergebnisdarstellung, S. 352). Von den Selbst- und Abwärtsbeurteilungen liegt naturgemäß jeweils eine pro fokussierter Führungskraft vor, so dass keine Werte aggregiert werden können. Nahezu alle in diesem Kapitel beschriebenen Ergebnisse nehmen Bezug auf Tabelle 64.

[178] Es wurde nicht ausschließlich die mittlere Korrelation angegeben, da der Range zusätzliche Aussagekraft hat. Insbesondere dann, wenn sowohl negative als auch positive Korrelation zu einem Mittel zusammengefasst werden, geht zuviel Information verloren.

Tabelle 64. Korrelativer Zusammenhang zwischen der Führungsbeschreibung (FzFü) und den Erfolgskriterien

Führungs-beurteilung (*N*) [a]	Aufwärtsbeurteilung									
	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB	EEF	EFF	SAT
Auf_i (203/133/52)	.42/.72 (.64) [b]	-.06/.24 (.15)	-.51/-.35 (-.39)	-.34/.25 (.19)	-.34/-.26 (-.30)	.32/.55 (.48)	-.35/-.48 (-.40)	.60/.69 (.67)	.60/.81 (.72)	.62/.81 (.75)
Auf_a (40/30)	.12/.70 (.51)	-.22/.10 (-.06)	-.74/-.09 (-.45)	-.31/.13 (-.06)	-.49/-.12 (-.39)	.16/.53 (.39)	-.18/-.51 (-.40)	.28/.74 (.54)	.34/.73 (.61)	.52/.82 (.73)
Auf_c (203/133/52)	.17/.25 (.20)	-.02/.10 (.00)	-.12/.02 (-.08)	-.09/.06 (.01)	-.23/-.13 (-.17)	.04/.16 (.13)	-.13/.10 (-.07)	.06/26 (.20)	.14/.26 (.23)	.23/.31 (.32)
Ab (32)	-.10/.31 (.14)	-.30/-.05 (-.14)	-.28/.25 (-.01)	-.07/.03 (-.01)	-.11/.07 (.01)	-.06/.36 (.17)	-.44/-18 (-.32)	.28/.47 (.38)	.14/.42 (.38)	.12/.44 (.29)
Selbst (40)	-.20/.01 (-.07)	-.07/.27 (.05)	.00/.31 (.19)	.07/.33 (.21)	-.10/.43 (.07)	.36/-.01 (-.19)	.03/.23 (.13)	-.25/.38 (.11)	-.22/.23 (.02)	-.15/.24 (.10)

Anmerkungen: grau hinterlegt: signifikante Korrelationen einzelner oder aller Führungsfaktoren mit dem entsprechenden Erfolgskriterium.
[a]Selbst: Selbstbeurteilung, Auf: Aufwärtsbeurteilung, Ab: Abwärtsbeurteilung. Die Selbst- und Abwärtsbeurteilung von Führung wurde jeweils mit dem Median der Aufwärts-Erfolgs-Schätzungen korreliert, die Aufwärtsbeurteilung mit den Aufwärtsmedianen (Auf_a; *N* = 40) sowie mit nicht aggregierten, individuellen Werten (Auf_i; *N* = 200). Auf_c: Criss-Cross-Auswertung.
N ist unterschiedlich für die FzFü-Dimensionen (E/F-P/FvF/S).
[b]Es ist jeweils die niedrigste und höchste Korrelation zwischen einem der sieben Führungsdimensionen und dem entsprechenden Erfolgskriterium angegeben. In Klammern wird die mittlere Korrelation (über die sieben Führungsdimensionen) berichtet.

Was die Korrelationen zwischen den einzelnen Beurteilungsperspektiven (FzFü) und den Erfolgsvariablen angeht, ist die **Aufwärtsbeurteilung** hervorzuheben. Es sind durchgehend für alle Erfolgskriterien deutliche – über die sieben Führungsdimensionen gemittelte – Effekte festzuhalten, für vier der zehn Kriterien sogar starke Effekte. Betrachtet man nur die Korrelationskoeffizienten, die auf der Basis der nicht aggregierten Aufwärtsbeurteilungen entstanden, dann bleiben von den 70 Koeffizienten nur fünf unter dem 5%-Niveau (vgl. Anhang M). Insgesamt gilt also: Je positiver das Führungsverhalten von den Aufwärtsbeurteilern bewertet wird, umso höher sind die Erfolge.

Der Aufwärtsbeurteilung darf aber nicht voreilig eine hohe Kriteriumsvalidität zugesprochen werden, da die hohen Korrelationen zu den Erfolgskriterien auch dadurch mitbedingt sein können, dass beide Konstrukte (Führung und Erfolg) von den gleichen Beurteilern eingeschätzt wurden. Um diese potentielle Single-Source-Problematik weiter aufzuklären, wurden die Daten mit einem Criss-Cross-Design analysiert (vgl. Seltzer & Bass, 1990).[179] Das bedeutet, dass die Führungsbeurteilung, die ein Mitarbeiter für seine Führungskraft abgibt, nicht zur Vorhersage seiner eigenen Erfolgseinschätzung herangezo-

[179] Dadurch kann nicht nur für die speziell vorliegende Fragestellung, sondern für verschiedene Fragestellungen der vorliegenden Studie fundierter argumentiert werden, wenn es um die Single-Source-Problematik geht.

gen wird, sondern dass die Führungsbeurteilung des einen Mitarbeiters mit den Erfolgseinschätzungen eines seiner Kollegen verknüpft wird. Wenn beispielsweise für eine Führungskraft A zwei Aufwärtsbeurteilungen (von Mitarbeiter B und C) vorliegen, dann wird die Führungsbeschreibung von B mit der Erfolgseinschätzung von C korreliert und umgekehrt. Analog wird für mehr Aufwärtsbeurteilungen (z.B. von Mitarbeiter B, C und D) vorgegangen. Die Führungsbeurteilung von B wird mit der Erfolgseinschätzung von C korreliert, die Führungsbeurteilung von C mit der Erfolgseinschätzung von D und die Führungsbeurteilung von D mit der Erfolgseinschätzung von B. Die Criss-Cross-Zuordnung erfolgt zufällig.

Die Analysen mittels des Criss-Cross-Designs erbringen folglich eine sehr konservative Schätzung des wahren Wertes, denn es werden nur die Effekte berücksichtigt, die zwischen der Führung und denjenigen Erfolgen bestehen, die sich auf der Ebene der Mitarbeiter-*Gruppe* zeigen. Alle Führungserfolge, die speziell die Dyade Führungskraft-Mitarbeiter auszeichnen, werden ignoriert. Voraussetzung für eine sinnvolle Anwendung dieser Methode ist daher, dass nicht nur *individuelle* Führungseffekte anzunehmen sind, sondern auch Gruppeneffekte. Diese Annahme findet sich beispielsweise wieder im Multi-Ebenen-Modell der Führung von Yammarino, Dansereau und Kennedy (2001). Für EFF sind vor diesem Hintergrund die stärksten Effekte zu erwarten, da zwei der vier Items der EFF-Skala ausdrücklich effektive Führung erfassen, die unabhängig vom einzelnen Mitarbeiter ist.[180] Alle anderen Erfolgskriterien beziehen sich stärker auf das individuelle Erleben des Einzelnen. Auch die relativ geringen Intraklassen-Koeffizienten (vgl. Kapitel 6.1) legten bereits nahe, dass personengebunden geführt wird und deshalb auch personengebundene Erfolge zu erwarten sind.

Tatsächlich kommt es zu einer beträchtlichen Reduktion der Korrelationen zwischen Führungsbeurteilung (Aufwärts) und Erfolgsvariablen, wenn man die „gekreuzten" (Criss-Cross-Auswertung) den „personengebundenen" Korrelationen gegenüberstellt (vgl. Tabelle 64). Doch auch nach der Criss-Cross-Auswertung weisen alle FzFü-Dimensionen außer Stellvertretung zu zwei bis sechs, durchschnittlich zu fünf der zehn Erfolgsvariablen mindestens auf dem 5%-Niveau signifikante Korrelationen auf (vgl. Anhang M). Fast alle Korrelationskoeffizienten erreichen mindestens geringe Stärke. Alle mittleren Korrelationen besitzen auch hier – im Unterschied zur Abwärts- und Selbstbeurteilung – das erwartete Vorzeichen. Der Zusammenhang zum Erfolgskriterium EFF stellt sich nur als zweithöchster – nicht wie erwartet als höchster – heraus ($\bar{r}$ = .23 vs. SAT $\bar{r}$ = .32). Jedoch steigt der Korrelationskoeffizient sprunghaft auf $\bar{r}$ = .66 an (Unterschied: q = .56), wenn man als Erfolgskriterium nicht die *gesamte* EFF-Skala heranzieht, sondern lediglich den Mittelwert der beiden oben herausgestellten, *gruppenbezogenen* Items (vgl. Fußnote 180). Für die personengebundene Auswertung unterscheidet sich der Zusammenhang zwi-

[180] Itemwortlaut: Meine direkte Führungskraft setzt sich effektiv für die Belange der Organisation ein bzw. kann eine Gruppe effektiv führen.

schen den FzFü-Führungsdimensionen und der gesamten EFF-Skala nicht von dem Zusammenhang zwischen den FzFü-Führungsdimensionen und der reduzierten EFF-Skala. Das lässt darauf schließen, dass für die erfassten Erfolgskriterien durch die Criss-Cross-Auswertung einerseits zwar potentielle Effekte verloren gehen. Dies lässt aber auch darauf schließen, dass dieser Verlust andererseits auf die Art der Erfolgskriterien und nicht auf die Beurteilerquelle – also den Single-Source-Bias – zurückzuführen ist.

Sowohl bei der „personengebundenen" als auch bei der „gekreuzten" Auswertung ergeben sich die höchsten Korrelationen für die MLQ-internen Erfolgsvariablen. Das ist ein Befund, der die Konstruktvalidität des neu konstruierten FzFü-Fragebogens stützt. Ähnlich stark sind die Zusammenhänge lediglich für AZ1.

Neben den FzFü-Dimensionen korrelieren auch die MLQ-Dimensionen stark mit den Erfolgsvariablen (Tabelle 65).[181] Die höchsten Korrelationen zeigen sich auch hier erwartungsgemäß für die MLQ-internen Erfolgsvariablen. Doch auch für die MLQ-externen Erfolgsvariablen liegen die Effekte für die transformationale Führung, CR und LF (invers) in einem mittleren bis hohen Bereich. Für MBA und MBP (invers) ergeben sich für die Mehrzahl der Erfolgskriterien mittlere Korrelationen.

Während AZ1, EEF, EFF und SAT am stärksten mit der transformationalen Führung, CR und LF (invers) zusammenhängen, kann zwar AZ2 insgesamt weniger gut prognostiziert werden, allerdings von allen MLQ-Skalen in ähnlichem Umfang. Das bedeutet, dass MBA und passive Führung (MBP, LF) zwar die Zufriedenheit mit Bedingungen (AZ2) ähnlich stark beeinflussen können wie aktive Führung, nicht aber die Zufriedenheit mit Faktoren, die durch Interaktionen mitbestimmt sind (AZ1). Das passt insofern zum Augmentationskonzept von Bass (vgl. Kapitel 2.5.3.1), dass weniger aktive Führung zwar als Basisvariable zu betrachten ist, dass durch aktivere Führung aber weiterführende Erfolge zu erzielen sind. So könnte die Zufriedenheit mit den Bedingungen als Basis-Erfolg konzeptualisiert werden, während AZ1 eine progressive Art von Arbeitszufriedenheit darstellt.

Für WOBE scheint neben der transformationalen Führung und CR vor allem MBA ein guter Prädiktor zu sein, während die passive Führung eher negativ mit dem Wohlbefinden zusammenhängt. Es ist also nicht die aktive und durchaus fordernde Führung, sondern die fehlende Führung, die bei den Mitarbeitern eher Stressempfinden hervorruft. Unklarheiten und Unsicherheiten könnten die kritische Variable in diesem Entstehungszusammenhang darstellen.

[181] Zur Erinnerung: Alle MLQ-Führungsbeurteilungen wurden aus der Aufwärtsperspektive abgegeben.

Tabelle 65. Korrelative Zusammenhänge zwischen der Führungsbeschreibung (MLQ) und den Erfolgskriterien

	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB	EEF	EFF	SAT
TF_i	.63	.21	-.42	.24	-.27	.44	-.45	.81	.84	.83
TF_a	.51	.06	-.39	.09	-.44	.56	-.43	.69	.69	.73
TF_c	.22	-.01	-.05	.06	-.14	.17	-.13	.17	.20	.23
CR_i	.68**	.27**	-.47**	.25*	-.22*	.43**	-.45**	.83**	.87**	.83**
CR_a	.47**	.04	-.42*	.25	-.19	.51**	-.55**	.60**	.67**	.74**
CR_c	.10	-.11	-.04	.03	-.04	.17	-.23*	.08	.17	.25*
MBA_i	.32**	.28**	-.12	.26*	.15	.13	-.11	.39**	.39**	.29**
MBA_a	.35+	.10	-.07	.31+	-.05	.12	.03	-.06	.02	-.04
MBA_c	.04	.05	.02	-.03	-.01	.00	-.01	-.04	-.06	-.07
MBP_i	-.47**	-.25**	.38**	-.03	.19+	-.31**	.34**	-.56**	-.55**	-.60**
MBP_a	-.23	-.12	.24	.03	.39*	-.52**	.35+	-.55**	-.56**	-.50**
MBP_c	-.10	-.03	-.07	-.07	-.04	-.15	.03	.00	-.17	-.14
LF_i	-.63**	-.29**	.34**	-.16	.18+	-.31**	.40**	-.75**	-.78**	-.75**
LF_a	-.52**	-.11	.31	-.16	.26	-.53**	.44*	-.62**	-.70**	-.65**
LF_c	-.16	-.02	-.01	-.09	.00	-.21	.11	-.14	-.21	-.21

Anmerkungen. TF: Mittlere Korrelation (daher auch keine Signifikanzangaben) aus den sechs Korrelationen zwischen den transformationalen MLQ-Führungsskalen (IIa, IIb, IM, IS, IC, AUS) und dem jeweiligen Erfolgskriterium, i: individuell (N = 85-91), a: aggregiert (N = 29-31), c: Criss-Cross (N = 85-91).

Auch für den Zusammenhang zwischen den MLQ-Führungsdimensionen und den Erfolgsvariablen wurden Criss-Cross-Analysen durchgeführt, um den Single-Source-Bias zu kontrollieren (Tabelle 65). Insbesondere verschiedene transformationale Subskalen (IIa, IIb, IC, AUS) korrelieren auch im gekreuzten Design mit verschiedenen Erfolgskriterien (AZ1, EF-WOBE, L, EEF, EFF, SAT) signifikant (bis zu r = .31, vgl. Anhang N). Auch CR korreliert mit den Erfolgseinschätzungen der jeweiligen Kollegen mit zwei Erfolgsmaßen in erwarteter Richtung signifikant (ZB, SAT, vgl. Tabelle 65). Dennoch zeigt sich insgesamt für alle MLQ-Führungsdimensionen – auch für die transformationalen Skalen sowie CR – eine deutliche Minderung der Korrelationskoeffizienten für die Criss-Cross-Ergebnisse im Vergleich zu den personengebundenen Ergebnissen.

Bei den MLQ-Analysen sprechen die Ergebnisse bezüglich der reduzierten EFF-Skala (vgl. Fußnote 180) genauso wie bei den oben berichteten FzFü-Analysen dafür, dass diese Minderung der Korrelationskoeffizienten allenfalls geringfügig am Single-Source-Bias liegt, sondern an der individuellen Ausrichtung der Erfolgsvariablen. Auch hier ergeben sich nämlich für die reduzierte EFF-Skala – also für die Items, die nicht personengebundenen Erfolg messen (vgl. Fußnote 180) – sehr hohe Korrelationskoeffizienten. Sie sind für alle MLQ-Führungsdimensionen sogar hochsignifikant (transformationale Skalen: $\bar{r}$ = .80, CR: r = .81, MBA: r = .38, MBP: r = -. 55, LF: r = -.75). Für die personengebundene Auswertung unterscheidet sich der Zusammenhang zwischen den MLQ-Führungsdimensionen und der reduzierten EFF-Skala wiederum nicht von dem Zusammenhang zwischen den MLQ-Führungsdimensionen und der gesamten EFF-Skala.

Vergleicht man die Kriteriumsvalidität der FzFü-Skalen mit der der MLQ-Skalen, dann stellt man fest, dass die FzFü-Skalen im Mittel soviel Erfolgsvarianz aufklären wie die transformationalen Skalen, CR oder LF. Alle Führungsdimensionen weisen für AZ1, EEF, EFF und SAT die höchste Validität auf. Nur MBA korreliert mit allen Erfolgskriterien außer AZ2 und WOBE tendenziell geringer als die anderen Führungsdimensionen. Hypothese 24) kann also für die Aufwärtsbeurteilungen der vorliegenden Studie bestätigt werden.

Auch die **Abwärtsbeurteilungen** weisen zu verschiedenen Erfolgskriterien bedeutsame Korrelationen auf. Wichtig ist anzumerken, dass hier generell kein Single-Source-Bias vorliegt, da die Führungsbeurteilungen von oben und die Erfolgsbeurteilungen von unten vorgenommen wurden. Die mittlere Abwärtsbeurteilung über alle sieben FzFü-Dimensionen hinweg erreicht für EEF, EFF, SAT und ZB Korrelationskoeffizienten mittlerer Stärke. Betrachtet man die Zusammenhänge zwischen einzelnen Führungsdimensionen und den Erfolgsvariablen, dann finden sich auch mittelstarke Korrelationen für AZ1 und L. Statistische Signifikanz erreichen einzelne Führungsdimensionen mit EEF, EFF, SAT und L in positiver sowie mit ZB erwartungsgemäß in negativer Richtung ($p < .05$). Aus der Abwärtsperspektive scheinen also eher die einsatz- und leistungsbezogenen Variablen prognostizierbar, weniger die Zufriedenheits- und Befindlichkeitsvariablen. Ein durchschnittlich kleiner Effekt in unerwarteter Richtung ergibt sich für den Zusammenhang zwischen der Abwärtsbeurteilung und AZ2. Je positiver Führungskräfte von oben beurteilt werden, umso unzufriedener sind ihre Mitarbeiter mit verschiedenen Kontextfaktoren. Möglicherweise spiegelt sich in den positiven Abwärtsbeurteilungen auch die Befürwortung eines zu fordernden oder aufgabenfokussierten Führungsstils wider.

Es sind insbesondere die Führungsdimensionen E/F, IB und S, die mit den genannten Erfolgsvariablen in statistisch bedeutsamem Zusammenhang stehen. Das sind auch genau die Führungsdimensionen, auf denen die Abwärts-Aufwärts-Korrelationen der Führungsbeurteilungen am stärksten sind (vgl. Tabelle 48 und Tabelle 49). Dies unterstreicht einmal mehr die Wichtigkeit interperspektivischer Übereinstimmung, sofern sie prognostische Funktionen übernehmen soll. Wenn durch die Abwärtsbeurteilung nicht für alle Erfolgsvariablen ein bedeutsamer Varianzanteil aufgeklärt werden kann, so gelingt dies doch für die zentralen Erfolgsvariablen. Hypothese 24) kann also auch für die Abwärtsbeurteilungen angenommen werden.

Vergleicht man die Höhe der Korrelationen, die die beiden Fremdbeurteilungsperspektiven zu den Erfolgskriterien aufweisen, dann liegen die Abwärts-Erfolgs-Korrelationen für die Mehrzahl der Erfolgsvariablen zwischen den personengebundenen und den gekreuzten Aufwärts-Erfolgs-Korrelationen. Dies ist deshalb zu erwarten, da die Führungsbeurteilungen aus der Abwärtsperspektive lediglich das generelle Führungsverhalten einer Führungskraft betreffen und nicht – wie aus der Aufwärtsperspektive – das Führungsverhalten, das

die beurteilte Führungskraft genau dem aufwärts beurteilenden Mitarbeiter gegenüber zeigt. Insofern wurde das Führungsverhalten aus der Abwärtsperspektive weniger individuumsbezogen erfasst und nähert sich daher dem Criss-Cross-Gedanken – wenn auch von der anderen Seite. In beiden Fällen reduziert sich der Zusammenhang zwischen Führung und Erfolg inhaltlich nämlich auf diejenigen Kernkomponenten, die die gesamte Mitarbeitergruppe und nicht nur einzelne Mitarbeiter/Aufwärtsbeurteiler betreffen. So können umgekehrt die Abwärtsbeurteilungen für kein einziges Erfolgskriterium bedeutsam mehr Varianz erklären als die gekreuzten Aufwärtsbeurteilungen.[182] Hypothese 25) wird vorläufig angenommen.

Dass nicht nur die personengebundenen Aufwärtsbeurteilungen, sondern auch die gekreuzten Aufwärtsbeurteilungen sowie die Abwärtsbeurteilungen Erfolgsvarianz erklären können, untermauert auch die Annahme über die *Art* des Zusammenhangs. Dadurch kann nämlich ausgeschlossen werden, dass implizite Annahmen über den Zusammenhang von Führung und Erfolg die Ergebnisse bestimmen. Aufgrund solcher Annahmen hätte angenommen werden können, dass Mitarbeiter ihren Führungskräften entweder dann eine bessere Führung zusprechen, wenn ihre Erfolgswahrnehmung hoch ist oder umgekehrt dann mehr Erfolg zuschreiben, wenn sie die Führung positiv beurteilen.[183] Wenn die Führungs- und Erfolgsbeurteilungen aber von verschiedenen Personen abgegeben werden (Abwärtsbeurteiler bzw. Kollege im Criss-Cross-Design), dann können sie als unabhängig voneinander betrachtet werden, so dass solche impliziten Annahmen über die Wirkung von Führung nicht mehr greifen.

Es wurde angenommen, dass der Zusammenhang zwischen den Fremdbeurteilungen und den Erfolgskennwerten dann stärker ausfällt, wenn man die Fremdratings aggregiert (vgl. Mount & Scullen, 2001). Die Korrelationen zwischen den Aufwärtsbeurteilungen von Führung (FzFü- sowie MLQ-Dimensionen) und den Erfolgsvariablen liegen für aggregierte und individuelle Aufwärtsratings indes in ähnlicher Höhe (vgl. Tabelle 64, Tabelle 65, Anhang M und N). Das gleiche gilt für die aggregierten Fremdratings insgesamt (Aufwärts- und Abwärtsratings hinsichtlich derselben Führungskraft), die Korrelationsindizes unterscheiden sich nicht von denen aggregierter Aufwärtsratings ($q < .10$).

Dass aggregierte Ratings nicht generell mehr Erfolgsvarianz aufklären, kann zum einen an der relativ hohen Varianzaufklärung der individuellen Aufwärtsratings liegen, zum anderen an der teilweise geringen Interrater-Übereinstimmung (vgl. Kapitel 6.1) oder der teilweise geringen Anzahl an Aufwärtsbeurteilern (durchschnittlich vier). Für Letzteres spricht, dass die aggregierten Aufwärtsbeurteilungen mit den Erfolgskriterien in den Fällen

[182] Am stärksten ist der Unterschied für die Zufriedenheit mit Beurteilungen (ZB), das sich aus den Abwärtsbeurteilungen tendenziell besser prognostizieren lässt als aus den gekreuzten Aufwärtsbeurteilungen ($z = 1.27$, $p = .10$).

[183] Letzteres ist aufgrund der Reihenfolge der Fragebogenitems wahrscheinlicher, da zunächst die Führungs-, dann die Erfolgsvariablen erhoben wurden. Implizite Annahmen könnten sein: „Bei so guter Führung kann ich doch gar nicht unzufrieden, gestresst etc. sein."

stärker korrelieren, in denen überdurchschnittlich viele – mindestens fünf (N = 11) vs. weniger als fünf (N = 29) – Aufwärtsratings vorlagen. Die mittlere Korrelation (über E/F, B/S, IB, K und P hinweg) unterscheidet sich zwischen diesen beiden Teilstichproben für AZ2 und EEF stark (AZ2: r = -.16 versus r = .47, z = -1.66, p = .04, q = .67 bzw. EEF: r = .62 versus r = .87, z = -1.50, p = .06, q = .60), für EF-AZ und L mittelmäßig (EF-AZ: r = -.38 versus r = -.61, z = .76, p = .22, q = .31 bzw. L: r = .42 versus r = .72, z = -1.13, p = .13, q = .47) und für EFF mit geringem Effekt (r = .65 versus r = .76, z = -.55, p = .29, q = .21). Im Mittel ergibt sich für alle Erfolgskriterien ein mittelstarker Effekt (r = .46 versus r = .62, z = -.56, p = .29, q = .22). Dann sind die aggregierten Aufwärtsratings auch im Vergleich zu den individuellen Ratings valider. Der Effekt wächst weiter an, wenn man nicht die dichotomisierten Teilstichproben, sondern stärkere Extremgruppen vergleicht. Es ist durchaus plausibel, dass bei nur mittelmäßiger Beurteilerübereinstimmung (vgl. Kapitel 6.1.2) aus der Aggregation *weniger* Aufwärtsratings kein so großer Reliabilitätsgewinn entstehen kann, dass er sich auf die Validität spürbar auswirkt. Folglich kann Hypothese 25) in Abhängigkeit von der Rateranzahl angenommen werden, ohne eine solche zusätzliche Bedingungsannahme ist sie jedoch abzulehnen. Die Annahme von Hypothese 27) konnte bestätigt werden.

Da die Validität der aggregierten Ratings aber auch in der Gesamtstichprobe mindestens so hoch ist wie die der individuellen Ratings, werden fortan aggregierte Ratings verwendet – auch um die Kehrseite der Medaille zu vermeiden. Würde man nämlich die Auswertungen, in denen neben Beurteilungen aus der Aufwärtsperspektive Beurteilungen aus der Selbst- oder Abwärtsperspektive integriert sind, mit den *individuellen* (statt aggregierten) Aufwärtsratings durchführen, dann hätte man unweigerlich aus anderem Grund Verzerrungen: Zu einer Gegenüberstellung mit der Aufwärtsperspektive müssten die Selbst- und Abwärtsbeurteilungen einer Führungskraft jeweils mehrfach in die Analysen eingehen.

Im Unterschied zu den Fremdbeurteilungen korreliert die **Selbstbeurteilung** nur sehr gering mit den erhobenen Erfolgskriterien. Von den 70 berechneten Korrelationskoeffizienten üben nur 42 einen geringen Effekt aus. Sucht man nach statistisch signifikanten Zusammenhängen, dann findet man lediglich erwartete signifikant positive Korrelationen zwischen der Selbstbeurteilung auf der Dimension Klarheit und dem Wohlbefinden der Mitarbeiter sowie zwischen der Selbstbeurteilung auf der Dimension Stellvertretung und Extra Effort (vgl. Anhang M). Erstgenannter Zusammenhang kann neben den Selbst-Aufwärts-Korrelationen auf der Dimension Klarheit als zweiter Hinweis dafür gewertet werden, dass die Selbstbeschreibung auf dieser Dimension besonders gut gelingt.

Erstaunlich ist die hohe Anzahl negativer, wenn auch überwiegend statistisch unbedeutender Korrelationen zwischen der Selbstbeurteilung und Erfolg. So zeigen sich für alle Erfolgsfaktoren mit Ausnahme von WOBE zu einigen Führungsdimensionen Zusammenhänge in nicht erwarteter Richtung. Die Selbstbeschreibung auf der Dimension Entwickeln/Fördern korreliert sogar signifikant negativ mit der subjektiven Leistung der

Mitarbeiter. Möglicherweise zeigt das, was die Führungskräfte selbst unter Entwickeln/Fördern verstehen, den Mitarbeitern hohe Verbesserungsmargen auf, so dass sie relativ dazu ihr aktuelles Leistungsniveau schlechter bewerten. In einem solchen Fall könnte im Übrigen ein adäquater Zielsetzungsprozess mit Zwischenzielen eine Lösung bieten (Carver & Scheier, 1998; Locke & Latham, 1990).

Hervorzuheben ist weiter die Selbstbeurteilung auf der Dimension Partizipation. Je partizipativer Führungskräfte nämlich ihre eigene Führung beurteilen, umso eher werden sie von ihren Mitarbeitern so wahrgenommen, als würden sie weniger für die Arbeitszufriedenheit und das Wohlbefinden der Mitarbeiter tun als sie könnten. Das deutet nicht unbedingt darauf hin, dass Führungskräfte ihre Wahrnehmung der eigenen partizipativen Führung stark verzerren. Es deutet eher darauf hin, dass die partizipativen Führungselemente als „Nicht-Kümmern“ anstatt als Mitgestaltungschance verstanden werden.

Erstaunlich ist, dass dieser Zusammenhang für das Wohlbefinden (EF-WOBE) besonders dann gilt, wenn die Mitarbeiter Partizipation als überdurchschnittlich wichtig (vs. unwichtig, Mediansplit) einstufen (r = .68 vs. r = .42, N = 12 vs. N = 20, z = .92, 1-p = .18, q = .38). Dieser unerwartete Zusammenhang könnte bedeuten, dass die Mitarbeiter kaum Möglichkeiten sehen, die Lücke selbst zu füllen, die von ihnen zwischen dem, was ihre Führungskraft für ihr Wohlbefinden tut und dem, was sie tun könnte, wahrgenommen wird. Partizipation wäre in diesem Fall also bei inadäquaten Bereichen gefordert. Auch Wohlbefinden selbst korreliert mit der Partizipation (Selbst) weniger stark, wenn die Mitarbeiter Partizipation als wichtig (vs. unwichtig, Mediansplit) betrachten (r = .22 vs. r = .47, N = 12 vs. N = 20, z = -.69, p = .24, q = .29). Das deutet darauf hin, dass Partizipation auch eine stressbegünstigende Komponente enthält. In der Führungspraxis müsste demzufolge verstärkt darauf geachtet werden, auf welche Art und Weise und bei welchen Aufgaben partizipative Führung eingeführt und kommuniziert wird.

Für alle anderen Erfolgskriterien gilt, dass diejenigen Mitarbeiter, denen Partizipation überdurchschnittlich wichtig ist (vs. unterdurchschnittlich wichtig, Mediansplit), ihren Führungskräften umso mehr Erfolg zusprechen, je partizipativer diese sich selbst beschreiben ($\bar{q}$ = .23). Dieser Effekt spricht im Sinne von kontingenztheoretischen Modellen der Führung dafür, dass Führung dann effektiver ist, wenn sie die Bedürfnisse der Mitarbeiter trifft (vgl. Kapitel 6.4.3).

Im Übrigen können die Analysen zur Partizipation dazu beitragen, die negative Interkorrelation zwischen Selbst- und Aufwärtsbeurteilung auf der Dimension Partizipation besser aufzuklären (vgl. Tabelle 49). Es ist plausibel, dass diejenigen Mitarbeiter, denen Partizipation nicht wichtig ist, die als partizipativ initiierten Führungselemente ihrer Führungskräfte gar nicht als solche verstehen. In der Folge können dann ihre Beurteilungen auch nicht übereinstimmen. Die Wichtigkeitsbeurteilung der Mitarbeiter zeigt auch hier einen leicht moderierenden Effekt. Die Selbst-Aufwärts-Übereinstimmung (Partizipation, Ist-Version) ist weniger negativ, wenn die Mitarbeiter Partizipation als überdurchschnitt-

lich wichtig (vs. unterdurchschnittlich wichtig) erachten (r = -.37 vs. r = -.16, N = 21 vs. N = 12, z = -.56, p = .29, q = .23).

Generell ist Folgendes festzuhalten: Die Selbstbeurteilung ist zur Vorhersage von Wohlbefinden in der vorliegenden Stichprobe besser geeignet als Fremdbeurteilungen, während für alle anderen Erfolgskriterien (im Mittel über die FzFü-Dimensionen) das Umgekehrte zutrifft (vgl. Tabelle 64). Das gilt nicht nur für die aggregierten, sondern auch für die personengebundenen Aufwärtsbeurteilungen. Dass die Selbstbeurteilungen also auch dann weniger stark mit subjektiven Leistungskriterien korrelieren, wenn sie *einzelnen* Fremdbeurteilungen gegenüber gestellt werden, ist insofern wichtig, weil beim Vergleich einzelner Selbst- mit einzelnen Aufwärtsratings der Einfluss der Rateranzahl konstant gehalten wird. Die Aufwärtsratings werden also nicht durch einen aus höherer Rateranzahl erwachsenden Reliabilitätsvorsprung bevorteilt. Sogar mit der konservativen Criss-Cross-Methodologie liegen die Aufwärts-Erfolg-Korrelationen für alle Erfolgsvariablen außer AZ2 und WOBE in einem tendenziell höheren Bereich als die Selbst-Erfolg-Korrelationen (AZ1: z = -1.52, p = .06, q = .27; AZ2: z = .28, p = .61, q = .05; EF-AZ: z = -1.52, p = .06, q = .27; WOBE: z = 1.13, p = .87, q = .20; EF-WOBE: z = -1.35, p = .09, q = .24; L: z = -1.80, p = .04, q = .32; ZB: z = -1.12, p = .13, q = .20; EEF: z = -.52, p = .30, q = .09; EFF: z = -1.20, p = .12, q = .21; SAT: z = -1.29, p = .10, q = .23). Auch für den Bereich der Personalführung in Industrieunternehmen kann demnach Hypothese 27) bestätigt werden, dass Fremdbeschreibungen valider sind als Selbstbeschreibungen.

Da nicht auszuschließen ist, dass die teilweise sehr hohen bivariaten Korrelationen einer multivariaten Auswertung nicht stand halten, ist es unerlässlich, die Beziehungen zwischen den Führungsdimensionen (FzFü- und MLQ-Skalen) und den Erfolgskriterien auch durch multiple Korrelationen (Regressionsanalysen) zu untersuchen (vgl. Kapitel 6.4.2). Da der Ausgangspunkt der vorliegenden Arbeit darin bestand, Führungserfolg nicht nur durch bestimmtes Führungsverhalten vorherzusagen, sondern insbesondere zu überprüfen, ob Diskrepanzen in der Führungsbeurteilung Erfolgsvarianz aufklären können, müssen auch auf die Ergebnisse von Kapitel 6.4.2 weitere Analysen aufbauen. Erst wenn sich bestätigt, dass nicht nur bestimmte Führungsbeurteilungen aus bestimmten Perspektiven, sondern auch die Selbst-Fremd-Kongruenz (Führungsbeurteilung) mit der Führungseffektivität einhergehen, könnte daraus ein Ansatz zur Erklärung von Führungserfolg entstehen, in dem sowohl die Art der Führung als auch die Selbst-Fremd-Übereinstimmung eine Rolle spielen (vgl. Kapitel 6.5).

Zusammenfassend zeigt sich, dass

- Aufwärtsbeurteilungen auf fast allen Führungsdimensionen und für fast alle Erfolgsmaße eine hohe Kriteriumsvalidität aufweisen,
- Abwärtsbeurteilungen eine insgesamt etwas geringere Validität besitzen als Aufwärtsbeurteilungen und vor allem leistungsbezogene Kriterien erklären können,
- aggregierte Fremdratings nicht prinzipiell mehr Varianz aufklären als nicht aggregierte (individuelle) Fremdratings (nur bei hoher Rateranzahl) und dass
- Selbstbeurteilungen zwar für einzelne Variablen (Wohlbefinden) eine bessere Vorhersageleistung haben, aber generell deutlich weniger Erfolgsvarianz erklären als Fremdbeurteilungen.

6.4.2 Inkrementelle Validität der Führungsbeurteilungen

Beim Einsatz multidimensionaler Instrumente ist zunächst zu fragen, ob die einzelnen Dimensionen redundante oder unterschiedliche Information erfassen. Für erstgenannten Fall müsste auf die Multidimensionalität verzichtet werden. Verschiedene Hinweise wurden bereits dafür berichtet, dass die einzelnen Faktoren nicht substitutiv füreinander zu betrachten sind. In diesem Kapitel soll weiter überprüft werden, ob die verschiedenen FzFü-Faktoren für unterschiedliche Kriterien unterschiedlich gute Prädiktorleistungen liefern und ob sie dazu imstande sind, inkrementelle Beiträge zu liefern. Aus der Literatur gewinnt man oft genug den Eindruck, dass Führung *an sich* prädiktiv ist. Spezifischere Untersuchungen sind notwendig, um Führungserfolg spezifisch auf konkrete Führungsdimensionen und -aufgaben zurückführen zu können.

Noch mehr als die inkrementelle Validität, die die einzelnen Führungsdimensionen leisten, interessiert in der vorliegenden Studie die Frage, ob die Beurteilungen aus den drei erhobenen Perspektiven (Selbst-, Aufwärts-, Abwärtsbeurteilungen) inkrementelle Beiträge liefern. Neben dem oben berichteten Ergebnis, dass Beurteilungen aus unterschiedlichen Perspektiven mit unterschiedlichen Erfolgsvariablen unterschiedlich gut korrelieren, würde der Nachweis einer inkrementellen Validität die multiperspektivische Führungsbeurteilung zusätzlich begründen.

Am Ende des Kapitels wird überprüft, ob die transformationale Führung inkrementelle Beiträge über die transaktionale Führung hinaus (Augmentationshypothese) respektive über die FzFü-Faktoren, die traditionelle Führungsaufgaben erfassen, hinaus liefern kann.

Zur Fragestellung 8 gehören also folgende Hypothesen:

29) Verschiedene Führungsdimensionen klären gemeinsam mehr Erfolg auf als einzelne Führungsdimensionen.
30) Die Selbst-, Aufwärts- und Abwärtsbeurteilung von Führung erklären gemeinsam mehr Erfolgsvarianz als die Beurteilung aus einzelnen Perspektiven.
31) Die transformationale Führung klärt gemeinsam mit der transaktionalen Führung mehr Erfolgsvarianz auf als die transaktionale Führung allein.
32) Die transformationale Führung klärt zusätzlich zu den klassischen Führungsdimensionen Erfolgsvarianz auf.

Zur Überprüfung von Hypothese 29) wurden schrittweise Regressionsanalysen für jedes erfasste Erfolgskriterium durchgeführt, wobei schrittweise jeweils diejenigen Führungsdimensionen aufgenommen wurden, die die kleinste F-Wahrscheinlichkeit und gleichzeitig eine F-Wahrscheinlichkeit unter .05 aufwiesen. In Tabelle 66 sind neben den FzFü-Dimensionen, die in das Regressionsmodell für das entsprechende Erfolgskriterium aufgenommen wurden, auch die jeweiligen Anteile an erklärter Varianz angegeben.

Die Führungsdimensionen der **Selbstbeurteilung** erbringen für Wohlbefinden, Zufriedenheit mit Engagement der Führungskraft (Wohlbefinden), Leistung, Zufriedenheit mit Beurteilung und Extra Effort signifikante Erklärungsmodelle. Mit Ausnahme von IB werden alle anderen sechs Variablen zur Vorhersage einzelner Führungskriterien herangezogen, so dass davon ausgegangen werden kann, dass die prognostische Leistung verschiedener Führungsparameter für verschiedene Erfolgskriterien differiert. Eine getrennte Erfassung sowie eine getrennte Analyse der Führungsdimensionen sind also geboten. Für die Zufriedenheit mit dem Engagement der Führungskraft (Wohlbefinden) bzw. für Extra Effort leisten zwei bzw. drei Prädiktoren eine bessere Vorhersage als ein einzelner Prädiktor. Weitere Prädiktoren erhöhen für kein Kriterium den Anteil erklärter Varianz.

Die Regressionsanalysen auf die Abwärtsbeurteilung erbringen keine Information, die über bereits oben berichtete bivariate Korrelationen hinausgeht. Für die fünf Erfolgskriterien, für die durch die Abwärtsbeurteilung ein signifikanter Beitrag geliefert werden kann, wird jeweils nur eine FzFü-Dimension aufgenommen – entweder IB oder E/F. IB und E/F sind auch die Faktoren, auf denen die Abwärtsbeurteilungen am stärksten mit den Aufwärtsbeurteilungen korrelieren (vgl. Tabelle 49). Es ist also von einer vergleichsweise reliablen Abwärtsbeurteilung auszugehen, was die Chancen erhöht, einen prädiktiven Beitrag aufzudecken. Aus dieser Perspektive erscheinen die anderen Führungsdimensionen also größeren Wert zu haben.

Sowohl für die Selbst- als auch die Abwärtsbeurteilung müssen die Modelle jedoch aufgrund des geringen Stichprobenumfangs als instabil bezeichnet werden. Für die **Aufwärtsbeurteilungen** wurden nicht nur Regressionsanalysen mit den aggregierten Führungs- und Erfolgsbeurteilungen durchgeführt, sondern auch mit den individuellen.

Tabelle 66. Erklärte Varianz und aufgenommene Variablen in den verschiedenen Beurteilungsperspektiven (schrittweise Regression auf einzelne FzFü-Dimensionen)

	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB	EEF	EFF	SAT
Selbst [a] (*N*=40)	-	-	-	R^2 = .09 K	R^2 = .29 P FvF	R^2 = .10 E/F	-	R^2 = .39 S P B/S	-	-
Abwärts [a] (*N*=32)	-	-	-	-	-	R^2 = .10 IB	R^2 = .16 E/F	R^2 = .19 E/F	R^2 = .15 IB	R^2 = .17 IB
Auf_a [b] (*N*=40)	R^2 = .47 E/F	-	R^2 = .29 K	-	R^2 = .22 E/F	R^2 = .26 IB	R^2 = .24 B/S	R^2 = .54 E/F	R^2 = .52 K	R^2 = .70 K E/F
Auf_i [b] (*N*=200)	R^2 = .56 E/F IB K	R^2 = .05 E/F	R^2 = .17 P IB	R^2 = .06 B/S	R^2 = .09 P	R^2 = .35 IB P B/S	R^2 = .23 E/F	R^2 = .54 K E/F B/S	R^2 = .71 K E/F IB	R^2 = .71 K E/F P IB

Anmerkungen. [a]Regressoren = alle sieben FzFü-Dimensionen, [b]Regressoren = E/F, B/S, IB, K und P (ohne FvF und S, weil dadurch für die Aufwärtsperspektive die Stichprobe drastisch verkleinert würde), Auf_a: aggregierte Aufwärtsratings (Median), Auf_i: individuelle Aufwärtsratings, R^2 = korrigiertes R^2.

Es zeigt sich, dass die individuellen Aufwärtsbeurteilungen tendenziell eine stärkere Vorhersagekraft besitzen als die aggregierten. Eine Ausnahme bilden die Effekte für engagierte Führung (Arbeitszufriedenheit und Wohlbefinden) und die Zufriedenheit mit Beurteilung. Für diese ändern sich auch die ins Modell aufgenommene Prädiktoren, was auf eine unreliable Messung der Erfolgskriterien hinweist. Das ist deshalb plausibel, weil die Erfolgsmaße als Differenzmaße operationalisiert sind, welche prinzipiell weniger reliabel sind als einfache Messungen (Fehlerkomponenten von zwei Variablen).

Während sich unter den Regressoren in denjenigen Analysen, die auf den Medianen beruhen, nie die Dimension Partizipation befindet, werden in die Analysen, die auf den individuellen Angaben beruhen, alle fünf zugrunde gelegten Variablen für einzelne Erfolgskriterien aufgenommen. Die Arbeitszufriedenheit (AZ1, AZ2, SAT) und Arbeitsmotivation (EEF) scheinen stark mit E/F zusammenzuhängen. Dieser Befund geht beispielsweise mit dem Job Characteristics Model (JCM) von Hackman und Oldham (1976, 1980) stimmig einher. Wenn ein Mitarbeiter entwickelt und gefördert wird, dann werden dadurch in der Regel Arbeitscharakteristika gestärkt, die nach diesem Modell das Motivationspotential und letztlich unter anderem die Arbeitszufriedenheit positiv beeinflussen.

So werden durch eine adäquate Entwicklung und Förderung insbesondere die Variabilität, die Ganzheitlichkeit und die Autonomie positiv beeinflusst. Da eine Förderung und erst recht eine Beförderung in der Regel auf bestimmtem Leistungsverhalten basiert, wird dadurch auch die Rückmeldung im Sinne des Modells gesteigert. Dass auch Leistung aus der *E/F*-Selbstbeurteilung am besten prognostiziert werden kann, stimmt zusätzlich mit den Modellannahmen überein, da Hackman und Oldham verschiedene positive Auswirkungen annehmen, nicht nur hohe Arbeitszufriedenheit. Die Relevanz der zentralen JCM-Variablen (Variabilität, Ganzheitlichkeit, Bedeutung, Autonomie, Rückmeldung) für erlebte Arbeitszufriedenheit konnte übrigens von Fried und Ferris (1987) meta-analytisch bestätigt werden.

Auch gemäß des Modells zur Arbeitsmotivation von Porter und Lawler (1968) ist der enge Zusammenhang zwischen E/F und Arbeitszufriedenheit zu erklären. So können Mitarbeiter es als Belohnung ihrer Leistung betrachten, wenn sie entwickelt und ge-/befördert werden – und Belohnungen resultieren nach den Annahmen des Modells in Zufriedenheit.

Dass Leistung am besten durch IB, P und B/S vorhergesagt werden kann, verwundert im Sinne der traditionellen Führungsstil-Forschung nicht. Während IB und P eher den mitarbeiterorientierten Führungsstil verkörpern, gestaltet sich B/S eher aufgabenbezogen. Dass beide Führungsstile erfolgsversprechend sind, dass aber der mitarbeiterbezogenen Führung mehr Gewicht zugesprochen wird als der aufgabenbezogenen Führung, geht beispielsweise mit Zwei-Faktoren-Modellen bzw. mit den Annahmen des Managerial Grid von Blake und Mouton (1964) einher (vgl. Kapitel 2.4, für einen Überblick Gabele et al., 1992).

Da sich für alle MLQ-internen Erfolgskriterien die Führungsdimensionen K und E/F als die gewichtigsten Prädiktoren herausstellen, ist von beträchtlichen Überschneidungen dieser Erfolgskonstrukte auszugehen. An einer besseren Differenzierung kann zukünftig gearbeitet werden. Dass gerade K und E/F zentrale Prädiktoren sind, scheint „im Kleinen“ die Augmentationshypothese nachzuzeichnen: Als grundlegender Prädiktor geht Klarheit – ein eher transaktionales Führungsverhalten – in die Modelle ein und klärt jeweils den mit Abstand größten Teil der Varianz auf. Erst als zusätzliche Prädiktoren werden dann Führungsdimensionen wirksam (E/F, IB), die mehr transformationale Anteile haben.

Abschließend kann für diese erste Teil-Fragestellung 8 festgehalten werden, dass die FzFü-Dimensionen für einzelne Kriterien unterschiedlich gute Prädiktoren darstellen. Dabei ist es keineswegs so, dass alle Führungsdimensionen für alle Kriterien als Prädiktoren mit Vorhersagekraft fungieren könnten. Da in aller Regel nicht nur ein spezifisches Kriterium interessiert, sondern verschiedene Indikatoren für Führungserfolg interessieren, ist vor dem Hintergrund der soeben berichteten Ergebnisse eine multidimensionale Führungsbeurteilung zu empfehlen.

Es zeigt sich, dass die Führungsbeurteilung aus der Selbstperspektive nur für vier Kriterien prognostische Validität aufweist und auch die Abwärtsbeurteilung nur für fünf

Kriterien statistisch bedeutsame Beiträge liefern kann. Hingegen liefert die Beurteilung aus der Aufwärtsperspektive für *alle* erfassten Erfolgsvariablen bedeutsame und zudem *bessere* Vorhersagen (Auf_i). Auch die Modelle auf die aggregierten Aufwärtsbeurteilungen – die aufgrund des gleichen Stichprobenumfangs eher mit denen für die anderen Perspektiven vergleichbar sind – erbringen für acht der zehn Kriterien mittlere bis starke Effekte ($f^2 > .15$). An dieser Stelle sei angemerkt, dass sich der Zusammenhang zwischen der Aufwärtsbeurteilung und den Erfolgskriterien nicht vollständig durch den Single-Source-Bias erklären lässt. Es werden auch bei der Verwendung eines Criss-Cross-Designs (Criss-Cross-Design vgl. S. 352) für fünf Kriterien Modelle mit statistisch wertvoller Erklärungskraft entwickelt (AZ1, EF-WOBE, EEF, EFF, SAT). Die Effekte fallen dann jedoch geringer aus (f^2 von .05 bis .11).

Weiter ist die Frage, ob verschiedene Beurteilungsperspektiven inkrementelle Beiträge zur Erklärung von Führungserfolg liefern können, für einen zielgerechten Einsatz von multiperspektivischen Führungsbeurteilungen von grundlegender Natur. Erst wenn inkrementelle Validität oder spezifische Vorhersageleistung für einzelne Kriterien bestätigt werden können, ist es aus methodischer Perspektive gerechtfertigt und sinnvoll, tatsächlich verschiedene Perspektiven zu erheben. Die Erfassung redundanter Information könnte allenfalls praxisrelevante Vorzüge bieten, beispielsweise könnte dadurch eine höhere Akzeptanz und Glaubwürdigkeit der Beurteilungen erzielt werden. Um überhaupt inkrementelle Beiträge entdecken zu können, dürfen die Perspektiven untereinander lediglich gering korrelieren (vgl. Tabelle 49). Sie müssen also einzigartige Information enthalten. Zur Überprüfung dieser Fragestellung werden schritt- und blockweise Regressionsanalysen durchgeführt.

In den schrittweisen Regressionsanalysen auf die Aufwärts-, Selbst- und Abwärtsbeurteilungen (jeweils fünf Führungsdimensionen)[184] wurden für die verschiedenen Erfolgsvariablen verschiedene Prädiktoren aufgenommen. Hervorzuheben ist, dass nur für L und EEF Prädiktoren aus zwei verschiedenen Beurteilungsrichtungen (Selbst und Aufwärts) gemeinsam die höchste Varianz erklären. Dies sind auch die Erfolgskriterien, für die jede einzelne Perspektive einen signifikanten Beitrag leisten kann (vgl. Tabelle 66). Gezielter kann die Fragestellung mit blockweisen Regressionsanalysen untersucht werden.

Da es unwahrscheinlich ist, dass diejenigen Perspektiven in einem „inkrementellen Modell“ einen Beitrag liefern, die nicht einmal ohne Hinzunahme der Führungsdimensionen der jeweils anderen Perspektiven signifikante Vorhersagen leisten, wurden in blockweisen Regressionen nur die Beurteilungen derjenigen Prädiktoren auf inkrementelle Beiträge geprüft, die aus ihrer jeweiligen Perspektive Effekte aufweisen (vgl. Tabelle 66). Tabelle 67 zeigt die Ergebnisse dieser hierarchischen Regressionsanalysen.

[184] Es wurden lediglich E/F, B/S, IB, K und P als potentielle Prädiktoren eingegeben. Auf FvF und S wurde verzichtet, um den Stichprobenumfang nicht weiter reduzieren zu müssen.

Tabelle 67. Inkrementelle Validität der Beurteilungsperspektiven

Kriterien	Prädiktoren[a]		*N*	*B*°	*SE B*°	*β*°	*p*°	*Korr. R*^{2} *(p)*	*f*^{2}
EF-WOBE	P	(selbst)	37	.73	.24	.53	(.004)		
	FvF	(selbst)		-.32	.16	-.33	(.057)	.29** (.009)	.49
	E/F	(aufwärts)		-.25	.12	-.31	(.046)	.36* (.046)	.69
L	IB	(aufwärts)	32	.33	.20	.33	(.125)	.12$^+$ (.058)	.19
	E/F	(selbst)		-.55	.18	-.53	(.006)	.36** (.008)	.72
	IB	(abwärts)		.27	.25	.22	(.299)	.36 (.299)	.82
ZB	B/S	(aufwärts)	32	-.32	.09	-.54	(.002)	.37** (<.001)	.65
	E/F	(abwärts)		-.19	.15	-.20	(.208)	.39 (.208)	.74
EEF	E/F	(aufwärts)	32	.33	.09	.52	(.001)	.59** (<.001)	1.53
	S	(selbst)		.25	.10	.33	(.024)		
	P	(selbst)		-.60	.21	-.48	(.012)		
	B/S	(selbst)		.33	.13	.38	(.019)	.76** (.004)	4.03
	E/F	(abwärts)		.19	.11	.22	(.090)	.78$^+$ (.090)	4.92
EFF	K	(aufwärts)	32	.44	.09	.68	(<.001)	.53** (<.001)	1.19
	IB	(abwärts)		.17	.16	.15	(.282)	.53 (.282)	1.28
SAT	K	(aufwärts)	32	.38	.12	.52	(.003)		
	E/F	(aufwärts)		.23	.11	.32	(.047)	.71** (<.001)	2.73
	IB	(abwärts)		.21	.14	.16	(.143)	.73 (.143)	3.05

Anmerkungen. [a] Die Prädiktoren aus einer Perspektive werden jeweils in einem Block in die Regressionsgleichung eingegeben, ein Perspektivenwechsel zeigt den Beginn eines weiteren Blocks an, °Kennwerte des Gesamtmodells (letzter Block), **/*/$^+$*Änderung* (Δ R^2) in Signifikanz von *F*: *p* < .01/.05/.10, Determinationskoeffizienten für die einzelnen Blöcke.

Im ersten Schritt wurden jeweils die Prädiktoren aus derjenigen Perspektive eingegeben, die das beste Modell erbringt, wenn man die Perspektiven getrennt voneinander betrachtet (vgl. Tabelle 66). Da für AZ1, AZ2 und EF-AZ ausschließlich Regressionsmodelle auf die Aufwärtsperspektive und für WOBE ausschließlich auf die Selbstperspektive[185] statistisch bedeutsam sind, wird hierfür nicht weiter überprüft, ob die anderen Perspektiven einen zusätzlichen Beitrag liefern können. Die Ergebnisse werden im Folgenden für die einzelnen Erfolgskriterien nacheinander dargestellt. Entsprechend Tabelle 66 werden für WOBE-EF, ZB, EFF und SAT jeweils zwei, für L und EEF alle drei Perspektiven aufgenommen.

Für EF-WOBE erweisen sich die in Tabelle 66 aufgeführten Prädiktoren allesamt für ein gemeinsames Modell als nützlich. In blockweisen Regressionen erbringen P und FvF aus der Selbstperspektive einen zusätzlichen Erklärungswert über E/F (Aufwärtsperspektive) hinaus und ebenso umgekehrt.

Zur Vorhersage von ZB kann die Abwärtsbeurteilung von E/F keinen statistisch bedeutsamen zusätzlichen Beitrag über die Aufwärtsbeurteilung von B/S hinaus leisten, umgekehrt jedoch schon. Die traditionelle Beurteilungsrichtung der Praxis (Abwärtsbeurteilung) ist also durchaus brauchbar. Doch der Befund legt nahe, die Mitarbeiter bei Führungsbeur-

[185] Regressionsanalysen, in die (unter anderem) Prädiktoren aus der Selbst- oder Abwärtsperspektive eingehen, können nicht mit individuellen Aufwärtsratings kombiniert werden. Alle anderen Werte müssten dann nämlich entsprechend mehrfach in die Analyse eingehen.

teilungen mit einzubeziehen. Es gilt, dass trotz fehlender Signifikanz der Effekt nicht unbedeutend ist.

Analog gilt für EFF und SAT: EFF (SAT) wird durch die Hinzunahme der Selbstbeurteilung der Dimension IB nicht besser vorhergesagt als durch die Aufwärtsbeurteilung der Dimension Klarheit (und E/F) alleine, umgekehrt allerdings schon.

Für das in Tabelle 67 angegebene Modell für Leistung fällt auf, dass die alleinige Vorhersage durch die Aufwärtsbeurteilung von IB das 5%-Niveau knapp verfehlt. In umgekehrter Reihenfolge, also zunächst E/F (Selbstperspektive) und dann IB (Aufwärtsperspektive) ergibt sich ein Modell mit zwei annähernd gleich starken Beta-Gewichten. Dies ist auch das Ergebnis einer schrittweisen Regressionsanalyse, für die die drei berichteten Prädiktoren (E/F Selbstperspektive, IB Aufwärts- und Abwärtsperspektive) berücksichtigt werden können.

Für L kann generell festgehalten werden, dass zwei der drei Perspektiven mehr Varianz aufklären als eine einzelne Perspektive, dass aber die Hinzunahme der jeweils dritten keinen statistisch bedeutsamen Beitrag liefert. Da die Selbstperspektive ein negatives Regressionsgewicht hat, ist für die Praxis anzuraten, auf die Fremdbeurteilungsperspektiven zurückzugreifen. Zumindest sollten mehrere Perspektiven berücksichtigt werden. Dies kann auch als Hinweis für ein Diskrepanzmodell gewertet werden (vgl. Kapitel 6.5).

Zur Vorhersage von EEF leisten alle drei Beurteilungsperspektiven bedeutende Beiträge. Dies gilt nicht nur für das Modell, das in Tabelle 67 dargestellt ist, in dem die Abwärtsbeurteilung nur auf dem 10%-Niveau einen zusätzlichen Beitrag über die Aufwärts- und Selbstperspektive hinaus leistet. Geht man von einer praxisüblichen Beurteilungstradition aus, dann wird zunächst von oben beurteilt. Allein die Abwärtsbeurteilung erreicht einen hochsignifikanten Determinationskoeffizienten (korr. $R^2 = .29$, $p < .01$). Entscheidet sich ein Anwender oder Forscher dazu, auch Selbst- oder Aufwärtsbeurteilungen hinzuzunehmen, erbringen in beiden möglichen Reihenfolgen beide Perspektiven einen großen Zugewinn.

Die Ergebnisse legen insgesamt nahe, sich nicht auf einzelne Perspektiven zur Vorhersage von Führungserfolg zu verlassen, sondern Beurteiler aus verschiedenen Perspektiven einzubeziehen. Diese Schlussfolgerung wird dadurch gestärkt, dass aufgrund der überwiegend geringen Korrelationen zwischen der Selbst- bzw. Abwärtsbeurteilung und den Erfolgskriterien (Tabelle 64) inkrementelle Beiträge schwer zu entdecken sind. So verweisen die Ergebnisse auch auf eine besonders effektvolle Aufwärtsperspektive. Ihr Effekt kann nur selten durch neue Informationen aus anderen Perspektiven verbessert werden.

Folgeuntersuchungen müssen verstärkt darauf fokussieren, welche Beurteilungsperspektiven auf welchen Beurteilungsdimensionen welche Erfolgskriterien vorhersagen können. Mit dieser Untersuchung wird jedenfalls weiter dazu angeregt, sich nicht auf allgemeine Führungsbeurteilungen, sondern gezielt auf inhaltlich spezifische Dimensionen zu stützen. Dafür sprechen zusätzlich zwei weitere Befunde der vorliegenden Untersuchung.

Erstens wird bei einer Analyse anhand der Gesamtskala, also bei Missachtung der Faktorisierung, für kein Erfolgskriterium mehr als eine Beurteilungsperspektive in eine schrittweise Regression aufgenommen wird. Auch wenn lediglich die Abwärts- und Selbstperspektive als Regressoren fungieren, also die „mächtige“ Aufwärtsperspektive außen vor bleibt, erreichen nie beide Beta-Gewichte Signifikanz. Es entspricht der üblichen Beurteilungspraxis, dass dann die Abwärtsbeurteilung und nicht die Selbstbeurteilung häufiger den einzigen Prädiktor darstellt.

Zweitens wurden Regressionsanalysen durchgeführt, die jeweils nur die Selbst-, Abwärts- und Aufwärtsbeurteilung *einer* (der gleichen) Führungsdimension als Prädiktoren aufnahmen. Ausschließlich zur Vorhersage von Leistung durch die Dimension Klarheit tragen *zwei* Perspektiven bedeutend bei (Selbst- und Aufwärtsbeurteilung). Das bedeutet, dass aus unterschiedlichen Perspektiven die Beurteilung *unterschiedlicher* inhaltlicher Dimensionen relevant wird. Wenn die Beurteilung der *gleichen* Führungsdimension aus verschiedenen Perspektiven für eine optimale Vorhersage genügen würde, dann müssten sich häufiger multivariate Modelle ergeben.

Die Ergebnisse weisen zudem darauf hin, dass auf die Selbstbeurteilung wegen mangelnder Übereinstimmung zu Fremdbeurteilern nicht prinzipiell zu verzichten ist. Sie hat für verschiedene Kriterien einen bedeutsamen prädiktiven Wert. Allerdings ist vor einer einfachen Anwendung zu warnen, da sie häufig negative Prädiktionsgewichte hat. Eine umgekehrt proportionale Interpretation zum Erfolg erlangt in der Praxis voraussichtlich keine Akzeptanz. Dass für EF-WOBE und ZB auch andere Perspektiven negative Prädiktoren erhalten, liegt daran, dass die Kriterien als Differenzen operationalisiert sind.

Insgesamt deuten die Ergebnisse daraufhin, dass Beurteilungen aus verschiedenen Perspektiven mehr Erfolgsvarianz erklären als einzelne Perspektiven (Hypothese 29). Außerdem zeigt entsprechend sich Hypothese 29), dass unterschiedliche Führungsdimensionen spezifische Informationen enthalten und inkrementelle Beiträge liefern. Dies wird durch das Ergebnis unterstützt, dass für verschiedene Kriterien *verschiedene* Führungsdimensionen aus verschiedenen Richtungen aufgenommen werden.

Um zu überprüfen, ob die **transformationale Führung zusätzlich zur transaktionalen Führung** einen Beitrag zur Vorhersage der Erfolgskriterien leisten kann, wurden hierarchische Regressionsanalysen durchgeführt. Die theoretische Basis für die Sequenzierung der Variablen – die Augmentationshypothese – wurde gemeinsam mit bereits vorliegenden empirischen Befunden in Kapitel 2.5.3 beschrieben. Werden dementsprechend im ersten Schritt der Regressionsanalyse die drei transaktionalen Skalen (CR, MBA, MBP) und im zweiten Schritt zusätzlich die fünf transformationalen Skalen (IIa, IIb, IM, IS, IC)[186] aufgenommen, dann finden sich zwar scheinbar erwartete Augmentationsef-

[186] Ausstrahlung (AUS) als sechste, in der deutschen Version hinzugenommene Subskala der transformationalen Skala wurde nicht aufgenommen, da jeder zusätzliche Prädiktor ohne gleichzeitige Erhöhung des Stichprobenumfangs dazu führt, dass die Ergebnisse der Regressionsanalyse weniger robust sind. Außerdem können die Ergebnisse ohne AUS besser mit international bereits publizierten Befunden verglichen werden.

fekte, doch weisen die Toleranzwerte auf eine Multikollinearität für die transformationalen Skalen hin. Da die Subskalen hoch interkorrelieren (vgl. Kapitel 6.1.1.2), ist dies nicht verwunderlich. Gleichzeitig rechtfertigt die Interkorrelation ihre Zusammenfassung zu einem Mittelwert.

In Tabelle 68 sind die Ergebnisse der blockweisen Regressionsanalysen dargestellt, in die zunächst die drei transaktionalen Skalen und anschließend der Mittelwert der fünf transformationalen Skalen eingingen.[187] Dadurch erreichen alle Variablen die festgesetzte Toleranzschwelle (vgl. Kapitel 5.4). Für alle Erfolgskriterien leistet wie erwartet die transaktionale Führung einen statistisch bedeutsamen Erklärungswert. Für AZ1, ZB, EEF, EFF und SAT liefert die transformationale Führung darüber hinaus einen signifikanten zusätzlichen Beitrag. Für L ergibt sich ein zusätzlicher Beitrag, der zwar die 5%-Signifikanzschwelle nicht erreicht, der aber dennoch als bedeutsam interpretiert werden kann ($\Delta f^2 = .05$). Das gleiche gilt für EF-WOBE ($\Delta f^2 = .04$).

Tabelle 68. Hierarchische Regressionsanalysen zur Vorhersage der Erfolgskriterien durch transaktionale und transformationale Führung ($N = 90$)

Kriterium	Schritt	R^2	Korr. R^2	ΔR^2	f^2
AZ1	1	.49	.47	.49**	.96
	2	.51	.49	.03*	1.04
AZ2	1	.13	.10	.13**	.15
	2	.13	.09	.00	.15
EF-AZ	1	.25	.22	.25**	.33
	2	.25	.21	.00	.33
WOBE	1	.11	.08	.11*	.12
	2	.11	.07	.01	.12
EF-WOBE	1	.11	.08	.11*	.12
	2	.14	.10	.02	.16
L	1	.20	.17	.20**	.25
	2	.23	.19	.03$^+$	.30
ZB	1	.23	.20	.23**	.30
	2	.28	.24	.05*	.39
EEF	1	.73	.72	.73**	2.70
	2	.81	.80	.07**	4.26
EFF	1	.78	.77	.78**	3.55
	2	.85	.84	.07**	5.67
SAT	1	.73	.72	.73**	2.70
	2	.83	.82	.10**	4.56

Anmerkungen. Schritt 1: CR, MBA, MBP (transaktionale Führung), Schritt 2: CR, MBA, MBP, transformationale Führung (*Mittel* der 5 Skalen), Schritt 1 und 2 stellen die beiden Blöcke der hierarchischen Regressionsanalyse dar, **/*/$^+$ *Änderung* (ΔR^2) in Signifikanz von F: $p < .01/.05/.10$, Determinationskoeffizient für die einzelnen Blöcke angegeben.

[187] Für die transaktionalen Skalen ergibt sich kein Bedarf zur Aggregation. Diese wäre aufgrund der geringen und teilweise negativen Interkorrelationen auch kontraindiziert.
Bei genauerer Betrachtung der Regressionsgleichungen stellt man zudem fest, dass für die unterschiedlichen Erfolgskriterien unterschiedliche transaktionale Skalen die stärksten Beta-Gewichte erbringen.

Es ist plausibel, dass die transformationale Führung auf die Arbeitszufriedenheit mit Arbeitsbedingungen, Gesundheitsschutz, Entlohnung und Arbeitsplatz an sich (AZ2) keinen zusätzlichen Einfluss hat. Eine Zufriedenheit mit solch strukturellen und klar umrissenen Maßnahmen scheint auch allein durch eine klare, transaktionale Führung erreichbar zu sein. Ähnliches scheint für die vorliegende Operationalisierung von Wohlbefinden zu gelten. Selbst die alleinige Vorhersage durch die transaktionalen Subskalen zeigt nur einen geringen Effekt.

Bei der Auswertung fällt auf, dass die aufgeklärte Varianz für MLQ-interne Erfolgskriterien (EEF, EFF, SAT) deutlich über der für die anderen Erfolgskriterien liegt. Dies ist ein Ergebnis, das in allen bisherigen Studien so gefunden wurde (z.B. Tartler, Goihl et al., 2003). Da nicht nur der zusätzliche Beitrag, sondern schon die Varianzaufklärung durch die transaktionalen Skalen höher ausfällt als für alle anderen Erfolgskriterien, ist nicht davon auszugehen, dass die transformationale Führung für diese spezifischen Erfolgskriterien besonderen prädiktiven Wert hat. Vielmehr könnte ein Fragebogenartefakt vorliegen, da die MLQ-internen Erfolgskriterien und die MLQ-Führungsbeurteilungen in einen gemeinsamen Fragenblock integriert waren, während MLQ-externe Erfolgskriterien in einem getrennten Fragenblock erhoben wurden. Möglicherweise werden auch implizite Führungstheorien der Aufwärtsbeurteiler erfasst, die dann kontrolliert werden könnten, wenn die Führungsbeurteilungen und die Erfolgseinschätzungen von unterschiedlichen Personen stammen würden.

Um diese Problematik weiter aufzuklären, wurden die Daten mit dem Criss-Cross-Design analysiert (vgl. S. 352). Allerdings zeigt sich dann, dass die transaktionale Führung weder für MLQ-externe noch für MLQ-interne Erfolgskriterien einen statistisch bedeutsamen Beitrag liefern kann und auch die transformationale Führung nur für AZ1 einen signifikanten zusätzlichen Beitrag leistet. Möglicherweise misst der MLQ im Vergleich zum FzFü, für den durchaus auch mit der Criss-Cross-Methodologie Zusammenhänge aufgedeckt werden konnten, eher Führung, die sich hauptsächlich auf den geführten Mitarbeiter und weniger auf das gesamte Mitarbeiter-Team auswirkt. Aufgrund der unterschiedlichen Führungskonzeptionen, die der Konstruktion von MLQ und FzFü zugrunde liegen, ist dies durchaus plausibel. Ersterer basiert auf einem Führungsmodell, das die Motivation und die Selbstwirksamkeit einzelner Mitarbeiter als kritische Variablen definiert, Letzterer basiert auf allgemeingültigen, konkreten Führungsaufgaben. Insofern ist der stärkere Augmentationseffekt für MLQ-interne Erfolgsmaße nicht abzustreiten.

Bei einem gründlichen Test der Augmentationshypothese muss ausgeschlossen werden, dass nicht auch in umgekehrter Reihenfolge ein Effekt auftritt. Entsprechend wurden hierarchische Regressionsanalysen durchgeführt, in die zuerst die transformationale Führung aufgenommen wurde und erst in einem zweiten Schritt CR, MBA und MBP hinzugenommen wurden. Lediglich für eines der zehn Erfolgskriterien (EFF) ergibt sich, dass die transaktionalen Skalen über die transformationale Führung hinaus einen signifikanten Beitrag leisten ($p < .01$). Dass für neun Erfolgskriterien kein Effekt auftritt, unterstützt die erwarte-

te Richtung der Augmentation. Insgesamt unterstützen die Ergebnisse also die Augmentations-Annahmen des Führungsmodells von Bass (1985) und damit Hypothese 30).

Geht man davon aus, dass die transformationale Führung nicht nur im Vergleich zur transaktionalen Führung, sondern auch im Vergleich zu den FzFü-Faktoren im „Full Range of Leadership“ weiter „rechts oben“ anzusiedeln ist (vgl. Kapitel 2.6), dann sollte sie auch über die FzFü-Faktoren hinaus einen inkrementellen Beitrag zur Vorhersage der Erfolgskriterien leisten. Aufgrund der in Kapitel 5.2.1 berichteten hohen Korrelationen zwischen den FzFü- und den MLQ-Faktoren ist dies allerdings nicht durchgängig zu erwarten.

Die Hypothese wurde mit hierarchischen Regressionsanalysen überprüft. Den ersten Block bildete das beste FzFü-Modell (vgl. Tabelle 66 und Tabelle 67), den zweiten Block die transformationale Führung (Mittelwert der fünf Subskalen). Nur für EEF ergibt sich ein signifikanter Zuwachs an erklärter Varianz durch die Hinzunahme der transformationalen Führung (korr. R^2 = .81, N = 21, f^2 = 6.19, ΔR^2 = .05, $p(\Delta F)$ = .048). Für Leistung ergibt sich ein Zuwachs, der das Signifikanzniveau knapp verfehlt (korr. R^2 = .47, N = 31, f^2 = 1.09, ΔR^2 = .06, $p(\Delta F)$ = .09). Der Unterschied der Effektgröße ist bedeutsam, aber gering. Testet man gezielter und überprüft man nur, ob die charismatische Führung (IIa und IIb) über die FzFü-Modelle hinaus prognostisch relevant ist, ergibt sich für SAT ein signifikantes (korr. R^2 = .86, N = 21, f^2 = 8.26, ΔR^2 = .11, $p(\Delta F)$ = .005) und für EFF ein nahezu signifikantes Modell (korr. R^2 = .68, N = 21, f^2 = 2.94, ΔR^2 = .10, $p(\Delta F)$ = .067) mit jeweils geringen Zusatzeffekten.

Ein umgekehrtes Bild ergibt sich für die umgekehrte Reihenfolge (erster Block: Transformationale Führung [Mittelwert], zweiter Block: FzFü-Modell): Für jedes Erfolgskriterium zeigt sich mindestens auf dem 10%-Niveau ein signifikanter Zuwachs durch den zweiten Block (vgl. Anhang O). Das kommt mit Ausnahme von AZ2 mittleren bis starken Effekten gleich. Ebenso hohe Effekte ergeben sich sogar dann, wenn der erste Block nicht nur aus der transformationalen Führung, sondern zusätzlich aus den transaktionalen Skalen (CR, MBA, MBP) besteht (Tabelle 69).

Während die transformationale Führung also keine prognostische Leistung über die FzFü-Faktoren hinaus erbringt, klären diese über die transformationale sowie sogar über die transformationale *und* transaktionale Führung hinaus Erfolgsvarianz auf – und zwar teilweise mit sehr großen Effekten. Die Ergebnisse weisen also auf einen zur Annahme von Hypothese 31) umgekehrten Zusammenhang hin. Ferner unterstützt der Befund für die Mehrzahl der Kriterien obige Schlussfolgerung, dass Führungsbeurteilungen auf verschiedenen Dimensionen und aus verschiedenen Perspektiven zur Steigerung der Erfolgsaufklärung beitragen (vgl. Hypothese 29) und 29).

Tabelle 69. Hierarchische Regressionsanalysen zur Vorhersage der Erfolgskriterien durch transformationale, transaktionale und FzFü-Führung

Kriterium	Prädiktoren		*N*	*B*°	*SE B*°	*β*°	*p*°	*Korr. R²*		*f²*
AZ1	TF		31	.05	.43	.04	.911	.26*	(.016)	.57
	CR			.16	.44	.09	.723			
	MBA			.37	.21	.28	.083			
	MBP			.28	.30	.20	.358			
	E/F	(aufwärts)		.55	.20	.63	.010	.42*	(.010)	1.06
AZ2	TF		31	-.38	.38	-.47	.335	.13	(.962)	.02
	CR			-.11	.39	-.10	.770			
	MBA			.15	.18	.16	.431			
	MBP			-.20	.27	-.21	.462			
	E/F	(aufwärts)		-.32	.17	.56	.076	.03+	(.076)	.16
EF-AZ	TF		31	.06	.45	.06	.896	.09	(.173)	.28
	CR			-.09	.44	-.06	.843			
	MBA			-.08	.21	-.07	.701			
	MBP			-.05	.31	-.04	.878			
	K	(aufwärts)		-.48	.24	-.59	.054	.19+	(.054)	.50
WOBE	TF		31	-.34	.27	-.49	.220	.07	(.216)	.24
	CR			.50	.30	.49	.104			
	MBA			.21	.14	.26	.154			
	MBP			-.09	.24	-.11	.703			
	K	(selbst)		.41	.20	.41	.045	.18*	(.045)	.46
EF-WOBE	TF		29	-1.13	.49	-.97	.032	.25*	(.027)	.55
	CR			1.09	.52	.64	.051			
	MBA			.09	.22	.07	.678			
	MBP			-.26	.36	-.20	.468			
	E/F	(aufwärts)		-.04	.23	-.05	.858			
	P	(selbst)		.69	.35	.47	.062			
	FvF	(selbst)		.06	.29	-.05	.848	.35	(.105)	1.06
L	TF		31	.35	.40	.37	.388	.28*	(.013)	.60
	CR			.00	.35	.00	.997			
	MBA			.04	.18	.04	.815			
	MBP			-.02	.26	-.02	.930			
	E/F	(selbst)		-.43	.16	-.40	.013			
	IB	(aufwärts)		.19	.26	.21	.483	.40*	(.039)	1.10
ZB	TF		31	.01	.32	.02	.963	.22*	(.033)	.48
	CR			-.21	.42	-.15	.625			
	MBA			.18	.18	.16	.311			
	MBP			.10	.26	.09	.705			
	B/S	(aufwärts)		-.39	.17	-.53	.026	.34*	(.026)	.81
EEF	TF		21	.06	.32	.06	.855	.56**	(.002)	1.82
	CR			.22	.26	.18	.417			
	MBA			-.07	.13	-.08	.603			
	MBP			-.30	.19	-.26	.140			
	E/F	(aufwärts)		.19	.12	.32	.127			
	E/F	(abwärts)		.13	.11	.15	.274			
	B/S	(selbst)		.43	.13	.51	.007			
	P	(selbst)		-.54	.25	-.43	.051			
	S	(selbst)		.22	.12	.29	.098	.80*	(.012)	8.26

EFF	TF		21	-.15	.52	-.14	.777	.52** (.003)	1.59
	CR			.02	.35	.01	.965		
	MBA			-.02	.20	-.01	.913		
	MBP			-.46	.27	-.37	.106		
	IB	(abwärts)		-.03	.21	-.03	.881		
	K	(aufwärts)		.54	.21	.78	.025	.62^{+} (.075)	2.76
SAT	TF		21	.03	.41	.02	.948	.77** (<.001)	4.38
	CR			.30	.26	.20	.284		
	MBA			-.22	.15	-.20	.173		
	MBP			-.40	.20	-.28	.070		
	IB	(abwärts)		.02	.16	.02	.905		
	K	(aufwärts)		.49	.17	.63	.015		
	E/F	(aufwärts)		-.03	.12	-.05	.784	.83^{+} (.077)	7.93

Anmerkungen. °Kennwerte des Gesamtmodells (alle Blöcke), erster Block der Regressionsanalyse: TF, CR, MBA, MBP, zweiter Block: bestes FzFü-Modell, **/*/$^{+}$*Änderung* (ΔR^2) in Signifikanz von F: p < .01/.05/.10, Determinationskoeffizient für die einzelnen Blöcke angegeben.

Dieses Ergebnis lässt bezweifeln, dass durch das Führungsmodell von Bass tatsächlich der Full Range of Leadership abgedeckt ist. Zumindest bleiben im MLQ Führungsfacetten unberücksichtigt, die zusätzlich zur Erklärung von Führungserfolg beitragen können. Vergleicht man die Items des MLQ und des FzFü, dann liegt nahe, dass konkrete Führungstätigkeiten im FzFü deutlicher formuliert sind als im MLQ. Für die psychologische Führungsforschung bedeutet dies, dass in der Theorienbildung möglichst verhaltensnah zu operationalisieren ist, welche Konstrukte als erfolgsrelevant betrachtet werden. Bereits auf Seite 370 wurden unterschiedliche Ausrichtungen der beiden Führungsskalen erwähnt. Ein Ziel der zukünftigen Führungsforschung sollte daher sein, Führungsaufgaben und Führungstheorien stärker miteinander zu verbinden.

Berechtigterweise muss daher die Frage aufgeworfen werden, ob es ratsam ist, die transformationale Führung verstärkt in Unternehmen einzuführen. Genau dies entspricht zwar dem aktuellen Trend. Doch meist wird nicht vorab geklärt, welcher Zusatznutzen davon zu erwarten ist.

Zusammenfassend wird festgehalten, dass

- die FzFü-Dimensionen für unterschiedliche Kriterien unterschiedlich gute Vorhersagekraft haben,
- für manche Kriterien mehrere Dimensionen gemeinsam bessere Prognoseleistungen erbringen als einzelne Kriterien,
- aus der Aufwärtsperspektive mehr Erfolgsvarianz und zudem von mehr Kriterien erklärt werden kann als aus der Abwärts- und Selbstperspektive,
- Führungsbeurteilungen aus verschiedenen Perspektiven für viele Erfolgskriterien mehr Varianz erklären als Führungsbeurteilungen aus einzelnen Perspektiven,

- die transformationale Führung über die transaktionale Führung hinaus zusätzliche Beiträge zur Erklärung des Führungserfolgs liefern kann (nicht umgekehrt) und dass
- die FzFü-Dimensionen über die transformationale und sogar über die transformationale *und* transaktionale Führung hinaus zusätzliche Beiträge zur Erklärung des Führungserfolgs liefern (nicht umgekehrt).

6.4.3 Moderatorvariablen für den Zusammenhang zwischen Führungsverhalten und Erfolg

Das vorliegende Kapitel beschäftigt sich mit potentiellen Moderatoreffekten auf den Zusammenhang zwischen Führungsverhalten und Führungserfolg (Fragestellung 9). Es werden drei potentielle Moderatorvariablen überprüft:

33) Je wichtiger die Aufwärtsbeurteiler die Führungsdimensionen bewerten, umso stärker ist der Zusammenhang zwischen diesen Führungsdimensionen und Erfolg.

34) Je transformationaler die Aufwärtsbeurteiler ihre Führungskraft beschreiben, umso stärker ist der Zusammenhang zwischen Führung und Erfolg.

Um die moderierende Wirkung der Wichtigkeitseinschätzung (Aufwärtsbeurteiler) sowie der transformationalen Führung auf den Zusammenhang zwischen Führungsverhalten und Führungserfolg zu überprüfen, wurden hierarchische moderierte Regressionsanalysen durchgeführt. Die einfachen Prädiktoren und Moderatoren wurden für diesen Zweck zentriert, um Multikollinearitäten zu vermeiden (vgl. Aiken & West, 1993). Die Interaktionsterme wurden jeweils durch die Multiplikation des zentrierten Prädiktors und des zentrierten Moderators gewonnen. Als Operationalisierung des Führungsverhaltens wurden hier die Führungsbeurteilungen aus den drei Perspektiven (Aufwärts, Abwärts, Selbst) herangezogen.

In den folgenden drei Tabellen (Tabelle 70 bis Tabelle 72) sind die Ergebnisse derjenigen hierarchischen moderierten Regressionsanalysen berichtet, die statistisch bedeutsame Interaktionseffekte zwischen der Führungsbeurteilung und der **Wichtigkeitseinschätzung** der entsprechenden Führungsdimension erbringen (Aufwärtsperspektive). Für alle nicht dargestellten Führungsdimensionen und Erfolgskriterien zeigen sich keine bedeutsamen Interaktionen.

Tabelle 70. Hierarchische moderierte Regression für verschiedene Erfolgsvariablen (AV) auf E/F-Ist und E/F-Wichtig (Aufwärts), $N = 106$

	Modell 1				Modell 2					
	B	*SE B*	*β*	*p*	*B*	*SE B*	*β*	*p*	*korr.* R^2	*p* (*ΔF*)
AV: EF-WOBE										
1. Block									.11	< .01
Auf	-.32	.09	.34	< .01	-.38	.09	-.41	< .01		
Wi	-.08	.13	-.05	.56	-.04	.13	-.03	.77		
2. Block									.14	.02
Auf x Wi					.25	.11	.22	.02		
AV: EEF										
1. Block									.48	< .01
Auf	.49	.05	.67	< .01	.52	.05	.72	< .01		
Wi	.13	.08	.12	.10	.11	.08	.10	.16		
2. Block									.50	.02
Auf x Wi					-.15	.06	-.17	.02		
AV: EFF										
1. Block									.53	< .01
Auf	.50	.05	.74	< .01	.53	.05	.79	< .01		
Wi	-.09	.07	-.09	.20	-.11	.07	-.11	.12		
2. Block									.55	.03
Auf x Wi					-.13	.06	-.15	.03		
AV: SAT										
1. Block									.53	< .01
Auf	.58	.05	.74	< .01	.62	.05	.80	< .01		
Wi	-.07	.08	-.06	.42	-.09	.08	-.08	.25		
2. Block									.56	< .01
Auf x Wi					-.18	.07	-.19	< .01		

Anmerkungen. *p* (Δ*F*): Änderung in Signifikanz von *F,* AV: Abhängige Variable, Auf: Aufwärtsbeurteilung des Führungsverhaltens (Entwickeln/Fördern), Wi: Wichtigkeitsbeurteilung (Aufwärts, Entwickeln/Fördern).

Im Sinne von kontingenztheoretischen Annahmen über die Wirkung von Personalführung wird erwartet, dass die Erfolgsvariablen umso stärker von Führungsverhalten abhängen, je wichtiger dieses für erfolgreiches Arbeiten betrachtet wird. Generell ist festzuhalten, dass die Aufwärts- sowie die Selbstbeurteilungen nur selten mit der Wichtigkeitsschätzung interagieren (sieben bzw. zwei der siebzig potentiellen Effekte erreichen Signifikanz). Aus der Abwärtsperspektive ergibt sich kein einziger Interaktionseffekt. Es scheint also so zu sein, dass die subjektive Relevanz von Führung in den meisten Fällen einen vernachlässigbaren Einfluss darauf hat, ob eine entsprechende Ausrichtung der Führung auch zu Erfolg führt oder nicht. Dieses Ergebnis, das den kontingenztheoretischen Modellen widerspricht, wird sogar noch dadurch untermauert, dass die Mehrzahl der

extrahierten Interaktionseffekte in die unerwartete Richtung weist. So gilt beispielsweise für die signifikanten Interaktionen zwischen der Aufwärtsbeurteilung von E/F und der Wichtigkeitseinschätzung von E/F, dass der Zusammenhang zwischen Führung und Erfolg umso geringer ist, je wichtiger E/F gehalten wird (Tabelle 70). Das bedeutet, dass die Wichtigkeitszuschreibung zunehmend weniger Einfluss auf den Führungs-Erfolgs-Zusammenhang nimmt, je positiver die Führung bewertet wird.

Für EF-WOBE, also die wahrgenommene Differenz zwischen dem, was eine Führungskraft für das Wohlbefinden ihrer Mitarbeiter tut und dem, was sie tun könnte, zeigen sich inkonsistente Ergebnisse. Während sich diese Führungslücke/Unzufriedenheit mit zunehmend besserer B/S-Führung bei hoher Wichtigkeit von B/S stärker verringert als bei geringer Wichtigkeit von B/S (Tabelle 72), verringert sie sich mit zunehmend besserer E/F- und IB-Führung bei hoher Wichtigkeit von E/F bzw. IB weniger als bei geringer Wichtigkeit von E/F bzw. IB (Tabelle 70 bzw. Tabelle 71) .

Tabelle 71. Hierarchische moderierte Regression für EF-WOBE auf IB-Ist und IB-Wichtig (Aufwärts), $N = 106$

	Modell 1				Modell 2					
	B	*SE B*	*β*	*p*	*B*	*SE B*	*β*	*p*	*korr.* R^2	*p* (*ΔF*)
1. Block									.16	< .01
Auf	-.46	.11	-.39	< .01	-.49	.11	-.42	< .01		
Wi	-.08	.19	-.04	.66	-.12	.19	-.06	.52		
2. Block									.21	.01
Auf x Wi					.48	.19	.23	.01		

Anmerkungen. *p*(Δ*F*): Änderung in Signifikanz von *F,* AV: Abhängige Variable, Auf: Aufwärtsbeurteilung des Führungsverhaltens (Individuelle Berücksichtigung), Wi: Wichtigkeitsbeurteilung (Aufwärts, Individuelle Berücksichtigung).

Anders verhält es sich für die Beurteilung von B/S hinsichtlich der Auswirkungen auf die subjektiv erbrachte Leistung (Tabelle 72).

Tabelle 72. Hierarchische moderierte Regression für EF-WOBE und L auf B/S-Ist und B/S-Wichtig (Aufwärts), $N = 106$

	Modell 1				Modell 2					
	B	*SE B*	*β*	*p*	*B*	*SE B*	*β*	*p*	*korr.* R^2	*p* (*ΔF*)
AV: EF-WOBE										
1. Block									.11	< .01
Auf	-.29	.10	-.31	< .01	-.31	.10	-.35	< .01		
Wi	-.09	.14	-.07	.49	.04	.15	.03	.78		
2. Block									.14	.03
Auf x Wi					.20	.09	.22	.03		
AV: L										
1. Block									.16	< .01
Auf	.38	.08	.47	< .01	.36	.08	.44	< .01		
Wi	-.23	.12	-.19	.06	-.12	.13	-.10	.35		
2. Block									.19	.04
Auf x Wi					.17	.08	.21	.04		

Anmerkungen. *p*(Δ*F*): Änderung in Signifikanz von *F,* AV: Abhängige Variable, Auf: Aufwärtsbeurteilung des Führungsverhaltens (B/S), Wi: Wichtigkeitsbeurteilung (Aufwärts, B/S).

Im ersten Block der Regressionsanalyse zeigt sich ein erwarteter Haupteffekt insofern, dass sich die Leistung steigert, je mehr B/S gezeigt wird. Der zweite Block bringt einen positiven Interaktionseffekt zwischen der Aufwärtsbeurteilung und der Wichtigkeitseinstufung zutage. Die Interpretation von Interaktionseffekten kann durch eine Darstellung der Regressionsgeraden auf verschiedenen Stufen der Moderatorvariable veranschaulicht werden (vgl. Aiken & West, 1993; O'Connor, 1998).

Entsprechend zeigt Abbildung 32 die Regressionsgerade für drei Ausprägungsstufen der Wichtigkeitsschätzung, erstens für die mittlere Wichtigkeitseinschätzung, zweitens für die um eine Standardabweichung nach oben bzw. drittens für die um eine Standardabweichung nach unten abweichende Wichtigkeitseinschätzung. Beurteilen die Mitarbeiter die B/S-Führung ihrer Führungskraft positiv (vs. negativ), dann profitiert ihre Leistung von einer hohen Wichtigkeitseinschätzung (vs. Leistung büßt ein). Das Ergebnis ist konsistent mit kontingenztheoretischen Modellen von Führung. Vor allem bei einer starken B/S-Führung ist es leistungsförderlich, wenn Mitarbeiter B/S für wichtig erachten.

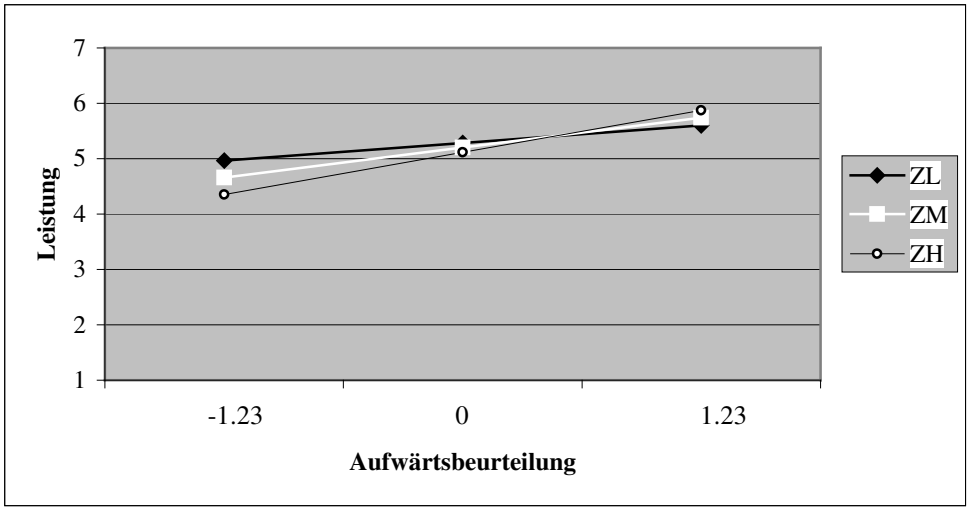


Abbildung 32. Interaktion von Führung (Aufwärtsbeurteilung) und Wichtigkeit von Führung auf der Dimension B/S als Prädiktor von Leistung

Anmerkungen. ZL: Wichtigkeitsbeurteilung (aufwärts) = Mittelwert (*M*) minus 1 Standardabweichung (*SD*), ZM: Wichtigkeitsbeurteilung (aufwärts) = *M* (der Wichtigkeitsbeurteilungen), ZH: Wichtigkeitsbeurteilung (aufwärts) = *M* plus 1 *SD*, Aufwärtsbeurteilung: zentrierte Werte (*SD* = 1.23).

Die vereinzelten Interaktionseffekte, die die *Selbstbeurteilung* mit der Wichtigkeitseinschätzung (Aufwärts) erbringt, sind in Anhang P dargestellt. Um tatsächlich erkenntnistheoretische oder gar praktische Relevanz zu erlangen, treten sie zu unsystematisch auf. Es ist von methodischen Artefakten auszugehen, so dass auf eine Interpretation verzichtet wird. Da auch die Wichtigkeitseinstufung, die die Führungskräfte (vs. Geführten) abgeben, nicht häufiger mit dem Führungsverhalten (verschiedene Perspektiven) interagiert, ist davon auszugehen, dass die Wichtigkeit entgegen der Annahme keinen konsistenten moderierenden Einfluss hat (Hypothese 33).

Als weitere potentielle Moderatorvariable wurde – ebenfalls mittels hierarchischer moderierter Regressionen – die **transformationale Führung** untersucht. Da sich lediglich ein einziger statistisch signifikanter Interaktionseffekt zeigt (auf den Zusammenhang zwischen FvF und SAT, Anhang Q), ist festzuhalten, dass die Ausführung von bestimmten Führungsaufgaben/-grundsätzen[188] vor dem Hintergrund einer transformationalen Führung nicht wie erwartet in größerem Erfolg resultieren als vor dem Hintergrund weniger stark ausgeprägter transformationaler Führung (Hypothese 34).

[188] Zur Erinnerung: Die Führungsdimensionen des FzFü stellen Operationalisierungen der Führungsgrundsätze des Unternehmens dar.

6.5 Validität der interperspektivischen Übereinstimmung

„Is multi-perspective measurement a means or an end?“
(Titel von Tornow, 1993)

Aufgrund der mangelhaften Übereinstimmung zwischen Selbst- und Fremd-, insbesondere zwischen Selbst- und Aufwärtsbeurteilungen, müssen Urteilsverzerrungen angenommen werden. Welcher Art diese sind, ob sie im Antwortprozess entstehen oder ob sie der alltäglichen Wahrnehmung entsprechen, muss erst noch erforscht werden. Es bleibt daher offen, ob die Selbst-Fremd-Übereinstimmung überhaupt als Erfolgsprädiktor gelten kann. Genau so wird sie (zumindest implizit) bei 360°-Feedbacks in der Praxis allerdings interpretiert. In Kapitel 3.5.1 wurden bereits korrelative Zusammenhänge zwischen der Selbst-Fremd-Übereinstimmung und Erfolgskriterien berichtet. Für die Automobilbranche, also eine Branche, die in Deutschland viel Geld und Zeit in 360°-Beurteilungen investiert, liegen keine Ergebnisse vor. Insgesamt existieren für den deutschsprachigen Raum kaum aktuelle Ergebnisse (Ausnahme: Arbeitsgruppe von Felfe). Es sind keine Studien aus dem deutschsprachigen Raum publiziert, deren Daten nicht mit dem MLQ ermittelt wurden. Auch international ist das Forschungsfeld stark von Arbeitsgruppen geprägt, die überwiegend mit dem MLQ arbeiten. Das Antwortverhalten auf den MLQ scheint sich zwischen Deutschland und den USA jedoch wesentlich zu unterscheiden, weshalb eine methodische Erweiterung in dieser Hinsicht notwendig ist. Praxisrelevanz erlangt die Selbst-Fremd-Übereinstimmung aber nur dann, wenn sie mit praxisrelevanten Erfolgskriterien kovariiert und diese prognostiziert. Das betrifft die Frage der Kriteriumsvalidität (vgl. Kapitel 6.5.1, 6.5.2 und 6.5.4).

Aus Gründen der „Sparsamkeit“, der sowohl Forschung als auch Praxis verpflichtet sind, muss darüber hinaus ein strengeres Kriterium erfüllt werden: Der Einsatz von Ähnlichkeitsmaßen als Erfolgsindikatoren ist aus methodischer Sicht nur sinnvoll, wenn diese entweder einen besseren Beitrag zur Erklärung der Erfolgsvarianz leisten als die Betrachtung der einzelnen Beurteilungsrichtungen alleine oder wenn sie einen zusätzlichen Beitrag über einfache Beurteilungsrichtungen hinaus leisten (Kapitel 6.5.2 und 6.5.3).

Bereits oben wurde die Kontroverse über die Operationalisierung von „Ähnlichkeit“ diskutiert, bis dato ist sie in der Forschungsliteratur nicht abgeschlossen (vgl. Kapitel 6.3.4). Verschiedenes spricht dafür, zunächst mehrere Maße zu verwenden, unter anderem um ihren jeweiligen Wert erkennen zu können. Die Praxis verwendet in aller Regel einfache algebraische Differenzen. So wird auch in der vorliegenden Untersuchung nicht gänzlich auf diese verzichtet, um deren Validität mit der anderer Differenzmaße vergleichen zu können (vgl. S. 401ff.).

Zudem wird nicht nur mit verschiedenen Maßen, sondern auch mit verschiedenen Methoden gearbeitet, um die Validität zu ermitteln. So wird im Anschluss an korrelations- und varianzanalytische Auswertungen (Kapitel 6.5.1) im Sinne von Brutus, Fleenor und Tisak

(1999) durch weiterführende Regressionsanalysen getestet, ob die Selbst-Fremd-*Diskrepanz* die kritische Variable im Entstehungszusammenhang zwischen Führung und Erfolg darstellt (Kapitel 6.5.2). Polynomiale Regressionsmodelle können beispielsweise verschiedene Nachteile überwinden, die die alleinige Verwendung von Differenz-Scores mit sich bringt (Edwards, 1994a, 1994b). Die Betrachtung *korrelativer* Selbst-Fremd-Übereinstimmung im Zusammenhang mit den erhobenen Erfolgskriterien (Kapitel 6.5.4) kann neben der Aufklärung der Validität von Selbst-Fremd-Übereinstimmung auch zu Fragestellung 5 beitragen, die den Vergleich verschiedener Ähnlichkeitsmaße ins Zentrum rückte (vgl. Kapitel 6.3.4).

6.5.1 Hängt die interperspektivische Übereinstimmung der Führungsbeurteilungen mit Erfolg zusammen?

Zunächst ist zu klären, ob überhaupt ein Zusammenhang zwischen der interperspektivischen Beurteilerübereinstimmung und den verschiedenen Erfolgskennwerten besteht. Folgende Hypothesen wurden formuliert:

35) Mit steigenden Niveau-Unterschieden zwischen der Selbst- und Fremdbeurteilung sinkt der Erfolg.

36) Hypothese 34) gilt stärker für die Diskrepanz zwischen der Selbst- und Aufwärts-Beurteilung als für die Diskrepanz zwischen der Selbst- und Abwärts-Beurteilung.

37) Mit steigenden Niveau-Unterschieden zwischen der Abwärts- und Aufwärtsbeurteilung sinkt der Erfolg.

38) Die Selbst-Fremd-Übereinstimmung kann mehr Erfolgsvarianz aufklären als die Abwärts-Aufwärts-Übereinstimmung.

39) Die Mitarbeiter von Führungskräften, die sich im Vergleich zu Fremdbeurteilern überschätzen, sind am wenigsten erfolgreich, während der Erfolg bei Übereinstimmung am höchsten ausfällt. Die Mitarbeiter der Unterschätzer nehmen eine Zwischenstellung ein.

Nach den Hypothesen 34) bis 37) werden für steigende Diskrepanzen abnehmende Erfolge erwartet.[189] In Tabelle 73 sowie in Anhang R sind die entsprechenden Korrelationen berichtet.

[189] Zur Erinnerung: Das bedeutet, dass die Diskrepanzen mit EF-AZ, EF-WOBE sowie ZB positiv korrelieren, mit allen anderen Erfolgskriterien negativ. Die gegenteilige Richtung liegt daran, dass die erstgenannten

Tabelle 73. Korrelativer Zusammenhang zwischen Diskrepanzen in der Führungsbeurteilung und den Erfolgskriterien

	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB	EEF	EFF	SAT
aggregiert[1]										
S-Auf (D) (40/30/27)[2]	-.64/.23 (-.49)[3]	-.02/.26 (.12)	.30/.53 (.43)	.02/.44 (.09)	.28/.42 (.38)	-.57/.09 (-.49)	-.15/.41 (.33)	-.56/.07 (-.52)	-.69/.09 (-.62) [a]	-.75/-.06 (-.63)
S-Auf (40/30/27)	-.66/-.15 (-.46)	-.02/.25 (.03)	.24/.56 (.48)	.06/.52 (.15)	.31/.47 (.45)	-.58/-.11 (-.46)	.12/.54 (.45)	-.70/-.12 (-.45)	-.71/-.09 (-.60) [a]	-.71/-.18 (-.62)
S-Ab (23)	-.43/-.01 (-.24)	.18/.44 (.23)	.12/.40 (.26)	-.17/.36 (.22)	-.14/.60 (.20)	-.57/-.11 (-.44)	.30/.55 (.42)	-.33/-.14 (-.25)	-.23/.00 (-.14)	-.46/-.04 (-.29)
Ab-Auf (32/23/19)	-.65/-.06 (-.58)	-.32/-.16 (-.23)	.33/.58 (.46)	-.35/-.03 (-.27) [a]	.33/.50 (.41)	-.55/-.14 (-.38)	.16/.55 (.40)	-.54/-.10 (-.47)	-.61/.04 (-.47)	-.65/-.24 (-.59)
personengebunden, individuell[4]										
S-Auf (140/76/35)	-.69/-.18 (-.64) [a]	-.25/.09 (-.17) [a]	.24/.40 (.33)	-.24/.24 (-.15)	.36/.46 (.39)	-.54/-.05 (-.46)	.09/.48 (.39)	-.65/-.29 (-.58) [a]	-.76/-.20 (-.68) [a]	-.76/-.30 (-.69) [a]
Ab-Auf (120/62/26)	-.68/.04 (-.58) [a]	-.32/-.05 (-.27) [a]	.16/.46 (.28)	-.32/.16 (-.22) [a]	-.06/.19 (.13)	-.50/-.14 (-.39)	-.08/.32 (.27)	-.64/-.28 (-.61) [a]	-.74/-.19 (-.66) [a]	-.69/-.25 (-.61) [a]
Criss-Cross (*N* = personengebundene Analyse)[5]										
S-Auf	-.31/-.14 (-.22) [b]	-.10/.11 (-.03)	.00/.20 (.18)	.03/.11 (.05) [b]	-.09/.30 (.26)	-.24/.12 (-.18) [b]	-.19/.18 (.11) [b]	-.23/.09 (-.17) [b]	-.28/-.04 (-.23) [b]	-.34/-.12 (-.29) [b]
Ab-Auf	-.43/-.18 (-.21) [c]	-.38/-.14 (-.17)	.03/.15 (.11)	-.27/-.03 (-.10)	.13/.33 (.20)	-.21/-.02 (-.06) [c]	.00/.15 (.08)	-.24/-.07 (-.15) [c]	-.26/-.11 (-.17) [c]	-.33/-.14 (-.20) [c]

Anmerkungen.
[1] „S-Auf (D)“: D-Statistik nach Cronbach. Für alle anderen Vergleiche: absolute Differenz zwischen den Perspektiven. Dafür wurden die Aufwärtsratings aller Mitarbeiter einer Führungskraft aggregiert (Median). Für die Selbst- und Abwärtsperspektive liegt naturgemäß jeweils nur eine Beurteilung vor. Die Selbst-Abwärts-Diskrepanz wird dennoch zu den aggregierten Analysen gezählt, da auch sie mit dem Median der entsprechenden Erfolgsratings (Aufwärtsbeurteiler) korreliert wurde.
[2] S: Selbstbeurteilung, Auf: Aufwärtsbeurteilung, Ab: Abwärtsbeurteilung.
N in Klammern, variiert über die sieben Führungsdimensionen hinweg, erste Zahl betrifft E/F, B/S, IB, K und P zweite Zahl betrifft FvF, dritte Zahl betrifft S.
[3] Es ist jeweils die niedrigste und höchste Korrelation zwischen einem der fünf Führungsfaktoren und dem entsprechenden Erfolgskriterium angegeben. In Klammern: mittlere Korrelation über folgende FzFü-Faktoren: E/F, B/S, IB, K, P (vgl. S. 385).
grau hinterlegt: signifikante Korrelationen einzelner oder aller Führungsdimensionen mit dem Erfolgskriterium.
[4] Differenz zwischen den Selbst- bzw. Abwärtsbeurteilungen und den individuellen (nicht aggregierten) Aufwärtsbeurteilungen. Die Differenzen wurden jeweils mit den individuellen Erfolgsratings der Aufwärtsbeurteiler korreliert.
[5] Differenz zwischen den Selbst- bzw. Abwärtsbeurteilungen und den individuellen (nicht aggregierten) Aufwärtsbeurteilungen. Die Differenzen wurden jeweils mit den individuellen Erfolgsratings der *Kollegen* korreliert (vgl. S. 352).
[a] Korrelationsunterschied zu den S-Ab-Korrelationen ($p < .05$).
[b] Korrelationsunterschied zwischen den personengebundenen und gekreuzten mittleren S-Auf-Korrelationen ($p < .05$).
[c] Korrelationsunterschied zwischen den personengebundenen und gekreuzten mittleren Ab-Auf-Korrelationen ($p < .05$).

Erfolgskriterien als Differenzmaße operationalisiert sind und geringe Differenz als höhere Zufriedenheit der Mitarbeiter mit dem Engagement der Führungskraft für deren Arbeitszufriedenheit und Wohlbefinden bzw. mit der Beurteilungshäufigkeit betrachtet werden kann.

Im Allgemeinen bestätigt sich die Erwartung und es zeigt sich, dass mit zunehmender interperspektivischer Diskrepanz in den Führungsbeurteilungen die Erfolge abnehmen. Dies gilt für alle untersuchten Diskrepanzen, also für die Selbst-Aufwärts-, die Selbst-Abwärts- und die Abwärts-Aufwärts-Diskrepanz.

Genauer: Der Varianzanteil, den die **Selbst-Aufwärts-Diskrepanz** mit den Erfolgskriterien teilt, erreicht für die aggregierten Werte bis zu 56%, für die individuellen Werte gar bis zu 58%. Dies übertrifft bisherige Befunde in starkem Ausmaß. Vergleichbare Werte liegen beispielsweise bei Church (1997b) nicht über 15.2%. Lediglich die Ergebnisse der Korrelationsanalysen von AZ2 und WOBE entsprechen in der aggregierten Analyse diesem Bild nicht in konsistenter Weise. Für diese Kriterien zeigt sich das erwartete Korrelationsmuster erst auf der individuellen Ebene (fünfte Datenzeile aus Tabelle 73). Möglicherweise machen sich diese schwachen Effekte erst dann bemerkbar, wenn die Selbst-Aufwärts-Abweichung auch negative Werte annimmt, also eine größere Varianz aufweist. Dies ist erst bei den individuellen Selbst-Aufwärts-Differenzen der Fall. Aggregierte Differenzmaße liegen äußerst selten im negativen Bereich, Distanzmaße (S-Auf [D]) berücksichtigen per definitionem nur das Ausmaß der Abweichung und liegen daher immer im positiven Bereich. Abgesehen davon herrscht sehr große Ähnlichkeit zwischen den Ergebnissen der aggregierten und der individuellen Analyse. Da sich auch die Ergebnisse der beiden verschiedenen aggregierten Analysen nicht unterscheiden (erste und zweite Datenzeile aus Tabelle 73), wird im Zuge der bivariaten Korrelationsanalysen darauf verzichtet, den Einfluss des puren Ausmaßes der Abweichung von dem der Richtung der Abweichung trennen zu wollen. Bei weiteren Analysen wird darauf differenzierter eingegangen.

Da es ausschließlich um Erfolgsmaße geht, die allein die Mitarbeiter der beurteilten Führungskräfte betreffen, ist davon auszugehen, dass die Übereinstimmung zwischen der Selbst- und Aufwärtsbeurteilung stärker mit den Erfolgsmaßen zusammenhängt als die Übereinstimmung zwischen der Selbst- und Abwärtsbeurteilung. Problematisch ist dabei, dass die Validität der Selbst-Aufwärts-Diskrepanz potentiell durch den Single-Source-Bias künstlich erhöht sein kann. Trotzdem wurde der Zusammenhang zwischen Selbst-Fremd-Diskrepanzen in der Führungsbeurteilung und Erfolgskriterien bislang nur selten überhaupt für Erfolgskriterien untersucht, die von Dritten abgegeben wurden. Wo dies groß angelegt versucht wurde, konnten keine Zusammenhänge festgestellt werden. So ergaben sich beispielsweise bei Church (1997b) keine Zusammenhänge zwischen der Selbst-Aufwärts-Kongruenz und Erfolgseinschätzungen (Qualität und Kreativität von Serviceleistungen) durch den Vorgesetzten, Peers oder Klienten. Lediglich zu den Erfolgseinschätzungen durch die Aufwärtsbeurteiler selbst ließen sich signifikante Korrelationen feststellen. Solche Befunde basieren allerdings nicht auf Analysen, die den Single-Source-Bias kontrollierten.

Deshalb werden in der vorliegenden Studie zwei Möglichkeiten der besseren Validierung von Selbst-Fremd-Diskrepanzen genutzt. Erstens wird der Single-Source-Bias mit Hilfe des Criss-Cross-Designs kontrolliert (vgl. S. 352). Zweitens können zur Validierung Selbst-Abwärts-Diskrepanzen Kriterien herangezogen werden, die von Dritten abgegeben wurden. Erst nachdem die entsprechenden Ergebnisse dargestellt wurden, wird die Validität der Selbst-Aufwärts-Diskrepanz mit der Validität der Selbst-Abwärts-Diskrepanz verglichen.

Um den Einfluss des Single-Source-Bias auf die Validität der Selbst-Aufwärts-Diskrepanz zu überprüfen, wurde auch an dieser Stelle das Criss-Cross-Design eingesetzt. Es zeigt sich, dass die Zusammenhänge zwischen der Selbst-Aufwärts-Differenz und den Erfolgskriterien insgesamt zwar geringer ausfallen, dass aber dennoch zu allen Kriterien mit Ausnahme von AZ2 und WOBE statistisch (hoch)signifikante, wenn auch geringe Korrelationen bestehen. Dies wäre bei einem Zusammenhang, der hauptsächlich auf dem Single-Source-Bias gründet, nicht zu erwarten. Die gefundenen gekreuzten Korrelationen bedeuten stattdessen, dass sich die Selbst-Aufwärts-Differenz, wie sie sich für einen Mitarbeiter ergibt, auch auf den Outcome seines Kollegen auswirkt. Durch das gekreuzte Design wird also keine individuelle Führungswirkung erfasst, sondern lediglich der Wirkungsanteil, der für das gesamte Team gilt. Da hingegen bei der personengebundenen Auswertung sowohl die Effekte auf individueller als auch auf Teamebene gemessen werden, muss die Stärke des Zusammenhangs bei Anwendung des Criss-Cross-Designs zwangsläufig sinken.[190] Hypothese 34) ist für die Selbst-Aufwärts-Diskrepanz also anzunehmen. Je weniger die Selbsteinschätzung der Führungskräfte also mit der Fremdeinschätzung der Geführten übereinstimmt, umso weniger erfolgreich arbeiten diese auch.

Zur Validierung der **Selbst-Abwärts-Diskrepanzen** werden ebenfalls die Erfolgskriterien herangezogen, die die Aufwärtsbeurteiler abgaben. Wenn man bedenkt, dass die Kriterien also von Dritten abgegeben wurden und für solche Designs in bisherigen Studien kaum ein Nachweis für die Validität von Beurteilungsdiskrepanzen entdeckt wurde (z.B. Church, 1997b), ist die Varianzaufklärung bis zu 33% im vorliegenden Datensatz vergleichsweise hoch (vgl. dritte Zeile aus Tabelle 73). Die korrelativen Ergebnisse geben also einen starken Hinweis darauf, dass mit steigender Selbst-Abwärts-Diskrepanz die Führungserfolge – eingeschätzt durch die geführten Mitarbeiter – sinken. Besonders stark mit der Selbst-Abwärts-Diskrepanz hängen subjektive Leistung und die Zufriedenheit mit der Leistungsbeurteilung zusammen. Da Leistungsparameter die Variablen darstellen, die von den Vorgesetzten am ehesten für Entscheidungen herangezogen werden, sind sie

[190] Es könnte aufgrund der geringen Korrelationshöhe eingewendet werden, dass diese Zusammenhänge nicht teamabhängig, sondern zufälliger Natur sind. Gegen diesen Einwand spricht jedoch eindeutig, dass die Selbst-Aufwärts-Differenzen überwiegend in unbedeutender Höhe ($|r| < .10$) mit den Erfolgsratings anderer Mitarbeiter (außerhalb ihres Teams) zusammenhängen. Um dies zu kontrollieren, wurde das Criss-Cross-Design über die gesamte Stichprobe verwendet.

offenbar besonders aufmerksam gegenüber leistungsförderlichem Führungsverhalten. Wenn die Abwärtsbeurteilung aber einen guten Indikator für die Leistung darstellt, dann muss eine Diskrepanz in der Führungsbeurteilung auch einen besonderen Einfluss auf die Leistung nehmen.

Dass sich für steigende Selbst-Abwärts-Diskrepanzen *bessere* Arbeitszufriedenheit (AZ2) und besseres Wohlbefinden (WOBE) der Mitarbeiter zeigt, ist nicht gänzlich auf die Diskrepanz, sondern auch auf den unerwarteten, negativen Zusammenhang zwischen der „einfachen" Abwärtsbeurteilung und AZ2 sowie WOBE zurückzuführen (vgl. Tabelle 64). Das gleiche gilt im Übrigen für die aggregierten Aufwärtsbeurteilungen.

Deshalb wird Hypothese 34) trotzdem auch für die Selbst-Abwärts-Diskrepanz beibehalten. Dies ist für sich genommen ein Erkenntnisgewinn, der auch der Praxis dient, indem es die gewöhnliche Vorgehensweise unterstützt. Dort sind Beurteilungen nämlich nach wie vor häufig „Sache der Vorgesetzten", so dass Abwärtsbeurteilungen mit weniger Aufwand und gegen weniger Widerstand zu erhalten sind als Aufwärtsbeurteilungen.

Weiter interessiert nun, ob der negative Zusammenhang zwischen den Selbst-*Abwärts*-Diskrepanzen und Erfolg geringer ausfällt als der zwischen den Selbst-*Aufwärts*-Diskrepanzen und Erfolg. Darauf deuten bereits die in Tabelle 73 berichteten Befunde hin, denn für alle Erfolgskriterien mit Ausnahme von ZB ist der Erfolgszusammenhang für die Selbst-Aufwärts-Diskrepanz tendenziell höher als für die Selbst-Abwärts-Diskrepanz.

Da eine Signifikanzprüfung auf Korrelationsunterschiede zwischen allen dargestellten Korrelationen nicht zielführend ist, wurden mittlere Korrelationen über die FzFü-Faktoren errechnet und verglichen.[191] Beispielsweise ergibt sich für das erste Feld in Tabelle 73 eine mittlere Korrelation von $\bar{r}$ = -.49, wenn man die jeweiligen Zusammenhänge der Selbst-Aufwärts-Diskrepanzen auf den Dimensionen E/F, B/S, IB, K und P mit AZ1 mittelt (vgl. Tabelle 73).[192] Eine statistische Prüfung auf Unterschiede zwischen den so gemittelten Selbst-Aufwärts-Korrelationen (individuelle Werte) und den entsprechenden Selbst-Abwärts-Korrelationen ergibt, dass die personengebundenen Selbst-Aufwärts-Diskrepanzen mit AZ1, AZ2, EEF, EFF und SAT signifikant stärker negativ zusammenhängen als die Selbst-Abwärts-Diskrepanzen ($p < .05$). Tendenziell zeigt sich ein derartiger Unterschied auch für weitere Kriterien (WOBE, EF-WOBE, EF-AZ). Lediglich die Leistungsparameter (L, ZB) scheinen mit der Selbst-Aufwärts- und der Selbst-Abwärts-Diskrepanz in ähnlich starkem Ausmaß zusammenzuhängen (q = .03 bzw. q = .04; zu Selbst-Abwärts-Diskrepanz und Leistung vgl. auch S. 302). Der durchschnittliche Korrelations-Unterschied zwischen den personengebundenen Selbst-Aufwärts- und den Selbst-

[191] Gewarnt wird davor, ausschließlich die mittleren Korrelationen zu betrachten. Probleme ergeben sich immer dann, wenn die primären – zu mittelnden – Korrelationskoeffizienten unterschiedliche Vorzeichen aufweisen. Dann nämlich nivellieren sie sich gegenseitig bis zu einem gewissen Grad.

[192] Die Dimensionen FvF und S mit jeweils geringerem Stichprobenumfang wurden ausgeschlossen, da für eine statistische Prüfung der Korrelationsunterschiede (siehe unten) ein einheitlicher Stichprobenumfang notwendig ist.

Abwärts-Differenzen und Erfolg liegt ohne Berücksichtigung von L und ZB im mittleren Bereich ($\bar{q}$ = .40). Somit ist Hypothese 36) für die meisten Erfolgskriterien zuzustimmen.

Es ist durchaus plausibel, dass die Selbst-Aufwärts-Übereinstimmung auf die Bereitschaft zu zusätzlicher Anstrengung, die Effektivitätseinschätzung der Führung und die (Arbeits)Zufriedenheit der Mitarbeiter mehr Einfluss hat als die Selbst-Abwärts-Übereinstimmung, denn es geht ja weder um die Erfolge (Anstrengung, Effektivität, Arbeitszufriedenheit) der Selbstbeurteiler noch um die der Abwärtsbeurteiler.

Da aber auch die Selbst-Abwärts-Diskrepanz negativ mit dem Erfolg der Mitarbeiter zusammenhängt, scheint auch der Mechanismus, der zur Selbstüberschätzung im Vergleich zur Abwärtsbeurteilung führt, erfolglose Führung mit zu bedingen. Dies kann andererseits als erster Hinweis dafür interpretiert werden, dass Selbstüberschätzung generell, nicht die unterschiedliche Einschätzung zwischen zwei *bestimmten* Perspektiven den Erfolg mindert. Unterstützt wird diese Annahme dadurch, dass sich die Validität der Selbst-Abwärts-Differenzen nicht von der Validität der *gekreuzten* Selbst-Aufwärts-Differenzen unterscheidet. Da die Validität jeweils nur für diejenigen Erfolgsanteile gilt, die das gesamte Team betreffen, scheint zumindest dafür eine generelle Selbstüberschätzung unabhängig von der jeweils gegenübergestellten Perspektive (Aufwärts oder Abwärts) zu sein.

Bisherige Forschungsbemühungen waren eindeutig auf die Untersuchung der *Selbst*-Fremd-Diskrepanzen ausgerichtet. Für Entwicklungsmaßnahmen im Bereich der Personalführung, die naturgemäß an der Führungskraft ansetzen, ist dies zweifellos ein wichtiger Zielgegen-stand. Dennoch ist denkbar, dass überdies **Fremd-Fremd-Diskrepanzen** auf generelle Unklarheiten, intransparente Ziele und Vorgehensweisen oder einen Art „Zick-Zack-Kurs“ der Führung hinweisen und folglich Erfolgsvarianz erklären können. Untersucht man die Abwärts-Aufwärts-Diskrepanz, stützen die Korrelationsanalysen (Tabelle 73) diese Annahme. So erklärt die Abwärts-Aufwärts-Diskrepanz bis zu 42% (aggregierte Werte) bzw. 55% (individuelle Werte) der Varianz. Generell sind die Effekte tendenziell höher als die der Selbst-Abwärts-Diskrepanzen und teilweise höher als die der Selbst-Aufwärts-Diskrepanzen. Vor allem das Wohlbefinden der Mitarbeiter kann durch die Fremd-Fremd-Diskrepanz im Vergleich zu den Selbst-Fremd-Diskrepanzen besser erklärt werden. Das spricht im Sinne allgemeinpsychologischen Wissens (z.B. Dörner, 1999) dafür, dass Mitarbeiter Unwohlsein und Stress eher aufgrund von Unbestimmtheit und weniger aufgrund negativer Konstanten erleben. Differieren nämlich verschiedene Fremdrater in ihrer Führungsbeurteilung stark, ist eher davon auszugehen, dass wenig konsistentes Führungsverhalten gezeigt wird, auf das sich Dritte, z.B. Mitarbeiter verlassen könnten. Hypothese 38), dass die Selbst-Fremd-Diskrepanzen bessere Erfolgsprognosen leisten als die Fremd-Fremd-Diskrepanzen, wird vor dem Hintergrund dieser Ergebnisse vorerst abgelehnt.

Wendet man das Criss-Cross-Design an, können zwar nur noch bis zu maximal 19% der Varianz erklärt werden. Durchschnittlich sind geringe Effekte auszumachen, wohingegen die personengebundene Analyse für vier Kriterien (AZ1, EEF, EFF, SAT) auch starke Effekte hervorbrachte. Doch dieser Rückgang entspricht den Erwartungen, wenn aus der Führungsbeurteilung von Mitarbeitern die Erfolge ihrer Kollegen prognostiziert werden. Er kann also allenfalls teilweise auf den Single-Source-Bias zurück geführt werden, so dass Hypothese 37) beibehalten wird.

Im Gegensatz zu den erwarteten negativen Zusammenhängen, die zwischen den interperspektivischen Niveau-Unterschieden und den Erfolgskriterien beschrieben wurden, kovariieren die Profil-Ähnlichkeiten zwischen zwei Beurteilern aus verschiedenen Perspektiven nicht beziehungsweise nicht konsistent mit den Erfolgskriterien. Dafür wurden τ-Korrelationen zwischen je zwei Beurteilern über die beurteilten Führungsdimensionen hinweg berechnet (bezüglich der gleichen Fokusperson), um diese mit den Erfolgskriterien in Beziehung zu setzen. Der inkonsistente Befund gilt sowohl für die beiden Selbst-Fremd-Übereinstimmungen als auch für die Abwärts-Aufwärts-Übereinstimmung (Ergebnistabelle in Anhang S) und verweist erneut darauf, dass verschiedene Übereinstimmungsmaße nicht austauschbar sind (vgl. Kapitel 6.3.4).

Aus den Korrelationsanalysen kann also festgehalten werden, dass

- die Selbst-Fremd- sowie die Fremd-Fremd-Diskrepanzen mit Führungserfolg negativ zusammenhängen,
- die Selbst-Aufwärts-Diskrepanzen stärker negativ mit Führungserfolg zusammenhängen als die Selbst-Abwärts-Diskrepanzen,
- der Zusammenhang zwischen der Selbst-Aufwärts-Diskrepanz und Erfolg nicht allein durch den Single-Source-Bias erklärt werden kann,
- der Zusammenhang zwischen der Abwärts-Aufwärts-Diskrepanz und Erfolg nicht allein durch den Single-Source-Bias erklärt werden kann,
- die Selbst-Fremd-Diskrepanzen nicht stärker mit Erfolg korrelieren als die Abwärts-Aufwärts-Diskrepanz (eher umgekehrt),
- für unterschiedliche Kriterien Diskrepanzen zwischen unterschiedlichen Perspektiven die besten Vorhersagen leisten, z.B. für die Abwärts-Aufwärts-Diskrepanz für das Wohlbefinden und die Selbst-Aufwärts-Diskrepanz für Extra Effort (also Motivation) und dass
- die korrelativen Selbst-Fremd- und Fremd-Fremd-Übereinstimmungen nicht mit Erfolg kovariieren.

Um die *Richtung* der Abweichung besser in Betracht ziehen zu können, wurden die Führungskräfte in Überschätzer, Übereinstimmer und Unterschätzer kategorisiert. Anschließend wurden für die einzelnen Erfolgskriterien einfaktorielle Varianzanalysen für dreifach gestufte Selbst-Aufwärts- bzw. Selbst-Abwärts-Diskrepanzen durchgeführt. In post hoc Vergleichen wurden die Erfolgskennwerte der jeweiligen Kategorien paarweise verglichen. Ein solches Vorgehen entspricht der üblichen Methode der aktuellen Forschungspraxis. Da bislang nicht abschließend geklärt ist, ob unterschiedliche Modelle zu unterschiedlichen Ergebnissen führen, ist es angebracht, quasi explorativ verschiedene Methoden einzusetzen und diese zu vergleichen. Dazu gehören auch unterschiedliche Differenzmodelle, auf deren Basis eine Kategorisierung vorgenommen wird. Deshalb wird zusätzlich zur Kategorisierung nach der *Richtung* der Abweichung auch eine Kategorisierung nach dem *Ausmaß* der Abweichung vorgenommen. Auch auf deren Grundlage wurden einfaktorielle Varianzanalysen – in diesem Fall für alle drei interperspektivischen Vergleiche – gerechnet.

Im Folgenden werden nun die Ergebnisse der Varianzanalysen für die Selbst-Aufwärts-, Selbst-Abwärts- und Abwärts-Aufwärts-Diskrepanzen dargestellt:

Die Befunde unterstützen die Annahme sowie die Ergebnisse der Korrelationsanalysen, dass das Ausmaß (unabhängig von der Richtung) der **Diskrepanz zwischen Selbst- und Aufwärtsbeurteilungen** einen Einfluss auf die Erfolgskriterien hat. Kategorisiert man – der Übersichtlichkeit zuliebe – das Ausmaß der Diskrepanz nach der Selbst-Aufwärts-Distanz der *Gesamtskala* (FzFü) in Terzile, dann finden sich für folgende Kriterien signifikante Unterschiede zwischen den drei Kategorien ($p < .01$): AZ1, EF-AZ, L, EEF, EFF und SAT. Für diese Kriterien sind die Effekte als stark zu interpretieren ($f = .61$ bis $f = .77$). Für die anderen Kriterien zeigen sich geringe bis mittlere Effekte, die Varianzaufklärung liegt für diese auch nur im Bereich zwischen 1% (AZ2, ZB) und 13% (EF-WOBE). Die durchschnittliche Varianzaufklärung beträgt 10%.

Post hoc Mehrfachvergleiche erbringen jeweils zwischen den beiden Extrem-Gruppen, teilweise auch zwischen mittlerer und einer der Extremgruppen signifikante Unterschiede (Tukey-HSD-Test bei Varianzhomogenität, Games-Howell-Test bei Varianzinhomogenität) (vgl. Abbildung 33, ohne EF-AZ). Diese stützen die Hypothese, dass mit zunehmender Diskrepanz zwischen Selbst- und Aufwärtsbeurteilungen die Erfolgskriterien negativer bewertet werden. Die Daten werden am besten durch einen linearen Trend beschrieben.

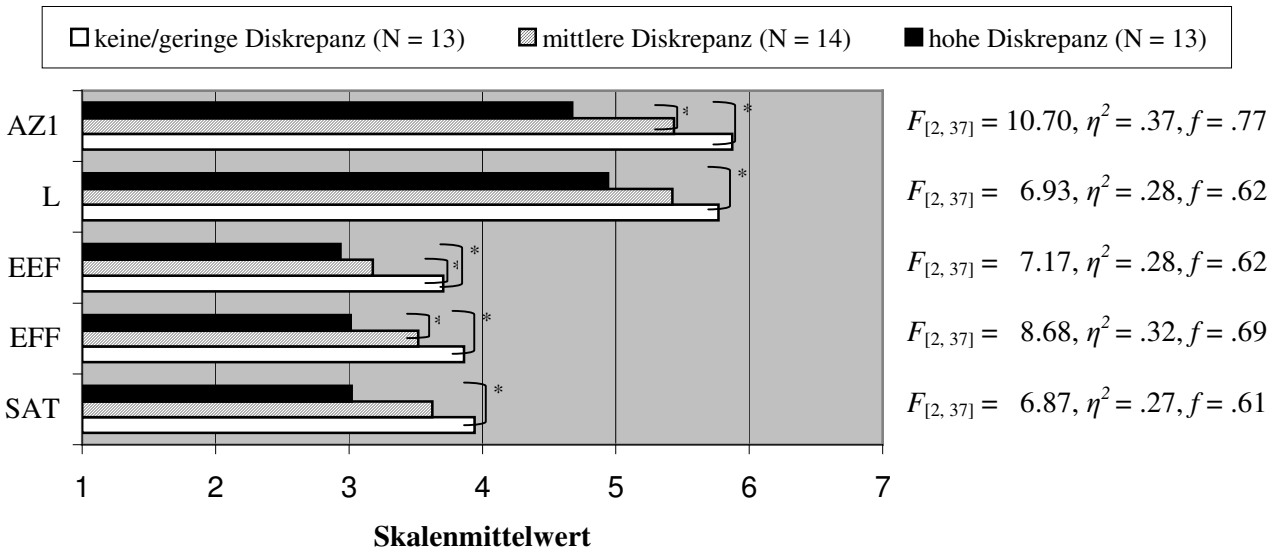


Abbildung 33. Einfaktorielle Varianzanalysen für dreifach gestufte Selbst-Aufwärts-Diskrepanz (FzFü-Mittelwert): Starke Effekte ($f \geq .40$) in erwarteter Richtung

Anmerkungen. Skala von 1 bis 7 für AZ1 und L, Skala von 1 bis 5 für EEF, EFF und SAT. EF-AZ ($F_{[2,37]}$ = 6.14, η^2= .26, f = .59) ist nicht in die Abbildung aufgenommen, da das Kriterium als Differenz operationalisiert ist, das geringe Werte annimmt und dadurch der Abbildung die Übersichtlichkeit nähme.

Tabelle 74 fasst zusammen, welche Erfolgsvariablen bei geringer Selbst-Aufwärts-Abweichung auf den *einzelnen* FzFü-Faktoren statistisch signifikant höher ausgeprägt sind als bei starker Abweichung ($p < .05$). Ein umgekehrter Zusammenhang ergibt sich für AZ2 (Erklärung vgl. S. 394). Für einige Variablen zeigt sich zusätzlich auf statistisch bedeutsamem Niveau, dass auch bei mittlerer Diskrepanz höhere Erfolge als bei hoher Diskrepanz und niedrigere Erfolge als bei niedriger Diskrepanz auftreten. Insgesamt sind häufig – auch für die nicht signifikanten Zusammenhänge – starke Effekte aufzudecken. Dafür ist in der Regel nur die lineare Trendkomponente statistisch bedeutsam. Für Ausnahmen erklären quadratische Terme zusätzliche Trendkomponenten; dann ist der Erfolg bei mittlerer Diskrepanz tendenziell höher als bei niedriger (z.B. Diskrepanz auf IB für die Mittelwertsunterschiede auf EEF).

Neben dem Ausmaß der Abweichung kann die *Richtung* der Selbst-Aufwärts-Abweichung betrachtet werden. Auch hierfür wird zunächst die Abweichung über die fünf FzFü-Faktoren hinweg analysiert, bevor anschließend die einzelnen Faktoren getrennt voneinander fokussiert werden.

Kategorisiert man die Führungskräfte in Unterschätzer, Übereinstimmer und Überschätzer, dann zeigen die Ergebnisse der Varianzanalyse mit anschließenden post hoc Mehrfachvergleichen zwischen diesen drei Faktorstufen genauso wie für das Ausmaß der Abweichung, dass für die Mehrzahl der Erfolgskriterien signifikante Unterschiede

zwischen Unterschätzern (N = 4), Übereinstimmern (N = 14) und Überschätzern (N = 22) bestehen. Es werden zwischen 4% (AZ2) und 31% (AZ1), im Durchschnitt 15% der Varianz erklärt. Starke Effekte konnten für AZ1, EEF, EFF und SAT ermittelt werden (f = .50 bis f = .67), für die anderen Kriterien liegen die Effektstärken im mittleren Bereich (f = .20 bis f = .39). Die Daten werden am besten durch einen linearen Trend beschrieben. Die höchsten Erfolge werden dabei den Unterschätzern, die geringsten Erfolge den Überschätzern bescheinigt. So unterscheiden sich für folgende Erfolgskriterien jeweils die beiden Extremgruppen (Über- und Unterschätzer) signifikant: AZ1, EF-AZ, EF-WOBE, EEF, EFF, SAT. Für AZ1 und EFF lassen sich zusätzlich Unterschiede zwischen Unterschätzern und Übereinstimmern sowie zwischen Übereinstimmern und Überschätzern festhalten.

Bei einer spezifischeren Analyse ergibt sich ein sehr ähnliches Bild für die Abweichungsrichtung auf den einzelnen Führungsdimensionen. Tabelle 74 gibt unter anderem an, welche Erfolgskriterien statistisch bedeutsame Mittelwertsunterschiede über die Kategorien der Abweichungs*richtung* hinweg aufweisen (p < .05).

Tabelle 74. Mittelwertsunterschiede (Erfolgskriterien) in Abhängigkeit des Ausmaßes (A) bzw. der Richtung (R) der Selbst-Aufwärts-Diskrepanz, N = 40

FzFü-Faktor[a]	AZ1	AZ2	EF-AZ	WO-BE	EF-WOBE	L	ZB	EEF	EFF	SAT
E/F	A/R					A/R	A/R	A	A/R	A/R
B/S	R						R	A	A/R	A/R
IB	A	(A)	A			A/R	R	A	A	A
K	R		A/R			A/R		R	A/R	A/R
P	R		A			A/R	A	A/R	A/R	A/R
E/F bis P	A/R		A/R		R	A		A/R	A/R	A/R
FvF (N=30)		(A)								
S (N = 20)		(A)								

Anmerkungen. einfaktorielle Varianzanalyse mit post hoc Mehrfachvergleichen (Tukey-HSD bei Varianzhomogenität, Games-Howell bei Varianzinhomogenität), [a]FzFü-Faktor, auf dessen Basis Diskrepanzkategorien gebildet wurden, S: keine Unterschätzer, so dass die Auswertung für die Richtung der Diskrepanz nicht vorgenommen wurde, A: Ausmaß der Abweichung (Kategorisierung basiert auf Terzilsplit), R: Richtung der Abweichung (Kategorisierung basiert auf Vorgehen von Atwater et al., vgl. S. 248), in Klammern: Vorhersage entgegen der erwarteten Richtung.

Für alle nicht signifikanten Mehrfachvergleiche zwischen Unterschätzern, Übereinstimmern und Überschätzern auf den Dimensionen E/F, B/S, IB, K und P ergibt sich zumindest die Tendenz, dass Unterschätzer höhere (oder ähnliche) Erfolge haben als (wie) Übereinstimmer und diese höhere Erfolge als Überschätzer. Mittlere Effektstärken für nahezu alle Kriterien unterstützen diese Annahme. Dass die Diskrepanz auf einem *einzelnen* Führungsfaktor nicht die Mittelwertsunterschiede auf *allen* Erfolgskriterien gleich gut erklären kann, ist keineswegs unerwartet und rechtfertigt im Grunde nur die multidimensionale Erfassung von Führungsverhalten. Dennoch können Diskrepanzen auf E/F, B/S, K sowie P die Mittelwertsunterschiede auf drei *gleichen* Kriterien erklären (AZ1, EFF, SAT).

Auch wenn das auf den ersten Blick einen stärkeren Einfluss der Diskrepanz für diese drei Kriterien vermuten lässt, spricht dagegen, dass die Aufwärtsbeurteilung allein mit diesen Erfolgsmaßen bereits höher als mit anderen Erfolgsmaßen korreliert.

Hervorzuheben ist, dass die Leistung – das explizit wichtigste Erfolgskriterium für Unternehmenserfolg – ebenfalls mit der Diskrepanz auf vier der fünf FzFü-Faktoren zusammenhängt. Erstaunlich ist, dass ausgerechnet zur Diskrepanz auf der Dimension B/S, also dem Führungsinstrument, das gewöhnlich explizit mit dem Ziel der Leistungsverbesserung eingesetzt wird, kein Zusammenhang besteht. Kreuzvalidierungen mit weiteren Stichproben sollten dies zukünftig überprüfen.

Für die Dimension S wurden keine post hoc Mehrfachvergleiche gerechnet, da sich keine einzige Führungskraft im Vergleich zum Urteil ihrer Mitarbeiter unterschätzte. In t-Tests, mittels derer die Mittelwerte zwischen Übereinstimmern und Überschätzern verglichen wurden, ergab sich ein überraschendes Ergebnis. Auf AZ2, WOBE und EFF erreichen die Überschätzer höhere Werte als die Übereinstimmer. Dies zeigt sich (für die Diskrepanz auf S) tendenziell auch für alle anderen Erfolgskriterien mit Ausnahme von EF-AZ und EF-WOBE. Möglicherweise lernen die Mitarbeiter, die aufgrund einer inadäquaten Stellvertretungsvorbereitung quasi „ins kalte Wasser“ geworfen werden, eher, sich selbst erfolgreich zu führen und zugleich die Führungsmängel der eigentlichen Führungskraft zu sehen.

Vergleicht man den Einfluss der Richtung der Selbst-Aufwärts-Diskrepanz mit dem Einfluss des puren Ausmaßes der Abweichung, dann stellt man keine prinzipiellen Unterschiede in der Wirkung auf die Erfolgskriterien fest. Die Richtung erweist sich insgesamt lediglich als leicht besserer Prädiktor. Das überrascht nicht, da die Einteilung der Führungskräfte nach dem Ausmaß bzw. der Richtung der Abweichung nicht unabhängig voneinander ist. Führungskräfte mit *hoher* Diskrepanz zu ihren Mitarbeiterratings sind fast ausschließlich *Überschätzer*. Zumindest gibt es nur auf B/S, IB, K und FvF Führungkräfte – je eine – mit hoher Diskrepanz, die zugleich als Unterschätzer kategorisiert wurden.

Auffällig ist bei einem Vergleich allenfalls, dass AZ1 bei mehr FzFü-Faktoren durch die Richtung der Diskrepanz beeinflusst ist, während EEF häufiger durch das Ausmaß beeinflusst ist. Außerdem scheint IB insbesondere dann Einfluss auf den Erfolg zu haben, wenn das Ausmaß der Differenz auf IB, nicht die Richtung der Differenz herangezogen wird. Einzelne Erfolgskriterien scheinen quasi nichts mit der Diskrepanz (Ausmaß oder Richtung) zu tun zu haben (WOBE, EF-WOBE). Dies deuteten die Korrelationsanalysen (vgl. Tabelle 73) bereits an. Da dies an den ungünstigen Skalenkennwerten dieser Erfolgsskalen liegen kann, wird auf eine inhaltliche Interpretation verzichtet.

Zusammenfassend kann festgehalten werden, dass sich mit zunehmendem Abstand der Selbstratings von den Mitarbeiterratings der Führungserfolg mindert. Eine mögliche Erklärung dafür kann in der Kommunikation zwischen Führungskraft und Mitarbeiter liegen. Tartler, Goihl et al. (2003) finden zum Beispiel, dass diejenigen Mitarbeiter, die in ihrer

Führungsbewertung nur gering (vs. stark) von ihrer Führungskraft abweichen, die Qualität der Kommunikation besser bewerten. Wenn die unterschiedlichen Bewertungen der Führungsverhaltensweisen auf einer unzureichenden Kommunikation basieren, dann liegt nicht fern, dass Kommunikationsprobleme auch die Arbeitszufriedenheit und –motivation, das Wohlbefinden, die Leistung und die Effektivität beeinflussen.

Für die *Richtung* der Abweichung zeigt sich deutlich, dass die Überschätzer am wenigsten erfolgreich sind und die Unterschätzer mindestens so erfolgreich wie die Übereinstimmer.[193] Dies könnte dadurch erklärt werden, dass die Unterschätzer eine permanente Motivation zur Verbesserung haben und deshalb auch Anstrengung in Führungsaufgaben investieren. Dafür spricht, dass die Unterschätzer die besten Aufwärts- und Abwärtsratings erhalten (Tabelle 76). Insbesondere die positiven Abwärtsratings, die unabhängig von der Kategorisierung als Unterschätzer sind, validieren quasi deren Führungserfolg. In vielen Studien (z.B. Fleenor et al., 1996; Furnham & Stringfield, 1994) sind es sogar gerade die Abwärtsratings, die prinzipiell als Erfolgskriterien herangezogen werden. Überschätzer fehldiagnostizieren hingegen ihre Stärken und Schwächen möglicherweise, so dass sie (wenn überhaupt) entweder Veränderungsmotivation in falsche Themen investieren oder sich und ihrem Team Ziele stecken, die auf der Basis ihrer Kompetenzen nicht erreichbar sind und zu Misserfolgen führen müssen. Für diese Erklärung spricht auch, dass die durchschnittliche Varianzaufklärung der mittleren Selbst-Aufwärts-Diskrepanz für die Richtung der Abweichung etwas höher liegt als für das Ausmaß der Abweichung (vgl. Tabelle 75, S. 397). Insgesamt liegt sie in einer Höhe, die im Einklang mit früheren Befunden steht. So können in den Stichproben von Church (1997b) und von Fleenor et al. (1996) bis zu 15% bzw. 10% aufgeklärt werden. Allein Tartler, Goihl et al. (2003) erklären durch die Selbst-Aufwärts-Diskrepanz bis zu 42% Varianz. Dies muss im Kontext der anderen Ergebnisse mindestens partiell auf einen Methoden-Bias zurückgeführt werden, denn sowohl die Führungs- als auch die Erfolgsratings entstammen dem MLQ. Für das MLQ-externe Kriterium der Arbeitszufriedenheit werden auch hier nur 11% der Varianz erklärt. Hypothese 38) wird nur teilweise bestätigt.

Kritisch bleibt auch hier anzumerken, dass die Mitarbeiter sowohl die Erfolgseinschätzungen als auch diejenigen Führungsbeurteilungen abgaben, die in die Berechnung der Selbst-Aufwärts-Diskrepanz eingingen. Deshalb werden die Selbst-Aufwärts-Diskrepanzen erstens zusätzlich an den Abwärtsbeurteilungen validiert (S. 397). Zweitens wird der Effekt der Führungsdiskrepanz insofern einem härteren Test unterzogen, als der Einfluss der Selbst-*Abwärts*-Diskrepanz auf die Erfolgskriterien, die von Aufwärtsbeurteilern eingeschätzt wurden, untersucht wird. Ein Effekt sollte sich zu dem Ausmaß zeigen, zu dem der Erfolg von einer prinzipiellen Fehleinschätzung der Führungskraft abhängt, die sich

[193] Wo der Zusammenhang keine Signifikanz erreicht, ist mit wenigen Ausnahmen (WOBE) zumindest die erwartete Tendenz erkennbar.

sowohl auf die Selbst-Abwärts- als auch auf die Selbst-Aufwärts-Diskrepanz auswirkt. Der Effekt muss allein deshalb geringer ausfallen als der der Selbst-Aufwärts-Diskrepanz, da Abwärtsbeurteiler das durchschnittliche Führungsverhalten einer Führungskraft beurteilen, während Aufwärtsbeurteiler das Führungsverhalten beurteilen, das sie persönlich erfahren.

Für die **Selbst-Abwärts-Diskrepanz** ergibt sich dementsprechend auch ein anderes Bild als für die Selbst-Aufwärts-Diskrepanz. Da der Stichprobenumfang sehr gering ist (N = 23) und die Selbst-Abwärts-Diskrepanz generell sehr gering ausfällt (vgl. Kapitel 6.3.2), sind nicht viele Zusammenhänge zu erwarten. Jeder statistisch bedeutsame Zusammenhang ist deshalb hoch zu gewichten.

Betrachtet man auch hier zunächst die *mittlere Diskrepanz* über die fünf FzFü-Faktoren hinweg (E/F, B/S, IB, K, P), dann zeigt sich, dass das reine *Ausmaß* der Diskrepanz auf kein Erfolgskriterium statistisch bedeutsamen Einfluss nimmt (Terzilsplit für Betrag der Differenz „Selbst minus Abwärts“, N = 9, 5, 9). Dennoch ist ein mittlerer Effekt von $\overline{f}$ = .30 auszumachen, durchschnittlich werden 9% der Varianz erklärt. Je höher das Ausmaß der Selbst-Abwärts-Abweichung ausfällt, umso geringer sind die Erfolgskennwerte der geführten Mitarbeiter.[194]

ZB und AZ2 sind Erfolgskriterien, die in statistisch bedeutsamem Ausmaß von der Diskrepanz *einzelner* Dimensionen beeinflusst werden. Bei einer geringen Selbst-Abwärts-Abweichung auf der Dimension P sind die Mitarbeiter zufriedener mit der Beurteilungshäufigkeit als bei einer starken Selbst-Abwärts-Abweichung. Bei geringer Abweichung auf S ist ebenfalls die Beurteilungszufriedenheit besser als bei mittlerer Abweichung. Weiter beschreiben die Mitarbeiter, deren Führungskräfte in ihrer Beurteilung auf S von der Abwärtsbeurteilung nur gering abweichen, höhere Arbeitszufriedenheit (AZ2) als bei mittlerer oder hoher Abweichung.

Betrachtet man die generelle *Richtung* der Abweichung (über die fünf FzFü-Faktoren hinweg), dann zeigt sich ein signifikanter Effekt für ZB. Die Mitarbeiter von Überschätzern sind weniger zufrieden mit der Beurteilungshäufigkeit als die Mitarbeiter der Unterschätzer (Tukey-HSD, p = .04). In der Regel gilt, dass die Überschätzer (N = 7) die geringsten und die Unterschätzer (N = 7) die höchsten Erfolge hervorbringen, wobei sich letztere kaum von den Übereinstimmern (N = 9) unterscheiden. Die Effekte reichen von f = .11 (für EFF) bis f = .64 (für ZB) und liegen für die zehn Kriterien im Mittel bei $\overline{f}$ = .37 (vgl. Abbildung 34). Die mittlere Varianzaufklärung beträgt 13%. Für alle Erfolgskriterien außer AZ2 und WOBE zeigt sich also mindestens ein kleiner Effekt in erwarteter Richtung, im Durchschnitt zeigt sich sogar ein mittlerer Effekt. Für AZ2 weist der Effekt in die umgekehrte Richtung: Mitarbeiter von Überschätzern sind am zufriedensten mit den

[194] Eine Ausnahme stellt das Kriterium L dar, das bei starker Selbst-Abwärts-Abweichung tendenziell stärker ausgeprägt ist als bei mittlerer oder schwacher Abweichung. Da sich die drei Gruppen jedoch insgesamt um weniger als einen halben Skalenpunkt unterscheiden, ist das Ergebnis praktisch unbedeutend.

Kontextbedingungen, während die der Unterschätzer angeben, am wenigsten zufriedensten zu sein ($F_{[2, 20]} = 1.89$, $p = .18$, $\eta^2 = .16$, $f = .44$) (Erklärung vgl. S. 394).

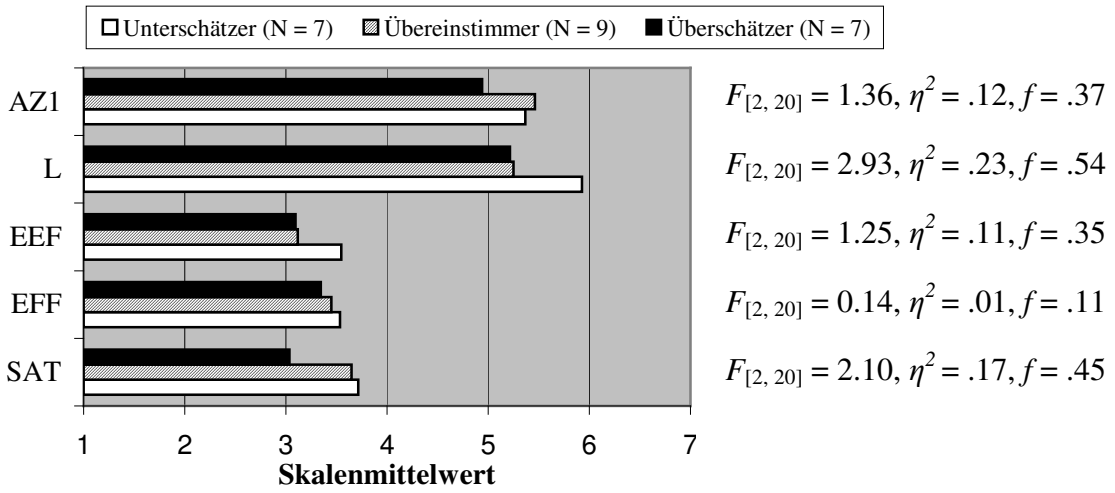


Abbildung 34. Einfaktorielle Varianzanalysen für dreifach gestufte Selbst-Abwärts-Diskrepanz (FzFü-Mittelwert): Bedeutsame Effekte ($f \geq .10$) in erwarteter Richtung

Anmerkungen. Skala von 1 bis 7 für AZ1 und L, von 1 bis 5 für EEF, EFF und SAT. EF-AZ ($F_{[2, 20]} = 1.07$, $\eta^2 = .10$, $f = .33$), EF-WOBE ($F_{[2, 20]} = .26$, $\eta^2 = .03$, $f = .16$) und ZB ($F_{[2, 20]} = 4.37$, $\eta^2 = .29$, $f = .64$) sind nicht in die Abbildung aufgenommen, da das Kriterium eine Differenz darstellt, die kleine Beträge annehmen und dadurch der Abbildung die Übersichtlichkeit nähmen.

Die Richtung der Selbst-Abwärts-Diskrepanzen auf *einzelnen Dimensionen* nimmt auf wenige Erfolgskriterien statistisch signifikanten Einfluss (Hypothese 39). Für fast alle Kriterien kann durch die Selbst-Abwärts-Abweichung auf fast jeder Führungsdimension jedoch soviel Varianz erklärt werden, dass durchschnittlich von mittleren Effekten zu sprechen ist. Nur die besonders starken, statistisch signifikanten Effekte werden im Folgenden genannt: So zeigt sich, dass die Mitarbeiter der Führungskräfte, die sich auf der Dimension E/F unterschätzen, höhere Leistung erbringen als die Mitarbeiter der Übereinstimmer. Außerdem sind die Mitarbeiter der Unterschätzer und der Übereinstimmer zufriedener mit Beurteilungen als die Mitarbeiter der Überschätzer. Für AZ2 zeigt sich ein teilweise umgekehrter Einfluss. Die Mitarbeiter der Überschätzer auf IB erreichen höherer AZ2-Werte als die Mitarbeiter der Übereinstimmer; die Mitarbeiter der Überschätzer auf FvF erreichen höhere AZ2-Werte als die Mitarbeiter der Unterschätzer. Nur die Mitarbeiter der Übereinstimmer auf S haben eine höhere Arbeitszufriedenheit (AZ2) als die Mitarbeiter der Überschätzer sowie der Unterschätzer.

Überraschenderweise scheint es generell eher so zu sein, dass die Überschätzung des eigenen Führungsverhaltens positiv mit der Zufriedenheit mit verschiedenen Kontextbedingungen (AZ2) zusammenhängt. Derartige Hinweise ergaben sich bereits für die Selbst-Aufwärts-Diskrepanz. Es ist denkbar, dass Führungskräfte, die sich selbst besonders posi-

tiv bewerten, zwar wenig Anstrengung in dyadische Führung zur individuellen Förderung ihrer Mitarbeiter investieren, aber in die Verbesserung der Kontextfaktoren. Dies sind Faktoren, die sich einer Führungskraft mit geringem Führungsengagement oder –können eher als verbesserungswürdig aufdrängen als subtilere Bedürfnisse einzelner Mitarbeiter. Eine derart ungleiche Investition in verschiedene Führungsaufgaben wäre auch dann stimmig, wenn Impression Management die Selbstüberschätzung des Führungsverhaltens herbeiführt. Kontextfaktoren sind – im Unterschied zu erfolgreichem und zufriedenem Arbeiten Einzelner – von Dritten gut beobachtbar oder insgesamt gut kommunizierbar und dienen somit der Selbstdarstellung. Sie können sogar Repräsentationsfunktion übernehmen und nach außen als vermeintliche Bestätigung der guten Führung angebracht werden.

Generell zeigt sich für die Selbst-Abwärts-Diskrepanz auf den einzelnen Faktoren weniger der Trend, dass die Unterschätzer die erfolgreichsten Führungskräfte darstellen, sondern dass die Übereinstimmer diesen Platz einnehmen. Die Daten werden häufig auch am besten durch quadratische Annahmen beschrieben. Auch hier besteht jedoch der Trend, dass die Überschätzer am wenigsten erfolgreich sind. Auch hier kann argumentiert werden, dass die Überschätzer keinen Anreiz sehen, sich in Führungsangelegenheiten besonders anzustrengen oder weiterzuentwickeln. So ist erklärbar, weshalb die Selbst-Abwärts-Diskrepanz den Erfolg Dritter – den Erfolg der Mitarbeiter – beeinflussen kann.

Wie für die Selbst-Aufwärts-Diskrepanz hat auch die Richtung der Selbst-Abwärts-Diskrepanz einen etwas höheren Erklärungswert als das Ausmaß der Selbst-Abwärts-Diskrepanz (vgl. Tabelle 75).

Ob auch die **Fremd-Fremd-Kongruenz** für Erfolgsprognosen valide ist, wurde – wie gesagt – bislang selten untersucht. Da es bei der Beantwortung dieser Frage nicht um Selbstüberschätzung der Führungskräfte geht, werden die Analysen lediglich für das Ausmaß der Diskrepanz, nicht für die Richtung der Abweichung genauer beschrieben. Die *mittlere Diskrepanz* über die fünf FzFü-Faktoren hinweg (E/F, B/S, IB, K, P), erbringt für AZ1, EF-AZ, EEF, EFF und SAT (je $p < .01$) statistisch hochsignifikante und starke Effekte (Terzilsplit für Distanzmaß zwischen der Abwärts- und den Aufwärtsbeurteilungen, N = 10, 11, 11) (vgl. Abbildung 35). Doch auch für alle anderen Erfolgskriterien lässt sich mindestens ein mittlerer Effekt (f_{min} = .31, AZ2) in erwarteter Richtung festhalten. Durchschnittlich werden 25.2% Varianz erklärt. Die Erfolgskennwerte fallen jeweils dann am niedrigsten aus, wenn die Abwärts-Aufwärts-Abweichung am höchsten ist. Die besten Ergebnisse erzielen die Mitarbeiter, die in ihrer Führungsbeurteilung der Vorgesetztenbeurteilung am ähnlichsten sind. Insofern unterscheiden sich die beiden Extremgruppen am deutlichsten.

Es gibt keine deutlichen Veränderungen, wenn man diejenigen Fälle ausschließt, für die die aggregierten Aufwärtsratings positiver ausfallen als die entsprechende Abwärtsbeurteilung (N = 5, 10, 10). Am veränderten Stichprobenumfang wird deutlich, dass sich nahezu alle dieser Fälle in der Gruppe mit geringer Abweichung befinden. Die durchschnittliche

Varianzaufklärung beträgt dann 19.1%. Dies kann als Hinweis dafür gewertet werden, dass – wie bei den Selbst-Fremd-Analysen – die Richtung der Abweichung einen geringen, zusätzlichen Erklärungswert über das Ausmaß hinaus besitzt.

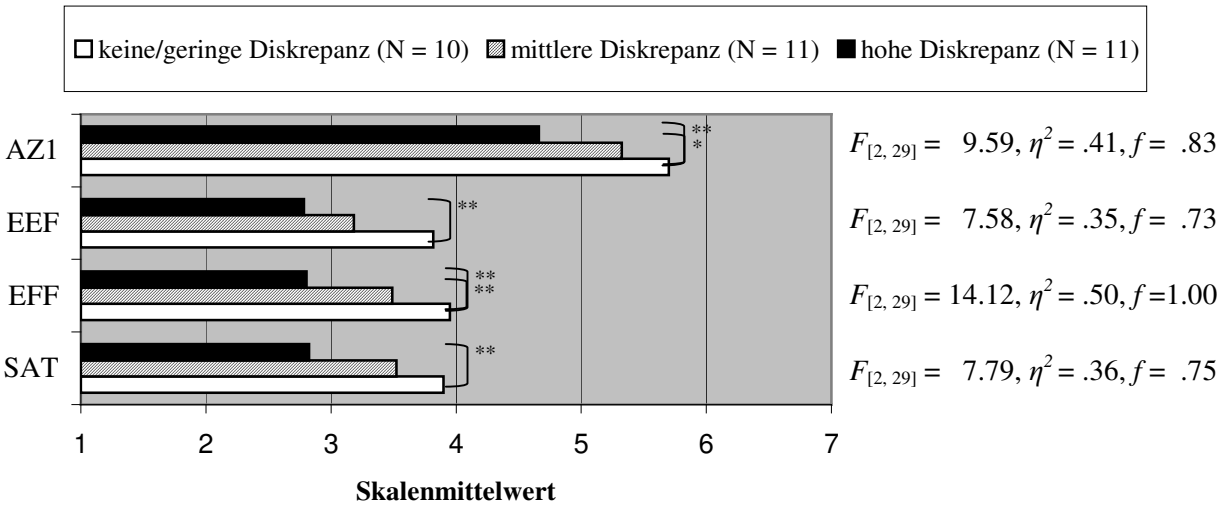


Abbildung 35. Einfaktorielle Varianzanalysen für dreifach gestufte Abwärts-Aufwärts-Diskrepanz (FzFü-Mittelwert): Starke Effekte ($f \geq .40$) in erwarteter Richtung

Anmerkungen. Skala von 1 bis 7 für AZ1, von 1 bis 5 für EEF, EFF und SAT. EF-AZ ($F_{[2,29]}$ = 6.06, η^2= .30, f = .65) ist nicht in die Abbildung aufgenommen, da das Kriterium eine Differenz darstellt, die kleine Beträge annehmen und dadurch der Abbildung die Übersichtlichkeit nähmen. ** $p < .01$, * $p < .05$.

Von den *einzelnen* Führungsdimensionen sind insbesondere B/S, K und P hervorzuheben. Die Abwärts-Aufwärts-Abweichungen auf diesen Dimensionen erklären – mit überwiegend starken Effekten von $f \geq .40$ – durchschnittlich 24%, 23% bzw. 22% der Erfolgsvarianz (über die zehn Erfolgskriterien hinweg). Für die anderen Führungsdimensionen sind überwiegend mittlere Effekte festzuhalten.

Im Vergleich zu den Selbst-Fremd-Diskrepanzen scheinen die Abwärts-Aufwärts-Diskrepanzen auf der *Dimension B/S* besonders valide zu sein. Dies geht stimmig mit der Annahme einher, dass Vorgesetzte die Führung ihrer unterstellten Führungskräfte vorwiegend dahingehend bewerten, ob diese leistungsfördernd wirkt. B/S ist freilich das Führungsinstrument, das am unmittelbarsten arbeitsunterstützende Effekte zeigen soll. Wenn aber die Wahrnehmung der Vorgesetzten auf dieser Dimension von der Wahrnehmung derjenigen abweicht, deren Arbeit gesteuert werden soll, ist von geringen Erfolgen auszugehen.

Da nicht nur Selbst-Fremd-, sondern auch Fremd-Fremd-Diskrepanzen mit den Erfolgskennwerte zusammenhängen, muss der Selbstüberschätzung der Führungskräfte ihre Rolle abgesprochen werden, als alleinige Wirkgröße zu fungieren (vorausgesetzt, es handelt sich

überhaupt um einen kausalen Zusammenhang). Neben der mangelnden Self-Awareness, die beispielsweise von Van Velsor et al. (1993) als kritische Variable für den Zusammenhang zwischen Diskrepanzen und Erfolgsvariablen betrachtet wird, müssen andere Mechanismen greifen. Plausibel scheint, dass unterschiedliche Fremdbeurteilungen auf einem inkonsistenten Führungskurs beruhen oder zumindest auf Führung, die von Außenstehenden nicht besonders gut „gegriffen" werden kann. Für diese Annahme spricht, dass die Dimension Klarheit – also das Gegenteil zu inkonsistentem Verhalten – zu den besonders validen zählt. Weiter gilt, dass ausschließlich die Diskrepanz auf der Dimension Klarheit einen nahezu starken Effekt auf Wohlbefinden ausübt ($F_{[2, 29]} = 2.17$, $p = .13$, $\eta^2 = .13$, $f = .39$). Durch die Abwärts-Aufwärts-Diskrepanzen auf den übrigen Führungsdimensionen können maximal 6% der Varianz an Wohlbefinden erklärt werden. Mehr Erfolgsvarianz kann auch durch keine Selbst-Fremd-Abweichung auf einer der Dimensionen erklärt werden.[195]

Die Abwärts-Aufwärts-Diskrepanz erklärt durchschnittlich mehr Varianz als die Selbst-Fremd-Diskrepanzen (vgl. Tabelle 75). Dies kann zum einen an den ohnehin reliableren und valideren Fremdbeurteilungsperspektiven liegen. Dies kann auch ein Hinweis darauf sein, dass von außen wenig greifbare Führung nicht erfolgreich ist.

Tabelle 75. Übersicht über die mittlere Varianzaufklärung durch die verschiedenen Beurteilungsdiskrepanzen

	Varianzaufklärung durch mittlere Diskrepanz[a] zwischen		
	Selbst und Aufwärts	Selbst und Abwärts	Abwärts und Aufwärts
Richtung der Abweichung	15.4%	13.1%	25.2% (19.1%)[b]
Ausmaß der Abweichung	10.0%	9.1%	

Anmerkungen. [a]mittlere Diskrepanz der fünf FzFü-Faktoren (E/F, B/S, IB, K, P) [b]Da die Führungskräfte bei den Analysen der Abwärts-Aufwärts-Diskrepanz nicht in Überschätzer, Übereinstimmer und Unterschätzer eingeteilt werden können, kann nicht analog zu den Selbst-Fremd-Diskrepanzen für Richtung und Ausmaß getrennt ausgewertet werden. Deshalb basiert die Kategorisierung auf einem Terzilsplit der euklidischen Distanz zwischen Abwärts- und Aufwärtsbeurteilern. Das Ergebnis in Klammern ergibt sich dann, wenn alle Fälle ausgeschlossen werden, für die der Median der Aufwärtsbeurteilungen über dem Wert der entsprechenden Abwärtsbeurteilung liegt.

Die verschiedenen Diskrepanzen wurden bislang an den erfassten Erfolgskriterien validiert. Eine andere Möglichkeit besteht darin, Führungsbeurteilungen anderer Perspektiven heranzuziehen.

Tabelle 76 gibt einen Überblick über entsprechende Ergebnisse. Sie zeigt zunächst, dass diejenigen Führungskräfte, die sich im Vergleich zu den Aufwärtsbeurteilungen überschätzen, die höchsten Selbst- und die niedrigsten Aufwärtsbeurteilungen erhalten. Das verweist

[195] Eine Ausnahme bildet die Selbst-Aufwärts-Abweichung auf der Dimensionen FvF. Hier werden 17% Varianz erklärt. Allerdings weist der Effekt in die unerwartete Richtung. Die Mitarbeiter, die auf FvF stark von der Selbstbeurteilung ihrer Führungskräfte abweichen, geben das beste Wohlbefinden an.

darauf, dass die *Selbst-Aufwärts-Diskrepanz* nicht allein durch überhöhte Selbstratings zustande kommt.[196] Zieht man die Abwärtsbeurteilungen zur Validierung heran, dann fällt auf, dass sich diese für die Überschätzer und Übereinstimmer nicht unterscheiden. Deutlich positiver ist das Abwärtsurteil für die Unterschätzer, die zudem die positivsten Aufwärtsbeurteilungen erhalten.[197] Dies geht stimmig damit einher, dass die Unterschätzer auch die höchsten Erfolge für sich verbuchen können (siehe oben). Da sie nicht nur aus den Fremdperspektiven positiver beurteilt werden als die Übereinstimmer, sondern sich selbst im Durchschnitt sogar bessere Noten geben, ist anzunehmen, dass unter den Übereinstimmern prozentual mehr Führungskräfte sind, die tatsächlich schlechtere Führungsleistung erbringen. So wird auch plausibel, weshalb die Gruppe der Übereinstimmer nicht die besten Erfolgsprognosen aufweist. Dies wäre bei vergleichbarem Leistungsniveau zu erwarten, da weniger Reibungsverluste durch Diskrepanzen verschiedener Art anzunehmen wären.

Tabelle 76. Mittlere Selbst- und Fremdbeurteilungen des Führungsverhaltens (FzFü) für Überschätzer, Übereinstimmer und Unterschätzer

	Selbst-*Aufwärts*-Kongruenz			Selbst-*Abwärts*-Kongruenz		
	Selbst-Rating	Aufwärts-Rating	Abwärts-Rating	Selbst-Rating	Aufwärts-Rating	Abwärts-Rating
Überschätzer	°5.56 (22)^{a*}	4.23 (22)^{ab*}	5.29 (14)	5.73 (11)^{ab*}	3.94 (7)	5.07 (11)^{b**}
	.50	.85	.55	.41	.94	.49
Übereinstimmer	5.06 (14)	5.12 (14)	5.33 (8)	5.17 (12)	4.87 (10)	5.18 (12)^{b**}
	.45	.35	.40	.48	.77	.26
Unterschätzer	5.26 (4)	5.75 (4)	6.05 (2)	5.14 (9)	4.80 (7)	5.95 (9)
	.35	.29	.19	.37	.83	.19

Anmerkungen. Kategorisierung in Überschätzer, Übereinstimmer und Unterschätzer (FzFü) nach Atwater (vgl. S. 248), °erste Zeile: Mittelwert, *N* in Klammern, zweite Zeile: Standardabweichung. ^{a}Unterschied zu Übereinstimmer, ^{b}Unterschied zu Unterschätzer, ** $p < .01$, * $p < .05$, $^{+}$ $p < .10$.

Unterstützt werden diese Befunde dadurch, dass den Überschätzern die geringste transformationale Führung und am wenigsten Contingent Reward zugesprochen wird, gleichzeitig aber am meisten passive Führung (MBP, LF). Außerdem ergibt sich auch hier die Tendenz, dass die Unterschätzer die aktivste Führung leisten (vgl. Tabelle 77).

Tabelle 77. Mittlere Aufwärtsbeurteilung des Full Ranges of Leadership (MLQ) für Überschätzer, Übereinstimmer und Unterschätzer

	Selbst-*Aufwärts*-Kongruenz					Selbst-*Abwärts*-Kongruenz				
	TF	CR	MBA	MBP	LF	TF	CR	MBA	MBP	LF
Überschätzer	3.03 (22)^{ab*}	3.15^{a*}	3.21	2.13^{a+}	2.55^{a*}	2.95 (7)	3.00	3.44	2.29^{b+}	2.61
Übereinstimmer	3.62 (14)	3.64	3.58	1.63	1.90	3.44 (10)	3.53	3.50	1.89	2.10
Unterschätzer	4.00 (4)	3.58	2.71	1.71	2.22	3.54 (7)	3.48	2.89	1.63	2.23

Anmerkung. *N* wo nicht anders vermerkt: Selbst-Aufwärts: 19, 9, 3. Selbst-Abwärts: 6, 8, 7.

[196] Diese starke „Schere“ zwischen Selbst- und Aufwärtsratings macht plausibel, weshalb oben für Überschätzer die geringste Selbst-Aufwärts-Korrelation festgehalten wurde (vgl. Tabelle 60).
[197] Aufgrund der extrem geringen Anzahl an Unterschätzern sind Signifikanztests nicht aussagekräftig.

Was die *Selbst-Abwärts-Diskrepanz* betrifft, gilt analog, dass die Überschätzer die höchsten Selbst- und die niedrigsten Abwärtsbeurteilungen auf sich vereinen. Auch diese Diskrepanz gründet also nicht allein auf milden Selbstbeurteilungen. Da die Selbstbeurteilungen der Überschätzer aber deutlich besser ausfallen als die der Übereinstimmer, während die Abwärtsbeurteilungen für die Überschätzer und Übereinstimmer nur minimal differieren, tragen sie einen verhältnismäßig großen Anteil an der Diskrepanz. Darauf verweisen außerdem die Aufwärtsbeurteilungen, die für die Überschätzer tendenziell negativer ausfallen als für die Übereinstimmer oder Unterschätzer. Das betrifft sowohl die Aufwärtsbeurteilungen anhand des FzFü als auch anhand des MLQ (Tabelle 76 und Tabelle 77). Die Konvergenz der beiden Instrumente unterstreicht das Ergebnis vor allem auch aufgrund des geringen Stichprobenumfangs. Zieht man also zur Validierung der Selbst-Abwärts-Differenz die Aufwärtsbeurteilungen heran, so deckt sich auch dies mit den oben berichteten Befunden, dass die Überschätzer die geringsten Erfolge aufweisen, während die Übereinstimmer und Unterschätzer ähnlich erfolgreich sind.

Im Unterschied zu den Selbst-Fremd-Abweichungen scheinen für das Ausmaß der Abwärts-Aufwärts-Abweichung vor allem die Aufwärtsratings verantwortlich zu sein (vgl. Anhang T). Es zeigt sich auch hier, dass die Selbstbeurteilung am geringsten ausfällt, wenn die Abwärts-Aufwärts-Abweichung am höchsten ist. Dies unterstützt oben berichtete negative Zusammenhänge zwischen der Abwärts-Aufwärts-Diskrepanz und den Erfolgsmaßen. Allerdings ist der Effekt weniger deutlich als bei den Selbst-Fremd-Diskrepanzen.

Insgesamt zeigt sich also:

- Die Richtung und das Ausmaß der Selbst-Aufwärts-Diskrepanz haben einen starken Effekt auf den Führungserfolg (gemessen an Erfolgseinschätzung durch Mitarbeiter).
- Es sind die Überschätzer und diejenigen Führungskräfte, die in ihrer Selbstbeurteilung besonders stark von der Beurteilung ihrer Mitarbeiter abweichen (unabhängig von der Richtung), die die geringsten Erfolge erbringen. Insgesamt ist am ehesten ein linearer Trend festzustellen. Das bedeutet, dass die Mitarbeiter der Unterschätzer die höchsten Erfolge angeben.
- Die Unterschätzer erhalten die besten Aufwärts- und Abwärtsratings ihrer Führung. Insbesondere die positiven Abwärtsratings, die unabhängig von der Kategorisierung als Unterschätzer sind, validieren deren Führungserfolg zusätzlich.
- Im Vergleich zur Selbst-Aufwärts-Diskrepanz erklärt die Selbst-Abwärts-Diskrepanz (Ausmaß und Richtung) weniger Erfolg.

- Auch hier sind es die Überschätzer und diejenigen Führungskräfte, die in ihrer Selbstbeurteilung besonders stark von der Beurteilung ihres Vorgesetzten abweichen (unabhängig von der Richtung), die die geringsten Erfolge erbringen. Es zeigt sich eher ein quadratischer als ein linearer Trend insofern, dass die Übereinstimmer am erfolgreichsten führen.
- Die Überschätzer erhalten auch hier die negativsten Aufwärtsbeurteilungen, Übereinstimmer und Unterschätzer unterscheiden sich kaum.
- Möglicherweise resultiert eine adäquate Kommunikation sowohl in einer Abgleichung zwischen Selbst- und Fremdbild als auch in erfolgreicher Führung. Führungskräfte, die sich ihrer eigenen Kompetenzen durch Austausch und Feedback selbst bewusst geworden sind, verfolgen und erreichen adäquate Ziele. Alternativ könnte es sein, dass sich erfolgreiche Führungskräfte nicht überschätzen „müssen", weil ihr Erfolg an sich selbstwertdienlich genug ist (siehe unten).
- Die höchste Varianzaufklärung erzielt wider Erwarten keine der Selbst-Fremd-Diskrepanzen, sondern die Abwärts-Aufwärts-Diskrepanz.
- Inkonsistentes oder von außen wenig greifbares Führungsverhalten scheint also eher in geringem Erfolg zu münden als konsistent wahrgenommene Selbst-Fremd-Diskrepanzen.
- Die Richtung der Abweichung weist jeweils einen etwas höheren Erklärungswert auf als das pure Ausmaß der Abweichung.

Das ist der methodische Punkt, an dem nahezu alle Studien „stehen bleiben" (z.B. Atwater & Yammarino, 1992; Bass & Yammarino, 1991; Church, 1997a; Tartler, Goihl et al., 2003). Vorsicht bei der Interpretation ist geboten, weil ein solcher Zusammenhang zwischen Beurteilungskongruenz und Erfolgskriterien nicht bedeutet, dass tatsächlich die Kongruenz der kritische Prädiktor ist (und nicht die einzelnen Beurteilungsperspektiven). Die Frage nach der prädiktiven Validität bleibt also bis dato unbeantwortet.

Die Befunde von Studien, die in Kapitel 3.5 zitiert wurden, werden in der Regel – weniger vorsichtig – so interpretiert, dass eine mangelnde Kongruenz zu einer Minderung der Erfolge führt. Der vermutete Mechanismus liegt in der Kommunikation zwischen den nicht übereinstimmenden Beurteilern. Seit Brutus, Fleenor und Tisak (1999) in ihrer Studie gezielt überprüften, ob wirklich die Diskrepanz zwischen Selbst- und Peer-Ratings und nicht die einzelnen Beurteilungsperspektiven den mangelnden Erfolg ausmachen, und dabei zu dem Schluss kamen, dass dem nicht so sei, sind korrelative Zusammenhänge indes noch vorsichtiger als vorher zu interpretieren. Vielmehr plädieren sie dafür, für andere Selbst-Fremd-Beurteilungen[198] zu untersuchen, wie stark die prädiktive Validität der Führungsdiskrepanzen tatsächlich ist.

[198] neben der von ihnen untersuchten Selbst-Peer-Diskrepanz

Im folgenden Abschnitt wird entsprechend überprüft, ob die Übereinstimmung zwischen der Selbst- und Aufwärtsbeurteilung die *kritische* Variable für Erfolg darstellt. Wo es methodisch erlaubt ist, werden auch die Selbst-Abwärts- sowie die Abwärts-Aufwärts-Diskrepanzen weitergehend untersucht. Nur weil die Plausibilität sehr hoch ist, darf nicht ungeprüft angenommen werden, dass die Selbst-Fremd-Diskrepanzen *bessere* Prädiktoren darstellen als die Beurteilungen aus den einzelnen Perspektiven.

Dieser Fragenbereich hat weit reichende Konsequenzen, da eine negative Antwort die aktuell üblichen Beratungskonzepte anzweifeln würde (z.B. Scherm & Sarges, 2002). So ist etwa nicht nur verbreitet, dass diejenigen Führungskräfte Pluspunkte für ihre Karriere sammeln, denen eine hohe Selbst-Fremd-Kongruenz bestätigt wird, sondern auch, dass Unternehmen in die Führungskräfte-Förderung finanziell investieren, um beispielsweise über Feedbackprozesse und (Team-)Coaching eine höhere Kongruenz zu erzielen. Von praktischer Seite aus lautet die vorliegende Fragestellung also: Lohnt diese Investition tatsächlich, um die organisationalen Kennwerte zu verbessern, oder wären Investitionen in andere Ziele nützlicher?

6.5.2 Interperspektivische Diskrepanz versus uniperspektivische Beurteilung

Eine zentrale Frage der vorliegenden Arbeit besteht darin, zu prüfen, ob die Übereinstimmung zwischen zwei Beurteilungsperspektiven zur Erklärung des Führungserfolgs erstens einen „kritischen“ (vorliegendes Kapitel 6.5.2) und wenn ja, zweitens einen größeren Beitrag als die einzelnen, „einfachen“ Perspektiven leisten kann (Kapitel 6.5.3).

Partialkorrelationen geben einen ersten Hinweis darauf, dass der Zusammenhang zwischen der Selbst-Fremd-Diskrepanz und den Erfolgskriterien stark durch die Aufwärts- bzw. Abwärtsbeurteilung an sich bestimmt ist. Wird diese nämlich jeweils auspartialisiert, dann erreichen nur noch wenige Korrelationen ein Signifikanzniveau von 5%.[199]

40) Es ist die Selbst-Aufwärts-Diskrepanz, die die kritische prognostische Leistung für Erfolg erbringt (vs. einfache Selbst- und Aufwärtsbeurteilungen).

41) Die Selbst-Aufwärts-Diskrepanz eignet sich zur Erfolgsprognose mindestens ebenso gut wie die einfachen Beurteilungen aus der Selbst- und Aufwärtsperspektive.

[199] Im Einzelnen (*N* siehe Tabelle 73*) für die Selbst-Aufwärts-Diskrepanz: Korrelation zwischen Diskrepanz auf IB und EF-AZ (r = .34, p = .04), IB/SAT (r* = -.37, *p* = .02), K/EFF (*r* = -.37, *p* = .02), K/SAT (*r* = -.32, *p* < .05), FvF/EF-AZ (*r* = .38, *p* < .05), FvF/WOBE (*r* = .37, *p* < .05), S/AZ1 (*r* = .42, *p* = .04). Die beiden zuletzt genannten Zusammenhänge fallen entgegen der Erwartung positiv aus.
Für die Selbst-Abwärts-Diskrepanz: E/F/L (*r* = -.51, *p* = .01), K/L (*r* = -.48, *p* = .03), P/SAT (*r* = -.47, *p* = .02), P/EF-AZ (*r* = .44, *p* < .05), P/ZB (*r* = .47, *p* = .02).

Nach Edwards (1994a) müssen Kongruenzmodelle (= Diskrepanzmodelle) – also Modelle, die auf Beurteilungsdiskrepanzen beruhen – vier Kriterien genügen, um nicht abgelehnt zu werden:

1. Das Kongruenzmodell erklärt einen statistisch bedeutsamen Varianzanteil. Das Beta-Gewicht der Diskrepanz ist signifikant.
2. Wenn eine Differenz als Regressor in einer Regressionsgleichung mathematisch aufgelöst wird, folgt daraus, dass ein Prädiktor ein positives, der andere ein gleich starkes, negatives Gewicht erhalten muss.[200] Selbstverständlich müssen die Gewichte statistisch bedeutsam sein, wenn sie das Modell stützen sollen.
3. Die Varianz, die das Kongruenzmodell aufklärt, darf nicht geringer sein als die Varianz, die ein Modell mit den „einfachen Prädiktoren" der Beurteilungsrichtungen aufklären kann. Unter einem „einfachen Modell" ist zu verstehen, dass die Fremd- und Selbstbeurteilung als zwei getrennte Regressoren („einfache Prädiktoren") in die Regressionsgleichung eingehen.
4. Es sollten keine signifikanten Terme höherer Ordnung für das spezifizierte Modell bestehen.

Ob diese Kriterien erfüllt sind und die Diskrepanzmodelle nicht zugunsten der entsprechenden einfachen Modelle verworfen werden müssen, wird für die Selbst-Aufwärts-Diskrepanz anhand folgender Schritte für jedes Erfolgskriterium überprüft (in Anlehnung an das Vorgehen von Brutus, Fleenor & Tisak, 1999).

1. Bestimmen des besten Kongruenzmodells (schrittweise Regression auf die Selbst-Fremd-Diskrepanzen der einzelnen FzFü-Faktoren)[201]
2. Errechnen der Regressionskoeffizienten für das einfache Modell und Überprüfen der Beta-Koeffizienten auf Bedeutsamkeit, ähnliche Stärke und Unterschiedlichkeit der Vorzeichen
3. Vergleich der erklärten Varianz zwischen Kongruenz- und einfachem Modell
4. Exploration von Prädiktoren höherer Ordnung (polynomiale Regression)

Parallel zum so geprüften Differenzmodell wird jeweils das entsprechende Distanzmodell betrachtet. Für diese gelten die Kriterien, die an die Beta-Koeffizienten gestellt sind, selbstverständlich nicht in dieser Weise, weil sich die Distanz nicht aus einer einfachen Subtraktion ergibt (vgl. Kapitel 5.2.1). Deshalb verfolgt eine Gegenüberstellung der erklär-

[200] $Z = b_0 + b_1(X - Y) + e$ kann aufgelöst werden nach $Z = b_0 + b_1X - b_1Y + e$.

[201] Nur diesen prognostisch nützlichen Kongruenzmodellen wird das jeweils entsprechende einfache Modell gegenüber gestellt. Dieser Schritt wurde vorangestellt, um sparsam mit Prädiktoren umzugehen, aber dennoch die zentralen Prädiktoren zu berücksichtigen. Als Regressoren wurden die Differenzen der z-standardisierten Selbst- und Aufwärtsbeurteilungen der fünf FzFü-Faktoren (E/F, B/S, IB, K, P) verwendet. Durch die Standardisierung kann verhindert werden, dass die Differenzscores durch die Aufwärtsbeurteilungen stärker beeinflusst werden als durch die Selbstbeurteilungen, da erstere eine höhere Varianz aufweisen (vgl. Kapitel 6.1.1.1).

ten Varianz zwischen Distanz- und einfachem Modell eher das Ziel praktischer Relevanz als das Ziel methodisch-mathematischer Exaktheit. Praktisch relevant ist die Gegenüberstellung deshalb, weil aus diesem Forschungsgebiet langfristig auch ein Modell entwickelt werden soll, das den Praktikern die beste Vorhersage erlaubt und damit den besten Ansatz für Entwicklungsprozesse liefert.

Da Regressionsanalysen, zumal polynomiale Regressionsanalysen, sehr powerintensive Methoden sind (Edwards & Parry, 1993), können sie für die Selbst-Abwärts- sowie die Abwärts-Aufwärts-Diskrepanz nicht durchgeführt werden, da für sie der Stichprobenumfang zu klein ist.

Der Analyseprozess für die Selbst-Aufwärts-Diskrepanz führte zu folgenden Ergebnissen:

Ad 1. Für alle Erfolgsprädiktoren mit Ausnahme von AZ2 und WOBE konnte die Selbst-Aufwärts-Diskrepanz auf mindestens einem Faktor einen signifikanten Beitrag leisten. Für AZ2 und WOBE wurden keine weiteren Analysen vorgenommen. Bereits einzelne Beurteilungsperspektiven zeigten für diese Kriterien wenig Einfluss (vgl. Tabelle 66). Alle weiteren Befunde beziehen sich also auf die verbleibenden acht Erfolgskriterien, die jeweils in statistisch bedeutsamem Ausmaß durch Diskrepanzen auf einzelnen Führungsdimensionen prognostiziert werden können. Die erklärte Varianz liegt zwischen R^2 = .16 (EEF) und R^2 = .36 (L) für die Modelle, in die die algebraischen Differenzen der z-standardisierten Selbst- und Aufwärtsbeurteilungen als Regressoren eingehen (vgl. Tabelle 78). Zum Vergleich: Die erklärte Varianz liegt zwischen R^2 = .09 (ZB) und R^2 = .55 (SAT), wenn als Diskrepanzmaß die euklidische Distanz zwischen Selbst- und Aufwärtsbeurteilern herangezogen wird.[202] Die Beta-Koeffizienten der Selbst-Fremd-Distanzen/Differenzen tragen jeweils das erwartete Vorzeichen. Diese Resultate waren aufgrund oben berichteter varianzanalytischer Ergebnisse anzunehmen.Damit ist das erste der vier Kriterien für nahezu alle Kriterien erfüllt. Im nächsten Schritt müssen die Kongruenzmodelle den einfachen Modellen gegenüber gestellt werden.

Ad 2. Die einfachen Modelle selbst erweisen sich allesamt als hochsignifikant. Tabelle 78 berichtet auch die Ergebnisse der einfachen Modelle. Die Regressionsgewichte der Selbst- und der Aufwärtsbeurteilungen erfüllen die Kriterien eines Differenzmodells indes nur für EF-WOBE und L (auf K): Nur hierfür erreichen beide Regressionsgewichte mindestens das 10%-Signifikanzniveau (was aufgrund des relativ geringen Stichprobenumfang als bedeutsam betrachtet wird), tragen die Regressionsgewichte *unterschiedliche Vor-*

[202] Im Einzelnen: Regression für AZ1 auf E/F (F[1,37] = 25.47**, *korr.* R^2 = .39). EF-AZ auf IB (F[1,36] = 14.03**, *korr.* R^2 = .26). EF-WOBE auf P (F[1,35] = 7.23*, *korr.* R^2 = .15). L auf IB (F[1,35] = 17.59**, *korr.* R^2 = .30). L auf K (F[1,35] = 11.92**, *korr.* R^2 = .22). ZB auf E/F (F[1,38] = 5.00*, *korr.* R^2 = .09). EEF auf E/F (F[1,38] = 15.73**, *korr.* R^2 = .27). EEF auf K (F[1,38] = 16.87**, *korr.* R^2 = .29). EFF auf K (F[1,38] = 34.55**, *korr.* R^2 = .46). SAT auf IB (F[1,38] = 48.86**, *korr.* R^2 = .55). SAT auf P (F[1,38] = 23.34**, *korr.* R^2 = .36).

zeichen und sind annähernd *gleich stark*. Die geforderte Vorzeichenverteilung, nicht jedoch die ähnliche Höhe der Beta-Gewichte, wird im Übrigen in allen Modellen außer in denen für EEF (auf E/F) und für SAT erfüllt.

Während bei Brutus, Fleenor und Tisak (1999) die Regressionskoeffizienten weder gleich stark sind noch unterschiedliche Vorzeichen haben (für die Selbst-Peer-Diskrepanz), sprechen die vorliegenden Daten weniger eindeutig für eine Zurückweisung des zweiten der oben aufgeführten Kriterien nach Edwards (1994a) und damit weniger eindeutig für eine Zurückweisung des Kongruenzmodells. Dennoch muss festgehalten werden, dass nahezu für alle Kriterien jeweils die Beta-Gewichte der Aufwärtsbeurteilung deutlich stärker sind als die der Selbstbeurteilung. Dies spricht generell nicht dafür, dass einfache algebraische Differenzen das beste Modell darstellen. Überprüfungen an weiteren und insbesondere größeren Stichproben sind unbedingt anzuraten.

Tabelle 78. Ergebnisse der Regressionsanalysen auf die einfachen Selbst- und Aufwärtsbeurteilungen sowie auf die Selbst-Aufwärts-Differenz, $N = 40$

	Einfaches Modell					Differenzmodell			
Kriterium	Prädiktoren		β	korr. R^2	f^2	Prädiktor	β	korr. R^2	f^2
AZ1	E/F	selbst	-.15			E/F-Differenz	-.60**	.34**	.56
		auf	.70**	.48**	1.02				
EF-AZ	IB	selbst	.24			IB-Differenz	.48**	.20**	.29
		auf	-.46**	.21**	.33				
EF-WOBE	P	selbst	.36*			P-Differenz	.52**	.25**	.37
		auf	-.30$^+$	.23**	.37				
L	IB	selbst	-.26$^+$			IB-Differenz	-.54**	.28**	.42
		auf	.55**	.31**	.54				
L	K	selbst	-.42**			K-Differenz	-.62**	.36**	.61
		auf	.60**	.38**	.70				
ZB	E/F	selbst	.23			E/F-Differenz	.49**	.22**	.32
		auf	-.54**	.27**	.44				
EEF	E/F	selbst	.04			E/F-Differenz	-.43**	.16**	.22
		auf	.74**	.53**	1.25				
EEF	K	selbst	-.24$^+$			K-Differenz	-.54**	.28**	.42
		auf	.78**	.50**	1.12				
EFF	K	selbst	-.20$^+$			K-Differenz	-.54**	.27**	.40
		auf	.80**	.54**	1.31				
SAT	IB	selbst	.04			IB-Differenz	-.48**	.21**	.30
		auf	.76**	.56**	1.40				
SAT	P	selbst	.06			P-Differenz	-.56**	.29**	.45
		auf	.78**	.56**	1.41				

Anmerkungen. ** $p < .01$, * $p < .05$, $^+p < .10$.

Ad 3. Bei einem Vergleich der Varianzaufklärung zwischen den Differenz- und einfachen Modellen kann für die vorliegende Studie festgestellt werden, dass das einfache Modell tendenziell größere Effekte zeigt. Für Leistung und EF-AZ liegt der Vorsprung der einfachen Modelle bei lediglich ΔR^2 = .01 bis ΔR^2 = .03. Große Unterschiede ergeben sich jedoch für EEF, EFF und SAT (max. ΔR^2 = .37).

Für die beiden Modelle, für die auch die ersten beiden Kriterien angenommen werden konnten (EF-WOBE auf P und L auf K), kann auch das dritte Kriterium nach Edwards als ausreichend erfüllt gelten, da die Varianzaufklärung der Differenzmodelle nicht unter der der einfachen Modelle liegt. Während für diese beiden Kriterien das Differenzmodell gehalten werden kann, muss es für die anderen Kriterien zugunsten der einfachen Beurteilungsperspektiven verworfen werden. Für diese stellen nämlich allein die Aufwärtsbeurteilungen die mit Abstand besten Prädiktoren dar. Ihr Gewicht entspricht mit unbedeutenden Abweichungen den bivariaten Korrelationskoeffizienten zwischen der Aufwärtsbeurteilung und dem entsprechenden Kriterium (vgl. Anhang M). Allenfalls die Korrelation zwischen Klarheit (Aufwärtsbeurteilung) und EEF liegt mit r = .69 merkbar unter dem Regressionsgewicht, so dass hier die Selbstbeurteilung nicht außer Acht zu lassen ist. Generell gilt, dass einfache Beurteilungen aus verschiedenen Perspektiven und auf verschiedenen Dimensionen inkrementelle Beiträge zu noch größerer Erfolgsaufklärung liefern können (vgl. Kapitel 6.4.2, z.B. Tabelle 67).

Geringer fallen die Unterschiede zwischen den einfachen Modellen und den entsprechenden Diskrepanzmodellen aus, die auf dem Distanzmaß anstatt der Differenz der z-standardisierten Beurteilungsrichtungen beruhen. Auch dann erklären die einfachen Modelle zwar für sieben der acht Kriterien tendenziell mehr Varianz, große Differenzen (ΔR^2 = .21 bzw. ΔR^2 = .26) zeigen sich indes nur für EEF auf K bzw. auf E/F (vgl. Fußnote 202). Auch das spricht dagegen, die einfachen algebraischen Differenzen uneingeschränkt vorzuziehen. Dennoch muss die Rolle der Diskrepanzen in weiterführenden Studien besser geklärt werden. Das dritte der angeführten Kriterien nach Edwards (1994a) trifft für die meisten Kriterien eher nicht zu.

Ad 4. Ob zu den jeweils favorisierten Modellen Prädiktoren höherer Ordnung hinzugenommen werden sollten, wurde durch polynomiale Regressionsanalysen exploriert. Das bedeutet, dass neben den einfachen Regressoren (Selbst- und Aufwärtsbeurteilung) auch deren Quadrate sowie deren Produkt („Kreuzprodukt“) in die Regressionsgleichung aufgenommen wurden. Ein signifikantes Interaktionsgewicht wäre ein Beleg dafür, dass keine lineare Beziehung zwischen den einzelnen Beurteilungsrichtungen und Erfolg besteht, sondern dass die Interaktion der beiden Beurteilungsrichtungen den Erfolg prognostiziert.

Bedeutsame quadrierte Prädiktoren sind aufgrund der bis dato berichteten Ergebnisse nicht für die Aufwärtsbeurteilung zu erwarten. Die Selbstbeurteilung könnte mit den Erfolgskriterien aber auch deshalb so niedrige Pearson-Koeffizienten aufweisen, weil sie nicht linear zueinander in Beziehung stehen. Plausibel wäre nämlich auch, dass die

Erfolgskriterien erst ab einem bestimmten Grad an Selbstbeurteilung mit dieser kovariieren. Das wäre anzunehmen, wenn man von einem „Grundstock“ an Führungskompetenz ausgeht, den sich jede Führungskraft zuschreibt, ob sie ihm gerecht wird oder nicht. Möglicherweise zeigt sich in diesem unteren Bereich der Führungsbeurteilung sogar eine umgekehrte Relation insofern, dass die Führungsbeurteilung negativ mit dem Führungserfolg zusammenhängt. Das wäre plausibel damit zu begründen, dass ein solches „minimales Führen“ negative Konsequenzen auf Seiten der Mitarbeiter hervorruft (ähnlich wie passive Führung, operationalisiert durch MBP oder LF im MLQ). Eine solche parabolische Beziehung kann durch polynomiale Regressionsanalysen aufgeklärt werden. Ein starker Beta-Koeffizient für die quadrierte Selbstbeurteilung gäbe einen wichtigen Hinweis darauf.

Um signifikante Effekte aufdecken zu können, sind dafür jedoch entweder große Stichprobenumfänge oder starke Effekte notwendig. Dies liegt daran, dass das Risiko der Multikollinearität erhöht ist, da die einfachen Prädiktoren gleichzeitig die Faktoren der Interaktions- und exponentiellen Terme darstellen. Die Multikollinearität stellt sich für die für diese Fragestellung angebrachten Regressionsanalysen als unhaltbar heraus (Toleranzwerte < .10). Um die Multikollinearität zu reduzieren, wurden auch die polynomialen Regressionsanalysen auf *z-standardisierte* Regressoren durchgeführt (vgl. Bobko, 1995).

Für kein Kriterium kann ein statistisch bedeutsamer Prädiktor höherer Ordnung ermittelt werden. Das bedeutet, dass für EF-WOBE und L, für die bereits die ersten drei Kriterien nach Edwards (1994a) gehalten werden können, das jeweilige Kongruenzmodell endgültig nicht verworfen werden muss. Für alle anderen Erfolgskriterien ist Hypothese 39) jedoch für die vorliegende Stichprobe abzulehnen. Die Auswertungsergebnisse verweisen weiter nicht durchgängig darauf, dass die einfachen Selbst- und Aufwärtsbeurteilungen tatsächlich schlechtere Erfolgsprognosen abgeben als deren Diskrepanz (Hypothese 41).

Festzuhalten ist:

- Nach den Kriterien von Edwards (1994a) können Selbst-Aufwärts-Differenzmodelle allenfalls für EF-WOBE und L gehalten werden. Für alle anderen Kriterien müssen sie eher zugunsten der einfachen Modelle aufgegeben werden. Es ist die Aufwärtsbeurteilung, die die besten Vorhersagen leistet.
- Möglicherweise sind nur bestimmte Kriterien durch bestimmte Beurteilungsdiskrepanzen vorhersagbar.
- Die Distanzmodelle sind möglicherweise besser zur Erfolgsprognose geeignet als die Differenzmodelle.
- Die Varianzaufklärung der Regressionsanalysen ist – nicht überraschend – deutlich höher als die der Varianzanalysen. Dies ist bei der Interpretation und bei einem Vergleich mit Befunden anderer Studien, die in der Regel auf Varianzanalysen basieren, zu berücksichtigen.

6.5.3 Leistet die interperspektivische Beurteilungsübereinstimmung inkrementelle Beiträge zur Erfolgsaufklärung?

Streng genommen wurde in Kapitel 6.5.2 der Frage nachgegangen, ob wirklich die Kongruenz den kritischen prognostischen Beitrag leistet und nicht die in die Kongruenzberechnung eingehenden einfachen Beurteilungsperspektiven. Die Antwort ist uneindeutig, für die meisten Kriterien fällt sie eher negativ aus.

Denkbar wäre allerdings, dass es darum geht, ein Diskrepanzmaß *zusätzlich* zu anderen Prädiktoren zu nutzen. Da bei einer multiperspektivischen Führungsbeurteilung und erst recht bei einer 360°-Beurteilung häufig mehr als zwei Perspektiven erfasst werden, ist zu untersuchen, ob sich die Vorhersagekraft erstens durch die gleichzeitige Betrachtung verschiedener Diskrepanzen und zweitens durch die gleichzeitige Betrachtung einfacher Beurteilungen *und* Diskrepanzen steigern lässt. Gerade Letzteres könnte Hinweise dafür erbringen, dass die Diskrepanzen im Vergleich zu den einfachen Perspektiven nicht nur redundante Informationen enthalten. Die beiden Fragen wurden regressionsanalytisch erforscht. Auch die Führungsqualität an sich kann berücksichtigt werden. Genauer wurden folgende Hypothesen aufgestellt:

42) Mehrere Diskrepanzen können mehr Erfolgsvarianz erklären als einzelne Diskrepanzen.

43) Das beste Diskrepanzmodell erklärt mehr Erfolgsvarianz als das beste einfache Modell.

44) Die Diskrepanz zwischen zwei Perspektiven leistet (über die einzelnen Beurteilungsperspektiven hinaus) einen inkrementellen Beitrag zur Erfolgsaufklärung.

45) Der Erfolg fällt für Übereinstimmer mit hoher Führungsqualität höher aus als für Unter- und Überschätzer mit hoher Führungsqualität.

46) Der Erfolg fällt für Übereinstimmer mit geringer Führungsqualität niedriger aus als für Unter- und Überschätzer mit hoher Führungsqualität.

Für AZ1, L und EFF können Modelle erstellt werden, die mit zwei mittleren Diskrepanzen (über E/F, B/S, IB, K und P) mehr Varianz erklären als mit einer Diskrepanz ($p < .05$). Für SAT kann zwar kein statistisch signifikanter Unterschied erzielt werden, dennoch ist die Hinzunahme einer zweiten Diskrepanz praktisch relevant, da sich ein starker Unterschied in der Effektgröße ergibt ($\Delta f^2 = .37$). Diese vier Modelle, die Hypothese 42) entsprechen, sind in Tabelle 79 berichtet.

Tabelle 79. Inkrementelle Validität verschiedener Diskrepanzen, $N = 24$

	Modell 1				Modell 2						
	B	*SE B*	*β*	*p*	*B*	*SE B*	*β*	*p*	korr. R^2	*p* (*ΔF*)	Δf^2
Kriterium: AZ1											
1. Ab-Auf	-.75	.16	-.70	< .01	-.79	.16	-.70	< .01	.42		
2. S-Ab					-.36	.14	-.37	.02	.58	.02	.58
Kriterium: L											
1. S-Ab	-.59	.18	-.57	< .01	-.64	.17	-.62	< .01	.30		
2. Ab-Auf					-.42	.19	-.35	.04	.40	.04	.33
Kriterium: EFF											
1. Ab-Auf	-.87	.14	-.80	< .01	-.60	.18	-.55	< .01	.62		
2. S-Ab					-.35	.16	-.35	< .05	.67	< .05	.59
Kriterium: SAT											
1. S-Auf	-.87	.15	-.79	< .01	-.64	.19	-.58	< .01	.60		
2. Ab-Auf					-.37	.21	-.30	< .10	.63	< .10	.37

Anmerkungen. *p*(Δ*F*): Änderung in Signifikanz von *F*, Ab/Auf/S: mittlere Abwärts-/Aufwärts-/Selbstbeurteilung (über fünf FzFü-Faktoren E/F, B/S, IB, K, P).

Es fällt auf, dass jeweils eine Selbst-Fremd-Diskrepanz mit der Abwärts-Aufwärts-Diskrepanz kombiniert wird und niemals zwei Selbst-Fremd-Diskrepanzen gemeinsam das beste Modell darstellen. Dies kann als weiterer Hinweis dafür gewertet werden, dass unterschiedliche, sich ergänzende Mechanismen zum Zusammenhang zwischen Beurteilerdiskrepanz und Erfolg beitragen. Dass zum einen jeweils eine Selbst-Fremd-Diskrepanz aufgenommen wird, spricht dafür, dass die Selbstüberschätzung der Führungskräfte eine Rolle im Entstehungszusammenhang des Führungserfolgs spielt. Dass zum andern jeweils die Abwärts-Aufwärts-Diskrepanz zusätzlich Prognosekraft besitzt, spricht dafür, dass ein weiterer Erklärungsmechanismus existiert, der nicht mit der Selbstüberschätzung zusammenhängt. Plausibel ist, dass eine eher inkonsistente oder undurchsichtige Art von Führung nicht nur die Abwärts-Aufwärts-Übereinstimmung, sondern gleichzeitig den Erfolg reduziert. Mangelnde Self-Awareness und ein Zick-Zack-Führungskurs addieren sich also in ihrem negativen Einfluss auf Erfolg.

Auch wenn durch die Hinzunahme einer zweiten Diskrepanz die Varianzaufklärung einiger Kriterien gegenüber einer einzelnen Diskrepanz erhöht werden kann, wären die Modelle *einfacher* Prädiktoren dann vorzuziehen, wenn letztere mehr Varianz erklärten. Zieht man einen Vergleich zwischen dem besten Diskrepanz- und dem besten einfachen Modell (vgl. Tabelle 66 für AZ1 und Tabelle 67 für L, EFF und SAT), dann stellt sich heraus, dass SAT besser durch die Beurteilungen auf den validesten Einzeldimensionen prognostiziert werden kann, AZ1, L und EFF hingegen nicht. Nur diese entsprechen den Annahmen von Hypothese 43).

Streng genommen müssten für einen solchen Vergleich auch die Differenzmodelle – nicht nur die einfachen Modelle – für einzelne Führungsdimensionen erarbeitet werden. Dies ist für weitere Studien auch dringend anzuraten. Aufgrund des Stichprobenumfangs können in der vorliegenden Arbeit jedoch keine stabilen Modelle generiert werden.

Denkbar ist außerdem, dass die Diskrepanzen über die einfachen Prädiktoren hinaus Validität aufweisen. Dieser Annahme entsprechend ergibt sich für vier Kriterien durch die gleichzeitige Berücksichtigung der validesten Einzeldimensionen (Tabelle 66 und Tabelle 67) und Diskrepanzen (Tabelle 79 bzw. valideste einzelne Diskrepanz, Ergebnis schrittweiser Regression) inkrementelle Validität. Das beste Diskrepanzmodell klärt für L, EEF, EFF uns SAT über das beste einfache Modell hinaus zusätzlich Varianz auf ($p[\Delta R^2] < .05$). So können für L 57%, für EEF 84%, für EFF 70% und für SAT 82% der Varianz erklärt werden (korrigiertes R^2).[203]

Weiter geht der Zugewinn an Erklärungskraft auch dann nicht verloren, wenn man auf die Analyseebene der einzelnen Führungs*dimensionen* geht. So zeigt sich in blockweisen Regressionsanalysen, dass beispielsweise die Selbst-Aufwärts-Distanz für die Vorhersage von EFF durch K ($\Delta R^2 = .05$, $\Delta F_{[1, 36]} = 4.16$, $p < .05$) sowie von SAT durch IB ($\Delta R^2 = .07$, $\Delta F_{[1, 36]} = 7.17$, $p = .01$) einen statistisch signifikanten Beitrag leistet. Doch auch für EF-AZ, L und ZB kann die Distanz zwischen 2 und 6% mehr Varianz erklären als die einfachen Prädiktoren. Dass auch auf dieser Analyseebene inkrementelle Validität erreichbar ist, gibt einen weiteren Hinweis für den Nutzen von Diskrepanzmaßen und wird an dieser Stelle nicht im Einzelnen inhaltlich interpretiert. Hypothese 43) kann also vorerst für verschiedene Kriterien also angenommen werden.

Dass verschiedene Diskrepanzen gemeinsam bzw. Diskrepanzen und einfache Prädiktoren gemeinsam mehr Varianz aufklären als einzelne Diskrepanzen bzw. einfache Modelle, spricht dafür, Diskrepanzen auch in der Praxis weiterhin als Erfolgsindikatoren heranzuziehen. Gewarnt wird allerdings davor, sie als alleinige Indikatoren zu betrachten. Gerade die inkrementelle Validität der Diskrepanzen über die einfachen Prädiktoren hinaus ist mehr als plausibel. Es gibt keinen vernünftigen Grund, der dafür spräche, dass eine übereinstimmend mangelhafte Führungsbeurteilung mit ähnlich starkem Erfolg einhergehen sollte wie übereinstimmend hervorragende Führungsbeurteilung. Würde die Kongruenz aber lediglich als substituierende – nicht als ergänzende – Alternative für die einfachen Prädiktoren gelten, wäre ausgeschlossen, dass die *Qualität* der Führung eine erfolgsunterstützende Rolle spielt. Letztere spiegelt sich vielmehr in den einfachen Beurteilungen wider, so dass diese unbedingt zu berücksichtigen sind. Dies zu unterlassen, wäre gleichsam ein Angriff auf die komplette Geschichte der Führungsforschung.

Tabelle 80 illustriert, wie sich die Richtung der Selbst-Fremd-Abweichung und die Führungsqualität auf den Erfolg auswirken. Da der Stichprobenumfang sehr gering ist, wurden keine Mittelwertsvergleiche auf statistische Signifikanz vorgenommen. Es wird jedoch deutlich, dass nicht nur die Übereinstimmer erfolgreicher beschrieben werden als die Über-

[203] Für die anderen Kriterien können die Diskrepanzen der Gesamtskala auch auf dem 10%-Niveau keinen zusätzlichen Beitrag über das beste einfache Modell hinaus liefern.

schätzer, sondern dass außerdem fast ausnahmslos diejenigen Führungskräfte den höheren Erfolg verzeichnen, die jeweils überdurchschnittlich (vs. unterdurchschnittlich) transformational führen. Im Sinne von Hypothese 45) zeigen dann erst recht diejenigen Führungskräfte, die sowohl als Übereinstimmer als auch als hochtransformational kategorisiert sind, die höchsten Erfolge.

Tabelle 80. Erfolgskennwerte (Skalenmittelwert) in Abhängigkeit von der Selbst-Fremd-Diskrepanz *und* der Führungsqualität (transformationale Führung)

	Selbst-Aufwärts-Vergleich				Selbst-Abwärts-Vergleich					
	Übereinstimmer		Überschätzer		Übereinstimmer		Überschätzer		Unterschätzer	
	A (N = 4)	B (N = 5)	A (N =11)	B (N = 8)	A (N = 4)	B (N = 4)	A (N = 4)	B (N = 2)	A (N = 2)	B (N = 5)
AZ1	5.67	5.80	4.77	4.95	5.08	5.67	4.87	4.84	4.84	5.58
AZ2	5.88	5.70	5.42	5.55	5.31	5.72	5.81	5.56	5.00	5.20
EF-AZ	.25	.00	.59	.38	-.25	.25	1.00	-.25	.50	.00
WOBE	4.63	5.10	4.82	4.69	4.63	5.06	5.06	4.50	4.25	4.70
EF-WOBE	-.13	.20	.68	.50	.25	.00	.50	.25	.00	.40
L	5.06	5.85	5.05	5.34	5.00	5.19	4.88	5.75	5.25	6.20
ZB	.88	.30	1.27	.69	.75	.50	1.63	1.25	1.00	.40
EEF	3.33	3.57	2.68	3.60	2.75	3.46	2.62	3.67	3.00	3.77
EFF	3.37	3.85	2.80	3.75	3.06	3.75	2.70	4.25	2.88	3.80
SAT	3.38	4.00	2.82	3.75	3.50	4.00	2.31	4.00	3.00	4.00

Anmerkungen. A: transformationale Führung unter Median, B: transformationale Führung über Median. Für die Selbst-Aufwärts-Diskrepanz sind keine Unterschätzer dargestellt, da lediglich für drei Führungskräfte Daten zur transformationalen Führung vorliegen, die allesamt als überdurchschnittlich zu kategorisieren sind.

Unerwartet ist das Ergebnis dahingehend, dass die Übereinstimmer auf niedrigem Qualitätsniveau auf manchen Erfolgskriterien (insbesondere Arbeitszufriedenheit) höhere Werte erlangen als die Über- und Unterschätzer, deren Führung als überdurchschnittlich positiv bewertet wurde (Hypothese 46). Angenommen wurde, dass Über- und Unterschätzer, denen überdurchschnittlich gute Führung bescheinigt wird, erfolgreicher sind. Letztlich geht es also um die Frage, ob die Übereinstimmung oder das Führungsniveau mehr Auswirkung auf den Führungserfolg hat. Es wäre durchaus denkbar, dass die Antwort darauf für unterschiedliche Kriterien unterschiedlich ausfällt. Möglicherweise gehört die Arbeitszufriedenheit zu den Kriterien, die weniger von einem hohen Führungsniveau, als vielmehr von ähnlichen Sichtweisen begünstigt werden.

Analog fallen die Ergebnisse dann aus, wenn die Führungsqualität nicht durch die transformationale Führung, sondern durch den FzFü operationalisiert ist (jeweils diejenige Fremdperspektive, die nicht in die Selbst-Fremd-Diskrepanz involviert ist) (vgl. Anhang U).

Festzuhalten ist:

- Die gleichzeitige Betrachtung verschiedener Diskrepanzen (Selbst-Aufwärts, Selbst-Abwärts, Abwärts-Aufwärts) trägt bei wichtigen Kriterien zur Validitätssteigerung bei.
- Das beste Diskrepanzmodell erklärt in der Regel mehr Erfolgsvarianz als das beste einfache Modell.
- Die gleichzeitige Betrachtung der validesten einfachen Prädiktoren und der validesten Diskrepanzen trägt bei wichtigen Kriterien zur Validitätssteigerung bei. Auf diesem Weg können bis zu 82% der Varianz erklärt werden (SAT). Dies spricht dafür, dass die Kongruenz mehr Informationen enthält als einfache Beurteilungen.
- Mittelwertsvergleiche sprechen dafür, dass die höchsten Führungserfolge bei Vorliegen von Selbst-Fremd-Übereinstimmung *und* hoher Führungsqualität zu verzeichnen sind. Für die Mehrzahl der Kriterien scheint die Führungsqualität ausschlaggebender zu sein als die Übereinstimmung.

6.5.4 Bedeutet Erfolg automatisch auch Übereinstimmung?

Die Praxis stellt häufig die Frage, wodurch sich erfolgreiche Führungskräfte im Unterschied zu weniger erfolgreichen Führungskräften auszeichnen. Die Beantwortung dieser Frage fordert einen – im Vergleich zum bisher berichteten Vorgehen – umgekehrten Analyseweg, quasi von einer Kategorisierung anhand der Erfolgskennwerte ausgehend. Im Zentrum steht die Frage, ob erfolgreiche Führungskräfte durch interperspektivische Übereinstimmung charakterisiert sind. Abschließend werden aber auch personale und organisationsbezogene Variablen auf ihre Erfolgsindikation geprüft:

47) Berichten Mitarbeiter überdurchschnittlichen (vs. unterdurchschnittlichen) Erfolg, sind die Diskrepanzen zwischen den Beurteilungsperspektiven geringer.

48) Berichten Mitarbeiter überdurchschnittlichen (vs. unterdurchschnittlichen) Erfolg, sind die Korrelationen zwischen den Beurteilungsperspektiven höher.

49) Berichten Mitarbeiter überdurchschnittlichen Erfolg, dann gilt im Vergleich zu Mitarbeitern mit unterdurchschnittlichem Erfolg:

a. Die Mitarbeiter sind älter.
b. Ihre Führungskraft ist älter.
c. Die Abteilung ist kleiner.
d. Die Beteiligungsquote an der Befragung ist höher.
e. Die Dauer der Zugehörigkeit zur Organisation und zum Team ist länger.

Entsprechend der Erwartungen konnte bestätigt werden, dass die Aufwärtsbeurteilungen (Mittel über E/F, B/S, IB, K und P) derjenigen Mitarbeitergruppen, die überdurchschnittlich (vs. unterdurchschnittlich) hohe Erfolge auf den einzelnen Kriterien aufweisen (Mediansplit), sowohl weniger stark von der Selbstbeurteilung ihrer Führungskraft ($\bar{d}$ = .81; vgl. Abbildung 36) als auch weniger stark von der Abwärtsbeurteilung abweichen ($\bar{d}$ = 1.26). Auf faktorbezogener Ebene zeigt sich ebenso, dass sich überdurchschnittlich hoher Erfolg insbesondere in geringen Selbst-Aufwärts- und Abwärts-Aufwärts-Abweichungen ausdrückt. Subjektiv hohe *Leistung* der Mitarbeiter ist zudem durch überzufällig geringe *Selbst-Abwärts*-Differenzen auf IB und K gekennzeichnet (IB: $t(37) = 4.30$, $p < .01$, $d = 1.12$ bzw. K: $t(37) = 2.94$, $p < .01$, $d = 1.06$). Mittlere Effekte in angenommener Richtung (Hypothese 47) zeigen sich für viele Differenzen auf anderen Dimensionen sowie für weitere Erfolgskriterien.

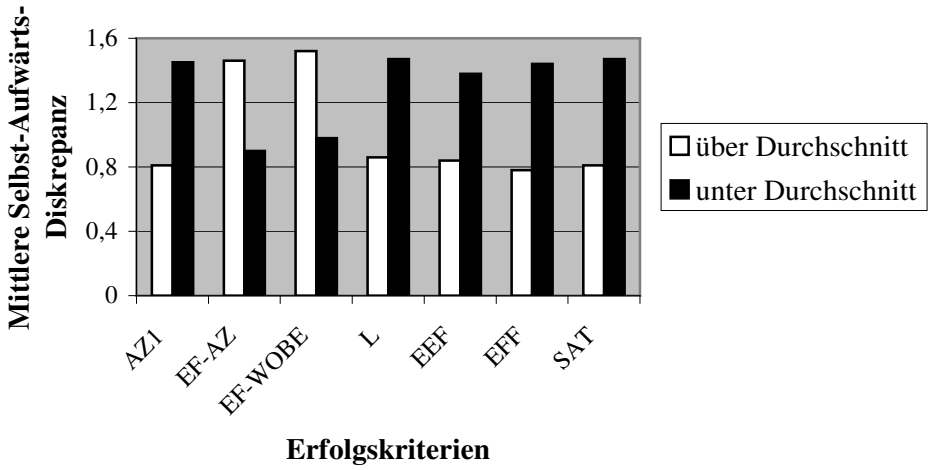


Abbildung 36. Mittlere Selbst-Aufwärts-Diskrepanz für über- und unterdurchschnittlich erfolgreiche Mitarbeiterteams bezüglich verschiedener Kriterien

Anmerkung. In der Abbildung sind nur die Kriterien dargestellt, für die große Effekte ($d > .80$) bestehen.

Erstaunlich ist nun, dass sich die überdurchschnittlich erfolgreiche von der unterdurchschnittlich erfolgreichen Teilstichprobe (Mediansplit der einzelnen Erfolgsvariablen) weder hinsichtlich der Selbst-Aufwärts- oder Selbst-Abwärts- noch hinsichtlich der Abwärts-Aufwärts-*Korrelation* konsistent in erwarteter Richtung unterscheidet. Dies gilt sowohl für die mittleren Korrelationen[204] als auch für die Korrelationen auf den einzelnen Führungsdimensionen. Sogar wenn man ausschließlich starke Korrelationsunterschiede betrachtet ($q \geq .50$),[205] fallen diese nicht wie erwartet überwiegend zugunsten der erfolgreichen Gruppe aus (Hypothese 48). Hingegen wirkt es eher willkürlich verteilt, ob die korrelative Übereinstimmung der Perspektiven in der erfolgreichen oder weniger erfolgreichen Teilstichprobe höher ist (vgl. Anhang V).

[204] über Items von E/F, B/S, IB, K und P.
[205] Eine vernachlässigbare Anzahl an Korrelationsunterschieden erreicht statistische Signifikanz ($p < .05$).

Es wurde kontrolliert, ob sich die überdurchschnittlich erfolgreiche Teilstichprobe von der unterdurchschnittlichen Teilstichprobe hinsichtlich verschiedener demografischer oder organisationsbezogener Variablen unterscheidet. Dies konnte für keine der untersuchten Variablen in statistisch bedeutsamer Stärke gefunden werden (Abteilungsgröße, Beteiligungsquote, Alter des Ratees, durchschnittliches Alter der Mitarbeiter, Dauer der Zugehörigkeit zu Organisation und Team [Ratee und Rater]). Auch multivariate Analysen (Regressionsanalysen), die diese Kriterien gleichzeitig betrachteten, erbrachten lediglich für SAT eine geringe Varianzaufklärung von 5%. Für alle anderen Kriterien konnten nicht mehr als 2% aufgeklärt werden, was den verschiedenen Annahmen von Hypothese 49) entgegen spricht.

Zusammenfassend spricht diese quasi umgekehrte Analyse – ausgehend von einer Erfolgsdichotomisierung der Stichprobe – also dafür, dass

- erfolgreiche Führung durch geringe Niveau-Unterschiede gekennzeichnet ist und dass
- die korrelative Übereinstimmung keinen hohen prognostischen Wert besitzt.
- Weiter konnten keine personalen oder organisationsbezogenen Charakteristika von erfolgreicher Führung aufgedeckt werden, was die Kovariation der Niveau-Unterschiede mit Erfolg zusätzlich aufwertet.

6.5.5 Spielt die interperspektivische Übereinstimmung eine moderierende Rolle für den Zusammenhang zwischen Führung und Erfolg?

Wenn die Rolle der Diskrepanz im gesamten Entstehungsgefüge von Führungserfolg diskutiert wird, dann schließt sich an oben berichtete Befunde[206] und insbesondere an die Ergebnisse aus Tabelle 80 die Frage an, ob die Diskrepanz möglicherweise eine moderierende Rolle auf den Zusammenhang zwischen Führung und Erfolg einnimmt. Es wurde überprüft, ob sich die prädiktive Validität von Führung in Abhängigkeit von der Diskrepanz unterscheidet:

50) Liegt Übereinstimmung zwischen den Perspektiven vor (vs. Unter- oder Überschätzung), fällt der Zusammenhang zwischen Führung und Erfolg höher aus.

[206] Es wurde gezeigt, dass die Diskrepanz in korrelativem Zusammenhang zu verschiedenen Erfolgsmaßen steht und für manche Erfolgskriterien inkrementelle Validität über die Einzelbeurteilungen hinaus aufweist.

Führung wurde jeweils durch die Ratings aus derjenigen Perspektive operationalisiert, die nicht in die Berechnung der Diskrepanz einging. Die genauen Ergebnisse hierfür sowie für die Führungsbeurteilungen aus allen anderen Perspektiven sind in Anhang W nachzulesen. Angenommen wurde, dass sich die Führung insbesondere dann in Erfolg niederschlägt, wenn Übereinstimmung zwischen den Beurteilungen verschiedener Perspektiven besteht.

Ob die Richtung der *Selbst-Aufwärts*-Diskrepanz eine moderierende Wirkung auf die prädiktive Validität der Abwärtsbeurteilungen (FzFü) hat, konnte nur anhand der Übereinstimmer und Überschätzer überprüft werden, da lediglich zwei Unterschätzer vorlagen (vgl. Anhang W, Tabellen W-1). Entgegen der Annahmen stellte sich allerdings für die Überschätzer ein höherer Zusammenhang zwischen der Abwärtsbeurteilung und den Erfolgsvariablen heraus als für die Übereinstimmer ($\bar{r}$ = .49 vs. $\bar{r}$ = -.04, q = .58)[207]. Auch wenn man die einzelnen Erfolgskriterien betrachtet, ist für die *Übereinstimmer* die Vorhersage durch die Abwärtsbeurteilungen lediglich für das Kriterium Leistung besser ($\bar{r}$ = .33 vs. $\bar{r}$ = .09, q = .25).

Möglicherweise können Abwärtsbeurteiler die *sich überschätzenden* Führungskräfte deshalb besser bewerten, weil sie häufiger beobachtbar sind. So ist denkbar, dass die Überschätzer aufgrund eines erhöhten Führungs-Selbstbewusstseins viel Führungsverhalten „preisgeben“ oder dass sie besonders häufig über Führung sprechen. Dazu passt, dass die Abwärtsbeurteilungen für die Überschätzer stärker streuen als die Abwärtsbeurteilungen für die Übereinstimmer (*SD* = .55 vs. *SD* = .40; *N* = 8 vs. *N* = 14). Rothstein (1990) stellte die Varianz von Ratings als wichtigen Faktor für deren Reliabilität heraus.

Ebenfalls entgegen der Erwartungen ist mit soeben berichteten Ergebnissen konsistent, dass die prädiktive Validität der Abwärtsbeurteilungen auch für diejenigen Führungskräfte am höchsten ist, deren Selbstbeurteilung unabhängig von der Beurteilungsrichtung am stärksten von den Aufwärtsbeurteilungen abweicht (Terzilsplit D-Statistik: $\bar{r}$ = -.39 vs. $\bar{r}$ = .08 vs. $\bar{r}$ = .36; *N* = 5, 10, 9). Auch hier streuen die Abwärtsbeurteilungen für die Hoch-Diskrepanten am stärksten (*SD* = .66 vs. *SD* = .43 für mittelmäßige sowie für nicht/gering Diskrepante).[208]

Betrachtet man die *Selbst-Abwärts*-Abweichung, ist die prädiktive Validität der Aufwärtsbeurteilungen auf den Dimensionen CR, MBA, MBP und LF erwartungsgemäß entweder für die Unterschätzer oder die Übereinstimmer am höchsten, auch wenn der

[207] Mittlere Abwärts-Erfolgs-Korrelation über folgende Erfolgskriterien hinweg: AZ1, AZ2, WOBE, L, EEF, EFF, SAT.

[208] Die Ergebnisse sind nicht konsistent vereinbar mit dem Befund von Atwater und Yammarino (1992), die für die Übereinstimmer den höchsten Zusammenhang zwischen Führungs- und Erfolgsvariablen (Vorgesetztenbeurteilung der vergangenen Jahre und Empfehlungen zur Beförderung) fanden. Doch da in der zitierten Studie die Fremdratings zur transformationalen Führung sowohl in die Korrelation mit den Erfolgsvariablen als auch in die kategoriale Einordnung der Führungskräfte als Übereinstimmer eingehen, stellen die Analysen der vorliegenden Studien eine härtere Überprüfung dar.

Unterschied zur prädiktiven Validität für die Überschätzer auf den Dimensionen CR und LF nur unbedeutend ist. Für die FzFü-Skala sowie die transformationale Führung (jeweils Mittelwert) tritt ein solcher moderierender Effekt nur für die MLQ-externen Erfolgskriterien ein, nicht jedoch für die MLQ-internen Erfolgskriterien EEF, EFF und SAT. Hierfür weisen die Überschätzer die höchsten Aufwärts-Erfolgs-Korrelationen auf (vgl. Anhang W, Tabellen W-2). Da Überschätzer nicht nur durch erhöhte Selbstbeurteilungen, sondern auch durch niedrigere Fremdbeurteilungen charakterisiert sind, könnte die prädiktive Validität auch deshalb verhältnismäßig hoch sein, weil die Aufwärtsbeurteilungen verhältnismäßig „ungeschönt" ausfallen und daher am ehesten dem tatsächlich erlebten Führungsverhalten entsprechen. Nur dieses kann natürlich Erfolge hervorrufen, nicht ein aus Angst oder anderen Gründen mild dargestelltes Führungsverhalten.
Ob das Ausmaß der *Abwärts-Aufwärts*-Abweichung einen moderierenden Effekt auf die prädiktive Validität der Selbstbeurteilung hat, wird mit Hilfe eines Terzilsplits der Abwärts-Aufwärts-Distanz untersucht (vgl. Anhang W, Tabellen W-3). Die Richtung der Abweichung bleibt in diesem Fall unberücksichtigt, da unabhängig davon eine geringe Abwärts-Aufwärts-Übereinstimmung als Zick-Zack-Kurs der Führung interpretiert wird. Es zeigt sich, dass die Selbstbeurteilung mit den Erfolgsmaßen durchschnittlich dann am höchsten korreliert, wenn Aufwärts- und Abwärtsratings am wenigsten voneinander abweichen. Allerdings unterscheidet sich die durchschnittliche prädiktive Validität der Selbstbeurteilung nicht bedeutsam zwischen diesen Niedrig-Diskrepanten und den Hoch-Diskrepanten ($\bar{r}$ = .30 vs. $\bar{r}$ = .27, q = .03). Die Validität für Letztere könnte auch deshalb verhältnismäßig hoch sein, weil die Selbstbeurteilungen für die Hoch-Diskrepanten am stärksten streuen (SD = .62 vs. SD = .29 bzw. SD = .48 für mittelmäßige bzw. nicht/gering Diskrepante).

Die soeben dargestellten Ergebnisse der Moderatoranalysen basieren auf Kategorisierungen, die sich aus den interperspektivischen Abweichungen der *Gesamt*-FzFü-Skala ergaben. Kategorisiert man gemäß der Abweichungen auf *einzelnen* Führungsdimensionen, dann erhält man keine grundlegend anderen Ergebnisse. Deshalb wird auf deren Darstellung verzichtet. Die Auswertungen ergeben über die Perspektiven-Vergleiche und Erfolgskriterien hinweg keine konsistente Auskunft über die Annahme von Hypothese 50). Generell sprechen die Ergebnisse zumindest gegen eine einfache moderierende Wirkung in dem Sinne, dass geringe Niveau-Unterschiede zwischen den Perspektiven die prädiktive Validität von Führung erhöhen würden.

Zusammenfassend ist zur Moderatorrolle der Diskrepanz festzuhalten:

- Es ist keine konsistent interpretierbare moderierende Rolle der interperspektivischen Diskrepanz auf die prädiktive Validität von Führung erkennbar.
- Die Selbst-Aufwärts-Diskrepanz moderiert die prädiktive Validität der Abwärtsbeurteilungen von Führung in unerwarteter Richtung: Am meisten Erfolgsvarianz kann für die *Überschätzer* aufgeklärt werden.

- Für die Selbst-Abwärts- und die Abwärts-Aufwärts-Diskrepanz lässt sich keine ähnlich eindeutige Wirkung festhalten. Vielmehr variiert der Einfluss über die Erfolgskriterien hinweg.

Die Ergebnisse dieser Studie zeigen auf, wie problematisch die (alleinige) Verwendung der Kongruenz-Kategorien ist. Forscher und Praktiker müssen sich dieser Problemlage bei der Interpretation der Daten und bei der Implementation von Beurteilungskonsequenzen (z.B. Entwicklungsmaßnahmen, Prämienzahlung) bewusst sein.

6.5.6 Zusammenfassende Diskussion zur Rolle der interperspektivischen Diskrepanz

Die vorangehend dargestellte, umfassende Erforschung der Rolle der interperspektivischen Diskrepanz im Entstehungszusammenhang zwischen Führung und Erfolg basierte – im Vergleich zum bisherigen Forschungsmainstream – auf grundsätzlichen Erweiterungen. So wurden erstens *verschiedene* Übereinstimmungsmaße und Analysemethoden für die Auswertung der gleichen Rohdaten eingesetzt, was auf der Suche nach einem validen Maß unabdingbar ist. Zweitens wurden Führungsbeurteilungen aus *drei Perspektiven* eingeholt, so dass die Übereinstimmung zwischen zwei Perspektiven an der dritten validiert werden konnte. Zudem konnten sie an Erfolgskriterien validiert werden, was insbesondere für die Selbst-Abwärts-Übereinstimmung starkes Gewicht erlangte, weil die Erfolgskriterien von den Aufwärtsbeurteiler abgegeben wurden. Drittens kamen zwei *verschiedene Fragebogen-Instrumente* zum Einsatz, der an Führungsaufgaben orientierte FzFü und der auf dem Führungsmodell von Bass (1985) beruhende MLQ. Diese Erweiterungen verbesserten die Interpretationsgrundlage gegenüber bisherigen Studien deutlich.

Nach dem Kriterium der Sparsamkeit müssten Übereinstimmungsmaße zugunsten einzelner Beurteilungsperspektiven dann als Erfolgsprädiktoren zurückgewiesen werden, wenn sie in ihrer Vorhersageleistung letztere nicht übertreffen oder wenn ihre Vorhersageleistung auf den einzeln eingehenden Beurteilungsperspektiven, nicht auf der Diskrepanz an sich, beruht. Falls sich Kongruenzmodelle aber als mindestens ebenbürtig zu einfachen Modellen erweisen, dann können sie in der Praxis zweifellos verwendet werden. Dies bietet sich aus Gründen der Plausibilität und Akzeptanz sogar eher an als die alleinige Verwendung einfacher Fremd- oder Selbstbeurteilungen. So werden Kongruenzmodelle ohnehin bereits als Erklärungsmodelle für Führungsentwicklungen herangezogen, da sie sich als Ansatzpunkt für Veränderungsprozesse als geeignet herausgestellt haben (vgl. Hossiep und Paschen, 1998a, 1998b). Nur wenn die Diskrepanz nicht zumindest Mit-Urheber der kovariierenden Misserfolge ist, dann würde eine praktische Anwendung, wie wir sie derzeit erleben, auf Luftschlössern basieren, seien sie noch so plausibel. Deshalb muss die Rolle, die die Diskrepanz spielt, in Folgestudien weiter untersucht werden.

Die Ergebnisse der vorliegenden Studie werfen trotz der oben genannten methodischen Verbesserungen im Hinblick auf eine umfassende Betrachtung der Rolle der interperspektivischen Diskrepanz, fast mehr Fragen als Antworten auf. Zumindest kristallisiert sich keine klare Antwort heraus, die zudem für die Mehrzahl der berücksichtigten Erfolgskriterien zutreffen würde. Der Leser mag dies in den vorangehenden Kapiteln als Komplexität oder Detail-Reichtum wahrgenommen haben. Im Folgenden werden deshalb – quasi zusammenfassend – diejenigen empirischen Befunde der vorliegenden Studie, die für die prognostische Validität der interperspektivischen Diskrepanz sprechen, denen gegenübergestellt, die gegen die prognostische Validität sprechen.

Für die Kriteriumsvalidität aller drei untersuchten interperspektivischen Diskrepanzen (Selbst-Aufwärts, Selbst-Abwärts, Abwärts-Aufwärts) sprechen zunächst deren korrelativen Zusammenhänge mit Erfolg, die auch nach Kontrolle des Single-Source-Bias weiter bestehen. Dies lässt vermuten, dass die Diskrepanz einen kritischen Erfolgsprädiktor darstellt. Unterschätzern und Übereinstimmern sowie Führungskräften mit geringer Abweichung von Fremdratern werden von ihren Mitarbeitern mehr Erfolge zugesprochen als Überschätzern.

Diese Argumentation ist aber irreführend, wenn man die Ergebnisse der Regressionsanalysen heranzieht, anhand derer die Vorhersagekraft der Selbst-Aufwärts-Diskrepanz untersucht wurde. Dann zeigt sich nämlich für die meisten Erfolgskriterien, dass sie am besten durch die „einfachen“, uniperspektivischen Beurteilungen (hauptsächlich die Aufwärtsbeurteilung) vorhergesagt werden können. Der *Differenz* zwischen zwei Beurteilungsperspektiven kann innerhalb dieses Ansatzes nur für wenige Kriterien ein prognostischer Wert bescheinigt werden. Dies ist das gewichtigste Argument gegen die Validität von Differenzen (vgl. auch Brutus, Fleenor & Tisak, 1999). Gängige Interpretationen, nach denen die korrelativen Zusammenhänge zwischen der Selbst-Fremd-Übereinstimmung und den Erfolgsvariablen auf den Ähnlichkeitsgrad der erhobenen Perspektiven zurückgeführt werden, wären damit nicht mehr haltbar.

Allerdings spricht die regressionsanalytische Überprüfung nicht eindeutig und eben nicht für alle Erfolgskriterien gegen die Verwendung von Kongruenzmodellen. Zusätzlich ergeben sich aus der Frage nach potentieller inkrementeller Validität der Diskrepanzen weitere Hinweise dafür, dass sie spezifische Informationen und Vorhersagekraft enthalten. So gilt für manche Erfolgskriterien nicht nur, dass mehrere Diskrepanzen gemeinsam mehr Erfolg aufklären als einzelne, sondern auch, dass Diskrepanzen zusätzlich zu den einfachen Perspektiven einen Beitrag leisten können. Beispielsweise klärt die Selbst-Aufwärts-Diskrepanz über die Aufwärtsbeschreibung hinaus Varianz auf (für EFF und SAT). Orientiert man sich lediglich an der Varianzaufklärung, ist außerdem anzumerken, dass die besten Diskrepanzmodelle für manche Kriterien einen höheren Varianzanteil aufklären als die besten einfachen Modelle. Dies verweist zumindest darauf, dass die Diskrepanzen mehr Informationen enthalten müssen als die einfachen Beurteilungen.

Gegen eine unbedingte prognostische Validität der Übereinstimmung spricht aber nicht nur, dass Diskrepanzmaße für viele andere Erfolgsvariablen keinen zusätzlichen Beitrag über eine uniperspektivische Beurteilung hinaus leisten. Zusätzlich unterscheiden sich die Führungskräfte, deren Mitarbeiter überdurchschnittliche Werte in den Erfolgsvariablen berichten (Mediansplit), hinsichtlich der Selbst-Aufwärts-, Selbst-Abwärts- und Abwärts-Aufwärts-*Korrelationen* nicht von denen, deren Mitarbeiter unterdurchschnittliche Werte in den Erfolgsvariablen angeben. Lediglich die Niveau-Unterschiede differieren insofern erwartungsgemäß, dass sie für die erfolgreiche Teilstichprobe am geringsten ausfallen.

Vorläufig sprechen die in der vorliegenden Studie erhobenen Daten dafür, in jedem Fall Fremdbeurteilungen, insbesondere Aufwärtsbeurteilungen von Führung einzuholen. Diskrepanzen können als *zusätzliches* Maß verwendet werden, sollten jedoch nicht alleinige Grundlage von Entscheidungen darstellen, erst recht nicht von weit reichenden, administrativen Entscheidungen.

Für die gleichzeitige Verwendung von uniperspektivischen Beurteilungen *und* von Diskrepanzen sprechen nicht nur die erwähnten inkrementellen Beiträge der Diskrepanzen über die einfachen Beurteilungen hinaus, sondern auch das Ergebnis, dass Übereinstimmer mit hoher Führungsqualität[209] erfolgreicher sind als Unter- und Überschätzer mit hoher Führungsqualität. Es ist durchaus plausibel, dass die Führungsqualität grundlegenden Einfluss auf den Erfolg nimmt – davon gehen ja auch alle Führungstheorien und -aufgaben aus – und dass die Übereinstimmung eine weniger direkte Rolle spielt.

Relativ unwahrscheinlich scheint auf der aktuellen Datenbasis eine moderierende Rolle der Diskrepanz auf die prädiktive Validität von Führung zu sein. Die Ergebnisse fallen zu inkonsistent aus. Allenfalls für die Selbst-Aufwärts-Diskrepanz zeigt sich mehrheitlich, dass für die Überschätzer die prädiktive Validität der Abwärtsbeurteilungen am höchsten ausfällt. Dieser unerwartete Befund wurde dadurch erklärt, dass Überschätzer ihren Vorgesetzten mehr Beobachtungsgrundlage für ihre Führung liefern, indem sie diese eher in den Mittelpunkt ihrer Selbstdarstellung rücken als Unterschätzer oder Übereinstimmer. Außerdem ist anzunehmen, dass Überschätzer ihr Führungsverhalten nicht in Abhängigkeit davon ändern, ob ihr Vorgesetzter anwesend ist oder nicht. In der Folge kann dieser das „wahre“ Führungsverhalten beobachten und letztlich beurteilen.

Vorstellbar ist hingegen auch eine partielle Mediatorfunktion insofern, dass Führung (erfasst durch uniperspektivische Beurteilungen) eine direkte Wirkung auf Erfolg hat, gleichzeitig aber auch einen über die Diskrepanz vermittelten Einfluss ausübt. Dies muss in Folgestudien anhand von größeren Stichproben mittels Pfadanalysen überprüft werden.

Dann kann streng genommen auch erst die bislang implizit angenommene kausale Richtung des Zusammenhangs besser kontrolliert werden. Es ist zwar insbesondere dadurch, dass erstens die Führung/Führungsdiskrepanzen mit den Erfolgskriterien auch im Criss-

[209] überdurchschnittlich positive uniperspektivische Führungsbeurteilung

Cross-Design negativ korrelieren und dass zweitens Führung und Führungsdiskrepanzen an Führungsbeurteilungen und Erfolgskennwerten aus einer *dritten* Perspektive validiert werden konnten, nicht anzunehmen, dass implizite Theorien zu einer umgekehrten Entstehungsrichtung führen. Es wäre unplausibel, eine positive Führungsbeurteilung deshalb abzugeben, weil Kollegen oder andere Dritte hohen Erfolg wahrnehmen. Es ist auch nicht von Drittvariablen auszugehen, die sowohl die Führungs- als auch die Erfolgsbeurteilungen in gegebener Weise beeinflussen, zumal die Erfolgsvariablen mit keiner der betrachteten personalen oder organisationsbezogenen Variablen bedeutsam zusammenhängen, die Führungsdiskrepanzen hingegen schon. Aber diese methodisch schwachen Argumente für eine Kausalinterpretation ersetzen die strengere Prüfung durch Pfadanalysen, Zeitreihen- und insbesondere experimentelle Untersuchungsdesigns nicht.

Entscheidet man sich für die Verwendung von Differenzmaßen, dann stellt sich die Frage nach der Art des verwendeten Maßes. Aus den regressionsanalytischen Ergebnissen ergab sich, dass die Vorhersage durch die Selbst-Aufwärts-*Distanz* besser gelingt als durch die algebraische Selbst-Aufwärts-Differenz. Außerdem hat sich sowohl für die Selbst-Aufwärts- als auch die Selbst-Abwärts-Diskrepanz gezeigt, dass die Distanz durchschnittlich weniger Varianz erklärt als die *Richtung* der Abweichung. Das ist kein Punkt, der für die Praxis unterstrichen werden muss. Hier ist die Richtung der Abweichung in der Regel in anschaulichen Grafiken leicht erkennbar. Es ist Aufgabe der Forschung, zu ermitteln, welche prognostischen Anteile auf das Ausmaß und welche auf die Richtung der Abweichung zurückgehen, um anschließend auch entsprechende Möglichkeiten vorzuschlagen, Ausmaß und Richtung gleichzeitig zu berücksichtigen. Abgesehen von den Niveau-Unterschieden ist zudem der prädiktive Wert der korrelativen Übereinstimmung ungeklärt. Die Befunde der vorliegenden Studie verweisen darauf, dass sich korrelative und absolute Übereinstimmung nicht gegenseitig ersetzen können. Insgesamt scheinen Korrelationen weniger stark mit Erfolg zu kovariieren als Niveau-Unterschiede.

Was die einbezogenen Beurteilungsrichtungen betrifft, sind die Abwärts-Aufwärts-Diskrepanzen aufgrund ihrer etwas höheren Varianzaufklärung den Selbst-Fremd-Diskrepanzen vorzuziehen. Aus praktischen Gesichtspunkten (z.B. Akzeptanz) muss allerdings in umgekehrter Reihenfolge präferiert werden. Wo Regressionsanalysen einen inkrementellen Beitrag von zwei Diskrepanzen erbringen konnten, war jeweils die entsprechende Abwärts-Aufwärts- mit einer Selbst-Fremd-Diskrepanz kombiniert. Dies setzt voraus, dass neben der Selbstperspektive mindestens zwei Fremdperspektiven in die Beurteilung einbezogen werden.

Da der parallele Einsatz verschiedenster Maße und Methoden einzelne Forschungsfragen sehr komplex gestaltet, muss unbedingt darauf hingearbeitet werden, optimale Maße und Methoden zu entwickeln. Nur durch eine simple Durchführung und Dokumentation von Untersuchungen kann dann gewährleistet, dass Erkenntnisse gewonnen und verbreitet werden können.

6.6 Entstehung und Reduktion von Selbst-Aufwärts-Diskrepanzen

Für eine Generierung von Ideen, wie Selbst-Aufwärts-Diskrepanzen zustande kommen können und wie sie reduziert werden können, wurden den Aufwärtsbeurteilern offene Fragen am Ende des Fragebogens gestellt. Da dieses Datenmaterial aus sehr kurzen Antworten besteht, entspricht die Analyseeinheit in der Regel der gesamten Äußerung einzelner Personen. Nur selten gaben einzelne Personen mehr als eine Erklärung bzw. Reduktionsstrategie an. Diese Angaben auf inhaltlich strukturierende Art zusammengefasst. Die Erstellung des Kategoriensystems war vollständig materialgeleitet. Die Begrenzungen der einzelnen Analyseeinheiten lagen durchgehend auf der Hand, da jeweils die Angaben einer Person hinsichtlich eines Erklärungs- oder Reduktionsansatzes als kodierbare Einheit (= Analyseeinheit) betrachtet wurden. Diesem eindeutigen Charakter und der Kürze der Antworten ist geschuldet, dass auf eine ausführliche Operationalisierung unter Angabe von Ankerbeispielen verzichtet wurde. Die Antworten der Befragten wurden nach ihrer Häufigkeit kodiert.

6.6.1 Hinweise auf die Entstehung von Selbst-Aufwärts-Diskrepanzen

In Tabelle 81 sind die subjektiven Erklärungen für die Selbst-Aufwärts-Diskrepanz dargestellt, die die befragten Führungskräfte und Mitarbeiter als Antworten auf die entsprechende offene Frage am Ende des Fragebogens genannt haben.

Die subjektiven Erklärungen der Führungskräfte verteilen sich fast gleichmäßig auf die drei Oberkategorien „Unterschiedlichkeit der Erfahrungen und Bewertungen von Führung“, „Kontakt und Kommunikation“ und „Persönliche Inkongruenz“. Die Mitarbeiter betrachten Missstände im Kontakt und der Kommunikation als Hauptursache für die Selbst-Aufwärts-Diskrepanzen (32 der 121 Nennungen). Daneben geben auch sie an, dass unterschiedliche Erfahrungen und Bewertungen von Führung sowie persönliche Inkongruenzen Diskrepanzen hervorrufen können. Dass sie auch Charakteristika der Führungskraft – kurz: mangelnde Führungskompetenz, zu wenig Zeit und überzogendes Selbstvertrauen – als Erklärungen für die Diskrepanzen heranziehen, unterscheidet die Äußerungen der Mitarbeiter von denen der Führungskräfte.

Tabelle 81. Subjektive Erklärungen der befragten Führungskräfte und Mitarbeiter für eine potentielle Selbst-Aufwärts-Diskrepanz

Subjektive Erklärung für Selbst-Aufwärts-Diskrepanz	Anzahl der Nennungen (gesamt) Führungskräfte (39)	Mitarbeiter (121)
Unterschiedlichkeit der Erfahrungen und Bewertungen von Führung		
a. Unterschiedliches/r Aufgabenspektrum/Verantwortungsbereich	1	7
b. Unterschiedliche Auffassungen von Führungsverhalten	1	2
c. Falsche Auffassung der Führungsgrundsätze	1	-
d. Unterschiedliche Prioritätensetzung (z.B. übergeordnete Führungsbelange)	7	5
e. Unterschiedliche Erwartungen	2	6
Kontakt und Kommunikation		
f. Zu wenig persönlicher Kontakt	2	11
g. Fehlender Informationsaustausch	2	5
h. Mangelnder Austausch zu Leistungsgrundsätzen und -erwartungen	2	2
i. Zu selten Mitarbeitergespräche/Feedback, keine kritische Offenheit gegenüber Führungskraft	2	5
j. Schlechte, missverständliche Kommunikation, Unehrlichkeit	1	5
k. Zu wenig Kenntnisse der Aufgaben des jeweils anderen	1	4
Persönliche Inkongruenz		
l. Unterschiedliche Sichtweise (z.B. wegen unterschiedlicher Sozialisation oder unterschiedlich langer Organisationszugehörigkeit)	5	5
m. Persönlichkeits- und Beziehungscharakteristika (Antipathie, Misstrauen, Vorurteile, Desinteresse, unterschiedlicher Arbeitsstil)	2	15
Charakteristika der Führungskraft		
n. mangelnde Führungskompetenz und -sensibilität		4
o. zu wenig Zeit für Führung/einzelne Mitarbeiter		3
p. Egoismus, Dominanz, überzogenes Selbstvertrauen		7
Rest		
q. Fragebogen (Itemformulierung zu kurz oder unkonkret)		2
r. Nicht kodierte Äußerungen[1]	10	33

Anmerkungen. [1]Nicht kodiert wurden diejenigen Äußerungen, die keine Erklärungen für die Diskrepanz boten, die auf die Dauer der Zusammenarbeit abzielten oder missverständlich formuliert waren. Die Kodierübereinstimmung zwischen zwei Kodierern beträgt für die Führungskräfteaussagen $\kappa = .94$ und für die Mitarbeiteraussagen $\kappa = .98$ (Vorgehen, vgl. Bortz & Döring, 2003).[210]

[210] Für die Durchführung der Kodierungen zur Bestimmung der Beurteilerübereinstimmung sei Dipl.-Psych. Iris Six-Materna herzlich gedankt.

Diese Antworten stimmen im Übrigen sehr gut mit denen überein, die eine selektive Stichprobe von jeweils zehn Führungskräften und Mitarbeitern auch in der Interviewsituation frei generierten (Senger, in Vorb.). Vergleicht man diese Erklärungsansätze der „Betroffenen“ ferner mit den theoretischen Überlegungen aus Kapitel 3.4, dann fällt auf, dass eine sehr gute Passung existiert. So gehen beispielsweise die

- Kategorien f, k und o mit der Erklärung über unterschiedliche Verhaltensstichproben (Kapitel 3.4.1.1),
- die Kategorien a, f, g, j, k und n mit der Erklärung über mangelnde Beobachtbarkeit und Transparenz (Kapitel 3.4.1.2),
- besonders die Kategorie b, aber auch die Kategorien c, f, g, h, i, k, l mit der Erklärung über unterschiedliche Führungsdefinitionen (Kapitel 3.4.1.3),
- besonders die Kategorien d, aber auch die Kategorien e, h und j mit der Erklärung über unterschiedliche Gewichtung von Führungsverhalten (Kapitel 3.4.1.4),
- die Kategorie l mit der Erklärung über unterschiedliche Differenzierungen (Kapitel 3.4.1.5),
- die Kategorie e und m mit einzelnen Urteilsfehlern (Kapitel 3.4.2.1),
- die Kategorien i und j mit der Erklärung über mangelndes Feedback (Kapitel 3.4.2.2) und
- die Kategorien j und p mit der Erklärung über Selbsttäuschung und Eindrucksmanagement (Kapitel 3.4.2.3) einher.

Insofern versprechen in diesem noch spärlich untersuchten Forschungsbereich qualitative Ansätze – vorliegend offene Befragungen – sogar auch für bereits angedachte Erklärungsansätze durchaus einen Erkenntnisgewinn. Sie sollten daher zukünftig verstärkt in ein multimethodales Vorgehen integriert werden. Jedenfalls dürfen die Erklärungen nicht auf der fast deskriptiven Ebene stehen bleiben. So klingen allerdings populäre Erklärungen aus der aktuellen Literatur, die das mangelnde Self-Awareness der Führungskräfte (Van Velsor et al., 1993) oder deren Inkompetenz, eigene Stärken und Schwächen zu diagnostizieren (Atwater & Yammarino, 1997), für die geringe interperspektivische Diskrepanz verantwortlich machen. Eine ausführliche Diskussion zu Entstehungsmechanismen von interperspektivischen Beurteilungsdiskrepanzen folgt in Kapitel 7.3 vor dem Hintergrund bereits berichteter Befunde der vorliegenden Studie. Im Sinne einer kumulativen Validierung können die Ergebnisse dann abgeglichen werden.

6.6.2 Ansatzpunkte zur Reduzierung der Beurteilungs-Diskrepanz

In diesem Abschnitt werden die Vorschläge zur Reduzierung der Selbst-Aufwärts-Diskrepanz dargestellt, die die befragten Mitarbeiter und Führungskräfte auf die entsprechende offene Frage im Fragebogen generierten. Die Kurzantworten fallen sehr einheitlich aus. 23 der 27 Antworten der Führungskräfte betreffen im weitesten Sinne die Kommunikation zwischen Führungskraft und Mitarbeiter. Die Vorschläge reichen vom häufigeren Führen persönlicher Gespräche und der Verbesserung des Informationsaustausches bis hin zum professionell moderierten Feedbackgespräch. Die anderen Antworten fokussieren darauf, das Führungsverhalten zu verändern – und zwar entweder gezielt die im Fragebogen auffälligen Führungsverhaltensweisen oder generell in Richtung kooperatives Führen. Sehr ähnliche Vorschläge zur Reduzierung potentieller Selbst-Aufwärts-Diskrepanzen bringen die befragten Mitarbeiter. Von den 74 Vorschlägen beziehen sich 49 auf die Kommunikation zwischen Führungskraft und Mitarbeiter, 17 auf das Führungsverhalten der Führungskraft und 8 auf strukturelle Bedingungen. Was die Kommunikation betrifft, reichen die Vorschläge von der Erhöhung des direkten, persönlichen Kontakts bis zu institutionalisierten Feedbackgesprächen. Hinsichtlich des Führungsverhaltens erwarten sich die Mitarbeiter dann eine bessere Selbst-Aufwärts-Kongruenz, wenn die Führung kooperativer im Sinne der Führungsgrundsätze ausgeführt wird und wenn Führungskräfte ihre Führungskompetenzen fortbilden. Diejenigen Vorschläge, die struktureller Natur sind, betreffen die hierarchische Organisationsstruktur, eine als ungleichmäßig wahrgenommene Arbeitsbelastung (über die Abteilungen hinweg), die mangelnden Möglichkeiten, mehr Einblick in den Handlungsspielraum der Führungskräfte zu erlangen und der häufige Wechsel von Führungskräften, was keinen Aufbau von Vertrauensverhältnissen zuließe.

Abschließend werden in Tabelle 82 noch einmal alle Hypothesen auf einen Blick dargestellt, bevor die Diskussion (Kapitel 7) mit einer Zusammenfassung der berichteten Ergebnisse beginnt.

Tabelle 82. Überblick über die bearbeiteten Hypothesen

Hypothese	Bestätigung?
Fragestellung 1:	
Unterschiede zwischen beobachtetem und erwartetem Führungsverhalten	
1) Die Bewertung des aktuellen Führungsverhaltens (Ist) fällt negativer aus als die Einschätzung des erwarteten Führungsverhaltens.	ja
2) Mitarbeiter nehmen eine größere Ist-Ideal-Diskrepanz wahr als Führungskräfte.	ja
3) Die Ideal-Einschätzung der Führungskräfte (Selbst-Ideal) stimmt stärker mit dem Unternehmens-Soll überein, wenn diese die Führungsgrundsätze kennen (vs. nicht/kaum kennen).	nein
4) Je höher die Hierarchieebene der fokussierten Führungskraft ist, umso besser stimmen Selbst-Ideal und Soll überein.	nein
Fragestellung 2:	
Korrelative Zusammenhänge zwischen Selbst-, Aufwärts- und Abwärtsbeurteilungen	
5) Wenn man die Aufwärts-Beurteilungen, die die gleiche Führungskraft betreffen, aggregiert (versus nicht aggregiert), korrelieren sie stärker mit Beurteilungen aus anderen Perspektiven (Selbst, Abwärts).	nur für einzelne Dimensionen, insb. Abwärts-Aufwärts
6) Die Beurteilungen verschiedener Perspektiven (Selbst-Ist, Aufwärts-Ist, Abwärts-Ist) korrelieren gering ($.10 \leq r < .30$).	für Selbst-Aufwärts und Abwärts-Aufwärts
7) Die Korrelationen zwischen Selbst- und Fremdbeurteilungen sind geringer als die Korrelationen zwischen den Fremdbeurteilungen.	ja
8) Die Korrelationen zwischen Selbst- und Aufwärtsbeurteilungen unterscheiden sich nicht von den Korrelationen zwischen Selbst- und Abwärtsbeurteilungen.	nein (Selbst-Aufwärts < Selbst-Abwärts)
9) Die Korrelationen zwischen den Perspektiven variieren in ihrer Stärke über die Dimensionen hinweg.	ja
Fragestellung 3:	
Niveau-Unterschiede zwischen Selbst-, Aufwärts- und Abwärtsbeurteilungen	
10) Zwischen allen erfassten Perspektiven (Selbst-, Abwärts-, Aufwärtsbeurteilung) bestehen Niveau-Unterschiede.	für Selbst-Aufwärts und Abwärts-Aufwärts

11)	Die Niveau-Unterschiede zwischen der Selbst- und Abwärtsbeurteilung sind geringer als die zwischen der Selbst- und Aufwärtsbeurteilung.	ja
12)	Die Selbstbeurteilungen fallen am positivsten, die Aufwärtsbeurteilungen am negativsten aus. Die Abwärtsbeurteilungen nehmen eine Zwischenstellung ein und unterscheiden sich von den Selbstbeurteilungen weniger als von den Aufwärtsbeurteilungen.	Ja, außer: Selbst und Abwärts ähnlich positiv
13)	Die Selbstüberschätzung der Führungskräfte im Vergleich zur Abwärtsbeurteilung ist unabhängig von deren Selbstüberschätzung im Vergleich zur Aufwärtsbeurteilung.	für Mehrzahl der Dimensionen
14)	Das Ausmaß der interperspektivischen Differenzen variiert über die Führungsdimensionen hinweg.	ja

Fragestellung 4:
Moderatorvariablen für die Beurteilungs-Übereinstimmung zwischen verschiedenen Perspektiven

15)	Die in Tabelle 14 aufgeführten Variablen moderieren in jeweils angegebener Richtung die Stärke des korrelativen Zusammenhangs und das Ausmaß der Niveau-Unterschiede zwischen den angegebenen Beurteilungsperspektiven. Die Übereinstimmung zwischen den Perspektiven steigt oder fällt also mit der Ausprägung der Moderatorvariablen.	teilweise (abhängig von Moderatorvariable und Perspektive)

Fragestellung 5 (Erwartung: *Ablehnung* der Hypothesen):
Zusammenhang zwischen verschiedenen Übereinstimmungsindizes

16)	Je geringer die Selbst-Abwärts-Niveau-Unterschiede sind, umso höher sind die Selbst-Abwärts-Korrelationen.	nein
17)	Je geringer die Selbst-Aufwärts-Niveau-Unterschiede sind, umso höher sind die Selbst-Aufwärts-Korrelationen.	nein
18)	Je geringer die Abwärts-Aufwärts-Niveau-Unterschiede sind, umso höher sind die Abwärts-Aufwärts-Korrelationen.	nein

Fragestellung 6:
Above-Average-Effekt

19)	Führungskräfte beurteilen ihr eigenes Führungsverhalten positiver als das anderer Führungskräfte.	ja

20)	Die Mehrzahl der Führungskräfte beurteilt ihr eigenes Führungsverhalten positiver als das anderer Führungskräfte.	ja
21)	Die Selbstüberschätzung nach oben fällt geringer aus als die Selbstüberschätzung nach unten.	nein
22)	Die Selbstüberschätzung nach oben ist unabhängig von der Selbstüberschätzung nach unten.	ja

Fragestellung 7:
Kriteriumsvalidität der Führungsbeurteilungen

23)	Die Fremdbeurteilungen von Führung korrelieren positiv mit Erfolgskennwerten.	ja
24)	Die Aufwärtsbeurteilungen von Führung korrelieren stärker mit den Erfolgskennwerten als die Abwärtsbeurteilungen.	ja
25)	Aggregierte Aufwärtsbeurteilungen korrelieren mit den Erfolgsvariablen stärker als nicht aggregierte (individuelle) Aufwärtsbeurteilungen.	nur bei hoher Rateranzahl
26)	Aggregierte Fremdbeurteilungen (Aufwärts- und Abwärtsbeurteilungen) korrelieren mit den Erfolgsvariablen nicht stärker als aggregierte Aufwärtsbeurteilungen alleine.	ja
27)	Selbstbeurteilungen korrelieren in geringerem Ausmaß mit Erfolg als Fremdbeurteilungen.	ja

Fragestellung 8:
Inkrementelle Validität der Führungsbeurteilungen

28)	Verschiedene Führungsdimensionen klären gemeinsam mehr Erfolg auf als einzelne Führungsdimensionen.	ja
29)	Die Selbst-, Aufwärts- und Abwärtsbeurteilung von Führung erklären gemeinsam mehr Erfolgsvarianz als die Beurteilung aus einzelnen Perspektiven.	ja
30)	Die transformationale Führung klärt gemeinsam mit der transaktionalen Führung mehr Erfolgsvarianz auf als die transaktionale Führung allein.	ja
31)	Die transformationale Führung klärt zusätzlich zu den klassischen Führungsdimensionen Erfolgsvarianz auf.	nein (umgekehrter Zusammenhang)

Fragestellung 9:
Moderatorvariablen für den Zusammenhang zwischen Führungsverhalten und Führungserfolg

32)	Je wichtiger die Aufwärtsbeurteiler die Führungsdimensionen bewerten, umso stärker ist der Zusammenhang zwischen diesen Führungsdimensionen und Erfolg.	nein (für Mehrzahl der Dimensionen)
33)	Je transformationaler die Aufwärtsbeurteiler ihre Führungskraft beschreiben, umso stärker ist der Zusammenhang zwischen Führung und Erfolg.	nein

Fragestellung 10:
Kriteriumsvalidität der Selbst-Fremd- und Fremd-Fremd-Übereinstimmung

34)	Mit steigenden Niveau-Unterschieden zwischen der Selbst- und Fremdbeurteilung sinkt der Erfolg.	ja
35)	Hypothese 34) gilt stärker für die Diskrepanz zwischen der Selbst- und Aufwärts-Beurteilung als für die Diskrepanz zwischen der Selbst- und Abwärts-Beurteilung.	ja
36)	Mit steigenden Niveau-Unterschieden zwischen der Abwärts- und Aufwärtsbeurteilung sinkt der Erfolg.	ja
37)	Die Selbst-Fremd-Übereinstimmung kann mehr Erfolgsvarianz aufklären als die Abwärts-Aufwärts-Übereinstimmung.	nein (umgekehrter Zusammenhang)
38)	Die Mitarbeiter von Führungskräften, die sich im Vergleich zu Fremdbeurteilern überschätzen, sind am wenigsten erfolgreich, während der Erfolg bei Übereinstimmung am höchsten ausfällt. Die Mitarbeiter der Unterschätzer nehmen eine Zwischenstellung ein.	teilweise (für Selbst-Aufwärts unterscheiden sich Übereinstimmer und Unterschätzer nicht)
39)	Es ist die Selbst-Aufwärts-Diskrepanz, die die kritische prognostische Leistung für Erfolg erbringt (vs. einfache Selbst- und Aufwärtsbeurteilungen)	uneindeutig (für manche Erfolgskriterien)
40)	Die Selbst-Aufwärts-Diskrepanz eignet sich zur Erfolgsprognose mindestens ebenso gut wie die einfachen Beurteilungen aus der Selbst- und Aufwärtsperspektive.	eher nein
41)	Mehrere Diskrepanzen können mehr Erfolgsvarianz erklären als einzelne Diskrepanzen.	teilweise (für manche Erfolgskriterien)

42)	Das beste Diskrepanzmodell erklärt mehr Erfolgsvarianz als das beste einfache Modell.	teilweise (für manche Erfolgskriterien)
43)	Die Diskrepanz zwischen zwei Perspektiven leistet (über die einzelnen Beurteilungsperspektiven hinaus) einen inkrementellen Beitrag zur Erfolgsaufklärung.	teilweise (für manche Erfolgskriterien)
44)	Der Erfolg fällt für Übereinstimmer mit hoher Führungsqualität höher aus als für Unter- und Überschätzer mit hoher Führungsqualität.	ja
45)	Der Erfolg fällt für Übereinstimmer mit geringer Führungsqualität niedriger aus als für Unter- und Überschätzer mit hoher Führungsqualität.	teilweise
46)	Berichten Mitarbeiter überdurchschnittlichen (vs. unterdurchschnittlichen) Erfolg, sind die Diskrepanzen zwischen den Beurteilungsperspektiven geringer.	ja
47)	Berichten Mitarbeiter überdurchschnittlichen (vs. unterdurchschnittlichen) Erfolg, sind die Korrelationen zwischen den Beurteilungsperspektiven höher.	nein
48)	Berichten Mitarbeiter überdurchschnittlichen Erfolg, dann gilt im Vergleich zu Mitarbeitern mit unterdurchschnittlichem Erfolg: a. Die Mitarbeiter sind älter. b. Ihre Führungskraft ist älter. c. Die Abteilung ist kleiner. d. Die Beteiligungsquote an der Befragung ist höher. e. Die Dauer der Zugehörigkeit zur Organisation und zum Team ist länger.	nein
49)	Liegt Übereinstimmung zwischen den Perspektiven vor (vs. Unter- oder Überschätzung), fällt der Zusammenhang zwischen Führung und Erfolg höher aus.	uneindeutig

7 Diskussion

Ziel der vorliegenden Arbeit war es, die Befundlage zu verschiedensten Forschungsfragen im Bereich der multiperspektivischen Führungsbeurteilung zu erweitern, wozu eine 270°-Beurteilungsstudie durchgeführt wurde. Im ersten Teil der Arbeit wurden die Unterschiede zwischen beobachtetem und erwartetem Führungsverhalten und insbesondere die Übereinstimmung zwischen den Selbst- und Aufwärts-, Selbst- und Abwärts- sowie Abwärts- und Aufwärtsbeurteilungen von Führung untersucht. Zur Erforschung der interperspektivischen Übereinstimmung gehörte ferner die Untersuchung einer Vielzahl potentieller Moderatorvariablen. Solche Analysen fehlten bis dato insbesondere für die Selbst-Aufwärts-, erst recht aber für die Abwärts-Aufwärts-Übereinstimmung. Moderierende Effekte zu erkennen, trägt zu einer zukünftigen Modellentwicklung zur Entstehung von interperspektivischen Diskrepanzen bei. Untersucht wurde in diesem ersten Teil der Arbeit außerdem eine Art Above-Average-Effekt sowie der Zusammenhang zwischen verschiedenen Übereinstimmungsmaßen.

Im zweiten Teil der Arbeit wurde die Kriteriumsvalidität erstens von Selbst- und verschiedenen Fremdbeurteilungen von Führung und zweitens von deren interperspektivischen Diskrepanz untersucht. Das zentrale Ziel dabei war ein besseres Verständnis der Rolle, die die interperspektivische Diskrepanz im Zusammenhangsgefüge zwischen Führung und Erfolg einnimmt. Abschließend wurden die Ideen der Befragten zu Entstehungsmechanismen und Reduzierungsmöglichkeiten von Selbst-Aufwärts-Diskrepanzen zusammengefasst.

Die grundlegenden Ergebnisse zu den bearbeiteten Fragestellungen werden in Kapitel 7.1 zusammengefasst. Im Anschluss daran werden aufgrund der Vielschichtigkeit der Fragestellung einzelne Punkte gezielt diskutiert. Dabei werden auch bereits Grenzen der vorgestellten Studie erläutert sowie daraus folgende Ansatzpunkte für weiterführende Forschungsarbeiten beschrieben (Kapitel 7.2 bis 7.6). Abschließend werden bis dahin nicht berichtete Problembereiche der Studie dargestellt und Anregungen dazu gegeben, wie diese zukünftig besser gehandhabt werden können.

7.1 Allgemeine Zusammenfassung

Zu Fragestellung 1, die sich mit den Unterschieden zwischen beobachtetem und erwartetem Führungsverhalten beschäftigte, ist festzuhalten, dass das beobachtete Führungsverhalten aus allen drei Beurteilungsperspektiven negativer bewertet wird als das erwartete Führungsverhalten. Im Vergleich zur Soll-Erwartung der Unternehmensleitung nehmen die Beurteiler der beiden Fremdperspektiven einen durchschnittlichen Unterschied von etwas mehr als einer Standardabweichung wahr ($\bar{d}$ = 1.17 bzw. 1.01) und die sich selbst beurteilenden Führungskräfte einen Unterschied von etwas weniger als einer Standardabweichung ($\bar{d}$ = .80). Die Unterschiede zwischen dem beobachteten Verhalten und den *Ideal*-Erwartungen fallen für Aufwärts- und Selbstbeurteiler ähnlich stark aus ($\bar{d}$ = 1.18 bzw. 1.28). Geht man davon aus, dass eigene Idealvorstellungen als (akzeptierte) Ziele fungieren und daher Motivation zur Erreichung dieser Ideale aufgebracht wird, dann müssen die Soll-Vorstellungen nicht zusätzlich mit Anreizen versehen werden, zumal die Ideale über dem Unternehmens-Soll liegen. Die Ideal-Vorstellung der Selbstbeurteiler nähert sich dem Soll-Wert weder in Abhängigkeit der Hierarchieebene der Selbstbeurteiler noch in Abhängigkeit davon an, ob die Selbstbeurteiler die Führungsgrundsätze kennen oder nicht. Generell gilt, dass die Ist-Ideal-Diskrepanzen für Mitarbeiter ohne eigene Führungsverantwortung größer sind als für Führungskräfte und Mitarbeiter mit Führungsverantwortung. Es ist wahrscheinlich, dass ohne eigene Führungserfahrung situationale Faktoren, die den Führungsprozess erschweren, nicht berücksichtigt werden und deshalb eine größere Lücke zwischen Ist und Ideal besteht. Auf diese potentielle Ursache für Unzufriedenheiten muss in der organisationalen Praxis reagiert werden.

Fragestellung 2 erbrachte folgende Erkenntnisse über die korrelativen Zusammenhänge zwischen den Beurteilungen verschiedener Perspektiven. Während die Selbst-Fremd-Korrelationen unbedeutend bis gering ausfallen, sind Abwärts-Aufwärts-Korrelationen mittlerer Stärke festzuhalten (vgl. Kapitel 7.2). Insbesondere Letztere können durch die Aggregation der Aufwärtsratings einer Arbeitsgruppe gesteigert werden. Dass die Selbst-Aufwärts-Korrelationen von der Aggregation nahezu unberührt bleiben, muss als Hinweis auf die geringe Reliabilität der Selbstbeurteilungen gewertet werden (vgl. Kapitel 7.2). Die interperspektivischen Korrelationen variieren über die Führungsdimensionen hinweg stark (vgl. Kapitel 7.6), unterscheiden sich im Durchschnitt freilich kaum von Befunden anderer Stichproben (vgl. Kapitel 7.2).

Damit gehen die gefundenen Niveau-Unterschiede zwischen den Beurteilungsperspektiven (Fragestellung 3) insofern einher, dass diese zwischen der Selbst- und Aufwärtsbeschreibung am größten sind. Erstaunlich geringe Mittelwertsunterschiede zeigten sich – auch im Vergleich zu bisherigen Befunden – zwischen der Selbst- und Abwärtsbeurteilung. Aus verschiedenen Gründen ist die Ursache dafür nicht nur in der Verzerrung einer Perspektive zu suchen (vgl. Kapitel 7.2). Wichtig ist, dass Selbstbeschreibungen nicht per se

inflationär sind. So ist immer auch die Vergleichsperspektive zu berücksichtigen, wenn von Überschätzung gesprochen wird, denn eine Überschätzung gegenüber den Mitarbeiterratings ist nicht zwangsläufig von einer Überschätzung gegenüber den Vorgesetztenratings begleitet. Außerdem ist eine positive Einschätzung stark von der zu beurteilenden Dimension abhängig, so dass auch die Anzahl der Unterschätzer oder Übereinstimmer über die einzelnen Führungsdimensionen hinweg variiert (vgl. Kapitel 7.6).

Fragestellung 4 zielte darauf ab, Moderatorvariablen für die interperspektivische Übereinstimmung zu extrahieren. Generell ist festzuhalten, dass für Niveau-Unterschiede mehr und stärkere Effekte aufgedeckt werden konnten als für die korrelative Übereinstimmung. Dies stimmt generell mit dem meta-analytischen Ergebnis von Heidemeier und Moser (2002) überein, auch wenn dort lediglich die Selbst-Abwärts-Übereinstimmung beruflicher Leistungsbeurteilung im Allgemeinen berücksichtigt wurde. Außerdem konnten für die Abwärts-Aufwärts-Übereinstimmung mehr und stärker moderierende Effekte ermittelt werden als für die Selbst-Fremd-Übereinstimmung. Auch dies könnte an einer geringen Reliabilität der Selbstbeurteilungen liegen. Folgende Variablen zeigen für alle interperspektivischen Übereinstimmungen moderierende Effekte (wenn auch zum Teil äußerst geringe) in die erwartete Richtung: Alter der Fremdrater, Führungsspanne, ergebnisbezogene Itemformulierung, Beteiligung an der multiperspektivischen Beurteilung (Fremd- und Selbstrater) und Kenntnis der Führungsgrundsätze (vgl. Kapitel 7.2).

Durch die Analysen bezüglich Fragestellung 5 stellt es sich als Notwendigkeit heraus, *verschiedene* Übereinstimmungsindizes zu verwenden. Insbesondere wenn das Ausmaß der Übereinstimmung nicht nur im Sinne von Mess-Ungenauigkeiten interpretiert wird, sondern als zusätzliche Information betrachtet wird, die erfolgsprognostisch valide sein kann, sollten verschiedene Indizes hinsichtlich ihres Wirkungsgrads verglichen werden. So zeigte sich, dass verschiedene Differenzmaße und Korrelationen nicht ausschließlich redundante Informationen enthalten.

Wenn die Betrachtungsrichtung umgedreht wird und solche Führungsbeurteilungen miteinander verglichen werden, die *eine* Führungskraft zur Beurteilung verschiedener Führungskräfte abgibt, bestätigt sich die sich abzeichnende Selbstüberschätzung (Fragestellung 6). So sind die Beurteilungen des eigenen Führungsverhaltens *durchschnittlich* positiver als die des Führungsverhaltens anderer – übergeordneter genauso wie unterstellter – Führungskräfte. Zudem schätzt die *Mehrzahl* der Führungskräfte ihr eigenes Führungsverhalten positiver ein als das anderer.

Als zentrales Ergebnis zu Fragestellung 7, die sich mit der Kriteriumsvalidität der Führungsbeurteilungen aus drei verschiedenen Perspektiven beschäftigte, ist festzuhalten, dass keine der Perspektiven invalide ist, auch wenn die Selbstbeurteilungen durchschnittlich sehr gering mit den Erfolgsvariablen korrelieren. Am höchsten ist die Varianzaufklärung durch die Aufwärtsbeurteilung (genauer vgl. Kapitel 7.4, 7.5 und 7.6). Weder die Wichtigkeitseinstufung noch das Ausmaß der transformationalen Führung oder die Interrater-

Übereinstimmung moderieren den Zusammenhang zwischen den Führungsbeurteilungen und Erfolg (Fragestellung 9).

Als wichtige Unterstützung der grundlegenden Annahmen von multiperspektivischen Führungsbeurteilungen sind die Ergebnisse zu Fragestellung 8 zu bewerten. Die Beurteilungen aus den verschiedenen Perspektiven besitzen inkrementelle Validität, so dass die höchste Varianzaufklärung durch die Betrachtung mehrerer Beurteilungsperspektiven erreicht werden kann (vgl. Kapitel 7.4 und 7.5). Unterschiedliche Führungs*dimensionen* spielen dabei eine entscheidende Rolle (vgl. Kapitel 7.6).

Neuere Untersuchungen sprechen der Übereinstimmung zwischen Selbst- und Fremdbeurteilungen von Führung einen spezifischen Informationswert zu, der Erfolgsvarianz erklären kann (z.B. Atwater et al., 1998). In der vorliegenden Studie wurde zusätzlich die Validität der Abwärts-Aufwärts-Diskrepanz untersucht, die tendenziell sogar höher ausfiel als die der Selbst-Fremd-Diskrepanzen (Fragestellung 10). Für alle drei interperspektivischen Diskrepanzen ergaben sich starke Hinweise für ihren prognostischen Wert. Sowohl die Richtung der Abweichung als auch das pure Ausmaß der Abweichung gehen mit den Erfolgskriterien insofern einher, dass letztere bei Überschätzung (vs. Unterschätzung und Übereinstimmung) sowie bei höchster Abweichung am geringsten ausfallen. Uneindeutig bleibt jedoch, ob die kritische Komponente in der Erfolgsprognose wirklich der interperspektivischen Übereinstimmung und nicht den einzelnen Perspektiven zukommt. Verschiedene Befunde sprechen dafür, andere dagegen. Zumindest ein einfaches Differenzmaß zwischen der Selbst- und Aufwärtsbeurteilung muss aufgrund regressionsanalytischer Verfahren für die Mehrzahl der erfassten Kriterien zugunsten der einfachen Perspektiven, in der Regel zugunsten der Aufwärtsperspektive, verworfen werden. Dass für manche Erfolgsmaße aber ein inkrementeller Beitrag der Diskrepanzen über die einfachen Perspektiven hinaus ermittelt werden konnte und dass die Varianzaufklärung durch Diskrepanzen für manche Erfolgskriterien höher ist als die Varianzaufklärung durch alle einfachen Perspektiven gemeinsam, spricht für die zusätzliche, spezifische Validität der Übereinstimmung. Plausibel ist das Ergebnis, dass diejenigen Führungskräfte die erfolgreichsten sind, die besonders positive Führungsbeurteilungen erhalten *und zugleich* eine hohe Übereinstimmung zwischen der Selbst- und Fremdbeurteilung aufweisen. Der korrelative Zusammenhang zwischen zwei Perspektiven unterscheidet sich – im Gegensatz zu den Niveau-Unterschieden – von über- und unterdurchschnittlich Erfolgreichen nicht. Dies verweist wie die Ergebnisse zu Fragestellung 5 darauf, dass verschiedene Übereinstimmungsindizes keine identischen Beiträge zur Erfolgsaufklärung liefern. Demografische und organisationsbezogene Variablen blieben ohne Einfluss auf die Erfolgskennwerte. Gerade auch in Relation dazu ist die Varianzaufklärung der interperspektivischen Übereinstimmung beachtlich.

Die befragten Führungskräfte nennen Erklärungsansätze für die Selbst-Aufwärts-Diskrepanzen, die entweder auf unterschiedliche Erfahrungen und Bewertungen von Führung, auf Kontakt- und Kommunikationsdefizite oder auf persönliche Inkongruenz

zwischen ihr selbst und ihren Mitarbeitern schließen lassen. Zusätzlich dazu geben Mitarbeiter Gründe an, die die Führungskraft als Person oder die Qualität ihres Führungsverhaltens betreffen (Fragestellung 11). Ebenfalls in offenen Fragen wurden subjektiv als nützlich bewertete Ansatzpunkte zur Reduzierung der Diskrepanzen exploriert. Die Vorschläge lassen sich überwiegend zwei Kategorien zuordnen, nämlich Verbesserungen der Kommunikation und Verbesserungen der Führungsqualität an sich (Fragestellung 12).

Die einzelnen Fragestellungen der vorliegenden Arbeit knüpften an unterschiedliche Forschungszweige und folglich natürlich auch an unterschiedliche Forschungsarbeiten an. Die Integration der neu gewonnenen Erkenntnisse in den bereits publizierten Wissensstand versuchen Kapitel 7.2 bis 7.7 zu schaffen. In diesem Rahmen werden einzelne Ergebnisse zusammenfassend dargestellt und Diskussionspunkte, die bereits in der Ergebnisdarstellung angedeutet wurden, fortgeführt. Für eine ausführlichere Diskussion zur Rolle der Diskrepanz im Entstehungsprozess zwischen Führung und Erfolg wird auf Kapitel 6.5.6 verwiesen (vgl. außerdem Kapitel 7.5).

7.2 Kann man sich auf Führungsbeurteilungen verlassen? Zu ihrer Güte im weitesten Sinne

Insbesondere in Kapitel 3.4 wurde eine Vielzahl an Motiven und Charakteristika sozialer Kognitionen angerissen, die bei Leistungsbeurteilungen aktiviert sein können und die in Kapitel 7.3 weiter diskutiert werden. Entsprechend enthalten Führungsbeurteilungen aus allen Beurteilungsperspektiven Informationen, die nicht das Ratee-Verhalten beschreiben. Moser, Zempel und Schultz-Amling (2000) sprechen von „strategischen" Verzerrungen, also von Verzerrungen der Ratings, die bewusst und zielgerichtet vorgenommen werden. Auch wenn solche strategischen und politischen Komponenten von Leistungsbeurteilungen nie vollständig auszuschalten sein werden, sollte das Ziel Bestand haben, möglichst genaue Leistungsbeurteilungen zu gewinnen. Denn Beurteilungen ziehen in der Regel positive oder negative Konsequenzen nach sich (Bezahlung, Beförderung, Karrieremöglichkeiten etc.) – und diese sind mindestens für den Betroffenen und häufig sogar zusätzlich für das Unternehmen von weit reichender Bedeutung. Das gilt umso mehr, je zentraler die organisationalen Rollen der Ratees sind. Zum einen sind dann strategische Verzerrungen wahrscheinlicher, zum anderen sind dann die Konsequenzen aus einer Organisationsperspektive gewichtiger. Für die *Führungskräfte*-Beurteilung sollte also besonderes Augenmerk auf die Gütekriterien gerichtet werden. Generell sollte freilich bereits die ureigene Struktur von multiperspektivischen Beurteilungen dazu beitragen, Verzerrungen durch die induzierte Validierungserwartung[211] zu reduzieren (identifizierte Moderatorvariable in den Meta-Analysen von Heidemeier & Moser, 2002 sowie von Mabe & West, 1982).

7.2.1 Gütekriterien im engeren Sinne

Die Güte von Beurteilungen – also von Selbst- sowie Fremdbeurteilungen – kann anhand verschiedener Kriterien bestimmt werden. Dies sind erstens die Anfälligkeit für Urteilstendenzen, zweitens die Akkuratheit und drittens die Reliabilität und Validität als psychometrische Maße (Murphy & Cleveland, 1995). Sie werden im Folgenden für die vorliegende Studie unter Rückgriff auf bestehende Kenntnisse diskutiert.

Was die Urteilstendenzen betrifft, weisen in der vorliegenden Studie die Selbstbeurteilungen im Vergleich zu den Fremdbeurteilungen eine höhere Mildetendenz, eine Tendenz zu reduzierter Streuung und eine geringere Korrelationstendenz (Skaleninterkorrelationen) auf. Diese Befunde stimmen mit denen überein, die aus der Beurteilung beruflicher Leistung im weitesten Sinne bekannt sind (Moser et al., 1994). Gerade die geringere Korrela tionstendenz von Selbstbeurteilungen stellt einen Vorteil dar, der bislang in der multiper-

[211] Erwartung, dass die abgegebenen Beurteilungen an Ratings anderer validiert werden

spektivischen Forschung zu Führungsbeurteilungen noch ungenutzt bleibt. Dadurch, dass in der Regel keine multidimensionalen Führungsinstrumente eingesetzt werden – und wenn doch, die Skalenmittelwerte zu einem Summenscore gemittelt werden (z.B. Atwater et al., 1998) – wird die bessere Differenzierung zwischen den Führungsdimensionen wieder zunichte gemacht (vgl. Kapitel 7.6 und 7.7).

Die Akkuratheit stellt ein Kriterium dar, das nur dann von Bedeutung sein kann, wenn ein „wahrer Wert“ messbar ist und als Referenzpunkt verwendet werden kann. Da für die Führungsleistung aber weder ein kriteriumsrelevantes Maß existiert, das objektiv und reliabel erfassbar ist, noch eine potentiell kontinuierliche Beobachtung von Führung in quantitative Maße transferiert werden kann, ist das Kriterium weder auf Selbst- noch auf Fremdbeurteilungen anwendbar (vgl. Moser, 1999). Wo dies versucht wird, werden in aller Regel Fremdbeurteilungen von Führung als „wahre“ Referenzpunkte zur Bestimmung der Genauigkeit der Selbstratings herangezogen. Das explizit formulierte Ziel vieler Studien, die Akkuratheit von Selbstbeurteilungen schätzen zu wollen (z.B. Randall, Ferguson & Patterson, 2000), muss also trotz des hehren Vorhabens aufgrund mangelnder objektiver Referenzwerte ad absurdum geführt werden. In der vorgestellten Arbeit wurde daher lediglich von Selbst-Fremd-„Übereinstimmung“ gesprochen, was die mangelnde „Wahrheit“ der Fremdbeurteiler in Betracht stellt. Die Selbst-Fremd-Korrelation wird im Übrigen häufig als Validität ausgegeben (z.B. Meta-Analyse von Conway & Huffcutt, 1997; Moser et al., 1994), was in den Kapiteln 7.4 und 7.5 weiter diskutiert wird.

Die Reliabilität der Führungsbeurteilungen muss differenziert betrachtet werden. Für die vorliegende Studie ergibt sich ein gemischtes Bild. Die interne Konsistenz der verschiedenen Führungssubskalen ist für beide Instrumente (FzFü, MLQ), alle Modusversionen (Wichtigkeit, Ideal, Ist) und alle Perspektiv-Versionen (Selbst, Abwärts, Aufwärts) zufriedenstellend bis sehr gut. Am geringsten sind die Alpha-Koeffizienten für die Ist-Version der Selbstbeurteilung. Dies geht Hand in Hand mit der oben erwähnten besseren Differenzierungsleistung der Selbstbeurteiler.

Lediglich zufriedenstellend sind die Interrater-Reliabilitäten der Mitarbeiter einer Führungskraft. Da sie eher als grenzwertig betrachtet wurden, wurden an gegebenen Stellen die Analysen mit aggregierten Aufwärtsratings um Analysen mit individuellen Aufwärtsratings ergänzt. Da in der Mehrzahl der publizierten Studien keine Interrater-Reliabilitäten angegeben sind (z.B. Fleenor et al., 1996), ist eine Bewertung ihrer Höhe in Relation zu Führungsbeurteilungen in anderen Stichproben nicht möglich. Sie ist vergleichbar mit der meta-analytisch geschätzten Interrater-Reliabilität von Aufwärtsratings beruflicher Leistung (Conway & Huffcutt, 1997; $r = .30$). Die relativ zu anderen Beurteilungsperspektiven geringe Interrater-Reliabilität der Aufwärtsbeurteiler (Conway & Huffcutt, 1997) könnte an der vergleichsweise geringen Beurteilerübung oder -erfahrung liegen. Mitarbeiter übernehmen zumindest in Unternehmen der traditionellen Industrie erst selten Beurteilungsaufgaben. Zudem ist nicht nur die interperspektivische Übereinstimmung, sondern auch die Interrater-Reliabilität für die Beurteilung von Management-Aufgaben geringer als für die

Beurteilung von Nicht-Management-Aufgaben (Conway & Huffcutt, 1997). In einer aktuellen Studie von Ostroff et al. (2004) konnten im Vergleich zu den vorliegenden Daten zur Aufwärtsbeurteilung dennoch höhere Indizes (.64) erreicht werden.

Insbesondere deshalb, weil in der Literatur fast ausschließlich aggregierte Analysen berichtet werden, die aber nur beim Vorliegen einer ausreichenden Interrater-Reliabilität berechtigt sind, sollte deren Dokumentation zukünftig stärker als bislang eingefordert werden. Denn auch wenn keine Garantie besteht, dass Führungskräfte ihre Leistung aufgrund der Rückmeldung reliabler Daten steigern, kann schon gar nicht erwartet werden, dass sie ihr Verhalten „auf Erfolgskurs“ bringen, wenn sie mit unreliablen Daten versorgt werden. Die Steigerung der Reliabilität ist also auch für Praktiker von immanentem Interesse.

Prinzipiell ist anzuraten, möglichst viele Aufwärtsrater zu befragen, um den Reliabilitätsgewinn für aggregierte Ratings, der aus der höheren Rateranzahl resultiert, zu nutzen (Conway & Huffcutt, 1997). Dass eine Steigerung der Ratingzahl für Aufwärtsbeurteiler leichter zu bewerkstelligen ist als beispielsweise für Abwärtsbeurteiler, liegt wegen der pyramidenförmigen Organisationsstrukturen auf der Hand. Es versteht sich von selbst, dass die Aggregation von Aufwärtsratings erst dann in dem erwünschten Validitätsvorteil münden kann, wenn die Ratings eine hohe Reliabilität aufweisen. Auch wenn in der vorliegenden Studie kein moderierender Effekt auf den Zusammenhang zwischen Führung und Erfolg ermittelt werden konnte, ist bei ausreichend hoher Reliabilität von einer Auswirkung auf die Validität der Führungsbeurteilungen auszugehen. Diese zu erhöhen, ist das grundlegende Anliegen, das ja auch den Forschungszweig der multiperspektivischen Beurteilungen erst rechtfertigt.

Bei starker Inhomogenität der Antworten ist auch aus inhaltlichen Gründen die Sinnhaftigkeit von aggregierten Ratings anzuzweifeln. Gerade im Kontext der Personalführung kann eine unterschiedliche Beurteilung durch unterschiedliche Mitarbeiter auch eine besonders erfolgreiche Führung charakterisieren. So könnte nämlich ein sehr individuumsbezogenes Führungsverhalten, wie es verschiedene Führungstheorien fordern, umgesetzt sein (z.B Path-Goal-Theorie von House, 1971, 1996; LMX-Ansatz der Führung von Graen & Scandura, 1987a, 1987b; IC-Baustein der transformationalen Führung von Bass, 1985).

Finden sich geringe Interrater-Reliabilitäten, bestehen für die Forschung sowie für die Praxis verschiedene Möglichkeiten, darauf zu reagieren. Wie soeben angerissen, kann die Inhomogenität als Information über eine differenzierte Führung gewertet werden.[212] Außerdem kann – wie in der vorliegenden Untersuchung – die aggregierte Auswertung durch individuumsbezogene Analysen ergänzt werden. Gerade auch der Vergleich der Ergebnisse aus aggregierten und individuellen Analysen lässt bessere Interpretationen zu. Für die Beratungspraxis empfiehlt es sich, bei geringer Übereinstimmung den Range der

[212] Eine solche Interpretation fände dann Unterstützung, wenn sich die Kriteriumsvalidität von Führung in Abhängigkeit davon stark unterscheiden würde, ob man personengebundene oder gekreuzte (Criss-Cross-Design) Analysen durchführt.

Aufwärtsbeurteilungen zurückzumelden, wenn nicht sogar die Werte einzelner Mitarbeiter und insbesondere die der „Ausreißer“. Allerdings muss diese Art der Rückmeldung *vor* dem Beurteilungsprozess angekündigt sein, da selbst ohne Namensnennung keine vollständige Anonymität mehr zugesichert werden kann. Die positive Kehrseite davon wurde in Kapitel 3.1.3 unter dem populär gewordenen Schlagwort „Accountability“ angerissen (vgl. London, Smither & Adsit, 1997). Der vollständige Verzicht auf aggregierte Analysen ist deshalb nicht anzuraten, weil auch Analysen mit individuellen Werten problematisch sind. So gehen alle Beurteilungen, die den Aufwärtsratings gegenübergestellt werden (z.B. Selbst-, Abwärtsbeurteilung) mehrfach (entsprechend der Anzahl der Aufwärtsratings) in die Analyse ein. Zudem ist der Beurteilungsgegenstand unterschiedlich. Während ein einzelner Mitarbeiter das Führungsverhalten beurteilt, das er persönlich erfährt, geben andere Beurteiler immer das Führungsverhalten an, das sie der Führungskraft generell zuschreiben, also bezüglich der Führung all ihrer Mitarbeiter. Erst durch eine Mittelung der Aufwärtsbeurteilungen können die unterschiedlichen Foki einander tendenziell angeglichen werden.

Möglicherweise sind vor einem „blinden“ Einsatz von Aufwärtsbeurteilungen gezielte Beurteilerschulungen angebracht. Zudem muss eine Diskussion über Möglichkeiten zur entsprechend einfachen Gestaltung der Beurteilungsinstrumente in Gang gebracht werden, da immer auch die Nützlichkeit betrachtet werden muss. Sollte sich die Interrater-Reliabilität nicht steigern lassen, ist die ohnehin empfehlenswerte Strategie noch dringender anzuraten: die Befragung von Ratern aus unterschiedlichen Perspektiven (vgl. Kapitel 7.5) – also auch aus Perspektiven, die mehr Beurteilererfahrung besitzen (Vorgesetzte, Peers).

Abschließend soll zur Interrater-Reliabilität gesagt sein, dass sich durch die Aggregation aller Fremdratings (Abwärts- *und* Aufwärtsratings) in der vorliegenden Studie erwartungsgemäß kein Vorteil ergab. Dies geht einher mit den Ergebnissen von Sala und Dwight (2002), steht aber im Widerspruch zu denen von Wohlers und London (1989) und zum Resumée von Hedge et al. (2001). Während Erstere durch die Aggregation keine erhöhte Kriteriumsvalidität finden, ist dies bei Letzteren der Fall. Eine solche Zusammenfassung der Ratings aus verschiedenen Perspektiven – mag sie sich auch in manchen Studien als aus psychometrischen Gründen gerechtfertigt erweisen – unterläuft indes die grundlegende Annahme von multiperspektivischen Beurteilungen. Denn gerade die jeweilige perspektivenspezifische Information ist es, von der inkrementelle Validität erwartet wird (vgl. Kapitel 7.5). Sie darf nicht „weg-gemittelt“ werden, ganz zu schweigen von mühelos prognostizierbaren Akzeptanzproblemen, die in der Beratungspraxis folgen würden.

Die Retest-Reliabilität konnte in der vorliegenden Studie lediglich für einen Auszug der Aufwärts-Ist-Version und lediglich an einer sehr geringen Teilstichprobe ermittelt werden. Auch wenn angesichts dessen keineswegs von gesicherten Ergebnissen zu sprechen ist, verweist der hohe Koeffizient (r = .90) auf eine ausreichende Stabilität. Zum Vergleich:

Salgado et al. (2003) schätzen die Stabilität von Abwärtsbeurteilungen von Führung auf r = .53. Damit scheint Führung weniger stabil beurteilbar zu sein als berufliche Leistung im weiteren Sinne (Abwärtsbeurteilungen: r = .79 bei Salgado et al., 2003; r = .81 in der Meta-Analyse von Viswesvaran et al., 1996). Dass Selbstbeurteilungen von Leistung am wenigsten stabil sind (Harris & Schaubroeck, 1988), ist plausibel, da Selbstbeurteiler Veränderungen eigener Verhaltensweisen wahrnehmen, bevor Fremdbeurteiler ihre Fremdbilder revidieren. Abschließend muss mit Salgado et al. (2003) unbedingt die geringe Anzahl an Studien zur Retest-Reliabilität bemängelt werden.

In Kapitel 5.5.1.10 wurde ausführlich begründet, weshalb von einer hohen Inhalts- und Konstruktvalidität auszugehen ist. Die Kriteriumsvalidität betrifft den Kern der vorliegenden Forschungsfragen und wird deshalb in den Kapiteln 7.4 bis 7.7 ausführlich diskutiert. Vorab sei kritisch angemerkt, dass keine prädiktive Validität im engeren Sinne erfasst wurde, sondern aufgrund des querschnittlichen Designs der Studie lediglich die konkurrente Validität geschätzt werden konnte (vgl. Kapiteln 7.7) und dadurch selbstverständlich auch keine kausalen Mechanismen aufgedeckt werden konnten.

Im Sinne der erweiterten Gütekriterien für angewandte Forschung spricht es durchaus für die Praktikabilität und Akzeptanz des Vorgehens, wenn in den abschließenden Fragen des Fragebogens 34 der 48 sich selbst beurteilenden Führungskräfte angaben, dass ein Fragebogen, wie er eingesetzt wurde, für die Beratung von Führungskräften – also für ihre eigene Beratung – nützlich sein könnte.

7.2.2 Interperspektivische Beurteilungsübereinstimmung

Wie bereits erwähnt, werden in vielen Untersuchungen Fremdbeurteilungen aus einer oder mehreren Perspektiven zur Validierung der Selbstbeurteilungen herangezogen (z.B. Stempfle et al., 2004; vgl. auch Moser et al., 1994). So existieren inzwischen auch mehrere Meta-Analysen zur Selbst-Fremd-, insbesondere zur Selbst-Abwärts-Übereinstimmung (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002; Mabe & West, 1982; vgl. Kapitel 3.2). Der Fremd-Fremd-Übereinstimmung wurde bislang wenig Forschungsinteresse zuteil, die Abwärts-Aufwärts-Übereinstimmung von beruflicher Leistungsbeurteilung wurde beispielsweise nur in der Meta-Analyse von Conway und Huffcutt (1997) berücksichtigt, allerdings nicht wie die Selbst-Fremd-Übereinstimmung für Management- und Nicht-Management-Tätigkeiten getrennt analysiert.

Dieses mangelnde Interesse an den Fremd-Fremd-Übereinstimmungen hängt damit zusammen, dass sich die Selbstbeurteilung als Anknüpfungspunkt für die Beratungspraxis besonders eignet und deshalb Beurteilungssysteme auf diese ausgerichtet sind. Das gilt vor allem in der derzeitigen Hochkonjunktur-Phase der Schlüsselvariablen des Selbst: Selbstmanagement, Selbstreflexion, Selbstlernfähigkeit. Auch zugrunde gelegte Erklärungs

ansätze (z.B. Johari-Fenster von Luft, 1961, 1963) oder theoretische Modelle (Ashford, 1989; Atwater & Yammarino, 1997) zentrieren sich um die Selbstwahrnehmung. So kann die seltene Untersuchung von Fremd-Fremd-Übereinstimmungen auch als Nebenprodukt einer einseitigen Forschungsfokussierung auf die Selbstwahrnehmung betrachtet werden. Wenn aber ein ernsthaftes Interesse daran besteht, die validesten Beurteilungsperspektiven zu finden, dann sind Fremdbeurteilungen zukünftig stärker zu untersuchen. In der vorliegenden Arbeit wurden daher drei Perspektiven und deren interperspektivische Übereinstimmung berücksichtigt (Selbst, Abwärts, Aufwärts).

Betrachtet man die interperspektivischen Diskrepanzen in der untersuchten Stichprobe, dann fallen nicht nur die hohen Niveau-Unterschiede zwischen den Beurteilungen aus der Selbst- und Aufwärts- sowie zwischen den Beurteilungen aus der Abwärts- und Aufwärtsperspektive auf, sondern auch die sehr geringen Niveau-Unterschiede zwischen den Beurteilungen aus der Selbst- und Abwärtsperspektive.

Die Selbst-Aufwärts-Diskrepanzen liegen in einem Ausmaß, wie es auch bei Mount und Scullen (2001) für Management-Stichproben beschrieben wurde. Für Abwärts-Aufwärts-Diskrepanzen bestehen kaum Vergleichsstudien im Bereich der Personalführung, so dass ihr Ausmaß nicht relativ zu den Befunden aus anderen Stichproben bewertet werden kann. In einer der wenigen diesbezüglichen Untersuchungen konnten keine Niveau-Unterschiede zwischen Abwärts- und Aufwärtsbeurteilungen von Management-Dimensionen festgestellt werden (Fleenor et al., 1996). In der vorliegenden Studie sind die gefundenen Diskrepanzen vor allem auf die durchgehend sehr positiven Abwärtsbeurteilungen zurückzuführen. Besonders starke Abwärts-Aufwärts-Abweichungen kommen entsprechend auch überwiegend durch herabgesetzte Aufwärtsratings[213] zustande, weniger durch weiter erhöhte Abwärtsratings. Da die Abwärts-Aufwärts-Diskrepanz jedoch stark mit Erfolg zusammenhängt und die Diskrepanz fast ausnahmslos von der Aufwärtsbeurteilung bestimmt wird, ist dieser allein deshalb eine höhere Validität zuzusprechen. Für die Abwärtsbeurteilungen sind verzerrende Einflüsse anzunehmen, die im Zusammenhang mit der geringen Selbst-Abwärts-Diskrepanz erklärt wurden.

Die Selbst-Abwärts-Diskrepanz fällt im Vergleich zu meta-analytischen Ergebnissen hinsichtlich beruflicher Leistungsbeurteilung im Allgemeinen (Harris & Schaubroeck, 1988), aber auch hinsichtlich Management-Tätigkeiten (Conway & Huffcutt, 1997; Heidemeier & Moser, 2002) in der vorliegenden Studie sehr gering aus. Verschiedene Erklärungen für die interperspektivischen Diskrepanzen werden in Kapitel 7.3 diskutiert.

Wichtig festzuhalten ist, dass die durchschnittlichen Niveau-Unterschiede zwischen den Perspektiven nicht dadurch zustande kommen, dass *einzelne,* sehr starke „Abweichler“ zu Buche schlagen, sondern dass sich die Befunde auch in den Häufigkeiten niederschlagen. So gehören mehr als die Hälfte der Führungskräfte in die Kategorie der Selbst-Aufwärts-Überschätzer (mehr als eine halbe Standardabweichung Differenz) und weniger als 10% in

[213] gegenüber geringen Abwärts-Aufwärts-Diskrepanzen

die Kategorie der Unterschätzer. Diese Zahlen stimmen mit denen überein, die Tornow (1993) in seinem Diskussionspapier berichtet. Die annähernde Drittelung der Führungskräfte in Überschätzer, Übereinstimmer und Unterschätzer, wie sie in anderen Studien berichtet wird (Atwater, Roush & Fischthal, 1995 [Studenten der Marineakademie]; Van Velsor, Taylor & Leslie, 1993 [Manager]), kann in der vorliegenden Stichprobe lediglich für die Selbst-*Abwärts*-Übereinstimmung gefunden werden. Da Atwater et al. (1995) sowie Van Velsor et al. (1993) ohnehin nur geringe Selbst-Aufwärts-Niveau-Unterschiede finden (wie für Selbst-Abwärts in der vorliegenden Studie), kann spekuliert werden, dass eine gleichmäßige Verteilung auf Überschätzer, Übereinstimmer und Unterschätzer dann nicht zu erwarten ist, wenn durchschnittlich sehr starke Beurteilungsdifferenzen bestehen.

Weiter sind die Selbst-Fremd-Diskrepanzen nicht ausschließlich den Selbstbeurteilungen anzulasten. Dass diese nicht per se als inflationär zu verurteilen sind, zeigt sich beispielsweise darin, dass die berichteten Diskrepanzen auf erhöhten Selbst- und erniedrigten Fremdbeurteilungen gründen. Der Anteil, den die Selbstbeurteilungen tatsächlich an den Diskrepanzen tragen, muss in weiteren Studien gezielt untersucht werden. Fest steht nur, dass auch die Güte der Fremdbeurteilungen zu relativieren ist. Gegebenenfalls ist in der Folge auch die Operationalisierung der *Selbst*reflexionsfähigkeit über die Selbst-*Fremd*-Übereinstimmung neu zu überdenken. Zumindest muss bei einer solchen Operationalisierung gewährleistet werden, dass die Fremdbeurteilung annähernd genau ausfällt (z.B. Beurteilersystem im Assessment-Center, vgl. Stempfle et al., 2004).

Die Milde der Selbst- oder Abwärtsbeurteilungen allein ist zudem nicht unmittelbar Beleg für ihre Invalidität. Erst dann, wenn sie nicht durch eine systematische, konstante Erhöhung des Niveaus zustande kommt, ist von geringer Validität auszugehen. Dass vor allem bei den Selbstbeurteilungen nicht von systematischen, konstanten Verzerrungen auszugehen ist, zeigen die korrelativen Übereinstimmungen zwischen den Perspektiven. Während die Selbst-Aufwärts-Korrelation quasi unbedeutend und die Selbst-Abwärts-Korrelation gering ist, erreicht die Abwärts-Aufwärts-Korrelation mittlere Stärke. Die interperspektivischen Korrelationen liegen damit im Rahmen der meta-analytischen Angaben hinsichtlich beruflicher Leistungsbeurteilung allgemein. Einem Vergleich der korrelativen Zusammenhänge mit denen anderer Studien im Bereich der Personalführung ist Kapitel 7.2.3 gewidmet.

Betrachtet man die verschiedenen *Führungsdimensionen*, dann sind einzelne Interkorrelationen hervorzuheben. Eine verhältnismäßig hohe Selbst-Aufwärts-Korrelation ergab sich für die Dimension Klarheit. Dies wurde damit begründet, dass die zu beurteilenden Inhalte eng mit einem „Aushandeln“ zwischen den Selbst- und den jeweiligen Aufwärtsratern verbunden sind, so dass wenig Spielraum für unterschiedliche Interpretationen der Iteminhalte bleibt. Der ermittelte Moderatoreffekt für die Variable „ergebnisbezogene Itemformulierung“ unterstützt diese Begründung und unterstützt die Forderung einfach handhabbarer Beurteilungsinstrumente. Dieser Befund kann außerdem als Hinweis dafür

betrachtet werden, dass Führungsdimensionen immer von denjenigen Personen beurteilt werden sollten, die besondere Nähe zum Beurteilungsgegenstand aufweisen.

Für die Selbst-Aufwärts-Korrelationen sticht außerdem der Faktor Partizipation hervor. Die negative Interkorrelation von $r = -.27$ gewinnt gerade vor dem Hintergrund Vehemenz, dass partizipativer Führung generell motivierende Wirkung zugesprochen wird und dass das „kooperative Führen selbständiger Mitarbeiter" im vorliegenden Unternehmen als Leitlinie verankert werden soll. Ein Teil dieses negativen Zusammenhangs scheint davon beeinflusst zu sein, wie wichtig den Mitarbeitern ihre eigene Partizipation ist. Wird diese nämlich als überdurchschnittlich wichtig erachtet, fällt der Zusammenhang weniger stark negativ aus. Es ist plausibel, dass diejenigen Mitarbeiter, denen Partizipation nicht wichtig ist, die als partizipativ initiierten Führungselemente ihrer Führungskräfte gar nicht als solche verstehen. Was von den Führungskräften mit partizipativer Absicht eingeführt wird, wird von den Mitarbeitern möglicherweise eher als undefinierter Raum wahrgenommen, in dem keine individuelle Förderung/Chance existiert, sondern eher Unsicherheit oder Überforderung.

Dass partizipative Führung (Selbstbeurteilungsperspektive) von den Mitarbeitern als Mangel an Engagement der Führungskräfte wahrgenommen wird, ist ein weiterer Beleg dafür, dass dieses Führungsverhalten anders verstanden als intendiert wird. Außerdem weist die Partizipation (Selbstbeurteilung) zu allen berücksichtigten Erfolgskriterien außer Wohlbefinden dann eine höhere Korrelation auf, wenn die Mitarbeiter die Dimension der Partizipation als überdurchschnittlich wichtig (vs. unterdurchschnittlich wichtig) erachten. Dies spricht generell für eine kontingenztheoretische Sicht der Führung. Hinsichtlich des speziellen Zusammenhangs zwischen Partizipation und Wohlbefinden verweist das Ergebnis aber auf eine zwiegespaltene Wirkung von Partizipation auf die Befindlichkeit. Anzunehmen ist, dass sie auch eine stressbegünstigende Komponente enthält, da Mitgestaltungsmöglichkeiten mitsamt der zusätzlichen Verantwortung auch wahrgenommen werden müssen.

In der Führungspraxis müsste demzufolge verstärkt darauf geachtet werden, bei welchen Aufgaben partizipative Führung eingeführt und auf welche Art und Weise sie kommuniziert wird. In Folgeuntersuchungen könnte die Rolle der Partizipation gezielter untersucht werden, indem auf Mitarbeiterseite der Wunsch nach partizipativer Führung genauer erfasst wird und die Prozesse, die in die Mitgestaltungsmöglichkeiten involviert sind, exploriert oder beobachtet werden. Auch Interventionsstudien sind in diesem Forschungsfeld denkbar. Auf diesem Wege könnten ferner die oben hypothetisch angenommenen, stressbegünstigenden Elemente partizipativer Führung eruiert werden.

Dass der höchste Selbst-Abwärts-Zusammenhang für die Dimension Entwickeln/Fördern gefunden wird, erscheint deshalb plausibel, weil gerade die Karriereentwicklung von Mitarbeitern abteilungs- und insbesondere ebenenübergreifend geplant werden muss. Bei einer guten Entwicklung und vor allem Förderung der Mitarbeiter ist also ein diesbezüglicher Austausch zwischen den Selbst- und Abwärtsbeurteilern gegeben, der für die anderen Führungsdimensionen weniger nahe liegt.

der für die anderen Führungsdimensionen weniger nahe liegt. Der Austausch kann aufgrund einer besseren Beobachtungsstichprobe, einer abgestimmten Definition des Beurteilungsgegenstands oder abgestimmter Wichtigkeitszuschreibung in besser übereinstimmenden Beurteilungen resultieren.

Analog kann für die vergleichsweise hohen Abwärts-Aufwärts-Korrelationen auf den Dimensionen Entwickeln/Fördern und Stellvertretung argumentiert werden. Beide beanspruchen einen abteilungsübergreifenden Handlungsradius, was bereits bei der Ergebnisdarstellung erläutert wurde. Dass ebenso für die Dimension Individuelle Berücksichtigung ein starker Abwärts-Aufwärts-Zusammenhang existiert, verwundert zunächst, geht es doch um die Führungsverhaltensweisen, die am engsten zwischen der Führungskraft und einzelnen Mitarbeitern vermitteln. Doch individuelle Profile können nach oben nur kommuniziert werden, wenn die Mitarbeiter auch individuell behandelt werden. Dass ausgerechnet das Fördern von Führungsverhalten eine Null-Korrelation aufweist, erstaunt in diesem Zusammenhang, da gerade die Führungsentwicklung mit übergeordneten Ebenen geplant sein sollte. Der Befund spricht deshalb eher dafür, dass das Führungsverhalten kaum spezifisch gefördert wird.

An diesen Ausführungen wird neben der konkret-inhaltlichen Interpretation deutlich, dass eine multidimensionale Erfassung von Führung unbedingt erforderlich ist (vgl. Kapitel 7.6), weil die interperspektivischen Zusammenhänge interdimensional stark variieren. Von einer einheitlichen Prognoseleistung für verschiedene Erfolgsvariablen ist dann schon gar nicht mehr auszugehen.

7.2.3 Vergleich der interperspektivischen Korrelationen mit denen anderen Stichproben

In diesem Abschnitt werden die berichteten korrelativen Befunde gemäß folgender Aspekte diskutiert: Sie werden zunächst mit Korrelationen verglichen, die meta-analytisch für die allgemeine berufliche Leistungsbeurteilung bereits geschätzt wurden. Anschließend wird der Vergleich zu Befunden aus anderen Einzelstudien gezogen, denen ebenfalls multiperspektivische Beurteilungen für *Management*-Stichproben zugrunde lagen. Auf diese Meta-Analysen und Einzelstudien wurde bereits in Kapitel 3.2 eingegangen.

Wie auch in der Meta-Analyse von Conway und Huffcutt (1997) zur berufliche Leistungsbeurteilung zeigt sich in der vorliegenden Automobil-Führungskräfte-Stichprobe, dass die Abwärts-Aufwärts-Korrelationen höher ausfallen als die Selbst-Fremd-Korrelationen, wobei die Selbst-Abwärts-Korrelationen höher liegen als die Selbst-Aufwärts-Korrelationen. Während die Selbst-Fremd-Übereinstimmungen der vorliegenden Stichprobe verhältnismäßig gering sind (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002), ist die Abwärts-Aufwärts-Übereinstimmung verhält-

nismäßig hoch (Conway & Huffcutt, 1997). Generell liegen die Koeffizienten jedoch in sehr ähnlicher Höhe (Conway & Huffcutt, 1997; Heidemeier & Moser, 2002). Insofern fügen sich die durchschnittlichen interperspektivischen Korrelationen der vorliegenden Studie in die durchschnittlichen Übereinstimmungen anderer Branchen und Leistungsdimensionen ein.

Da allerdings Hinweise darauf bestehen, dass die interperspektivische Übereinstimmung in der Beurteilung von Führungsverhalten nicht ohne Weiteres mit der von anderen Leistungsdimensionen einhergeht,[214] wird im Folgenden ein gezielter Vergleich mit Ergebnissen aus Studien zur Führungsbeurteilung angestellt. Im Zentrum des Interesses steht der Selbst-Aufwärts-Zusammenhang, der bislang erst in *einer* Meta-Analyse zur interperspektivischen Übereinstimmung berücksichtigte wurde (Conway & Huffcutt, 1997). Dass ein diesbezüglicher Erkenntnisgewinn aber erforderlich ist, zeigt die Beratungsrealität, aus der gerade Mitarbeiterbefragungen (Aufwärtsbeurteilungen) nicht mehr wegzudenken sind und in der sich implizite – aber ungenügend empirisch begründete – Annahmen verbreitet haben. So gibt es beispielsweise inzwischen Konsens darüber, dass ein ungenügender Selbst-Fremd-Zusammenhang viele potentielle Gefahren und Misserfolge birgt, die es abzuwenden gilt (vgl. z.B. Kapitel 6.5).

Um den Vergleich gezielt vornehmen zu können, sind in Tabelle 83 solche Studien aufgelistet, die mindestens den Zusammenhang zwischen Selbst- und Aufwärtsbeurteilungen von *Führungsbeurteilungen* angeben. Für Church (1997a, S. 284) ist ein solcher Vergleich zu früheren Studien im Übrigen Rechtfertigung genug bzw. die Rechtfertigung schlechthin, überhaupt Selbst-Fremd-*Korrelationen* zu berechnen. Da Korrelationen standardisierte Maße darstellen, sind sie im Unterschied zu Differenzmaßen – von denen eine Vielzahl existiert – tatsächlich sehr gut für einen Vergleich von Befunden aus verschiedenen Studien geeignet.

[214] Während sich beispielsweise für die berufliche Leistungsbeurteilung im Allgemeinen eher die Tendenz abzeichnet, dass die Selbst-Abwärts-Korrelation höher ist als die Selbst-Aufwärts-Korrelation (Meta-Analyse von Conway & Huffcutt, 1997), gibt es für Management-Stichproben gegenteilige Hinweise (Fleenor et al., 1996).

Tabelle 83. Selbst-Aufwärts-Korrelation der vorliegenden Studie (FzFü) im Vergleich zu anderen Studien mit multiperspektivischer Führungsbeurteilung

Referenzstudie					Vergleichs-parameter			
Studie[a]	r[b]	N	Branche	Beurteilungsdimensionen	Ri[c]	z[d]	p	q
(1)	.12	158	Militär (Marine)	Transformationale Führung	↖	.22	.59	.04
(1)	.19	91	Militär (Ausbildung)	Transformationale Führung	↖	.57	.72	.11
	-.09	91			↙	-1.08	.14	.23
	.35	91			↕	.23	.59	.06
(2)	.25	1.326	Management (öff. und privater Sektor)	Management, Führung, Interaktion	↖	1.05	.85	.18
(3)	.28	92	Führungskräfte (Metallindustrie, Dienstleistung)	Consideration, Initiating Structure, Produktionsorientierung, Handlungsspielraum	↖	1.06	.86	.21
(4)	.13	155	Militär (Marine)	Transformationale & transaktionale Führung	↖	.28	.61	.05
(5)	.25	504	Technologische, pharamzeutische & Flugzeugindustrie	Führung (z.B. Zusammenarbeit, Mitarbeiter- und Aufgabenorientierung)	↖	1.03	.85	.18
(6)	.08	152	Dienstleistung, Beratung	Geschäftssinn, Problemlösen, Prozessverbesserung, Zusammenarbeit	↖	.00	.50	.00
	.19	152			↙	.25	.60	.05
	.27	152			↕	-.14	.44	.03
(7)	.24	2.056	Management (öff. und privater Sektor)	Management, Führung, Interaktion	↖	.99	.84	.17
	.19	2.056			↙	.27	.61	.05
	.41	2.056			↕	.57	.72	.13
(8)	.27	390	Fluglinie	Anregen zu Innovation, Führen/Motivieren, Commitment, Entscheiden, Planen, General Management	↖	1.14	.87	.20
	.04	316			↙	-.52	.30	.10
	.23	318			↕	-.34	.37	.08
(9)	.24	86	Management (Fortune 100)	Mitarbeiterbeziehung	↖	.83	.80	.17
(10)	.39	12	Feuerwehr	Führung[e]	↕	.26	.60	.10
(11)	.21	63	IT-Branche	Transformationale Führung	↖	.64	.74	.13

Anmerkungen. N der Fokus-Führungskräfte, [a](1) Atwater & Yammarino, 1992, (2) Atwater et al., 1998, (3) Baril et al., 1994, (4) Bass & Yammarino, 1991, (5) Church, 1997a, (6) Church, 1997b, (7) Fleenor et al., 1996, (8) Furnham & Stringfield, 1994, (9) London & Wohlers, 1991, (10) Riggio & Cole, 1992, (11) Sosik & Megerian, 1999, [b]mittlere Korrelation, falls mehr als eine Führungsdimension beurteilt wurde, [c] Vergleichsrichtung: ↖ Selbst-Aufwärts-Korrelation wird verglichen, ↙ Selbst-Abwärts-Korrelation wird verglichen, ↕ Abwärts-Aufwärts-Korrelation wird verglichen, [d]z-Wert für Unterschiedstest für unabhängige Korrelationen: $z = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$ (vgl. Bobko, 1995, S. 53f.). Aus der vorliegenden Studie geht die gemittelte Korrelation über die fünf Führungsfaktoren E/F, B/S, IB, K und P ein (identisches N): Selbst-Aufwärts: $\bar{r} = .08$ ($N = 40$), Selbst-Abwärts: $\bar{r} = .14$ ($N = 32$), Abwärts-Aufwärts: $\bar{r} = .30$ ($N = 24$). Negative (positive) z-Werte bedeuten, dass die Korrelation in der Referenzstudie niedriger (höher) ausfällt als in der vorliegenden Studie, [e]die Abwärts-Aufwärts-Korrelation beträgt $\bar{r} = .32$, wenn man alle erfassten Dimensionen (z.B. zusätzlich Planen, Entscheiden) berücksichtigt.

Der Selbst-Aufwärts-Zusammenhang der vorliegenden Studie stimmt am besten mit den Koeffizienten überein, die an militärischen Stichproben (Atwater & Yammarino, 1992;

Atwater et al., 1998; Bass & Yammarino, 1991) bzw. an der Dienstleistungsstichprobe von Church (1997b) gewonnen wurden. Alle anderen angeführten Studien erbringen Selbst-Aufwärts-Zusammenhänge, die bedeutend höher liegen als die der vorliegenden Stichprobe (q = .13 bis q = .21). Zu erwarten waren die größten Ähnlichkeiten zu Koeffizienten aus Industrie-Stichproben (Baril et al., 1994; Church, 1997a; Fleenor et al., 1996; London & Wohlers, 1991). Doch diese liegen trotz unterschiedlicher Beurteilungsinstrumente und Managementdimensionen zwischen r = .24 und r = .28. Sie sind also sehr homogen und deutlich stärker als die durchschnittliche Selbst-Aufwärts-Korrelation der vorliegenden Industrie-Stichprobe.

Möglicherweise ist die Beurteilungsoffenheit und Verantwortlichkeit für Verbesserungsprozesse in deutschen Industrie-Unternehmen wie im US-amerikanischen *Militär* so gering ausgeprägt, dass Führungskräfte (Selbstbeurteiler) oder Mitarbeiter (Aufwärtsbeurteiler) ihre Urteile stark verzerren. Für Letzteres spricht die mehrfach geäußerte Beurteilungsangst vor dem Hintergrund historischer Erfahrungen. Dass generell dennoch eher die Selbstbeurteilungen und weniger die Fremdbeurteilungen „beeinträchtigt“ sind, zeigt sich an der verhältnismäßig hohen Abwärts-Aufwärts-Korrelation. Wenn diese Fremd-Fremd-Korrelation sogar höher ausfällt als in vielen anderen Studien/Stichproben, dann spricht das gegen allgemeine Beurteilungsungenauigkeiten und für Verzerrungen der Selbstbeurteilungen.

Denkbar ist auch, dass die Ähnlichkeit der Korrelationsindizes zu denen des amerikanischen Militärs durch die strukturelle Einbettung der deutschen Automobilindustrie in ein relativ starres interessenpolitisches Korsett begünstigt wird. Im Vergleich dazu scheinen amerikanische und internationale Industriestichproben wie „aus dem Haifischbecken gezogen“, wo sich Veränderungen des Marktes unmittelbarer auf die Beschäftigten auswirken und damit auch deren Beurteilungsverhalten mit anderen Konnotationen verknüpft. Denkbar sind in diesem Zusammenhang aufgrund verschiedener Charakterisitika der Automobilbranche (vgl. Schaller & Jonas, 2004; Wilkens, 1998) also auch branchenspezifische Indizes. Dies wird im Rahmen der Stichprobenselektivität auf Seite 484 erneut diskutiert.

Die Unterschiede zu amerikanischen/internationalen Industrie-Stichproben können auch an den unterschiedlichen Methoden zur Stichprobenrekrutierung liegen. So verwenden beispielsweise Fleenor et al. (1996) nur Daten von Managern, die beim Center for Creative Leadership an einem multiperspektivischen Feedbackprozess teilgenommen haben. Eine derart externe und zugleich renommierte Beratung genießt in der Regel uneingeschränktes Vertrauen, so dass Selbstbeurteilungen nicht aufgrund von befürchteten negativen Konsequenzen der Beurteilung verzerrt werden „müssen“. Das bedeutet außerdem, dass die Manager (Selbstbeurteiler) selbst dafür verantwortlich waren, Fremdbeurteiler zu akquirieren. Hingegen verlief das Vorgehen in der vorliegenden Arbeit umgekehrt, denn es wurden alle Führungskräfte und deren direkten Mitarbeiter eines Unternehmens angefragt. Da die Akquise der Militärstichproben ähnlich durchgeführt wurde, könnte aufgrund der dadurch bedingten Selektion sowie der dadurch bedingten unterschiedlichen Rahmenbedingungen

auch darin eine Erklärung für die ähnlich geringen Selbst-Aufwärts-Korrelationen liegen. Insofern sollte die Akquise und Auswahl der Befragten als potentielle Moderatorvariable für die interperspektivische Übereinstimmung in zukünftigen Studien untersucht werden.

Was zudem in der vorliegenden Untersuchung zu einer Reduzierung der Korrelationen führen konnte, ist die Konfundierung der Beurteilerrollen insofern, dass die Führungskräfte im Extremfall als Aufwärts-, Selbst- und Abwärtsbeurteiler fungierten (vgl. Abbildung 22). In anderen Studien tritt dies kaum auf, da dort entweder nur zwei Beobachtergruppen einander gegenüber gestellt werden oder keine hierarchieübergreifenden Führungskräfte-Stichprobe gezogen wird. Durch eine solche Konfundierung werden jedoch zusätzliche Motive angeregt, die die Beurteilungen verzerren können, beispielsweise das Motiv, Distinktheit zu schaffen bzw. zu vermeiden. Letzteres könnte dazu führen, dass die Selbstbeurteilung tendenziell stärker mit der Abwärtsbeurteilung (Rater jeweils Führungskräfte) als mit der Aufwärtsbeurteilung korreliert.

Doch auch die Art der berücksichtigten Beurteilungsdimensionen kommt als Erklärung dafür in Frage, weshalb die Selbst-Aufwärts-Korrelationen über die Stichproben hinweg schwanken. Beispielsweise setzten Atwater et al. (1998) und Fleenor et al. (1996) „Benchmarks“ ein, also ein Instrument, das in einem von drei Teilen Beziehungsaspekte misst. Eine Beziehung zwischen Führungskraft und Mitarbeiter ist aber von beiden Beteiligten unmittelbarer wahrzunehmen und in der Folge zu beurteilen als Management-Verhaltensweisen, die lediglich die Führungskraft ausführt. Church (1997b), der Verhaltensweisen beurteilen ließ, die nur zu einem kleinen Teil Aspekte der Zusammenarbeit betrafen, findet dem entsprechend auch ähnlich niedrige Korrelationen wie sie hier oder in den Militärstichproben gefunden wurden. Die Erklärung, dass auf unterschiedlichen Dimensionen unterschiedlich gute Beurteilungsübereinstimmungen erzielt werden können, geht auch mit den Daten der vorliegenden Studie einher. Die Korrelationen unterscheiden sich in der vorliegenden Studie über die sieben Führungsdimensionen hinweg (vgl. Tabelle 49). Umgekehrt kann die Ähnlichkeit zu den Militärstichproben aber nicht damit begründet werden, dass der FzFü zum dort verwendeten Instrument (MLQ) mehr Ähnlichkeiten aufweist als zu anderen Instrumenten, weil auch mit dem MLQ prinzipiell höhere Selbst-Aufwärts-Korrelationen zu erreichen sind (z.B. Sosik & Megerian, 1999).

Da die überwiegende Mehrheit der vorliegenden Befunde an US-amerikanischen Stichproben gewonnen wurde, ist auch eine interkulturelle Erklärung denkbar. Nicht unplausibel ist, dass in Abhängigkeit von unterschiedlichen Kulturen und Organisationskulturen auch unterschiedliche Feedback-Kulturen bestehen. So bestätigte die Meta-Analyse von Heidemeier und Moser (2002) die moderierende Wirkung von Kultur auf die Höhe der Selbst-Fremd-Diskrepanzen bei beruflicher Leistungsbeurteilung allgemein (vgl. Kapitel 3.3.1). Studien an deutschen Führungskräftestichproben lassen Koeffizienten für die Selbst-Fremd-Korrelation vermissen (z.B. Tartler, Goihl et al., 2003). Der Fokus liegt dann auf

den Niveau-Unterschieden.[215] Gerade vor dem Hintergrund der zunehmend international agierenden Unternehmen, was bereits der Diversity-Forschung zu einem starken Aufschwung verholfen hat, sollte einer solchen Hypothese in zukünftigen Fragestellungen nachgegangen werden. Möglicherweise muss nämlich das, was Führungskräfte als Indikatoren für ihre Führungsausrichtung heranziehen können, die Reaktionen der Fremdbeurteiler, in unterschiedlichen Kulturen unterschiedlich interpretiert werden. Ohne ein Wissen über interkulturelle Unterschiede und entsprechende Anpassungsstrategien sind bei einem Auslandseinsatz jedoch nicht die gewohnten Erfolge zu erreichen. Allein wegen dieses praxisrelevanten Anliegens sind vermehrt interkulturelle Vergleiche der interperspektivischen Beurteilerübereinstimmung im Bereich der Personalführung dringend erforderlich.[216]

Betrachtet man alle drei Perspektiven-Vergleiche gemeinsam, stimmen die Befunde der vorliegenden Studie am besten mit denen von Church (1997b) überein. Bei einer maximalen Abweichung von $r = .03$ (betrifft die Abwärts-Aufwärts-Korrelation) kann festgehalten werden, dass sich keine der drei interperspektivischen Korrelationen zwischen den beiden Studien unterscheidet. Im Vergleich zum in Kapitel 3.2 dargestellten Forschungsstand kann insbesondere die über die Studien hinweg relativ homogene Selbst-Aufwärts-Korrelation im Bereich von $r = .25$ nicht gefunden werden. Die Selbst-Abwärts- und Abwärts-Aufwärts-Korrelationen scheinen generell stärker zu variieren.

In jedem Fall sind weitere Studien notwendig, um mehr Erkenntnisse über die Generalisierbarkeit der interperspektivischen Übereinstimmung von Führungsbeurteilungen zu erhalten.

7.2.4 Moderatorvariablen für die interperspektivische Übereinstimmung

Die interperspektivische Übereinstimmung selbst ist von verschiedenen Faktoren abhängig, die bis dato für die Übereinstimmung von *Führung*sbeurteilungen kaum untersucht wurden. Lediglich die Tätigkeit als solche – Management- versus Arbeitertätigkeit – wurde als Moderatorvariable mehrfach bestätigt. Da multiperspektivische Feedbacks zur Erhöhung der Übereinstimmung jedoch fast ausschließlich für Führungskräfte angeboten werden, lohnt es aus praktischer Perspektive besonders, genau diese Stichprobe gezielt zu untersuchen. In einer eigenen Fragestellung wurden verschiedene Variablen auf ihre potentiell moderierende Wirkung hin untersucht. Es wurde nicht nur die Beeinflussung der Selbst-Aufwärts- und Selbst-Abwärts-Übereinstimmung, sondern auch die der Abwärts-

[215] Mit solchen Fragestellungen verzichten natürlich auch andere Studien auf die Angabe korrelativer Zusammenhänge (z.B. Van Velsor et al., 1993).

[216] Für andere Beurteilungsbereiche konnte ein interkultureller Effekt auf die interperspektivische Beurteilungsübereinstimmung bereits ermittelt werden. So unterscheidet sich die Selbst-Fremd-Übereinstimmung hinsichtlich der Einschätzung von (nicht-klinischer) Depressivität und Ängstlichkeit zwischen asiatischen und weißen Amerikanern (Okazaki, 2002).

Aufwärts-Übereinstimmung überprüft. Dabei wurden jeweils die korrelative Übereinstimmung und die Niveau-Diskrepanz betrachtet.

Grundsätzlich muss man sich aber auch die Frage stellen, inwieweit Moderatoranalysen für Einzelstudien tauglich sind. So wurden in der vorliegenden Studie bereits verschiedene Variablen umgesetzt, die die Selbst-Fremd-Übereinstimmung erhöhen sollten (Anonymität, Validierungserwartung, verhaltensbezogene Itemformulierung, Ausschließen administrativer Konsequenzen).[217] Geht man davon aus, dass diese Variablen robuste moderierende Bedingungen darstellen (z.B. Heidemeier & Moser, 2002; Mabe & West, 1982), dann erscheint es schwierig, einen Moderatoreffekt weiterer einzelner Variablen zu erkennen. Moser et al. (1994) diskutieren diesen Sachverhalt und kommen zu dem Schluss, dass es in Einzelstudien zu einer Unterschätzung von Moderatoreffekten kommt. Diese Problematik betrifft selbstverständlich beide gängigen Verfahren zur Moderatoranalyse: Mediansplits der potentiellen Moderatorvariable mit anschließendem Vergleich des fokussierten Zusammenhangs zwischen den beiden erstellten Teilstichproben sowie moderierte Regressionen. Deshalb wurden die Moderatorvariablen in Anlehnung an Moser et al. (1994) im Aggregat ausgewertet. Lediglich solche Effekte, die entweder über die verschiedenen Moderatoranalysen hinweg sehr konsistent sind oder die besonders hoch ausfallen, werden im Folgenden behandelt.

McClelland und Judd (1993) führen aus, dass es beispielsweise aufgrund weniger gut kontrollierbarer Bedingungen in Feldstudien schwieriger ist, reliable Moderatoreffekte aufzudecken als in Laborexperimenten – und selbst in Laborexperimenten bewege sich die erklärte Varianz oft nur bei 1%. Vor allem wenn einzelne Prädiktoren eine geringe Varianz aufweisen (hier: Selbstbeurteilung), dann fällt auch die Varianz des Produkts geringer aus. Sie empfehlen, trotz aller Schwierigkeiten moderierte Regressionen durchzuführen und gegebenenfalls höhere Typ-I-Fehler zu akzeptieren. Zumindest plädieren sie dafür, nicht uneingeschränkt Analysen mittels Mediansplit durchzuführen, weil dadurch nicht unbedingt extreme Kategorien entstehen. Nur die Extremisierung, nicht die pure Reduktion der betrachteten Kategorien, ist nämlich vorteilhaft beim Entdecken moderierender Effekte.

Dennoch wurde in der vorliegenden Arbeit überall dort auf Mediansplits zurückgegriffen, wo keine Annahme darüber vorlag, welche der Variablen, deren Zusammenhang fokussiert wurde, als Kriterium und welche als Prädiktor fungiert. Für die Untersuchung, ob die Stärke des Selbst-Fremd- oder Fremd-Fremd-Zusammenhangs von der Ausprägung einer Drittvariable beeinflusst wird, erscheint es sowohl für die Selbst- als auch Fremdbeurteilung unbegründbar, sie als Prädiktor für die jeweils andere Beurteilungsperspektive zu konzeptualisieren. Einem Regressionsmodell liegt eine solche Annahme aber in aller Regel zugrunde, auch wenn es natürlich nicht dazu imstande ist, sie im kausalen Sinne zu prüfen. Mediansplits haben zusätzlich den Vorteil, dass sie die Richtung der Effekte anschaulicher illustrieren. Für die vorliegende Fragestellung nach Moderatorvariablen für den Selbst-

[217] Ein solches Design ist in der angewandten Forschung dem Auftraggeber geschuldet.

Fremd-Zusammenhang interessiert außerdem nicht nur, *ob* ein signifikanter moderierender Effekt vorliegt, sondern es interessiert auch die Stärke des Selbst-Fremd-Zusammenhangs in den beiden Gruppen. Es wird dabei in Kauf genommen, dass durch die Dichotomisierung in zwei Subgruppen der Kontinuitätscharakter der Moderatorhypothesen deutlich reduziert wird, was zu einer Verringerung der Power führt (vgl. Stone-Romero & Anderson, 1994).

Für folgende Variablen ergab sich in allen Analysen die erwartete Richtung: Alter der Fremdrater, Führungsspanne, ergebnisbezogene Itemformulierung, Beteiligung an der multiperspektivischen Beurteilung (Fremd- und Selbstrater) und Kenntnis der Führungsgrundsätze. Deren moderierende Wirkung kann auf mehr Beurteilungserfahrung, eine bessere Beobachtungsgrundlage, weniger Interpretationsspielraum aufgrund höherer Klarheit des Beurteilungsgegenstandes, ein generell offenes Beurteilungsklima beziehungsweise eine ähnlichere Definition von Führung zurück geführt werden.

Überdies konnten insbesondere für die Abwärts-Aufwärts-Übereinstimmung starke moderierende Effekte für einzelne Variablen aufgedeckt werden. Da die Abwärts-Aufwärts-Übereinstimmung gleichzeitig besonders stark mit den Erfolgskriterien kovariiert (vgl. Kapitel 7.5), ist die Betrachtung von Variablen, die die Übereinstimmung verbessern, auch für die Praxis ein Gewinn. Dies gilt jedenfalls dann, wenn sich zukünftig herausstellt, dass der negative Zusammenhang zwischen der Fremd-Fremd-Diskrepanz und Erfolg zumindest teilweise darauf beruht, dass die Diskrepanz den Erfolg schmälert (und nicht umgekehrt).

Betrachtet man die Abwärts-Aufwärts-Übereinstimmung, dann gilt zum einen, dass für sie stärkere Moderatoreffekte festgehalten werden konnten als für die Selbst-Fremd-Übereinstimmung. Für die beiden untersuchten Selbst-Fremd-Zusammenhänge unterscheiden sich im Übrigen die moderierenden Einflüsse in der vorgestellten Studie kaum. Auch bei Ostroff et al. (2004) ergaben sich für die drei untersuchten Selbst-Fremd-Übereinstimmungen keine bedeutenden Unterschiede, moderierende Einflüsse auf Fremd-Fremd-Zusammenhänge wurden dort nicht untersucht.

Zum anderen wurden für den *korrelativen* Abwärts-Aufwärts-Zusammenhang mehr Effekte extrahiert als für die Abwärts-Aufwärts-*Niveau*-Diskrepanzen. Dies läuft entgegen der Verteilung bei den Selbst-Fremd-Zusammenhängen. Dafür konnten – wie auch bei Heidemeier und Moser (2002) – mehr und stärkere Effekte für die Niveau-Unterschiede als für die Korrelationen aufgedeckt werden. Dieser Unterschied zwischen den Einflussbedingungen auf Selbst-Fremd- und Fremd-Fremd-Übereinstimmungen kann so interpretiert werden, dass sich die beiden Fremdbeurteilungen stärker durch eine *systematische* Verzerrung unterscheiden. Dies zeigen auch die im Vergleich zu den Selbst-Fremd-Korrelationen höheren Fremd-Fremd-Korrelationen. Dies würde dafür sprechen, dass die Selbst-Fremd- und die Fremd-Fremd-Übereinstimmungen durch unterschiedliche Mechanismen zustande kommen. Wenn sie deshalb aber auch jeweils neue Information zur Erfolgsaufklärung lie-

fern können, ist es umso wichtiger, künftig auch die Fremd-Fremd-Übereinstimmung zu betrachten.

Für folgende Variablen konnten (neben dem Führungsverhalten, dazu Seite 452) große Effekte ($q > .50$) auf die Abwärts-Aufwärts-Korrelation ermittelt werden: Alter (Abwärtsrater), Dauer der Organisationszugehörigkeit (Aufwärtsrater), Dauer der Abteilungszugehörigkeit (Aufwärtsrater), Dauer der Zusammenarbeit von Selbst- und Aufwärtsrater, Beteiligung des Ratees und Kenntnis der Führungsgrundsätze (beide Fremdrater). Dass insbesondere die Zugehörigkeitsvariablen Effekte zeigen, verweist darauf, dass im Laufe der Zugehörigkeit Organisationsnormen übernommen werden, vor deren Hintergrund ähnliche Führungsinterpretationen und folglich -beurteilungen abgegeben werden. Die Selbstbeurteilungen scheinen dadurch weniger stark beeinflusst zu werden, da für die Selbst-Fremd-Übereinstimmung kaum entsprechende Effekte ausgemacht werden konnten. Wenn man in der Praxis die Wahl hat, sollte man vor diesem Hintergrund also verstärkt die „alten Hasen“ zur Fremdbeurteilung heranziehen.

Weiter fällt auf, dass die Variablen nicht ausschließlich die beteiligten Rater betreffen (Abwärts und Aufwärts), sondern bei zwei Variablen auch die Ratees involviert sind. So ist anzunehmen, dass die Dauer der Zusammenarbeit zwischen *Selbst*- und Aufwärtsrater deshalb auf die Übereinstimmung zwischen Abwärts- und Aufwärtsbeurteilungen Einfluss nimmt, weil die Aufwärtsbeurteilungen aufgrund der besseren Beobachtungsgrundlage reliabler sind. Außerdem kann die Beteiligung des Ratees (neben Aufwärts- und Abwärtsrater) so interpretiert werden, dass in der Abteilung Offenheit für Feedback besteht und deshalb Fremdratings nicht aus Angst vor Konsequenzen strategisch verzerrt werden „müssen“. Zudem ist anzunehmen, dass es unter solchen Bedingungen generell mehr Austausch innerhalb der Abteilung gibt, so dass Bedürfnisse, Erwartungen und damit auch Definitionen und Wichtigkeiten von Führung besser aneinander angeglichen sind.

Von besonderem Interesse in der vorliegenden Arbeit war die moderierende Wirkung der Führung selbst. Dazu wurden verschiedenste Facetten von Führungsverhalten durch den MLQ (Felfe & Goihl, 2003) erfasst. Es zeigte sich zwar, dass die Abwärts-Aufwärts- sowie die Selbst-Aufwärts-Diskrepanz bei hoher transformationaler und aktiver Führung geringer und bei hoher passiver Führung größer sind, für die Selbst-Abwärts-Diskrepanz gilt das aber nur ansatzweise, für die korrelativen Zusammenhänge gar nicht. Bei genauerer Betrachtung stellte sich heraus, dass die Niveau-Unterschiede nur deshalb bei stark transformationaler Führung sinken, weil die Aufwärtsbeurteilungen nicht nur im MLQ, sondern auch im FzFü positiver ausfallen. Nach den hier erbrachten Ergebnissen ist allenfalls von einem geringen Effekt auszugehen, mit dem die transformationale Führung den interperspektivischen Zusammenhang von Führungsbeurteilungen (FzFü) moderiert. Vor diesem Hintergrund sind auch Interpretationen anderer Studien (z.B. Tartler, Goihl et al., 2003) zu überdenken, in denen die geringeren Selbst-Aufwärts-Diskrepanzen der transformationalen Führung bei gleichzeitig stärkerer transformationalen Führung auf eben letztere zurückgeführt werden. Eine deutlicherer Effekt auf die interperspektivische Übereinstim-

mung zeigt sich für das Ausmaß der partizipativen Führung. Generell muss in diesem Zuge angeprangert werden, dass die transformationale Führung keine „Wunderdroge“ zur Heilung aller „offenen Wunden“ der Führungsforschung darstellt.

Da die Erkenntnis zentraler Moderatorvariablen für die Führungsforschung gleichwohl von Belang ist, da gerade robuste moderierende Effekte dazu beitragen, Erklärungsmechanismen für die geringen Übereinstimmungen zu stützen, zu integrieren oder zu verwerfen, werden im Folgenden weitere konkrete Forschungsfragen für diesen Bereich aufgeworfen.

Möglicherweise hat das Geschlecht der Führungskraft einen Einfluss auf die interperspektivische Übereinstimmung. Eine solche Annahme steht deshalb nicht im Widerspruch zum Befund von Heidemeier und Moser (2002), die in ihrer Meta-Analyse keinen Einfluss auf die Selbst-Abwärts-Übereinstimmung bei beruflicher Leistungsbeurteilung allgemein fanden, weil es gerade die *Führungstätigkeit* ist, die von Frauen anders ausgeübt wird als von Männer. Meta-analytischen Befunden zufolge führen Frauen demokratischer und partizipativer als Männer (Eagly & Johnson, 1990), was über einen intensiveren Austausch mit den Geführten zu erhöhter Beurteilungsübereinstimmung führen sollte. Eine Überprüfung dieses potentiellen geschlechtsspezifischen Moderatoreffekts war aufgrund der geringen Anzahl an weiblichen Führungskräften nicht möglich (N = 5 Selbstbeurteilerinnen, wovon für 4 Aufwärtsbeurteilungen abgegeben wurden).

Während in der vorgestellten Arbeit lediglich einzelne moderierende Effekte fokussiert wurden, ist es durchaus plausibel, dass verschiedene Moderatoreffekte gemeinsam wirken. Ihre multiple Wirkung ist zukünftig an größeren Stichproben zu untersuchen. Außerdem ist zu ergründen, durch welche Variablen eher die korrelative, interperspektivische Übereinstimmung gesteigert beziehungsweise die Niveau-Unterschiede verringert werden können. Spezifische Einflussbedingungen – wie sie sich in der vorliegenden Studie genauso wie bei Heidemeier und Moser (2002) andeuteten – bergen besonderen Erklärungsgewinn für die Entstehung der geringen Übereinstimmung.

Verstärktes Augenmerk kann zudem auf diejenigen Variablen gelegt werden, die an sich bereits eine Ähnlichkeit beziehungsweise Passung zwischen den Beurteilern ausdrücken. Dabei kann an die Arbeit von Ostroff et al. (2004) angeschlossen werden, die als Moderatorvariablen die Gleichheit der Beurteiler hinsichtlich Geschlecht, Alter und Rasse untersuchten. Dass diese globalen Ähnlichkeitsmaße fast durchgängig ohne Auswirkung auf die Beurteilerübereinstimmung blieben, verwundert zumindest rückblickend nicht. Auf einer Ebene, die sich auch im Führungsprozess selbst auswirkt (z.B. Ähnlichkeit von Kompetenzen oder beruflichen Zielen), ist durchaus eine stärkere moderierende Wirkung vorstellbar. Darauf verweisen auch die Aussagen der Befragten, die persönliche Inkongruenzen für die Selbst-Aufwärts-Diskrepanzen mitverantwortlich machen (vgl. Kapitel 6.6.1 und Interviewstudie von Senger, in Vorb.).[218]

[218] Ziel dieser qualitativen Studie war es, Erklärungsansätze für die Entstehung der Selbst-Aufwärts-Diskrepanzen zu generieren und zu vertiefen. In halbstrukturierten Interviews wurden viele der in Kapitel 3.4

Desweiteren sind verschiedene Variablen, die im Kontext beruflicher Leistungsbeurteilung Wirkung zeigen (vgl. Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002; Mabe & West, 1982), auch auf ihre Wirkung im Kontext der Führungsbeurteilung hin zu überprüfen. So kann beispielsweise das Wissen, ob umfeld- oder aufgabenbezogene Leistungen (vgl. Borman & Motowidlo, 1993) übereinstimmender beurteilt werden können, für die Fragebogenkonstruktion genauso nützlich sein wie für die Ableitung von Entwicklungsinterventionen.

7.2.5 Above-Average-Effekt

Wenn an dieser Stelle die Güte der Führungsbeurteilungen diskutiert wird, dann muss auch zur Sprache kommen, wie Führungskräfte ihr eigenes Führungsverhalten im Vergleich zum Verhalten anderer Führungskräfte beurteilen. Damit ist der Above-Average-Effekt angeführt, der durch die bekannt gewordene Arbeit von Meyer (1980) in die Führungsforschung transferiert wurde. Ein direkter Vergleich der Ergebnisse der vorgestellten Studie mit den Befunden traditioneller Above-Average-Studien (z.B. Meyer, 1980) ist jedoch aufgrund unterschiedlicher Skalenformate nicht möglich. Während normalerweise ein Vergleich zur *durchschnittlichen* Führungskraft gezogen werden muss – also zu einem sehr abstrakten Referenzpunkt – wurde in der vorliegenden Studie eine konkrete, bekannte Person aus dem unmittelbaren Arbeitsumfeld beurteilt, und zwar ohne direkte Vergleichsinstruktion und in einem eigenen Fragebogen. Damit ist zwar nicht mehr der Above-Average-Effekt im klassischen Sinne untersucht. Doch die Ergebnisse, die zwar weniger imposant ausfallen als bei Meyer (1980), aber dennoch eine relativ zu anderen bessere Selbstbewertung abbilden, untermauern dann umso mehr, dass auch im Führungsbereich von einer Selbstüberschätzung auszugehen ist. Denn im Unterschied zur Verwendung relationaler Skalen kann durch die getrennten Skalen in der vorliegenden Studie gewährleistet werden, dass allein die relative Abwertung anderer die „Selbstüberschätzung“ nicht ausmacht. Insgesamt liegt also eine deutlich konservativere Schätzung des Above-Average-Effekts vor, zumal aus Laborstudien bekannt ist, dass ein Vergleich mit konkreten und bekannten Personen dazu führt, dass die Selbstüberschätzung abnimmt (Alicke et al., 1995).

Ungeklärt bleibt, weshalb es trotz dieser Bedingungen zu einer deutlichen Selbstüberschätzung der Führungskräfte kommt. Ob auf einer weniger expliziten Ebene Vergleichsprozesse angestoßen werden, könnte durch eine zeitversetzte Bearbeitung der Fragebögen zur Beurteilung der eigenen und der fremden Führungsverhaltensweisen besser kontrolliert werden. Darin, dass die Vergleichsintention bei nicht-relationalen Skalen nicht derart offensichtlich ist, liegt eine Chance, die potentielle Selbstüberschätzung auf anderem Wege

vorgestellten potentiellen Erklärungsansätze weiterführend exploriert. Für die Durchführung der Interviews im Rahmen ihrer Diplomarbeit wird Nadja Senger ausdrücklich gedankt (Ergebnisse vgl. Senger, in Vorb.).

zu untersuchen. Die Erfassung von zusätzlichen Kontextbedingungen auf Perönlichkeits-, Gruppen- und Organisationsebene könnte Aufschluss über begünstigende Bedingungen geben. Die Forschung zum Above-Average-Effekt im Bereich des Führungsverhaltens kann zudem dadurch vorangetrieben werden, dass die Vergleichsinstruktionen manipuliert werden.

Vor dem Hintergrund, dass bei Führungsbeurteilungen immer mit Verzerrungen gerechnet werden muss, ist zu fragen, inwiefern sie Kriteriumsvalidität besitzen. Gerade weil sich die Verzerrungen in der mangelnden Übereinstimmung zwischen Beurteilungen aus verschiedenen Perspektiven ausdrücken, wurde die interperspektivische Übereinstimmung lange Zeit nur unter dem Gesichtspunkt ihrer Fehlerbehaftetheit betrachtet. Erst in den letzten 10 Jahren kam es insofern zu einem Umdenken, dass das Ausmaß der Übereinstimmung als zusätzliche Information gewertet wurde (vgl. Tornow, 1993). Bevor die Kriteriumsvalidität in den Kapiteln 7.4 und 7.5 diskutiert wird, folgt eine Auseinandersetzung mit potentiellen Erklärungsmechanismen für das Zustandekommen der mangelnden Selbst-Fremd-Übereinstimmung (Kapitel 7.3).

7.3 Wie kann die geringe interperspektivische Übereinstimmung in der Führungsbeurteilung erklärt werden?

Auch wenn die vorliegende Studie nicht explizit darauf ausgerichtet war, Erklärungsmechanismen zu überprüfen, und das Design der Studie keine Prüfung kausaler Mechanismen ermöglicht, erbringen die Befunde *indirekte Hinweise* für einzelne Erklärungen. So wird in diesem Kapitel vor allem unter Rückgriff auf die Ergebnisse von Moderatoranalysen zur interperspektivischen Übereinstimmung diskutiert, welche potentiellen Entstehungsmechanismen für die geringen interperspektivischen Beurteilerübereinstimmung *plausibel* sind. Dazu gehört im Grunde auch die Untersuchung, durch welche Perspektiven die beobachteten Diskrepanzen hauptsächlich bedingt werden und ob die Daten überhaupt für perspektivenbezogene Rating-Anteile (vs. alleinig idiosynkratische Anteile) sprechen.

Es hat sich gezeigt, dass sich die Selbst-Fremd-Diskrepanzen nicht allein durch verzerrte Beurteilungen aus einer Perspektive ergeben, sondern jeweils beide beteiligten Beurteilungsperspektiven für hohe Diskrepanzen mitverantwortlich sind (vgl. Tabelle 76). Deshalb sind diejenigen Erklärungsansätze zumindest ergänzungsbedürftig, die allein auf die Verzerrung von Selbstbeurteilungen abzielen. In Interviews (vgl. Fußnote 218) wurde sehr deutlich beschrieben, dass insbesondere die Aufwärtsbeurteilungen strategisch eingesetzt werden. Dabei können unterschiedliche Motive aktualisiert werden. Sie variieren zwischen den extremen Polen der Strenge und Milde, also beispielsweise den Fragebogen einerseits als eine Möglichkeit zu nutzen, seiner Unzufriedenheit Ausdruck zu verleihen und ihn andererseits als Möglichkeit zu nutzen, Lorbeeren an die Führungskraft zu verteilen, um weiterhin in ihrer Gunst zu stehen. Auch die Moderatorwirkung der Wichtigkeitseinschätzung der Aufwärtsbeurteiler spricht für die Verzerrung der Aufwärtsratings. Je wichtiger die Einschätzung ausfällt, umso geringer sind die Selbst-Aufwärts- und Abwärts-Aufwärts-Korrelationen und umso höher die Selbst-Aufwärts-Niveau-Unterschiede (vgl. Kapitel 6.3.3).

Als allgemeiner Hinweis auf das Übergewicht perspektivenbezogener gegenüber idiosynkratischer Verzerrungen kann gewertet werden, dass die Beurteilungen aus einer Perspektive stärker miteinander korrelieren als die aus verschiedenen Perspektiven. Allerdings ist auch von idiosynkratischen Verzerrungen auszugehen, da die korrelative Übereinstimmung nur gering (negativ) mit dem Differenzmaß einhergeht.

Es fällt auf, dass die Aufwärts- und Abwärtsratings für das Führungsverhalten von Überschätzern (sowohl Selbst-Aufwärts- als auch Selbst-Abwärts-Diskrepanz) eine deutlich höhere Varianz aufweisen als die Fremdratings für das Führungsverhalten von Übereinstimmern und Unterschätzern (vgl. Tabelle 76). Dies verweist darauf, dass Überschätzer ein Führungsverhalten zeigen, das besonders schlecht bewertet werden kann. Möglicherweise ist es sehr inkonsistent über die Zeit oder über Personen hinweg. Sie sind jedenfalls

weniger „greifbar“ als Unterschätzer, für die ein ähnliches Ausmaß an Diskrepanz vorliegt. Hier könnten Beobachtungsstudien aus der „work activity“-Forschung (vgl. Kapitel 2.3.6.2) mehr Aufschluss bringen.

Im Folgenden werden die in Kapitel 3.4 dargestellten, potentiellen Erklärungsmechanismen auf ihre Plausibilität hin untersucht (Reihenfolge siehe dort).

Unterschiedliche Verhaltensstichproben, mangelnde Beobachtbarkeit und Intransparenz. In Abbildung 37 ist dargestellt, dass die Selbst-Aufwärts-Diskrepanz dann deutlich geringer ausfällt, wenn die Aufwärtsbeurteiler eigene Erfahrungen in der Führungsrolle haben, wenn sie also kontinuierlich Führungsverantwortung tragen oder als Stellvertreter fungieren und dann zumindest vorübergehend Führungsverantwortung übernehmen (vgl. Kapitel 6.3.3). Für Ersteres zeigt sich ein sehr starker Effekt.

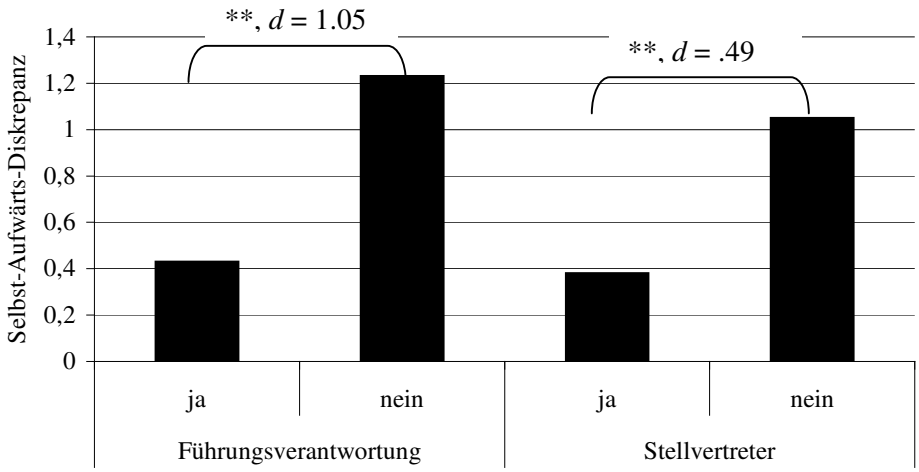


Abbildung 37. Selbst-Aufwärts-Diskrepanz in Abhängigkeit von Erfahrungen des Aufwärtsbeurteilers in der Führungsrolle (eigene Führungsverantwortung bzw. Stellvertreterfunktion)

Dass aber auch der Effekt für die Stellvertretung stabil ist, zeigt die dimensionale Auswertung. So liegt das Urteil derjenigen Mitarbeiter, die ihre Führungskraft bei deren Abwesenheit vertreten (vs. nicht vertreten), für alle sechs Führungsdimensionen[219] näher an der Selbstbeurteilung. Für E/F ($t_{[139]}$ = -2.57, p = .01, d = -.44), B/S ($t_{[139]}$ = -2.17, p = .03, d = -.37), IB ($t_{[139]}$ = -1.80, p = .07, d = -.31), K ($t_{[139]}$ = -2.51, p = .01, d = -.43), P ($t_{[139]}$ = -3.59, p < .01, d = -.61) und FvF ($t_{[139]}$ = -1.60, p = .11, d = -.27) ergeben sich geringe bis mittlere Effekte. Dass eigene Erfahrungen als Führungskraft die Beurteilungen von Führungsverhalten also genauer ermöglichen, kann so interpretiert werden, dass Stellvertreter einen guten Einblick in die Führungsaufgaben bekommen. Dies macht die Erklärung plausibel, dass die *Verhaltensstichprobe*, die sie ihrer Aufwärtsbeurteilung zugrunde legen,

[219] Für die siebte Führungsdimension – die Stellvertretung – existiert naturgemäß keine Teilstichprobe ohne Stellvertretungserfahrung.

besser mit der ihrer Führungskraft (Selbstbeurteiler) übereinstimmt. Außerdem sind Entscheidungsprozesse für sie *transparenter*. Generell ist also durch dieses bessere Verständnis der Führungsaufgaben eine bessere *Beobachtbarkeit* der Führung gegeben. Es ist plausibel, wenn dadurch eine geringere Diskrepanz entsteht. Für diese Erklärung spricht zudem, dass diejenigen Mitarbeiter, die die Führungsdimension Stellvertretung unterdurchschnittlich (vs. überdurchschnittlich) bewerten,[220] auf den anderen Führungsfaktoren höhere Diskrepanzen zur Selbstbeurteilung aufweisen. Im Unterschied zu den Niveau-Unterschieden fallen die korrelativen Zusammenhänge nur für B/S, P und FvF tendenziell höher aus, wenn der Aufwärtsbeurteiler die Stellvertreterrolle erfahren hat vs. nicht erfahren hat. Dieser Fragestellung kann in Folgeuntersuchungen gezielter nachgegangen werden, indem spezifischere Informationen zu den Stellvertreter-Erfahrungen erfasst werden.

Um Führungsverhalten überhaupt so umfassend *beobachten* zu können, dass eine genaue Beurteilung abgegeben werden kann, ist persönlicher Kontakt zur Führungskraft notwendig (vgl. auch Tabelle 81). Als grobes Maß für ausreichend persönlichen Kontakt kann die Führungsspanne herangezogen werden, die mit zunehmender Größe dyadische Interaktionen einschränkt. Plausibel wird die Wirkung dieser Erklärung, da sowohl für die Führungsspanne des Ratees als auch für die des Abwärtsraters starke moderierende Effekt auf verschiedene interperspektivische Diskrepanzen und Korrelationen festgehalten wurden (Kapitel 6.3.3). Auch die stark moderierende Wirkung durch ergebnisbezogene Itemformulierungen kann in die Beobachtbarkeits-Erklärung integriert werden, denn diese Items beinhalten Führungssequenzen, deren Ergebnisse unmittelbar beobachtbar sind. Die Wirkung der ergebnisbezogenen Itemformulierung spricht genauso wie die hohe Selbst-Aufwärts-Korrelation auf dem Faktor Klarheit aber nicht nur für die Erklärung der besseren Beobachtbarkeit generell, sondern auch dafür, dass verschiedene Beurteiler die Möglichkeit haben, ihren Beurteilungen auf ähnlichere *Verhaltensstichproben* zu stützen.

Dass die Selbst-Aufwärts-Diskrepanzen abnehmen, je zufriedener die Mitarbeiter (Aufwärtsbeurteiler) mit ihrer Führungskraft (Selbstbeurteiler) sind, kann damit erklärt werden, dass dann auch Entscheidungsprozesse besser nachvollziebar sind.[221] Der Zusammenhang zur Zufriedenheit mit der Führungskraft ($\bar{r}$ = -.66) ist auch deutlich höher als der Zusammenhang zu den beiden gesamten Arbeitszufriedenheitsskalen (AZ1 [ohne Item zur Zufriedenheit mit Führung]: $\bar{r}$ = -.52, AZ2: $\bar{r}$ = -.14).[222]

[220] In diesem Fall ist tatsächlich von einer negativen Handhabung der Stellvertretung auszugehen (und in der Folge von weniger Gelegenheit, Führung zu erleben), da die Aufwärtsbeurteilungen vor allem auf der Dimension Stellvertretung mit den Abwärtsbeurteilungen hoch korrelieren.

[221] Entscheidungsprozesse sollten nur dann keinen Einfluss auf die Arbeitszufriedenheit nehmen, wenn fast alle Entscheidungen aufgrund hoher Ressourcen positiv ausfallen.

[222] Die Zufriedenheit mit der Führungskraft kann als Einzelitem aus der Skala AZ1 operationalisiert werden (N = 141/79/36) oder als Gesamtmittelwert der Skala SAT (N = 136/77/34). Die mittlere Korrelation mit den Selbst-Aufwärts-Differenzen beträgt jeweils $\bar{r}$ = -.66, was im Übrigen die Konstruktvalidität unterstreicht.

Dadurch, dass für die Dauer der Zusammenarbeit, in deren Verlauf Führungsverhalten zunehmend häufiger beobachtet werden kann, allerdings kaum erwartete Moderatoreffekte gefunden wurden, wird die Wirkungsannahme der Beobachtbarkeits-Erklärung abgeschwächt.

Unterschiedliche Definitionen von Führung. Betrachtet man die Idealvorstellung von Führung als eine Art subjektive Definition von Führung, dann kann die Moderatorwirkung, die die Diskrepanz der Idealvorstellungen auf die Selbst-Aufwärts-Übereinstimmung ausübt, als Beleg dafür gewertet werden, das Zustandekommen der mangelnden Übereinstimmung durch unterschiedliche Führungsdefinitionen erklären zu können.

Genauso passen zu dieser Erklärung die Moderatorwirkungen der Zugehörigkeitsvariablen (Dauer der Zugehörigkeit zu Organisation, Abteilung bzw. Zusammenarbeit mit Führungskraft), denn im Laufe der Zeit sollten sich die Vorstellungen über Führung einander angleichen. Entsprechende Moderatoreffekte wurden für die Selbst-Abwärts- und die Abwärts-Aufwärts-Übereinstimmung, nicht jedoch für die Selbst-Aufwärts-Übereinstimmung festgestellt (Kapitel 6.3.3). Dies ist deshalb plausibel, weil es mehr Zeit und Kontakt braucht, bis sich Führungsvorstellungen über verschiedene (vs. nur eine) Hierarchien hinweg angenähert haben.

Durch die Beurteilungserfahrung (Selbst- und Aufwärtsbeurteiler), mit der unter anderem Führungserwartungen ausgetauscht werden, wird auch die Selbst-Aufwärts-Übereinstimmung moderiert. Zudem kann die bereits oben beschriebene Moderatorwirkung der Kenntnis der Führungsgrundsätze im Sinne dieses Erklärungsmechanismus interpretiert werden, da dann davon auszugehen ist, dass die beteiligten Beurteiler ähnliche Führungstheorien mit dem mehr oder weniger expliziten Referenzwert des Unternehmens-Solls verfolgen.

Dass die Selbst-Abwärts-Korrelationen höher und die Selbst-Abwärts-Niveau-Unterschiede geringer ausfallen als die entsprechenden Werte der Selbst-Aufwärts-Übereinstimmung kann ebenfalls vor dem Hintergrund dieser Erklärung betrachtet werden, da Führungskräfte (unabhängig von ihrer Hierarchieebene) mit hoher Wahrscheinlichkeit einander ähnlichere Definitionen von Führung haben als Führungskräfte und Mitarbeiter.

Unterschiedliche Gewichtungen von Führungsverhaltensweisen. Die Frage danach, ob unterschiedliche Gewichtungen von Führungsverhaltensweisen zu geringer interperspektivischer Übereinstimmung führen, geht zunächst davon aus, dass überhaupt unterschiedliche Gewichtungen vorliegen. Es zeigen sich für die fünf Führungsfaktoren E/F, B/S; IB, K und P Effekte dahingehend, dass sie von den Führungskräften durchgehend als wichtiger erachtet werden als von den Aufwärtsbeurteilern ohne eigene Führungsverantwortung

(d = .20 bis d = .38).[223] Neben diesen geringen Effekten auf den einzelnen Führungsdimensionen zeigt sich, dass die Führungsdimensionen selbst unterschiedlich stark gewichtet werden. In Regressionen für den Gesamtscore (Ist-Beurteilung) auf die einzelnen Führungsfaktoren (Ist-Beurteilung) erhalten diese in Abhängigkeit von der Beurteilungsrichtung unterschiedlich starke Beta-Gewichte. Relativ ähnlich ist die Gewichtung der Faktoren aus der Selbst- und Abwärtsperspektive. Während E/F und B/S jeweils die stärksten Gewichte erhalten, erhält P das geringste. Hingegen erhält aus der Aufwärtsperspektive B/S ein extrem starkes Gewicht, die Gewichte für E/F und P liegen etwa halb so hoch und die für IB und K sehr gering. Dies kann auch eine Erklärung dafür liefern, weshalb die Selbst-Abwärts-Übereinstimmung höher ist als die Selbst-Aufwärts-Übereinstimmung.

Dass die Wichtigkeitseinschätzung aber überhaupt Einfluss auf die Ist-Beurteilung hat, zeigen die Korrelationen zwischen der Ist- und Wichtigkeitsbeschreibung. Die Wichtigkeitsbeschreibungen einer Person korrelieren mit ihren Führungsbeurteilungen (Ist). Die Korrelationen für verschiedene Führungsdimensionen liegen bei den Selbstbeurteilungen überwiegend im mittleren und hohen Bereiche (r = .14 bis r = .64), bei den Aufwärtsbeurteilungen überwiegend im geringen und mittleren Bereich (r = .15 bis r = .58). Diese positiven Korrelationen können beispielsweise durch den Ankereffekt oder die Verfügbarkeitsheuristik erzeugt werden. Im Rahmen der Selbstwerttheorien werden sie auf Seite 462 erneut aufgegriffen.

Unterschiedliche Differenzierungen von Selbst- und Fremdbild. Dass die Skaleninterkorrelationen der Selbstbeurteilungen geringer sind als die der Fremdbeurteilungen (vgl. Kapitel 6.1.1.1), kann als Beleg dafür herangezogen werden, dass die Selbst- und Fremdwahrnehmung unterschiedlich stark zwischen verschiedenen Führungsverhaltensweisen differenziert. Es ist naheliegend, dass unterschiedlich differenzierte Angaben nicht besonders hoch miteinander korrelieren. Dass die Interkorrelationen der Abwärtsbeurteilungen fast so hoch wie die der Aufwärtsbeurteilungen ausfallen, könnte eine höhere Abwärts-Aufwärts-Übereinstimmung mitbedingen. Allerdings ist dann nicht erklärbar, weshalb die Selbst-Abwärts-Übereinstimmung besser gelingt als die Selbst-Aufwärts-Übereinstimmung.

Für Items, die eine starke Vorgabe des Differenzierungsgrads geben (ergebnisbezogene Formulierung) und dadurch stärker ausschließen, dass die Selbstbeurteiler Informationen aus spezifischen Führungs- oder sogar anderen Lebensbereichen abrufen, wurde entsprechend der Erwartung ein höherer interperspektivischer Zusammenhang festgestellt. Umgekehrt kann auch für die Variablen, die die Fremdbeurteilungen aufgrund besserer Informationsgrundlage spezifischer werden lassen (Stellvertretung, eigene Führungsverantwortung, geringe Führungsspanne), eine Moderatorwirkung festgestellt werden.

[223] Zur Erinnerung: Die Wichtigkeitseinschätzung wurde lediglich in die Aufwärtsbeurteilungs-Version integriert.

Akteur-Beobachter-Verzerrung. Dass die Fremdbeurteiler untereinander stärker korrelieren als sie jeweils mit der Selbstbeurteilung korrelieren, ist am besten durch den Actor-Observer-Bias erklärbar. Auch die Moderatorwirkung von Stellvertretung und eigener Führungserfahrung der Aufwärtsbeurteiler (auf den Selbst-Aufwärts-Zusammenhang) kann auszugsweise so erklärt werden, da die Fremdbeurteiler dann neben den Informationen aus der Beobachterrolle auch Informationen aus der Akteurrolle abrufen können. Analog dazu wird auch innerhalb dieses Erklärungsansatzes plausibel, weshalb die Selbst-Abwärts-Übereinstimmung höher ist als die Selbst-Aufwärts-Übereinstimmung. Jedoch müsste die Selbst-Abwärts-Übereinstimmung dann auch höher sein als die Abwärts-Aufwärts-Übereinstimmung, was nur für die Niveau-Unterschiede, nicht für die korrelative Übereinstimmung zutrifft.

Sozialer Vergleich und soziale Identität. Im Sinne des sozialen Vergleichs kann die positive Selbstbewertung und die relativ dazu negativere Bewertung des Führungsverhaltens anderer Führungspersonen erklärt werden – vorausgesetzt, die sozialen Vergleichsprozesse entspringen einem Motiv nach Selbstwertschutz. Weiter können Assimilationseffekte dafür verantwortlich gemacht werden, dass die Abwärtsbeurteiler ähnlich milde Ratings abgeben wie die Selbstbeurteiler (und die Selbst-Abwärts-Diskrepanz deshalb gering ist), da davon auszugehen ist, dass sie sich in den zu beurteilenden Dimensionen als ähnlich zu den Selbstbeurteilern wahrnehmen.

Legt man den Fokus nicht auf interpersonale, sondern auf intergruppale Vergleiche, dann ist zunächst festzuhalten, dass in der vorliegenden Untersuchung alle Selbst- und alle Abwärtsbeurteiler *Führungskräfte* sind. Demgegenüber werden Aufwärtsbeurteilungen aus der Geführtenperspektive erbracht. *Soziale Kategorisierungsprozesse* (Führungskräfte vs. Mitarbeiter) können also den Befund erklären, dass die Selbst-Abwärts-Übereinstimmung höher ist als die Selbst-Aufwärts-Übereinstimmung. Wenn eine Führungskraft die ihr unterstellten Führungskräfte positiv beurteilt, wertet sie sich dadurch in zweifacher Hinsicht selbst auf. Zum einen stärkt sie ihre soziale Identität, zum anderen unterstützt sie das Bild eines kompetenten Führungshandelns, da sie gemäß der Führungsgrundsätze für die Förderung ihr unterstellter Führungskräfte verantwortlich ist. Dass Aufwärtsbeurteilungen, die von Stellvertretern (vs. keine Stellvertreterfunktion) abgegeben werden, besser mit den Selbstbeurteilungen übereinstimmen, könnte zumindest teilweise auf einer Aufweichung der Distinktheit zwischen Führungskräfte und Mitarbeitern beruhen. Da Mitarbeiter mit eigener Führungsverantwortung bereits der Kategorie der Führungskräfte angehören, ist im Sinne der sozialen Kategorisierungstheorie erklärbar, dass für sie der Moderatoreffekt stärker ausfällt als für Personen mit vorübergehender Stellvertretungsrolle (vgl. Abbildung 37).

Neben einer motivationalen Erklärung der Ingroup-Favorisierung wird diese in neueren Untersuchungen auch kognitiv erklärt: Die Ingroup wird so konstruiert, dass sie dem Selbst ähnlich ist ("self-anchoring"). Wenn das Selbst aber wie im vorliegenden Fall eher positiv

beschrieben wird, dann muss auch die Ingroup positiv beschrieben werden. Die Outgroup hingegen, die ja nicht mit dem Selbst assoziiert ist, wird schlechter beurteilt. Erste Bestätigung dieses so genannten Self-Anchoring bei Ingroup-Favorisierung konnte in minimalen Gruppen hinsichtlich der Zuschreibung positiver und negativer Traits gefunden werden (Cadinu & Rothbart, 1996; Otten & Wentura, 2001).

Mangelndes Feedback. Da durch Feedback die interperspektivische Übereinstimmung erhöht werden kann (vgl. Kapitel 3.6), liegt nahe, dass eine geringe interperspektivische Übereinstimmung zumindest teilweise darauf zurückzuführen ist, dass kein oder nur wenig/inadäquates Feedback gegeben wird (vgl. auch Tabelle 81). Für diese Erklärung sprechen auch die Moderatoreffekte der Beteiligungsquote insgesamt und der Beteilung des Ratees, da diese mit hoher Wahrscheinlichkeit mit mehr Feedback und Offenheit für ehrliche Rückmeldung innerhalb der Abteilung verbunden sind. Wenn im Arbeitsalltag bereits Feedback gegeben wird, dann können sich zum einen die verschiedenen Selbst- und Fremdbilder aneinander angleichen, zum andern müssen dann auch in einer Befragungssituation die Antworten nicht verzerrt werden. Die moderierende Wirkung der ergebnisbezogenen Itemformulierung ist ebenso über Feedback erklärbar, weil die Items sich gerade dadurch auszeichnen, dass sofort Ergebnisse des Führungshandelns (also Feedback) wahrnehmbar sind. Dass mit höherer Erfahrung in der Rolle des Beurteilers und in der Rolle des Beurteilten die interperspektivische Übereinstimmung nicht bedeutend ansteigt, schwächt die Feedbackhypothese ab, da gerade das Beurteiltwerden als Feedback verstanden werden kann.

Da Feedback auf das Führungsverhalten der beurteilten Führungskraft (Selbstbeurteiler) zielt, also an den Selbstbeurteilungen ansetzt, ist plausibel, dass die Fremd-Fremd-Korrelation generell höher ist als die Selbst-Fremd-Korrelationen.

Selbsttäuschung und Eindrucksmanagement. Hinterfragt man den Erklärungsmechanismus, der an der Selbsttäuschung (vor allem Self Enhancement) der Führungskräfte ansetzt, kann zunächst festgehalten werden, dass Personen mit hohem Selbstwert auf berufsrelevanten Dimensionen erfolgreicher sind als Personen mit niedrigem Selbstwert, dass also in der Gruppe der Führungskräfte (erfolgreiche Berufsgruppe) ein überdurchschnittlich hoher Prozentsatz an Personen mit hohem Selbstwert zu finden sind.[224] Korman (2001) bringt hohen Selbstwert wiederum mit der Motivation zu Self-Enhancement in Verbindung.

[224] Drei indirekte Belege dafür, dass sich in der Gruppe der Führungskräfte eher Personen mit hohem Selbstwert finden als in der Gruppe der Nicht-Führungskräfte. 1. Hoher Selbstwert ist mit erfolgreicherem Jobsuche-Verhalten verbunden als niedriger Selbstwert (Ellis & Taylor, 1983). 2. Personen mit hohem Selbstwert nehmen eher Schritte in Angriff, die zur Lösung von berufsbezogenen Problemen führen als Personen mit niedrigem Selbstwert (Pierce, Gardner, Dunham & Cummings, 1993). 3. Meta-analytisch ergibt sich insbesondere für Erwachsene ein positiver Zusammenhang zwischen sozioökonomischem Status und Selbstwert (d = .21 bis d = .25; Twenge & Campbell, 2002).

Betrachtet man die vorliegenden Daten, können drei indirekte Belege für eine Erklärung der interperspektivischen Diskrepanzen über die Theorie des Selbstwertschutzes ausgemacht werden. Erstens besteht die Tendenz, dass Führungskräfte das Führungsverhalten ihrer unterstellten Führungskräfte auf denjenigen Dimensionen positiver bewerten, auf denen sie sich selbst positiver bewerten (vgl. Anhang X). Das kann so interpretiert werden, dass andere umso weniger abgewertet werden müssen, je zufriedener (hoher Selbstwert) Führungskräfte mit ihrem eigenen Führungsverhalten sind. Dass dieser Effekt für übergeordnete Führungskräfte nicht eintritt, ist damit zu begründen, dass diese sich ohnehin weniger gut eignen, zugunsten des eigenen Selbstwerts abgewertet zu werden. Dieser Unterschied spricht im Übrigen gegen die Self-Anchoring-Hypothese (S. 461).

Zweitens kann der Befund, dass Übereinstimmer positivere Fremdbeurteilungen erhalten als Überschätzer und in geringerem Ausmaß auch als Unterschätzer, im Sinne der Selbstwert-Theorien interpretiert werden. Es wäre nämlich plausibel, wenn sich diejenigen Führungskräfte, die gut führen (positive Fremdbeurteilung), nicht mehr selbst täuschen und positiver bewerten müssen als sie sind.

Drittens kann als selbstwertdienliche Verzerrung interpretiert werden, dass sich Führungskräfte auf denjenigen Dimensionen selbst besonders positiv beurteilen, die sie als besonders wichtig erachten (N = 25, $\bar{r}$ = .58). Dieser starke Zusammenhang könnte aber auch so erklärt werden, dass Führungskräfte vor allem auf wichtig erachtete Führungsverhaltensweisen Wert legen, sich dafür Zeit nehmen, sich darin fortbilden und letztlich auch besser sind. Das kann deshalb ausgeschlossen werden, weil die Einschätzung der Wichtigkeit durch die Führungskraft keine Zusammenhänge zu Fremdbeurteilungen der entsprechenden Verhaltensweisen aufweist (Aufwärts: N = 82, $\bar{r}$ = -.10; Abwärts: N = 17, $\bar{r}$ = .03). Aufgrund der geringen Zusammenhänge zwischen der Wichtigkeitseinschätzung der Führungskräfte und der Fremdbeurteilungen ist ferner als Alternativerklärung abzulehnen, dass Führungskräfte – quasi in umgekehrter Reihenfolge – diejenigen Dimensionen, auf denen sie gut sind, als besonders wichtig beschreiben. Vor diesem Hintergrund ist nicht erstaunlich, dass die Wichtigkeitseinschätzung der Führungskräfte auch die Selbst-Aufwärts-Diskrepanz insofern moderiert, dass diese bei hoher Wichtigkeit größer ist (vgl. Kapitel 6.3.3). Auf diese Moderatorwirkung wiesen bereits die Daten von Church (1997a) hin, der entgegen seiner Annahmen fand, dass die itembezogenen Selbst-Aufwärts-Differenzen positiv mit Selbst-Wichtigkeits-Einschätzungen korrelieren.

Dass Führungskräfte in Abhängigkeit von der Wichtigkeitsbeurteilung auch ihre Ist-Beschreibung (Selbst) verzerren, ist nicht nur über Selbsttäuschung, sondern auch über *Eindrucksmanagement* erklärbar. Es ist plausibel, dass man bestrebt ist, auf den persönlich wichtigen Dimensionen einen besonders guten Eindruck zu erzeugen und sich von anderen abzugrenzen. Mit der Selbsttäuschungstheorie und der Theorie des Eindrucksmanagements kann zudem der als Above-Average-Effekt bezeichnete Befund erklärt werden, dass Füh-

rungskräfte ihr eigenes Führungsverhalten deutlich besser beurteilen als das nachgeordneter und vorgesetzter Führungskräfte (vgl. Kapitel 6.3.5).

Sowohl die Selbsttäuschungs- als auch die Eindrucksmanagement-Erklärung setzt an der Verzerrung der Selbstbeurteilungen an, so dass sie die im Vergleich zu den Selbst-Fremd-Korrelationen höheren Fremd-Fremd-Korrelationen erklären können. Betrachtet man zusätzlich zur Eindruckssteuerung der Selbstbeurteiler die der Fremdbeurteiler, dann wird plausibel, weshalb die Selbst-Abwärts-Diskrepanzen deutlich geringer ausfallen als die Selbst-Aufwärts- oder Abwärts-Aufwärts-Diskrepanzen. So ist es wahrscheinlich, dass die Abwärtsbeurteiler dazu motiviert sind, ihre nachgeordneten Führungskräfte positiv zu bewerten, da deren Führungsleistung mit in ihrem eigenen Verantwortungsbereich liegt.

Die Befunde, die zu den Selbst-Theorien angeführt wurden, sprechen also für folgende Aussage: „High ratings may be a signal that the rater wants to *look* good; low ratings may be a signal that the rater *is* good.“ (Murphy & Cleveland, 1995, p. 249).

Zusammenfassend kann festgehalten werden: Dass die Selbst-Aufwärts- sowie die Abwärts-Aufwärts-Diskrepanz höher ist als die Selbst-Abwärts-Diskrepanz, ist am besten durch Annahmen der sozialen Kategorisierungstheorie erklärbar, aber auch durch Annahmen der Akteur-Beobachter-Verzerrung sowie durch Annahmen unterschiedlicher Differenzierungen oder positionsabhängiger Führungsdefinitionen oder Gewichtungen. Auch das Verständnis für Führungsprobleme und für entsprechende Handlungsspielräume ist bei Selbst- und Abwärtsbeurteilern im Vergleich zu Aufwärtsbeurteilern ähnlicher. Die Akteur-Beobachter-Verzerrung sowie die idiosynkratischen Erklärungsansätze (mangelndes Feedback, Selbsttäuschung, Eindrucksmanagement) können am besten erklären, weshalb die Selbst-Fremd-Korrelationen geringer ausfallen als die Abwärts-Aufwärts-Korrelationen. Die Befunde zum Above-Average-Effekt gehen am stimmigsten mit der sozialen Kategorisierungstheorie und der Annahme von Selbsttäuschung oder Eindrucksmanagement einher.

Das zeigt, dass verschiedene Annahmen auch die geringe Selbst-Aufwärts-Übereinstimmung plausibel erklären können und keine der Annahmen vor dem Hintergrund der vorliegenden Daten ausgeschlossen werden kann. Dies geht mit den Antworten der Befragten einher, auch wenn sich hier vor allem bei den Mitarbeitern eine starke Gewichtung derjenigen Erklärungen ergab, die mit Kontakt und Kommunikation zusammenhängen (vgl. Kapitel 6.6.1). Versucht man die beiden methodischen Zugänge zu integrieren, dann sollte den Erklärungen, die mit unterschiedlichen Verhaltensstichproben und mangelnder Beobachtbarkeit sowie mit unzureichendem Feedback arbeiten, Priorität eingeräumt werden. Die Antworten auf die offenen Fragen insgesamt stimmen ferner sehr gut mit denen überein, die eine selektive Stichprobe von jeweils zehn Führungskräften und Mitarbeitern auch in der Interviewsituation generierten (vgl. Fußnote 218).

Stellt sich Kontakt und Kommunikation tatsächlich als kritischer Variablenbereich heraus, dann ist im Zuge der zunehmenden Globalisierung und damit einhergehender räumlicher Trennung von kooperierenden Teams zu untersuchen, wie die Kommunikation befriedigt befriedigend gestaltet werden kann. Auch neue Arbeitsformen, z.B. das „home office“, wären dann entsprechend zu modifizieren.

Im Sinne einer Triangulation müssen zukünftig verschiedene methodische Zugänge zur Erforschung der Entstehungsmechanismen von interperspektivischen Diskrepanzen integriert werden (Flick, 1991). Möglicherweise stellen sich auch Typen von Situationen oder Personen heraus, für die unterschiedliche Erklärungen optimal passen. Am Ende muss allerdings eine experimentelle Überprüfung der auf verschiedenem Wege generierbaren Hypothesen stehen.

7.4 Wie erfolgsrelevant ist Führung eigentlich? Zur Kriteriumsvalidität der einfachen Führungsbeurteilungen

Nach wie vor strittig ist die Frage, wieviel Unternehmenserfolg überhaupt auf Personalführung zurück geht. In Kapitel 2.2 wurde diese Diskussion angerissen. Die vorliegende Arbeit ist in demjenigen Bereich der Führungsforschung verankert, der die Prognose von *Führungs*erfolg fokussiert, welcher – so die implizite Annahme – zum Unternehmenserfolg insgesamt beizutragen vermag. Bei der Leistungsbeurteilung generell, erst recht aber bei der Führungsbeurteilung, die ungleich komplexer ist als die Beurteilung vieler anderer Tätigkeiten, bestehen weder etablierte Messinstrumente zur Erfassung der Leistung und Leistungserfolge noch Kriterienkataloge, die durch „gute Leistung" standardmäßig erfüllt werden.

In der Praxis sind Leistungsbeurteilungen noch heute oft alleinige Sache der Vorgesetzten, die dafür methodisch nicht überprüfte Beurteilungsinstrumente verwenden. Es ist nicht schwer ersichtlich, dass ein solcher Beurteilungsprozess fehlerbehaftet sein kann. In der vorgestellten Arbeit wurde herausgearbeitet, dass Abwärtsbeurteilungen ähnlich milde wie Selbstbeurteilungen ausfallen, im Unterschied zu diesen aber mit anderen Fremdbeurteilungen (Aufwärtsperspektive) deutlich korrelieren. Verschiedene Führungsdimensionen werden aus der Abwärtsperspektive jedoch weniger stark differenziert als aus der Selbstbeurteilung. Zur Verbesserung der Erfolgsprognose wurde Führung in der vorliegenden Untersuchung multiperspektivisch (vgl. Kapitel 7.5) und multidimensional (vgl. Kapitel 7.6) erfasst – auch um zu explorieren, welche der Perspektiven sich durch die Beschreibung welcher Führungsdimensionen für die Prognose welcher Erfolgskriterien besonders eignen.

Führungs*erfolg* wurde in der vorliegenden Arbeit durch mehrere Kriterien operationalisiert, welche im Einstellungs-Verhaltens-Modell von Fishbein (1967) unterschiedlichen Stufen zuzuordnen sind. Verschiedene Zufriedenheitsparameter können den Einstellungen zugeordnet werden, Extra Effort kann als Verhaltensintention betrachtet werden und Wohlbefinden, Leistung sowie Effektivität sind als Verhalten respektive Verhaltensergebnisse zu bewerten. Eine derart breite Betrachtung von Führungserfolg bot die Möglichkeit, Führungseffektivität spezifischer und zugleich umfassender als gewöhnlich zu beleuchten. Dass die Führungsbeurteilungen keine systematisch sinkende Kriteriumsvalidität aufweisen (von den Einstellungs- zu den Verhaltensindizes), muss auf die subjektive Erfassung aller Erfolgskriterien zurück geführt werden. Dennoch muss die zukünftige Führungsforschung auch darauf ausgelegt werden, die Wirkungsebenen der verschiedenen Führungsverhaltensweisen zu erkennen. Wie es die Ergebnisse der Criss-Cross-Analysen bereits andeuteten, ist erstens durch transformationale Führung von einer stärkeren Wirkung auf individueller Ebene auszugehen und zweitens durch Führungsaufgaben, wie sie in den Führungsgrundsätzen geregelt sind, von einer quasi übergreifend generellere Wirkung auf

Gruppenebene auszugehen. Dies kann aus dem stärkeren Unterschied zwischen der individuellen und gekreuzten Auswertung für die transformationale Führung (vs. FzFü-Führung) geschlossen werden.

Erfolg wurde also ausschließlich so operationalisiert, dass er sich auf der Seite der geführten Mitarbeiter zeigt (vgl. Neuberger, 1976). Deshalb wurden alle Erfolgskriterien mittels Ratings durch die Mitarbeiter erfasst. Im Sinne einer interaktionistischen Führungsbetrachtung sollte zukünftig auch in Betracht gezogen werden, Führungseffektivität an Merkmalen der Führungskraft selbst festzumachen. Gerade vor dem Hintergrund, dass Berufseinsteiger nicht mit Samthandschuhen zu guten Führungskräften ausgebildet werden und demzufolge der Ruf nach Selbstmanagement und Psychohygiene in den Chefetagen lauter wird, sollte eine Führung auch dann als effektiv bewertet werden, wenn die Führungskraft selbst sowohl ihre eigenen Führungskompetenzen als auch ihre Führungsmotivation kontinuierlich aufrecht erhalten oder steigern kann. Entsprechende Studiendesigns sind bislang nicht bekannt.

Außerdem ist in zukünftigen Studien zu erforschen, welche Auswirkung die Ursachenzuschreibung verschiedener Erfolgsparameter auf das Bedingungsgefüge von Führung und Erfolg nimmt. Da Führungskräfte die Ursache für Arbeitszufriedenheit, Fehlzeiten oder Fluktuation (vgl. Steers & Mowday, 1981) eher den Mitarbeitern selbst als ihrem eigenen Führungsverhalten zuschreiben – da sie also ihren eigenen Einfluss nicht erkennen – ist anzunehmen, dass sie erstens ihr eigenes Führungsverhalten besser bewerten als es tatsächlich ist und zweitens entsprechende Führungsmöglichkeiten zur Erreichung dieser Parameter nicht wahrnehmen.

Als Minimalanforderung an die Validität von Führungsbeurteilungen aus einer Perspektive ist anzusetzen, dass diese mit denen aus anderen Perspektiven korrelieren. Das wurde meta-analytisch für berufliche Leistungsbeurteilung inzwischen bestätigt (z.B. Conway & Huffcutt, 1997, vgl. Kapitel 7.2). Explizit erwähnt sein muss, dass auch Beurteilungen aus anderen Perspektiven als Validierungskriterien defizient und kontaminiert sind. Da Führung in verschiedenen Erfolgskennwerten münden soll, sind ebendiese als Validierungskriterien ungleich bedeutender.

Für die einzelnen Beurteilungsperspektiven ist festzuhalten, dass sie – durchschnittlich über die Führungsdimensionen hinweg betrachtet – unterschiedlich gute Vorhersageleistung erbringen. So klärt insbesondere die Aufwärtsbeurteilung von Führung viel Erfolgsvarianz auf. Im Vergleich zu bereits vorliegenden Studien ist insbesondere für die Arbeitszufriedenheit eine höhere Vorhersagekraft festzuhalten. Für die FzFü- und die MLQ-Dimensionen gilt gleichermaßen, dass sich speziell für die Arbeitszufriedenheit und Zufriedenheit mit Führung, für Extra Effort und Effektivität starke Effekte ergaben. Diese verringerten sich zwar, sobald der potentielle Single-Source-Bias durch die Anwendung des Criss-Cross-Designs kontrolliert wurde. Sie erreichten aber auch dann noch mittlere Stärke, so dass die Reduktion weniger auf den Single-Source-Bias als vielmehr auf den

veränderten Kriteriumsbereich zurückzuführen ist. Durch das Criss-Cross-Design sind nämlich lediglich die teambezogenen, nicht die individuellen Erfolge erfasst. Entsprechend teilen die gekreuzte und die personenbezogene Analyse auch nur denjenigen Teil der Varianz, der die *Team*effektivität betrifft.

Dass durch die MLQ-Führungsdimensionen die MLQ-internen Erfolgskriterien besonders gut prognostiziert werden können, wurde wiederholt berichtet (Tartler, Goihl et al., 2003). Es ist aufgrund erster empirischer Hinweise davon auszugehen, dass dies daran liegt, dass im MLQ die Items zur transformationalen Führungsbeurteilung und zur Erfolgseinschätzung sehr ähnlich formuliert sind (Tartler, Liepmann, Felfe & Nettelnstroth, 2004; Tartler, Liepmann & Nettelnstroth, 2003). Mit einer solchen MLQ-immanenten Verzerrung ist jedoch weder zu erklären, weshalb die Beurteilungen auf den FzFü-Dimensionen mit den MLQ-internen Erfolgskriterien noch weshalb die Beurteilung der transformationalen Führung (MLQ) mit der Arbeitszufriedenheit (AZ1) so eng zusammenhängen. Deshalb ist trotz MLQ-immanenter Verzerrungspotentiale von einer tatsächlich hohen Varianzaufklärung verschiedener Erfolgsmaße durch Führungsbeurteilungen auszugehen ist.

Die Erfolgsaufklärung durch die Abwärtsbeurteilung liegt in einer vergleichbaren Höhe wie die der gekreuzten Aufwärtsbeurteilungen. Dies ist deshalb plausibel, weil die Abwärtsbeurteilung per se keine individuumsbezogene Führung beschreibt, sondern das generelle Führungsverhalten, das dann auch nur den generellen Führungserfolg prognostizieren kann. Gerade für die einsatz- und leistungsbezogenen Variablen, also die Kennwerte, auf die aus der Vorgesetztenperspektive das stärkste Augenmerk gelegt wird, kann die Abwärtsbeschreibung von Führung sogar mehr Erfolgsvarianz erklären als die gekreuzte Aufwärtsbeschreibung, nicht jedoch für die Zufriedenheits- und Befindlichkeitsvariablen.

Dass nicht nur die personengebundenen Aufwärtsbeurteilungen, sondern auch die gekreuzten Aufwärtsbeurteilungen sowie die Abwärtsbeurteilungen Erfolgsvarianz aufklären, untermauert auch die Annahme über die Art des Zusammenhangs zwischen Führung und Erfolg. Aufgrund dessen kann nämlich ausgeschlossen werden, dass implizite Annahmen dergestalt wirken, dass Mitarbeiter ihren Führungskräften entweder dann eine bessere Führung zuschreiben, wenn ihre Erfolgswahrnehmung hoch ist oder umgekehrt dann mehr Erfolg zusprechen, wenn sie die Führung positiv beurteilen.[225] Die Annahme eines derart unbegründet starken Zusammenhangs zwischen Personalführung und Erfolg wurde von Meindl (1993) unter dem Stichwort „Romantisierung der Führung“ populär gemacht. Wenn aber die Erfolgsbeurteilung von Person A (ein Mitarbeiter), die Führungsbeurteilung hingegen von Person B abgegeben werden (Abwärtsbeurteiler bzw. Kollege im Criss-Cross-Design), dann können die beiden Urteile als unabhängig voneinander betrachtet werden. Die angesprochenen impliziten Annahmen über die Wirkung von Führung sollten

[225] Letzteres ist aufgrund der Reihenfolge der Fragebogenitems wahrscheinlicher, da zunächst die Führungs-, dann die Erfolgsvariablen erhoben wurden. Implizite Annahmen könnten sein: „Bei so guter Führung kann ich doch gar nicht unzufrieden, gestresst etc. sein.“

dann nicht mehr greifen. Dennoch sind weiterführende Studien mit einem Längsschnittdesign (vgl. Keller, 1992) und vor allem mit einer experimentellen Manipulation notwendig, um die kausale Richtung des Zusammenhangs zwischen Führung und Erfolg – auch im Kontext verschiedener anderer Variablen – genauer ermitteln zu können.

Im Unterschied zu den Fremdbeurteilungen von Führung weist die Selbstbeurteilung lediglich geringe Effekte auf – und auch das nur für manche Erfolgskriterien. Hervorzuheben ist das Kriterium Wohlbefinden, das aus der Selbstperspektive sogar besser erklärt werden kann als aus den Fremdperspektiven. Auch die Validität der einzelnen Führungsdimensionen unterscheidet sich für die Selbstbeschreibung stärker als für die Fremdbeschreibungen. Vor allem die Selbstbeurteilung auf der Dimension Klarheit trägt zur Erfolgsaufklärung bei, was neben der besonders deutlichen Selbst-Aufwärts-Korrelation auf dieser Dimension (vgl. Seite 442) ein weiterer Hinweis darauf ist, dass die diesbezügliche Selbstbeschreibung besonders gut gelingt. Solche Dimensionen sind zukünftig verstärkt zu extrahieren, um die Validität der Selbstbeschreibung erhöhen zu können.

Bei einer nicht-dimensionalen Analyse gewinnt man schnell den Eindruck, dass auf Selbstbeschreibungen verzichtet werden kann, da sie im Durchschnitt über die Dimensionen hinweg tatsächlich keine ausreichende Validität besitzt. Dem sind jedoch einige Argumente entgegen zu setzen. Neben den vorliegenden validen Selbstbeurteilungen auf einzelnen Dimensionen (insbesondere Klarheit) existieren bereits Befunde, die ebenfalls auf einen validen Anteil schließen lassen (z.B. Fleenor, 1996; Schuler, 1989). Zudem wurden in der vorliegenden Studie sowie in Meta-Analysen zu allgemeiner beruflicher Leistungsbeurteilung Moderatorvariablen ermittelt, durch die die Validität von Selbstbeschreibungen weiter gesteigert werden kann (überwiegend im Sinne der Selbst-Fremd-Korrelationen; Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Heidemeier & Moser, 2002; Mabe & West, 1982).

Neben diesen methodischen Argumenten liegen für die organisationspsychologische Praxis einige weitere Argumente vor: Die Durchführung von Selbstbeurteilungen sensibilisiert die Führungskräfte für die Inhalte und Ziele ihrer Tätigkeit, was sich unter anderem darin zeigt, dass Bewertungserfahrung zu akkurateren Einschätzungen führt (Mabe & West, 1982). In diesem Sinne können Selbstbeurteilungen die persönliche Entwicklung fördern, indem sie die Selbstbewertung im Rahmen des Selbstregulierungsmodells von Carver und Scheier (1998) verbessern. Wenn zusätzlich Mitarbeiterratings erhoben werden, kann ein Austausch zu einer besseren Kommunikation zwischen Führungskraft und Mitarbeitern führen (Walker & Smither, 1999). Vor allem bei komplexen Aufgaben erscheint es denkbar, dass kein Außenstehender in der Lage ist, alle Facetten adäquat zu erfassen, so dass Führungskräfte als Beurteiler ihrer eigenen Führung aufgrund ihres unersetzbaren Blickwinkels eher trainiert als ignoriert werden sollten. Das gilt umso mehr, je stärker die räumliche Trennung von Führenden und Geführten voranschreitet. Wenn Führungskräfte als Selbstbeurteiler in den Beurteilerprozess einbezogen werden, dann wird ihnen implizit auch eine neue Rolle zugewiesen. So erhalten sie mehr Verantwortung für

ihre eigene Führung, während im Gegenzug die traditionellen Abwärtsbeurteiler ihre bis dato fast unumwerfliche Richterrolle verlieren (vgl. Meyer, 1991). Letztlich wird dadurch das zwingend notwendige Selbstmanagement der Führenden gefördert.

Zur Verbesserung der Validität der Selbstbeurteilung sollten erstens Führungsdimensionen, die von den Ratee-Personen selbst valide beurteilbar sind, ermittelt werden (z.B. Klarheit), zweitens sollten Erfolgskriterien, die besonders gut aus der Selbstbeschreibung prognostizierbar sind, erforscht werden (z.B. Wohlbefinden) und drittens sollten Kontextbedingungen aufgedeckt werden, unter deren Einfluss die Reliabilität und Validität der Selbstbeurteilungen steigen.

Da den Führungsbeschreibungen aus allen drei berücksichtigten Perspektiven Kriteriumsvalidität zugeschrieben wurde, kann die grundlegende Voraussetzung für deren potentielle inkrementelle Validität als erfüllt gelten. Denn nur wenn gewährleistet ist, dass durch die verschiedenen Ratingquellen kriteriumsrelevante Anteile erfasst werden, sind sie prinzipiell in der Lage, gemeinsam mehr Erfolg zu erklären als einzeln. Um inkrementelle Beiträge liefern zu können, ist dann wichtig, dass sie *unterschiedliche* Kriteriumsanteile vorhersagen. Darauf deuteten erstens die geringen Interkorrelationen zwischen den Urteilen verschiedener Perspektiven hin (siehe oben) sowie zweitens die unterschiedlich gute Varianzaufklärung durch unterschiedliche Führungsdimensionen (vgl. Kapitel 7.6) und drittens die unterschiedlich gute Varianzaufklärung für die verschiedenen Erfolgskriterien.

Die inkrementelle Validität der berücksichtigten Beurteilungsperspektiven ist die grundlegende Rechtfertigung für den vergleichsweise aufwändigen *multiperspektivischen* Beurteilungsprozess. Daneben besteht nur noch ein zweiter methodischer Grund für die Aufwandsrechtfertigung, nämlich ein prognostischer Wert der interperspektivischen Übereinstimmung selbst. In der vorliegenden Arbeit wurden diese beiden Annahmen überprüft, im folgenden Kapitel werden die Ergebnisse weiter diskutiert.

7.5 Wieso der Aufwand? Zum Nutzen von multiperspektivischen Führungsbeurteilungen

Ein wichtiges Ziel der vorliegenden Arbeit war es, die Befunde zur inkrementellen Validität verschiedener Beurteilungsperspektiven auch auf den Bereich der Führungsbeurteilung zu übertragen. Conway, Lombardo und Sanders (2001) erzielten in ihrer Meta-Analyse zur inkrementellen Validität von Aufwärts- und Peerbeurteilungen hinsichtlich allgemeiner beruflicher Leistung – über Abwärtsbeurteilungen hinaus – bedeutsame, wenn auch geringe Beiträge zur Varianzaufklärung objektiver Leistungskriterien (maximal 3% zusätzliche Varianzaufklärung).

In der vorgestellten Studie erklärte die Aufwärtsbeurteilung für fast alle Erfolgskriterien den größten Varianzanteil. Dieser konnte durch die Hinzunahme der Abwärts- und/oder Selbstbeurteilung teilweise mit mittleren Effekten ($f^2 > .15$) erhöht werden. Dass dieses Ergebnis nur dann erzielt werden konnte, wenn nicht die Summenscores über alle Führungsdimensionen hinweg, sondern die aus der jeweiligen Perspektive validesten Führungsdimensionen in die Analyse eingingen, spricht stark für eine multidimensionale Betrachtung der Führung (vgl. Kapitel 7.6). Zudem belegt es, dass eine Führungsforschung, die mit wenigen Variablen Führungserfolg im Allgemeinen versucht zu erklären, auf dem Holzweg ist. Angesichts des hohen Aufwands, der mit der multiperspektivischen Beurteilung von Führung einher geht, ist auch für deren praktischen Einsatz unbedingt das Optimum an Validität anzustreben.

Dass tatsächlich alle drei erhobenen Perspektiven inkrementelle Beiträge liefern können, unterstützt die Verwendung des Sandwich-Modells einer 270°-Beurteilung. So erbringen sowohl die Geführten als auch die traditionellen (Abwärts-)Beurteiler und die Führungskräfte selbst (Selbstbeurteiler) spezifische, erfolgsrelevante Information. Dass die Abwärtsbeurteilungen über die Aufwärtsbeurteilungen hinaus Erfolgsvarianz erklären können, kann im Übrigen als weiterer Hinweis dafür gewertet werden, die Urteile verschiedener Fremdbeurteiler nicht, wie es immer wieder vorgeschlagen wurde, zu aggregieren (z.B. Wohlers & London, 1989).

Der Verzicht auf die Peer-Perspektive, die bei einer Vollbefragung zur Mehrfachbelastung der entsprechenden Führungskräfte führt, kann also verschmerzt werden. Untermauert wird dies durch Befunde, nach denen Peerratings ohnehin keine perspektivenspezifischen Informationen liefern können (Scullen et al., 2000; vgl. Kapitel 3.4.3). Dies gilt unter der Annahme, dass die perspektiven-bezogenen im Unterschied zu den idiosynkratischen „Verzerrungen“ auf tatsächlichen, interessierenden Verhaltensaspekten beruhen, die eben nur aus der jeweiligen Perspektive wahrgenommen werden können (vgl. Borman, 1997). Da die Selbst- und Abwärtsbeurteilungen auf den FzFü-Dimensionen auch inkrementelle Beiträge über die Aufwärtsbeurteilungen des MLQ hinaus liefern, kann verstärkt davon

ausgegangen werden, dass tatsächlich die Perspektive ein entscheidender Punkt ist und kein Methodenartefakt durch Eigenheiten eines einzelnen Instruments vorliegt.

Allein aufgrund des beschriebenen inkrementellen Nutzens der einzelnen Beurteilungsperspektiven muss befürwortet werden, dass Führungsbeurteilungen aus mehreren Perspektiven eingeholt werden. Betont soll werden, dass aber genau dieser inkrementelle Nutzen bei interperspektivisch sehr ähnlichen Beurteilungen nicht entstehen könnte. Dann wären nämlich aufgrund redundanter Information kaum perspektiven-spezifische Beiträge zu erwarten. Zukünftig könnte sich die multiperspektivische Führungsbeurteilung also nach der immanenten Logik dieser Argumentationslinie selbst überflüssig machen. Wenn die Bedingungen erforscht sowie umgesetzt sein sollten, die zu einer hohen interperspektivischen Übereinstimmung führen, dann enthalten die Beurteilungen zunehmend soviel redundante Information, dass sie sich nicht mehr ergänzen, sondern substituieren.

Das Wissen um die inkrementelle Validität hat eine weitere methodische Implikation. Denn dadurch, dass verschiedenen Perspektiven spezifische kriteriumsrelevante Information zur Erfolgsaufklärung zugeschrieben wird, ist die gängige Forschungspraxis anzuzweifeln, die Selbstbeurteilung von Leistung an Fremdbeurteilungen zu validieren. Es sind also unbedingt externe, andere Validierungskriterien heranzuziehen.

Noch besser könnte der Einsatz multiperspektivischer Beurteilungssysteme für die Praxis dann begründet werden, wenn die Diskrepanz selbst spezifische Kriteriumsvalidität aufweist. Diese Annahme wäre aufgrund ihrer hohen Plausibilität in der Organisationsberatung besonders anschlussfähig. Einer diesbezüglichen Diskussion der interperspektivischen Übereinstimmung wurde bereits Kapitel 6.5.6 gewidmet, so dass an dieser Stelle lediglich einzelne Punkte aufgegriffen und fortgeführt werden.

In der vorliegenden Studie wurde Forderungen der Führungsforschung nachgegangen, die Selbst-Fremd-Kongruenz von Führungsbeurteilungen mit verschiedenen Erfolgskriterien in Beziehung zu setzen (Church, 1997a; Greguras & Robie, 1998). Dabei konnten erwartungsgetreu sehr starke negative korrelative Zusammenhänge zwischen Selbst-Aufwärts- beziehungsweise Selbst-Abwärts-Diskrepanzen und den Erfolgskriterien festgehalten werden. Zieht man die Richtung der Abweichung in Betracht, sind es insbesondere diejenigen Führungskräfte, die sich selbst besser bewerten als sie von Fremdbeurteilern bewertet werden, die am wenigsten Erfolg mit sich bringen. Darüber hinaus erwiesen sich die Abwärts-Aufwärts-Diskrepanzen als mindestens ebenso stark mit Erfolg kovariierend wie die Selbst-Fremd-Diskrepanzen. Die durchschnittliche Varianzaufklärung der Abwärts-Aufwärts-Diskrepanzen (25%) liegt sogar über der, die die Selbst-Fremd-Diskrepanzen in der vorliegenden Studie erbrachten (15% bzw. 13%) und ist auch im Vergleich zu verschiedenen Selbst-Fremd-Diskrepanzen aus der Literatur als hoch einzustufen. Damit wurde eine Forschungsfrage angerissen, die bislang in der Literatur keine Berücksichtigung fand. Allenfalls wurden Zusammenhänge zwischen Aufwärts- und Abwärtsbeurteilungen untersucht, die erhobenen Erfolgskriterien (z.B. Arbeitszufriedenheit) dann

jedoch nur mit den Beurteilungen aus den einzelnen Perspektiven in Beziehung gesetzt, nicht mit dem Ausmaß der Abwärts-Aufwärts-*Kongruenz* (z.B. Riggio & Cole, 1992).

Zur Erklärung des Effekts der *Selbst-Fremd*-Diskrepanzen kann auf Ansätze, die die Genauigkeit der Selbsteinschätzung als grundlegende Variable für einen erfolgreichen Führungsprozess betrachten (vgl. Ashford, 1989) und auf erste Befunde zurück gegriffen werden, die eine gute Kommunikation im Zusammenhang mit hoher interperspektivischer Übereinstimmung sehen (z.B. Tartler, Goihl et al., 2003). Allerdings lassen die Zusammenhänge zwischen der interperspektivischen Kongruenz in der Beurteilung transformationaler Führung auf der einen Seite und wahrgenommener Qualität der Kommunikation auf der anderen Seite offen, wie die Beeinflussungs*richtung* aussieht. Das Gleiche gilt für den Zusammenhang zwischen transformationaler Führung und wahrgenommener Qualität der Kommunikation (Torjus, Tartler & Liepmann, 2004).[226]

Während derart „angedachte" Erklärungen für den Effekt von Selbst-Fremd-Diskrepanzen weiter untersucht werden müssen, müssen für den Effekt von Fremd-Fremd-Diskrepanzen erste Erklärungsansätze überhaupt erst generiert werden. Geht man nicht davon aus, dass ausschließlich die ohnehin reliableren und valideren Fremdbeurteilungen ausschlaggebend sind, kann dieser Effekt auf eine Führung hinweisen, die von außen wenig „greifbar" ist. In zukünftigen Studien ist deshalb zu überprüfen, ob hohe Fremd-Fremd-Diskrepanzen vor allem deshalb auftreten, weil die Führungskraft eine diffuse und inkonsistente Kommunikationspolitik betreibt. Inkonsistentes, also auch unberechenbares Führungsverhalten verhindert – so die Annahme – auf der Seite der Mitarbeiter reibungsloses und letztlich motiviertes und effektives Arbeiten. Dass durch die hervorgerufene Unsicherheit auch Zufriedenheits- und Wohlbefindensparameter negativ beeinflusst werden, liegt auf der Hand. Insbesondere das Kriterium Wohlbefinden konnte in der vorliegenden Studie durch die Fremd-Fremd-Diskrepanz besser erklärt werden als durch die Selbst-Fremd-Diskrepanzen. Dies kann nicht nur als Hinweis darauf gewertet werden, dass die Fremd-Fremd-Diskrepanz mehr/andere Informationen enthält als die Selbst-Fremd-Diskrepanzen enthalten, sondern auch als erster Hinweis für den soeben angerissenen Erklärungsmechanismus. Denn nach allgemeinpsychologischen Kenntnissen ist Unsicherheit primär mit Stress verbunden (z.B. Dörner, 1999). Erst in zweiter Linie sollte sich dann ein Rückgang der Effektivität ergeben.

Daraus ergibt sich eine wichtige Implikation für die Interpretation multiperspektivischen Führungsfeedbacks. Es muss davon abgesehen werden, die Auswirkungen auf Erfolg ausschließlich auf die Selbstüberschätzung der Führungskräfte zurückzuführen. Auch auf theoretischer Ebene lohnt es, in Erklärungsmodelle zur Entstehung von Führungserfolg andere Facetten des Führungsverhaltens einzubeziehen, möglicherweise vermittelt über die

[226] Es ist bislang nicht untersucht, ob transformationale Führung die Qualität der Kommunikation beeinflusst oder ob gute Kommunikation vielleicht erst transformationale Führung zulässt.

Fremd-Fremd-Übereinstimmung (zur Diskussion der Kausalität vgl. z.B. Kapitel 7.7). Dass hohe interperspektivische Niveau-Unterschiede in der Führungsbeurteilung nicht nur mit geringem Erfolg einhergehen, sondern insgesamt ein negativeres Führungsverhalten ausdrücken, belegen die Zusammenhänge zu Führungsbeurteilungen aus anderen Perspektiven beziehungsweise auf anderen Instrumenten. Hohe Niveau-Unterschiede zwischen den Beurteilungen aus zwei Perspektiven gehen nämlich mit entsprechend negativer Führungsbeurteilung aus der jeweils dritten Perspektive einher sowie entsprechend wenig transformationaler Führung (MLQ). Der Vorteil von drei befragten Perspektiven und zwei verschiedenen Instrumenten zeigt sich also darin, dass keine Führungsbeurteilung in die Berechnung der interperspektivischen Diskrepanz eingeht und zugleich als deren Validierungskriterium herangezogen wird.

Umgekehrt betrachtet fallen für die überdurchschnittlich erfolgreichen Führungskräfte die Niveau-Unterschiede geringer aus als für die unterdurchschnittlich erfolgreichen, was andere Befunde aus der Organisationsforschung bestätigt (z.B. Zimolong, 2001). Der Wirkungszusammenhang scheint auch nicht spezifisch für Führungsverhalten und berufliche Leistungsvariablen zu sein, da analoge Ergebnisse zum Beispiel auch für Schülerverhalten vorliegen. Nicht die Gruppe der gemochten, sondern die Gruppe der von den Peers zurückgewiesenen Schüler weist die höchste Selbst-Lehrer-Differenz in der Beurteilung verschiedener Dimensionen von Interaktionsverhalten auf (Cillessen & Bellmore, 1999). Dass sich erfolgreiche von nicht erfolgreichen Führungskräften allerdings nicht in der korrelativen Selbst-Fremd- oder Fremd-Fremd-Übereinstimmung unterscheiden, war unerwartet. Auch die interperspektivischen τ-Korrelationen über die Führungsdimensionen hinweg hängen nicht systematisch mit den Erfolgskriterien zusammen. Dadurch wird die geringere Erfolgsrelevanz der korrelativen Übereinstimmung unterstrichen. Möglicherweise reagieren Korrelationen weniger sensibel als Niveau-Unterschiede.

Korrelative Zusammenhänge zwischen Führungsdiskrepanzen und Erfolg wurden in letzter Zeit immer wieder erbracht. In der Folge wurde auch immer stärker auf die interperspektivische Übereinstimmung selbst fokussiert und nicht mehr nur auf die Verwendung mehrerer einfacher, sich ergänzender Beurteilungsperspektiven. Häufig wurden die Daten allerdings zu leichtfertig in einem kausalen Sinne interpretiert. Entsprechend wurde als wichtiges Interventionsziel proklamiert, die Übereinstimmung zwischen Selbst- und Fremdbeurteilungen zu erhöhen (Yammarino & Atwater, 1997). Die zugegebenermaßen sehr plausible Annahme, dass mit zunehmender Selbst-Fremd-Diskrepanz der Führungsbeurteilungen schlechtere Kommunikation und weniger Erfolg einher geht, ist sogar so verbreitet, dass Conway und Huffcutt (1997) diesen Forschungsbereich als einen von drei Forschungsbereichen der multiperspektivischen Führungsbeurteilungen klassifizierten (neben Grundlagenforschung und Personalentwicklung).

Trotz der Plausibilität und praktischen Akzeptanz ist es aus methodischer Sicht dringend angezeigt, zu untersuchen, ob wirklich die Diskrepanz (und nicht die Beurteilungen aus den einzelnen Perspektiven) die kritische Variable im Führungs-Erfolgs-Zusammenhang darstellt. Dazu knüpfte die vorliegende Untersuchung an die Arbeit von Brutus, Fleenor und Tisak (1999) an. Dort wurde aus den Ergebnissen verschiedener regressionsanalytischer Prozeduren resümiert, dass die Selbst-Peer-Diskrepanz keine eigene prädiktive Validität hinsichtlich der Führungseffektivität besitzt. Analog wurde in der vorgestellten Studie die Selbst-Aufwärts-Diskrepanz untersucht. Dabei ergab sich für die meisten Erfolgskriterien ein ähnliches, wenn auch nicht ganz so eindeutiges Ergebnis. Die Selbst-Abwärts- und die Abwärts-Aufwärts-Diskrepanzen konnten aufgrund der geringeren Fallzahlen nicht auf diese Weise getestet werden.

Verschiedene Regressions- und varianzanalytische Auswertungen lassen den Schluss zu, dass der Führungserfolg umso höher ist, je positiver die Führungsbeurteilungen, insbesondere die Aufwärtsbeurteilungen ausfallen und je besser die Selbst- und Aufwärtsbeurteilungen zusätzlich übereinstimmen. Dies stimmt auch mit der Interpretation der Oberflächenanalyse von Atwater, Ostroff, Yammarino und Fleenor (1998) überein.
Letzendlich sind nicht nur für die hier überprüfte Selbst-Aufwärts-Diskrepanz weitere Untersuchungen und Kreuzvalidierungen notwendig, sondern auch für die bei Brutus, Fleenor und Tisak (1999) fokussierte Selbst-Peer-Diskrepanz, auch wenn dort eine große Stichprobe vorlag (N = 2.022) und eine Kreuzvalidierung (N = 1.053) stattgefunden hat. Es zeigten sich in der vorliegenden Studie jedoch starke Unterschiede zwischen den einzelnen Erfolgskriterien, so dass denkbar ist, dass bestimmte Kriterien durch die interperspektivischen Diskrepanzen stärker beeinflusst werden als andere. Bevor die Ergebnisse aber als stabil betrachtet beziehungsweise generalisiert werden können, sind Replikationsstudien notwendig sowie Studien mit weiteren Erfolgskriterien, an größeren Stichproben, in unterschiedlichen Organisationsformen und Industriezweigen und zudem Studien, die mit unterschiedlichen Methoden arbeiten.

Vorläufig können interperspektivische Diskrepanzen aus verschiedenen Gründen in der Beratungspraxis genutzt werden, falls sie für die dortigen Projekte bessere Ansatzpunkte versprechen als einfache Beurteilungsperspektiven. So korrelieren die Diskrepanzen beispielsweise nicht weniger mit Erfolg als die einfachen Beurteilungsperspektiven und das beste Diskrepanzmodell (mehrere Diskrepanzen gemeinsam) erklärt sogar mehr Varianz als das beste einfache Modell. Dadurch, dass verschiedene Diskrepanzen außerdem unterschiedliche Erfolgsanteile aufklären, wird ihre inkrementelle Validität ermöglicht. So kann sich die gleichzeitige Betrachtung von Selbst-Fremd- *und* Fremd-Fremd-Diskrepanzen als nützlich erweisen.

Vereinzelt deuteten Studien auf eine moderierende Rolle der Diskrepanzen insofern hin, dass ein positiver Zusammenhang zwischen Führung und Erfolg insbesondere für Übereinstimmer auftritt (z.B. Atwater & Yammarino, 1992). In der vorgestellten Studie ergaben sich diesbezüglich sehr inkonsistente Ergebnisse. Die Ergebnisse variieren nicht nur

unsystematisch über die Erfolgskriterien hinweg, sondern schwanken zudem über die verschiedenen Führungsdimensionen hinweg. Mehr Hinweise bestehen tatsächlich auf eine (partiell) mediierende Wirkung der Diskrepanz im Bedingungszusammenhang von Führung und Erfolg. Diese muss mit geeigneteren Methoden zukünftig genauer analysiert werden.

7.6 Was bringt die Multidimensionalität? Zum Facettenreichtum von Führung

In der Forschung zu multiperspektivischen Führungsbeurteilungen wird Führung in der Regel sehr global erfasst oder durch einen Mittelwert einzelner Führungsfaktoren repräsentiert. Da ohnehin Beurteilungen aus unterschiedlichen *Perspektiven* zu dokumentieren sind, wird die Komplexität durch eine zusätzliche Dokumentation unterschiedlicher Führungs*dimensionen* häufig vermieden (z.B. Atwater et al., 1998). Nur durch eine solche Komplexität ist indes ein Facettenreichtum zu erfassen, ohne den die Führungsaufgabe nicht ausreichend zu beschreiben ist.

Dass Skalen zu verschiedenen Führungsdimensionen Unterschiedliches messen, untermauern nicht nur die entsprechenden Faktorenanalysen dieser Arbeit. So berichten Salgado et al. (2003) beispielsweise auch unterschiedlich hohe Retest-Reliabilitätskoeffizienten für unterschiedliche Leistungsdimensionen von Führungskräften, was auf unterschiedliche Veränderbarkeit der Dimensionen hinweist.

In der vorgestellten Studie variiert keineswegs nur die interperspektivische Beurteilerübereinstimmung (Ist-Beurteilung) über die erfassten Führungsdimensionen hinweg, sondern auch die wahrgenommene Abweichung zwischen beobachtetem und erwartetem Führungsverhalten sowie die Kriteriumsvalidität der Führungsbeurteilungen. Ersteres wurde bereits auf den Seiten 442ff. dargestellt. Da für Entwickeln/Fördern die höchste Diskrepanz zwischen beobachtetem und erwartetem Verhalten berichtet wird, sollte das entsprechende Führungsverhalten bei den Führungskräften des Unternehmens besonders gefördert werden. Der wohl wichtigste Grund für eine multidimensionale Betrachtung von Führung ist, dass unterschiedliche Erfolgskriterien durch verschiedene Führungsdimensionen am besten prognostizierbar sind (vgl. auch Seiten 466ff.). So zeigt sich beispielsweise, dass die Arbeitszufriedenheit (AZ1, AZ2, SAT) und Arbeitsmotivation (EEF) hochsignifikant stärker mit der Dimension Entwickeln/Fördern als mit Partizipation zusammenzuhängen oder Effectiveness deutlich stärker mit Klarheit als mit Beurteilen/Steuern. Das gleiche gilt auch für interperspektivische Diskrepanzen. Während beispielsweise die interperspektivischen Diskrepanzen auf der Dimension Individuelle Berücksichtigung nicht mit Wohlbefinden einhergehen, weisen sie auf der Dimension Klarheit bedeutende Effekte auf. Auch in schrittweisen Regressionsanalysen werden für die verschiedenen Erfolgskriterien verschiedene Dimensionen aufgenommen.

Aufgrund solcher Erkenntnisse ist dringend anzuraten, nicht nur verschiedene Perspektiven in ein umfassendes Beurteilungssystem einzubeziehen, sondern auch jeweils multiple Verhaltensdimensionen zu erfassen und getrennt zu analysieren. Solche Studien müssen zukünftig klären, welche Erfolgskriterien durch welche Führungsdimensionen, beurteilt aus welchen Perspektiven, am besten prognostiziert werden können. Dass die Führungs-

beurteilung auf einem *einzelnen* Führungsfaktor nicht alle Erfolgskriterien gleich gut erklären kann, ist unmittelbar plausibel.

Die fortführende Evaluation des MLQ war ein untergeordnetes Ziel dieser Arbeit und soll aufgrund der regelmäßig kritisierten Multidimensionalität des MLQ an dieser Stelle diskutiert werden. Dass die Subskalen des MLQ, insbesondere die transformationalen Subskalen, hohe Interkorrelationen aufweisen und sie deshalb auch kaum spezifische Wirkungen auf Erfolgsvariablen ausüben (Bycio et al., 1995), wurde bereits in Kapitel 2.7.4 ausführlich diskutiert. Genau diese methodischen Schwierigkeiten wurden auch für die neue deutschsprachige Version bereits berichtet (Felfe & Goihl, 2003; Tartler et al., 2004) und sie wiederholten sich in der vorgestellten Arbeit. Da keine spezifischen Wirkzusammenhänge zwischen einzelnen transformationalen Skalen und bestimmten Erfolgsmaßen bestehen, wurden die transformationalen Subskalen zugunsten der Übersichtlichkeit meist zusammengefasst. Als Erklärung für die mangelnde Spezifität können die verhältnismäßig abstrakten Itemformulierungen herangezogen werden sowie die inhaltliche Nähe der verschiedenen transformationalen Führungsdimensionen. Transformationale und transaktionale Führungsdimensionen erbringen hingegen gleichwohl auch in der vorliegenden Studie unterschiedliche Erfolgsprognosen. Um künftig mehr Nutzen durch die Multidimensionalität zu gewinnen, sollte über eine erneute Itemkonstruktion nachgedacht werden, in der spezifische Konstruktdefinitionen berücksichtigt sind.

Möglicherweise sind die einzelnen Subskalen nur für die „weichen“ Erfolgskriterien, die freilich durch die transformationale Führung gewöhnlich prognostiziert werden sollen (z.B. Extra Effort), nicht ausreichend spezifisch. Bei Maaß (in Vorb.) zeigen sich nämlich für ein objektiv messbares Kriterium sehrwohl differenzielle Effekte: Während IIa, IS und IC starke Zusammenhänge mit der Quantität von patentfähigen Erfindungsmeldungen aufweisen, sind diese für IIb und IM gering.

Von einer ausschließlichen Rückkehr zur Zweidimensionalität, wie es regelmäßig gefordert wird (z.B. den Hartog et al., 1997), wird zum aktuellen Zeitpunkt Abstand genommen. Vielmehr sollte eine Verbesserung der Subskalen die Grundlage dafür bieten, endlich in eine spezifischere Führungsforschung einzusteigen, die Führungs-Erfolgs-Zusammenhänge differenziert erarbeiten kann. Der MLQ eignet sich dafür aufgrund seiner hohen Kriteriumsvalidität und aufgrund des breiten Spektrums an Führungsdimensionen.

Unter verschiedenen Gesichtspunkten können multivariate Effekte im Sinne von „Augmentationseffekten“ diskutiert werden. Zunächst konnten erwartungsgemäß für nahezu alle Erfolgskriterien inkrementelle Beiträge der transformationalen Führung über die transaktionalen Führung hinaus festgehalten werden. Da dies umgekehrt nicht gilt, wird die Modellannahme bestätigt, dass durch transformationale Führung der Erfolg, der allein durch transaktionale Führung hervorgerufen werden kann, weiter gesteigert werden kann. Außer-

dem sprechen die Ergebnisse nicht dafür, Contingent Reward (transaktionale Subskala) mit den transformationalen Subskalen zu einem Faktor zusammenzufassen, wie es in vorgeschlagenen zweidimensionalen Konzeptualisierungen angestrebt wird. Ein anderer Augmentationseffekt unterstützt nicht nur die Nützlichkeitsannahme multipler Führungsdimensionen, sondern bietet auch einen Ausgangspunkt zur Überarbeitung des MLQ. Dass nämlich einzelne Führungsdimensionen des FzFü für fast alle Erfolgskriterien inkrementelle Beiträge über die transaktionale und transformationale Führung hinaus liefern (nicht umgekehrt), spricht dafür, dass der MLQ nicht – wie von seinen Urhebern intendiert – den „Full Range of Leadership“ abdeckt.

Die Befunde, die für eine multidimensionale Führungsbeurteilung sprechen, geben denjenigen Organisationsberatern recht, die von Beurteilern verschiedener Perspektiven auch verschiedene Führungsdimensionen bewerten lassen. Doch werden die Dimensionen in der Praxis nicht auf der Basis empirischer Erkenntnisse ausgewählt, sondern vielmehr auf der Basis intuitiver Wahrnehmung. Eine multidimensionale Führungsforschung könnte dazu beitragen, für Selbst-, Abwärts- und Aufwärtsbeurteiler rollenspezifische Instrumente zur Messung des gleichen Konstrukts zu entwickeln (vgl. Bozeman, 1997). Denn möglicherweise gehen die geringen interperspektivischen Beurteilungen tatsächlich „nur“ auf die Naivität der Annahme zurück, dass Beurteiler aus verschiedenen Perspektiven das gleiche Verhalten gut beurteilen können. Voraussetzung ist allerdings, dass zunächst die gleichen Dimensionen aus verschiedenen Perspektiven beurteilt werden. Erst dann sind die besten Passungen zwischen Beurteilungsgegenstand und -perspektive zu erkennen.

7.7 Grenzen der vorliegenden Studie und Ausblick: Eine Baustelle mit Wegweisern für Forschung und Praxis

Die grundlegende Annahme von multiperspektivischem Führungsfeedback ist, dass „ein fortwährender Abgleich zwischen Selbst- und Fremdbild [...] die besten Chancen [bietet], sich hinsichtlich beruflicher wie auch privater Anforderungen und persönlicher Zielsetzungen weiterzuentwickeln“ (Hossiep & Paschen, 1998b, S. 11). Dies wird deshalb angenommen, weil eine akkurate Selbsteinschätzung gemeinhin als wichtiger Ausgangspunkt für erfolgreiche Selbstregulationsprozesse betrachtet wird (Ashford, 1989; Tisdale, 1998). Auch wenn die Analysen der vorliegenden Studie erstens verschiedene Bedingungen aufdeckten, unter denen die Selbst-Fremd- sowie die Fremd-Fremd-Übereinstimmung höher sind, und zweitens starke Zusammenhänge zwischen wahrgenommener interperspektivischer Übereinstimmung und verschiedenen Erfolgskriterien fand, ist die Aussagekraft der Studie insofern begrenzt, als dass sie keine unmittelbaren Erkenntnisse über die zugrunde liegenden Regulationsprozesse erbringt. Es stellt sich die Frage, ob wirklich mit einem Erfolgszuwachs gerechnet werden kann, sobald die interperspektivische Übereinstimmung – beispielsweise durch Feedbackprozesse – verbessert wird?

Um diese Frage, die die Kausalität des Zusammenhangs zwischen Führung beziehungsweise Führungsdiskrepanzen und Erfolgskriterien betrifft, besser beantworten zu können, sind unbedingt experimentelle Studien durchzuführen. Die interperspektivische Übereinstimmung selbst kann nur schwer manipuliert werden. Doch durch die Manipulation von Feedback könnte beispielsweise nicht nur untersucht werden, ob Feedback zur Erhöhung interperspektivischer Übereinstimmung beiträgt,[227] sondern auch, ob dies gleichzeitig in einer Erfolgssteigerung resultiert. Besonders relevant ist eine Überprüfung der kausalen Richtung vor allem vor dem Hintergrund, dass auch eine umgekehrte Annahme plausibel ist. So könnte die Wahrnehmung von Erfolg positiv auf die Fremdbeurteilungen von Führung wirken, wenn die Beurteiler der impliziten Annahme folgen, dass hoher Erfolg nur durch positive Führung erreicht werden kann. Auf die Selbstbeurteilung könnte die Wahrnehmung von Erfolg insofern wirken, als Erfolg an sich selbstwertdienlich genug ist und deshalb die Motivation zu positiver Selbstdarstellung (milden Selbstratings) reduziert ist. In der Folge wäre bei hohem Erfolg dann auch die interperspektivische Übereinstimmung höher.

Da in diesem Bereich der angewandten Forschung jedoch selten die Möglichkeit besteht, experimentell zu arbeiten, müssen gleichzeitig andere Designs und Methoden verbessert und unter den jeweils notwendigen Bedingungen angewendet werden. Quasi-

[227] Dass die interperspektivischen Diskrepanzen durch Feedback verringert werden können, wurde bereits mehrfach gezeigt (z.B. Johnson & Ferstl, 1999; vgl. Kapitel 3.6).

experimentelle Studien bieten sich dann an, wenn sich ohnehin eingeführte Veränderungen in der Organisation als Experimentalbedingung eignen.

Eine Zweiterhebung, welche zumindest erste Hinweise auf die Beeinflussungsrichtung geben könnte, war für die vorliegende Studie geplant, konnte aber bedauerlicherweise nicht durchgeführt werden. Positive Ergebnisse erbrachte diesbezüglich bereits die Arbeitsgruppe von Zimolong (2001), in der sich über mehrere Jahre hinweg überdurchschnittlich erfolgreiche Betriebe (bezüglich Arbeits- und Gesundheitsschutzes) durch eine besonders gute Selbst-Fremd-Übereinstimmung auszeichneten. In der vorliegenden Studie ist allenfalls auszuschließen, dass die berichteten Zusammenhänge allein auf einem Single-Source-Bias beruhen. Längsschnittdesigns würden im Übrigen auch eine Bestimmung der Retest-Reliabilität der interperspektivischen Übereinstimmung zulassen. Dies ist schon allein deshalb von praktischer Relevanz, da multiperspektivische Beurteilungsergebnisse in weit reichende Entscheidungen eingehen, wie etwa in Gehaltsfestlegungen. In entsprechenden Längsschnittdesigns wäre zudem überprüfbar, ob der Erfolg von selbstwertdienlichen Strategien davon abhängt, wie *lange* diese eingesetzt werden. Plausibel ist, dass selbstwertdienliche Strategien und damit die Überschätzung eigener Fertigkeiten zwar kurzfristig erfolgreich, langfristig aber eher schädlich sind, da keine situationsadäquate Verhaltensausrichtung erfolgt, z.B. keine Unterstützung ersucht wird (vgl. Jeserich, 1996). Aus dem Bereich der psychosozialen Anpassung ist bekannt, dass sich Diskrepanzen zwischen Selbst- und Fremdbild von Persönlichkeitsattributen langfristig negativ auswirken (Colvin, Block & Funder, 1995).

Betrachtet man die Analysemethoden, die in einem Querschnittsdesign zur Untersuchung des Zusammenhangs zwischen den interperspektivischen Diskrepanzen und den Erfolgskriterien zur Anwendung kommen, dann zeigt sich auch hier Verbesserungsbedarf. Varianzanalysen sind bei kategorialen Prädiktoren (Überschätzer, Übereinstimmer, Unterschätzer) zwar anschaulich, doch kann durch Regressionsanalysen statistische Power gewonnen werden (höhere Varianzaufklärung). Vor allem um das spezifische Prädiktionsgewicht der Diskrepanzen gegen das der einfachen Beurteilungsperspektiven zu testen, sind (polynomiale) regressionsanalytische Prozeduren hilfreich. In der vorgestellten Studie lag der diesbezügliche Schwerpunkt auf den Selbst-Aufwärts-Diskrepanzen, in weiteren Studien müssen auch andere Diskrepanzen entsprechend untersucht werden. Nicht zu vergessen ist dabei die Abwärts-Aufwärts-Diskrepanz, die sich hier durch eine besonders hohe Varianzaufklärung auszeichnete (Varianzanalyse). Auch wenn diese mit durchschnittlich 25% auf den ersten Blick nicht gerade besticht, kann sie dennoch als hoch bewertet werden, wenn man bedenkt, dass keine der erfassten demografischen oder organisationsbezogenen Variablen einen bedeutsamen Beitrag zum Erfolg leisten kann. Dabei wäre durchaus plausibel, wenn das Alter mit dem Wohlbefinden oder die Abteilungsgröße mit der Arbeitszufriedenheit zusammenhängen würde.

Generell muss jedoch auch über Regressionsanalysen hinaus gegangen werden. Da Hinweise darauf bestehen, dass viele Variablen im Bedingungsgefüge von Führung, interperspektivischen Diskrepanzen und Erfolg eine Rolle spielen, sollten Pfadanalysen und kanonische Korrelationsanalysen verwendet werden. So könnte beispielsweise besser untersucht werden, welche Rolle die Kommunikation im Führungsprozess spielt – ob sie indirekt über eine bessere interperspektivische Übereinstimmung oder direkt erfolgsrelevant ist. Beide Methoden erfordern allerdings große Stichproben. Jedenfalls sind langfristig Modelle zu entwickeln, die das Zustandekommen interperspektivischer Diskrepanzen und/oder deren Zusammenhang zu Erfolg erklären können.

Desweiteren ist zu untersuchen, ob unterschiedliche Maße der Übereinstimmung (Ausmaß der Abweichung, Richtung der Abweichung, Korrelation) unterschiedlich gute Erfolgsprognosen leisten können. Aus den vorgestellten Ergebnissen deutet vieles darauf hin, dass die korrelative Übereinstimmung weniger ausschlaggebend für Erfolg ist und dass sie von verschiedenen Variablen moderiert werden.

Da die interperspektivischen Diskrepanzen jeweils durch verhältnismäßig milde Ratings aus der einen und verhältnismäßig strenge Ratings aus der anderen Perspektive entstehen, kann aus den Befunden der vorgestellten Studie für praktische Interventionsvorhaben mit dem Ziel, die interperspektivische Übereinstimmung zu verbessern, abgeleitet werden, dass der größtmögliche Erfolg dann zu erzielen ist, wenn die Beurteiler beider Perspektiven in die Intervention einbezogen werden. Diese Empfehlung wird durch die Validität der einfachen Beurteilungsperspektiven untermauert. Ein alleiniger Feedback- oder Coachingprozess mit der Führungskraft würde gewichtiges Veränderungspotential auf dem Weg zu einer höheren interperspektivischen Übereinstimmung also ungenutzt lassen.

Um zu untersuchen, ob wirklich die interperspektivischen Diskrepanzen einen spezifischen, kritischen Beitrag über die einfachen einzelnen Beurteilungsperspektiven hinaus zur Erfolgsaufklärung leisten, müssen in jedem Fall größere Stichproben verwendet werden. Auf aggregierter Ebene (Zusammenfassung der Mitarbeiterratings eines Teams) lagen in der vorgestellten Studie nur geringe Stichprobenumfänge für die Selbst-Fremd- und Fremd-Fremd-Diskrepanzen vor. Dies schränkt die statistische Aussagekraft ein, auch wenn der Problematik dadurch begegnet wurde, dass zusätzlich zu den Signifikanzwerten Effektstärken berechnet wurden.

Die durchschnittliche Rücklaufquote über alle Hierarchieebenen hinweg (43%) lag im unteren Normbereich dessen, was für Mitarbeiterbefragungen im Allgemeinen zu erwarten ist. Allein die Führungskräfte beteiligten sich in hohem Ausmaß (64% vs. beispielsweise 53% bei Becker et al., 2002). Generell sind Rücklaufquoten freilich lediglich vor dem Hintergrund aktueller Kontextfaktoren zu bewerten. Diese waren, angefangen von der Brisanz des Beurteilungsgegenstands (Vorgesetztenbeurteilung) bis hin zu historischen Erfahrungen aus der ehemaligen DDR, für die vorgestellte Studie eher ungünstig. An dieser Stelle sollen – quasi als Checkliste zur Vorbereitung weiterer Befragungen – die wichtigsten

Gründe genannt werden, die von Angeschriebenen für ihre Nicht-Beteiligung an der vorliegenden Studie vorgebracht wurden:[228]

- Angst vor persönlichen Konsequenzen oder vor Rache des Vorgesetzten (basierend auf Misstrauen der Anonymität)
- Subjektiv erlebte Irrelevanz der Thematik:
  "Man ist zufrieden, einen Arbeitsplatz zu haben. Alles andere interessiert nicht."
  "In den Köpfen ist das Autobauen – es geht um nichts anderes. Man nimmt sich nicht einmal Zeit, um sich Notizen zu machen."
- Umfassende Ablehnung von Beurteilungen vor dem Hintergrund von DDR-Erfahrungen
- Feedback-Kultur: „Die Meinung von Mitarbeitern wird nicht geschätzt." Untermauert wird diese Annahme dadurch, dass die Mitarbeiter, die verhältnismäßig kurz im Unternehmen oder bei der aktuellen Führungskraft arbeiten, in der vorliegenden Befragung deutlich schneller geantwortet haben als „alte Hasen".
- Erfahrungen: Eindruck mangelnden Datenschutzes und ausbleibende Veränderungsmaßnahmen im Anschluss
- Bericht starker Teamnormen: Wenn eine Führungskraft gegen die Beteiligung eintritt, antworten auch weniger Mitarbeiter.
- Unterschätzung des Unternehmensinteresses und gleichzeitige Überbewertung des wissenschaftlichen Projekts durch Einzelne
- Zeitaufwand

Die aufgelisteten Gründe stimmen gut mit den Punkten überein, die Jöns und Mataja (1998) als hinderliche Bedingungen für eine Teilnahme an Befragungen diskutieren. Auch die Fokussierung auf *ein* Unternehmen kann im Übrigen für die verhältnismäßig geringe Beteiligungsquote der Aufwärtsbeurteiler mitverantwortlich gemacht werden. Dadurch konnten sich Zweifel an der Anonymität der Angaben länger halten. Für weitere Studien, die ihre Erhebung auf ein Unternehmen beschränken, wird deshalb empfohlen, auf eine Kompletterhebung zu verzichten und stattdessen dem Betriebsrat eine Zufallsziehung zu befragenden Personen zu überlassen.

Durch eine derartige Abgabe von Kontrolle könnte außerdem Dateneinsicht in objektive Erfolgskriterien (bzw. deren individuelle Zuordnung zu den abgegebenen Ratings) eher gewährleistet werden. Für Gesundheits- und Innovationsdaten war dies für die vorliegende Studie geplant, aber in letzter Konsequenz nicht umsetzbar. So konnten also weder indirekte, z.B. Fluktuation, noch direkte „harte Erfolgskriterien", z.B. Leistungsdaten, berücksichtigt werden, die die Common-Method-Problematik deutlich bis gänzlich verringert

[228] Zur Sprache kamen diese Gründe in Teamgesprächen vor Ort, in Form von Notizen auf leer zurück gesendeten Fragebögen oder in Mails, die zu diesem Zweck verfasst wurden.

hätten.[229] Es sei erwähnt, dass auch solche „objektiven“ Kriterien aus ganz pragmatischer Sicht problembehaftet gewesen wären. Um Analysen mit der Gesamtstichprobe durchführen zu können, hätten daraus nämlich erst abteilungsübergreifende Maßzahlen entwickelt werden müssen, die für die Produktion genauso gelten wie für den Angestelltenbereich im Personalwesen. Während dies für Krankheitstage über eine Relativierung am Abteilungsmittel noch machbar ist, sind für Leistungsmaße die Möglichkeiten stark begrenzt. Aus dem Blickwinkel der transformationalen Führung ist ohnehin zu bezweifeln, ob harte Daten die angestrebten Führungserfolge besser abzubilden vermögen als weiche. Denn erfolgreich ist eine Führung demnach dann, wenn sie eine veränderte Werthaltung sowie eine Bewertung der Unternehmensziele und damit auch der eigenen Tätigkeiten als erstrebenswert hervorruft.

Generell ist die Selektivität der Stichprobe zwiespältig zu betrachten. Dass die Daten ausschließlich von Führungskräften und Mitarbeitern eines einzigen Unternehmens (zwei Standorte) stammen, schränkt auf der einen Seite die Generalisierbarkeit der Befunde auf andere Stichproben und insbesondere auf andere Branchen ein. Auf der anderen Seite war es gerade ein Anliegen der vorliegenden Studie, Befunde aus der empirischen Führungsforschung, die entweder in hierzulande ökonomisch irrelevanten und hochspezialisierten Branchen (z.B. Militär/Marine) oder an US-amerikanischen beziehungsweise international agierenden Industriestichproben gewonnen wurden, an traditionellen Industriezweigen zu replizieren. Im Wesentlichen ist dies auch gelungen. Da die in der vorliegenden Stichprobe gefundenen Übereinstimmungsindizes aber eher denen aus Militär- als denen aus Industriestichproben ähnlich sind (Kapitel 7.2.3) und da vergleichsweise sehr geringe Selbst-Abwärts-Diskrepanzen gefunden wurden, ist von einer Generalisierung der Befunde solange Abstand zu nehmen, bis Kreuzvalidierungen stabilere Aussagen zulassen.

Denkbar ist, dass ein Teil dieser unerwarteten Verzerrungen am „Standort Ost“, also in Beurteilungserfahrungen der ehemaligen DDR, begründet liegt (vgl. Bungard, 1997). Möglicherweise leisten auch die strukturellen und berufspolitischen Rahmenbedingungen der deutschen Automobilbranche stärkeren Verzerrungen Vorschub als dies bei internationalen Industriebetrieben der Fall ist. Nicht unplausibel erscheint, dass sowohl die interperspektivischen Übereinstimmungen als auch deren Erfolgskorrelate branchenabhängig sind. So können branchenspezifische Erwartungen und Normen durchaus Führungswahrnehmungen, Antwortverhalten und den potentiellen Wirkungskreis von Personalführung insgesamt beeinflussen. Dem Fahrzeugbau wird eine Branchenkultur zugeschrieben, die eher durch langsames Feedback und niedriges Risiko, Hierarchie und Formalismus charakterisiert ist (Deal & Kennedy, 1982; Grieb, 1997). Denkbar ist, dass eine solche Branchenkultur dem Zeitgeist „hinterher hinkt“ und erst in einigen Jahren wirklich bereit für ehrliche multiperspektivische Führungsbeurteilung sein wird.

[229] Einteilung in indirekte und direkte Kriterien nach Lehner (1995)

Die Selektivität der Stichprobe bringt gemeinsam mit dem gewählten Sandwich-Design bei einer Komplettbefragung mit sich, dass eine Mehrfachbelastung der Befragten eintritt. Viele Fragestellungen hätten zwar ohnedies gar nicht beantwortet werden können. Dennoch ist aber zu überprüfen, ob die Ist-Beurteilungen als unabhängig von den Beurteilungen auf den zusätzlich bearbeiteten Modus-Versionen zu betrachten sind. Dabei zeigt sich, dass das Niveau der Ist-Beurteilungen von der Bearbeitung der Ideal- oder Wichtigkeitsversion unbeeinflusst bleibt. Dies gilt für beide davon potentiell betroffenen Beurteilungsperspektiven (Aufwärts- und Selbstperspektive) und für alle sieben Führungsdimensionen (Aufwärtsperspektive/Ideal: $\bar{d}$ = .14, Selbstperspektive/Ideal: $\bar{d}$ = .13, Aufwärtsperspektive/Wichtigkeit: $\bar{d}$ = .11). Insgesamt ist also auszuschließen, dass das gewählte Design – insbesondere die Vielzahl an Fragebogenversionen – die Güte der Beurteilungen beeinträchtigt.

Bereits in Kapitel 7.2.1 wurde resümiert, dass die Akkuratheit von Ratings aufgrund mangelnder „wahrer“ Referenzwerte nicht gemessen werden kann. Diesem Ziel könnte zukünftig beispielsweise dadurch besser Rechnung getragen werden, dass die fokussierten Führungskräfte nicht nur ihr eigenes Führungsverhalten beurteilen, sondern zusätzlich einschätzen, wie ihre Führung von den einzelnen Mitarbeitern wahrgenommen wird. Wenn gleichzeitig Mitarbeiterbeurteilungen dieser Führung vorliegen, dann kann durch den Vergleich dieser beiden Ratings erfasst werden, ob Führungskräfte die Fremdbilder akkurat wahrnehmen. Kenny und DePaulo (1993) sprechen in diesem Fall von „dyadic meta-accuracy“. Mehr Erkenntnisse über die Sensibilität der Führungskräfte, also darüber, wie gut sie zu einem Perspektivenwechsel in der Lage sind, könnte zum besseren Verständnis der Selbst-Fremd-Diskrepanzen und in der Folge auch des Entstehungszusammenhangs zwischen Führung und Erfolg beitragen. Wenn nämlich bei einer genauen Vorstellung über die Fremdbilder, also einer hohen dyadischen Meta-Akkuratheit, auch eine hohe Selbst-Fremd-Übereinstimmung vorliegt, dann ist davon auszugehen, dass der Prozess der Rückmeldung zentral für ein genaues Selbstbild ist.

Rückmeldung bedeutet dann, dass Mitarbeiter und andere Dritte ihr Fremdbild von der Führung der fokussierten Führungskraft so an diese vermitteln, dass es bei ihr auch möglichst unverzerrt ankommt. Dass dies im vorliegenden Kontext äußerst selten der Fall ist (vgl. Longenecker & Gioia, 1992; Senger, in Vorb.), muss nicht ausführlich wiederholt werden. Sollten sich die angerissenen Prognosen allerdings bestätigen, fänden all die Praktiker empirische Unterstützung, die genau dies seit langem fordern (vgl. Bracken et al., 2001). Ein solcher Ansatz kommt der ursprünglichen Theorie zur Entwicklung des Selbst von Cooley (1902), dem symbolischen Interaktionismus, sehr nahe. Wenn aber die Genauigkeit der Selbstwahrnehmung eine Reflexion der Selbst-Fremd-Interaktion darstellt, dann sollten erfolgreich Führende positive Führungsqualitäten genauer einschätzen können und nicht erfolgreich Führende entsprechende negative Führungsverhaltensweisen.

Ein Zusammenhang zwischen der dyadischen Meta-Akkuratheit und verschiedenen Erfolgsvariablen ist entweder indirekt über eine hohe Selbst-Fremd-Übereinstimmung (siehe oben) oder auch direkt plausibel. Denn wer die Fremdbilder seiner Mitarbeiter genau kennt, kann seine Führung auch entsprechend ihrer Bedürfnisse ausrichten – und umgekehrt. Ergebnisse aus dem Bereich der Führungsforschung existieren bisher nicht. Aus einem anderen Bereich liegen erste Hinweise auf einen Zusammenhang zwischen der Meta-Akkuratheit und relevanten Outcomevariablen vor (Cillessen & Bellmore, 1999). So sind diejenigen Schüler weniger einsam, die sich auf verschiedenen Verhaltensdimensionen akkurat im erläuterten Sinne einschätzen. Bereits an diesem Befund wird schnell deutlich, dass beide Kausalrichtungen plausibel sind.

Was die Varianzaufklärung *verschiedener Erfolgskriterien* betrifft, muss fast durchgängig für alle Analysen festgehalten werden, dass AZ2 (Zufriedenheit mit Kontextbedingungen) und Wohlbefinden am wenigsten durch die Führung zu beeinflussen sind. Da diese beiden Erfolgsskalen mangelhafte Gütekriterien aufweisen, sollte der Befund mit verbesserten Instrumenten repliziert werden. Auch ohnedem sind die Befunde, die auf Erhebungen der eigens für diese Studie konstruierten Fragebögen beruhen, deshalb unbedingt zu replizieren, weil die Konstruktionsstichprobe gleichzeitig die Erhebungsstichprobe darstellte.

Prinzipiell ist erstrebenswert, in diesen Forschungsbereich andere Methoden als Fragebögen einzuführen, denn der Fragebogen als solcher kann unerwünschte Antwortreaktionen hervorrufen (vgl. Schwarz & Hippler, 1987 zu sozial-kognitiven Verzerrungen von Fragebogenantworten). Was für die Operationalisierung der Erfolgskriterien durch den Zugang von „harten“ Erfolgskriterien noch realisierbar erscheint, ist für die Operationalisierung des Führungsverhaltens schwierig. Denkbar sind Interviewstudien (beispielsweise standardisierte Befragungen zu „critical incidents“) oder Simulationsstudien, die eine strukturierte Beobachtung ausgewählter Führungsverhaltensweisen zulassen.

Der Einsatz von Führungssimulationen eröffnet die Möglichkeit, experimentelle Manipulationen vorzunehmen. So könnte die Schwierigkeit der Führungsaufgaben graduell manipuliert werden, um zu überprüfen, ob dadurch der Zusammenhang zwischen Führung bzw. interperspektivischer Übereinstimmung und Erfolg beeinflusst wird. Erste Hinweise darauf erbrachten eine Laborstudie zu führungsfremder Leistung (Grammatik-Übungen; Kruger & Dunning, 1999) und eine Studie zur Validität von Assessment Centern (Stempfle et al., 2004). Es kann angenommen werden, dass bei schwierigen Aufgaben nur noch die Übereinstimmer erfolgreich sind. Plausibel ist nämlich, dass erst bei sehr hohen Anforderungen (sehr schwierigen Führungsaufgaben) die anderweitig erfolgreichen Strategien von Überschätzern im Misserfolg münden, weil die Bewältigungsstrategien aufgrund der Überschätzung eigener Fähigkeiten nicht genau genug auf die Anforderungen abgestimmt werden.

Solche Studiendesigns haben den Vorteil, dass das Führungsverhalten im Unterschied zum hier eingesetzten Fragebogen auch eher situationsabhängig erfasst werden kann. Aufgrund der klassischen Theorien zur Personalführung erscheint es durchaus lohnenswert, eine kontingenztheoretische Sichtweise auch in den Forschungsbereich der multiperspektivischen Führungsbeurteilung einzuführen. Dann müssen auch Augmentationseffekte erneut überprüft werden, da möglicherweise die berücksichtigten Führungsverhaltensweisen unter verschiedenen Bedingungen besonders effektiv sind.

Um den verschiedenen Fragestellungen Rechnung zu tragen, wurden also im Vergleich zu bisher publizierten Forschungsarbeiten in mehrfacher Hinsicht Erweiterungen vorgenommen. So wurden erstens Ergebnisse aus der Automobilindustrie gewonnen, um empirische Anhaltspunkte darüber zu erhalten, inwiefern die bereits existierenden Befunde zu interperspektivischen Übereinstimmungen branchenübergreifende Gültigkeit besitzen. Zweitens ging es um eine Erweiterung der verwendeten Erfolgsmaße. Auch wenn keine „harten“ Daten vorlagen, ist die Verwendung multipler Erfolgskriterien innerhalb einer Studie deshalb anzuraten, weil dadurch die Varianzaufklärung direkt vergleichbar ist. Drittens ließ die *multidimensionale* Erfassung von Führung eine spezifischere Analyse und Interpretation der interperspektivischen Übereinstimmung sowie ihrer Einflussnahme auf Erfolg zu. Dazu gehört auch eine weitere Validierung der MLQ-Subskalen. Durch den Einsatz verschiedener Instrumente und Methoden konnte viertens der (nicht kausal zu interpretierende) prognostische Nutzen der interperspektivischen Diskrepanz von Führungsbeurteilungen weiter erforscht werden. Im Mittelpunkt stand dabei die Selbst-Aufwärts-Diskrepanz. Doch es wurde auch untersucht, ob die Selbst-Abwärts- und die Abwärts-Aufwärts-Diskrepanz einen prognostischen Beitrag zum Erfolg leisten können. Dabei geht Letzteres grundlegend über bisherige Forschungsarbeiten hinaus. Letztlich kann die Betrachtung einer Vielzahl an potentiell moderierenden Variablen (auf die interperspektivische Übereinstimmung sowie die prädiktive Validität von Führungsbeurteilungen) und die Diskussion einer Vielzahl von Erklärungsmechanismen als Ausgangspunkt für spezifischere Fragestellungen mit „strengeren“ Untersuchungsdesigns genutzt werden. Erst dann sind nämlich Aussagen über zugrunde liegende kausale Mechanismen zu treffen. Langfristig sollten diese ein Mosaik ergeben, aus denen Modelle über das Zustandekommen von Erfolg entwickelt werden können.

Abschließend soll festgehalten werden, dass die Selbst-Fremd-Übereinstimmung ein sehr zukunftsweisendes Konstrukt ist, das in verschiedenen Anwendungsfeldern der Psychologie eingesetzt werden kann. So ist die Problematik der Selbst-Fremd-Übereinstimmung und ihrer Auswirkung auf andere Variablen (hier Erfolgsvariablen) keineswegs genuin organisationspsychologischer Natur. In der Klinischen Psychologie ist beispielsweise die „Einsicht“ (Awareness) eine Schlüsselvariable, die bei vielen Störungen Veränderungs-

prozesse verhindert. Methoden wie das Biofeedback werden gezielt zur Verbesserung der Selbstwahrnehmung angewandt (Rief & Birbaumer, 2000).

Betrachtet man die angewandte Organisationspsychologie, dann sind multiperspektivische Beschreibungen, häufig in Form von 270°-Beurteilungen, zum Zwecke multiperspektivischen Feedbacks am gebräuchlichsten. Für verschiedene andere personen- und organisationsbezogene Maßnahmen könnte sich die Verwendung multiperspektivischer Führungsbeurteilungen in der organisationspsychologischen Praxis aber auch als hilfreich erweisen (Überblick bei Rathgeber & Jonas, 2002). Zugegeben: Es sind noch viele Forschungserkenntnisse erforderlich, allein, um beispielsweise die beste Maßzahl für Übereinstimmung oder gar für Selbstreflexionsfähigkeit und „Self-Awareness“ entwickeln zu können – Konstrukte, die derzeit durch die Übereinstimmung selbst operationalisiert sind (Stempfle et al., 2004; Van Velsor et al., 1993). Unabhängig von der „Übersetzung“ der Selbst-Fremd-Übereinstimmung in organisationspsychologische Schlüsselvariablen, zeichnen Stempfle et al. (2004) einen erfolgsversprechenden Weg vor, die Selbst-Fremd-Übereinstimmung als diagnostisches Kriterium in Assessment Center zur Erhöhung deren Validität zu integrieren. Dort zeigte sich – wenn auch zunächst an einem geringen Stichprobenumfang – zusätzlich die Tendenz, dass Selbstreflexion eine Voraussetzung für Selbstmodifikation, also für eine zielgerichtete Veränderung des eigenen Verhaltens, darstellt. Insofern kann die Selbst-Fremd-Übereinstimmung einen wichtigen Prädiktor für die Lernfähigkeit von Personen darstellen.

Vielleicht sollten Unternehmen von der Offenheit des Tanzes lernen –
legen wir Musik auf!

8 Literaturverzeichnis

Aiken, L. S. & West, S. G. (1993). *Multiple regression. Testing and interpreting interactions* (3rd ed.). Newbury Park: Sage.

Aitkenhead, M. (1984). Impression-management and consistency effects in the processing of feedback. *British Journal of Social Psychology, 23*, 213-222.

Albach, H. (1983). Zum Einfluss von Führungsgrundsätzen auf die Personalführung. In R. Wunderer (Hrsg.), *Führungsgrundsätze in Wirtschaft und öffentlicher verwaltung* (S. 2-16). Stuttgart: Poeschel.

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L. Yurak, T. J. & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology, 68,* 804-825.

Allerbeck, M. (1978). Questionnaires on leadership behavior: Problems and results. *Psychologie und Praxis*, *22* (2)*,* 69-83.

Allison, S. T., Messick, D. M. & Goethals, G. R. (1989). On being better but not smarter than others: The Muhammad Ali effect. *Social Cognition, 7,* 275-296.

Ambrosch, R., Lippert, E. & Schneider, P. (1978). Die LPC-Skala als Messinstrument. *Psychologie und Praxis, 22,* 34-44.

Antonakis, J., Avolio, B. J. & Sivasubramaniam, N. (2003). Context and leadership: an examination of the nine-factor full-range leadership theory using the Multifactor Leadership Questionnaire. *The Leadership Quarterly, 14,* 261-295.

Antonioni, D. (1994). The effects of feedback accountability on upward appraisal ratings. *Personnel Psychology, 47,* 349-356.

Argyris, Ch. & Schön, D. A. (1999). *Die lernende Organisation. Grundlagen, Methode, Praxis*. Stuttgart: Klett-Cotta.

Ashford, S. J. (1989). Self-assessment in organizations: A literature review and integrative model. *Research in Organizational Behavior, 11,* 133-174.

Astin, H. S. & Astin, A. W. (1996). *A social change model of leadership development guidebook.* Los Angeles, CA: Higher Education Research Institute.

Atwater, L., Ostroff, Ch., Yammarino, F. & Fleenor, J. (1998). Self-other agreement: Does it really matter? *Personnel Psychology, 51,* 577-598.

Atwater, L., Roush, P. & Fischthal, A. (1995). The influence of upward feedback on self- and follower ratings of leadership. *Personnel Psychology, 48*, 35-59.

Atwater, L. E., Waldman, D., Atwater, D. & Cartier, P. (2000). An upward feedback field experiment: Supervisors' cynicism, follow-up and commitment to subordinates. *Personnel Psychology, 53,* 275-297.

Atwater, L. & Yammarino, F. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology, 45*, 141-164.

Atwater, L. & Yammarino, F. (1997). Self-other rating assessment. In G.R. Ferris (Ed.), *Research in personnel and human resources management* (Vol. 15, pp. 121-174). Greenwich: JAI Press.

Avolio, B. J. (1999). *Full leadership development. Building the vital forces in organizations.* Thousand Oaks: Sage.

Avolio, B. J. & Bass, B. M. (1988). Transformational leadership, charisma, and beyond. In J. G. Hunt, B. R. Baliga, H. P. Dachler & C. A. Schriesheim (Eds.), *Emerging leadership vistas* (pp. 29-50). Lexington, MA: Lexington Books.

Avolio, B. J. & Bass, B. M. (1995). Individual consideration viewed at multiple levels of analysis: A multi-level framework for examining the diffusion of transformational leadership. *Leadership Quarterly, 6,* 199-218.

Avolio, B. J. & Bass, B. M. (1998). You can drag a horse to water but you can't make it drink unless it is thirsty. *The Journal of Leadership Studies, 4* (1)*,* 1-17.

Avolio, B. J. & Bass, B. M. (1999). Re-examining the components of transformational and transactional leadership using the Multifactor Leadership Questionnaire. *Journal of Occupational and Organizational Psychology, 72,* 441-462.

Avolio, B. J. & Bass, B. M. (2002) (Eds.). *Developing potential across a full range of leadership: cases on transactional and transformational leadership.* Mahwah, NJ: Lawrence Erlbaum Associates.

Avolio, B. J., Bass, B. M. & Jung, D. I. (1999). Re-examining the components of transformational and transactional leadership using the Multifactor Leadership Questionnaire. *Journal of Occupational and Organizational Psychology, 72*, 441-462.

Avolio, B. J. & Gibbons, T. C. (1988). Developing transformational leaders: A life span approach. In J. A. Conger & R. N. Kanungo (Eds.), *Charismatic leadership. The elusive factor in organizational effectiveness* (pp. 276-308). San Francisco: Jossey-Bass.

Avolio, B. J., Waldman, D. A. & Einstein, W. O. (1988).Transformational leadership in a management game simulation. *Group and Organization Studies, 13*, 59-80.

Avolio, B. J., Yammarino, F. J. & Bass, B. M. (1991). Identifying common methods variance with data collected from a single source: An unresolved sticky issue. *Journal of Management, 17,* 571-587.

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2000). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (9. Aufl.). Berlin: Springer.

Bailey, R. C. & Bailey K. G. (1974). Self-perceptions of scholastic ability at four grade levels. *Journal of Genetic Psychology, 124*, 197-212.

Barbuto, J. E. (1997). Taking the charisma out of transformational leadership. *Journal of social behavior and personality*, *12* (3), 689-697.

Baril, G. L., Ayman, R. & Palmiter, D. J. (1994). Measuring leader behavior: Moderators of discrepant self and subordinate descriptions. *Journal of Applied Social Psychology, 24*, 82-94.

Barling, J., Loughlin, C. & Kelloway, E. K. (2002). Development and test of a model linking safety-specific transformational leadership and occupational safety. *Journal of Applied Psychology, 87,* 488-496.

Barling, J., Weber, T. & Kelloway, E. K. (1996). Effects of transformational leadership training on attitudinal and financial outcomes: A field experiment. *Journal of Applied Psychology, 81*, 827-832.

Bass, B. M. (1985). *Leadership and performance beyond expectations.* New York: Free Press.

Bass, B. M. (1990). From transactional to transformational leadership: Learning to share the vision. *Organizational Dynamics, 18,* 19-31.

Bass, B. M. (1994). (Ed.). *Improving organizational effectiveness through transformational leadership*. Thousand Oaks, CA: Sage.

Bass, B. M. (1997). Does the transactional-transformational paradigm transcend organizational and national boundaries? *American Psychologist, 52* (2)*,* 130-139.

Bass, B. M. (1998). *Transformational leadership: Industrial, military, and educational impact.* Mahwah: Lawrence Erlbaum Associates.

Bass, B. M. (1999). Two decades of research and development in transformational leadership. *European Journal of Work and Organizational Psychology*, 8 (1), 9-26.

Bass, B. M. & Avolio, B. J. (1990). Training and development of transformational leadership: Looking to 1992 and beyond. *European Journal of Industrial Training, 14,* 21-27.

Bass, B. M. & Avolio, B. J. (1993). Transformational leadership: A response to critiques. In M. M. Chemers & R. Ayman (Eds.), *Leadership theory and research. Perspectives and directions* (pp. 49-80). San Diego: Academic Press.

Bass, B. M. & Avolio, B. J. (1994) (Eds.). *Improving organizational effectiveness through transformational leadership.* Thousand Oaks: Sage.

Bass, B. M. & Avolio, B. J. (1995). *MLQ Multifactor Leadership Questionnaire. Technical report, leader form, rater form, and scoring key for MLQ form 5x-short*. Binghampton: Mind Garden.

Bass, B. M. & Avolio, B. J. (1997). *Full range leadership development. Manual for the Multifactor Leadership Questionnaire.* Redwood City: Mind Garden.

Bass, B. M. & Avolio, B. J. (1998). Improving organizational effectiveness through transformational leadership. Introduction. In G. R. Hickman (Ed.), *Leading organizations: Perspectives for a new era* (pp. 135-140). Thousand Oaks: Sage.

Bass, B. M. & Avolio, B. J. (2000). *MLQ Multifactor Leadership Questionnaire. Technical report, leader form, rater form, and scoring key for MLQ form 5x-short* (2nd ed.). Binghampton: Mind Garden.

Bass, B. M., Avolio, B. J., Jung, D. I. & Berson, Y. (2003). Predicting unit performance by assessing transformational and transactional leadership. *Journal of Applied Psychology, 88,* 207-218.

Bass, B.M. & Steidlmeier (1999). Ethics, character, and authentic transformational leadership behavior. *Leadership Quarterly, 10,* 181-218.

Bass, M. B., Waldman, D. A., Avolio, B. J. & Bebb, M. (1987). Transformational leadership and the falling dominoes effect. *Group and Organization Studies*, *12*, 73-87.

Bass, M. B. & Yammarino, F. J. (1991). Congruence of self and others' leadership ratings of naval officers for understanding successful performance. *Applied Psychology: An International Review, 40*, 437-454.

Bauer, T. N. & Green, S. G. (1996). Compliment for the attentive reader. *Psychological Reports, 73*, 563-573.

Baum, R. J., Locke, E. A. & Kirkpatrick, S. A. (1998). A longitudinal study of the relation of vision and vision communication to venture growth in entrepreneurial firms. *Journal of Applied Psychology*, *83*, 43-54.

Baumeister, R. F. (1989). Motives and costs of self-presentation. In R. A. Giacalone & P. Rosenfeld (Eds.). *Impression management in the organization* (pp. 57-71). Hillsdale, N. J.: Lawrence Erlbaum Associates.

Becker, J., Ayman, R. & Korabik, K. (2002). Discrepancies in self/subordinates' perceptions of leadership behavior: Leader's gender, organizational context, and leader's self monitoring. *Group and Organization Management, 27,* 226-244.

Beckmann, J., Zimolong, B., Stapp, M. & Elke, G. (2001). Personalmanagement erfolgreicher Betriebe. In B. Zimolong (Hrsg.), *Management des Arbeits- und Gesundheitsschutzes. Die erfolgreichen Strategien der Unternehmen (S. 49-82).* Wiesbaden: Gabler.

Beehr, T. A., Ivanitskaya, L., Hansen, C. P., Erofeev, D. & Gudanowski, D. M. (2001). Evaluation of 360 degree feedback ratings: Relationships with each other and with performance and selection predictors. *Journal of Organizational Behavior, 22,* 775-788.

Bennis, W. (1976). *The unconcious conspiracy.* New York: Amacon Press.

Bennis, W. (1989). *On becoming a leader.* London: Business Books.

Bennis, W. & Biederman, P. W. (1998). *Geniale Teams. Das Geheimnis kreativer Zusammenarbeit.* Frankfurt: Campus.

Bennis, W. & Nanus, B. (1985). *Leaders. The strategies for taking charge.* New York: Harper & Row.

Bennis, W. & Nanus, B. (1990). *Führungskräfte. Die vier Schlüsselstrategien erfolgreichen Führens.* Frankfurt: Campus. (Übersetzung der Originalausgabe „Leaders. The strategies for taking charge" von 1985).

Bernardin, H. J. (1978). Effects of rater training on leniency and halo errors of student ratings of instructors. *Journal of Applied Psychology, 65,* 60-66.

Bernardin, H. J. & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work.* Boston: Kent.

Bernardin, H. J., Dahmus, S. A. & Redmon, G. (1993). Attitudes of first-line supervisors toward subordinate appraisals. *Human Resource Management, 32,* 315-324.

Beyer, S. & Bowden, E. M. (1997). Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin, 23,* 157-172.

Binning, J. F., Zaba, A. J. & Whattam, J. C. (1986). Explaining the biasing effects of performance cues in terms of cognitive categorization. *Academy of Management Journal, 29,* 521-535.

Blake, R. R. & McCanse, A. A. (1995). *Das GRID-Führungsmodell* (3. Aufl.). Düsseldorf: ECON.

Blake, R. R. & Mouton, J. S. (1964). *The managerial grid.* Houston: Gulf.

Bobko, P. (1995). *Correlation and regression. Principles and applications for industrial/organizational psychology and management*. New York: McGraw-Hill, Inc.

Bodenhausen, G. V. & Wyer, R. S. (1987). Social cognition and social reality: Information acquisition and use in the laboratory and the real world. In H. J. Hippler, N. Schwarz & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 6-41). New York: Springer.

Böhnisch, W. (1991). *Führung und Führungskräftetraining nach dem Vroom/Yetton-Modell.* Stuttgart: Poeschel.

Bommer, W. H., Johnson, J. L., Rich, G. A., Podsakoff, P. M. & MacKenzie, S. B. (1995). On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology, 48*, 587-605.

Bono, J. E. & Judge, T. A. (2004). Personality and transformational and transactional leadership: A meta-analysis. *Journal of Applied Psychology 89*, 901-910.

Borg, I. (2000). *Führungsinstrument Mitarbeiterbefragung: Theorien, Tools und Praxiserfahrungen*. (2. Aufl.). Göttingen: Verlag für Angewandte Psychologie.

Borg, I. (2001). Mitarbeiterbefragungen. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 373-396). Göttingen: Hogrefe.

Borg, I. (2003). Affektiver Halo in Mitarbeiterbefragungen. *Zeitschrift für Arbeits- und Organisationspsychologie, 47,* 1-11.

Borman, W. C. (1974). The rating of individuals in organizations: An alternative approach. *Organizational Behavior and Human Perfomance, 12,* 105-124.

Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. *Journal of Applied Psychology, 60,* 556-560.

Borman, W. C. (1993). Job behavior, performance, and effectiveness. In M. D. Dunnetee & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., pp. 271-326). Palo Alto, CA: Consulting Psychologists Press.

Borman, W. C. (1997). 360° ratings: An analysis of assumptions and a research agenda for evaluating their validity. *Human Resource Management Review, 7,* 299-315.

Borman, W. C. & Motowidlo, S. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco, CA: Jossey-Bass, Inc.

Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5. Aufl.). Berlin: Springer.

Bortz, J. & Döring, N. (2003). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (3. Aufl.). Springer: Heidelberg.

Bortz, J., Lienert, G. A. & Boehnke, K. (1990). *Verteilungsfreie Methoden in der Biostatistik*. Berlin: Springer.

Boski, P. (1983). Egotism and evaluation in self and other attributions for achievement related outcomes*, European Journal of Social Psychology, 13,* 287-304.

Bozeman, D. P. (1997). Interrater agreement in multi-source performance appraisal: A commentary. *Journal of Organizational Behavior, 18,* 313-316.

Bracken, D. W. (1997). Maximizing the uses of multi-rater feedback. In D. W. Bracken, M. A. Dalton, R. A. Jako, C. D. McCauley & V. A. Pollmann (Eds.), *Should 360 degree feedback be used only for developmental purposes?* (pp. 11-17). Greensboro, NC: Center for Creative Leadership.

Bracken, D. W., Timmreck, C. W. & Church, A. H. (Eds.). (2001). *The handbook of multisource feedback.* San Francisco: Jossey-Bass.

Bracken, D. W., Timmreck, C. W., Fleenor, J. W. & Summers, L. (2001). 360° feedback from another angle. *Human Resource Management, 50,* 3-20.

Brandstätter, H. (1970). Die Beurteilung von Mitarbeitern. In A. Mayer & B. Herwig (Hrsg.), *Handbuch der Psychologie, Bd. 9: Betriebspsychologie* (S. 668-734). Göttingen: Hogrefe.

Brehm, J. W. (1966). *A theory of psychological reactance*. New York: Academic Press.

Brenner, D. & Brenner, F. (2002). *Beurteilungsgespräche souverän führen.* Köln: Deutscher Wirtschaftsdienst.

Brett, J. F. & Atwater, L. E. (2001). 360° feedback: accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology, 86,* 930–942.

Brodbeck, F. C., Frese, M., Akerblom, S. (2000). Cultural variation of leadership prototypes across 22 European countries. *Journal of Occupational and Organizational Psychology, 73,* 1-29.

Brose, P. (1983). *Partizipation in der Unternehmung*. München: Florenz.

Brown, J. D. (1986). Evaluations of self and others: self-enhancement biases in social judgements. *Social Cognition, 4*, 353-376.

Brown, S. & Eisenhardt, K. (1995). Product development: Past research, present findings, and future directions. *Academy of Management Review, 20,* 343-378.

Brunner, H. (2001). Die Automobilindustrie – eine Säule der deutschen Wirtschaft. *Dresdner Transferbrief, 9* (3). Verfügbar unter: http://www.tu-dresden.de/vd51/trabrief/032001/Leitart.html [02.02.2004].

Brutus, S. & Derayeh, M. (2002). Multisource assessment programs in organizations: An insider's perspective. *Human Resource Development Quarterly, 13,* 187-203.

Brutus, S., Fleenor, J. W. & London, M. (1998). Elements of effective 360-degree feedback. In M. London & W. Tornow, W (Eds.), *Maximizing the value of 360-degree feedback: a process for successful individual and organizational development* (pp. 11-27). San Francisco: Jossey-Bass.

Brutus, S., Fleenor, J. W. & Tisak, J. (1999). Exploring the link between rating congruence and managerial effectiveness. *Canadian Journal of Administrative Science, 16,* 308-322.

Brutus, S., London, M. & Martineau, J. (1999). The impact of 360 degree feedback on planning for career development. *Journal of Management Development, 18,* 676-693.

Bryman, A. (1992). *Charisma and leadership in organizations*. London: Sage.

Bryman, A. (1996). Leadership in organizations. In S. P. Clagg, C. Hardy & W. R. Nord (Eds.), *Handbook of organization studies* (pp. 276-292). Thousand Oaks: Sage.

Bungard W. (1984). Sinn und Unsinn der LPC-Skala – Zur Kritik des Kontingenzmodells von Fiedler. *Gruppendynamik, 15,* 59-74.

Bungard W. (1997). Mitarbeiterbefragungen in den neuen Bundesländern. In W. Bungard & I. Jöns (Hrsg.), *Mitarbeiterbefragung. Ein Instrument des Innovations- und Qualitätsmanagements* (S. 407-417). Weinheim: PVU.

Bungard, W., Fettel, A. & Jöns, I. (1997). Mitarbeiterbefragungen: Verbreitung, Einsatzformen und Erfahrungen bei den 100 umsatzgrößten Unternehmen in der Bundesrepublik Deutschland. In W. Bungard & I. Jöns (Hrsg.), *Mitarbeiterbefragung. Ein Instrument des Innovations- und Qualitätsmanagements* (S. 246-263). Weinheim: PVU.

Bungard, W. & Jöns, I. (1997) (Hrsg.). *Mitarbeiterbefragung. Ein Instrument des Innovations- und Qualitätsmanagements*. Weinheim: PVU.

Bungard, W., Jöns, I. & Schultz-Gambard, J. (1997). Sünden bei Mitarbeiterbefragungen – Zusammenfassung der wichtigsten Fehler und Fallgruben. In W. Bungard & I. Jöns (Hrsg.), *Mitarbeiterbefragung. Ein Instrument des Innovations- und Qualitätsmanagements* (S. 441-455). Weinheim: PVU.

Burns, J. (1978). *Leadership*. New York: Harper & Row.

Bycio, P., Hackett, R. D. & Allen, J. S. (1995). Further assessments of Bass's (1985) conceptualisation of transactional and transformational leadership. *Journal of Applied Psychology, 80*, 468-478.

Cadinu, M. R. & Rothbart, M. (1996). Self-anchoring and differentiation processes in the minimal group setting. *Journal of Personality and Social Psychology, 70*, 661-677.

Calder, B. J. (1977). An attribution theory of leadership. In B. Staw & G. Salancik (Eds.), *New directions in organizational behavior* (pp. 179-204). Chicago: St. Clair.

Carver, C. S. & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality, social, clinical, and health psychology. *Psychological Bulletin, 92,* 111-135.

Carver, C. S. & Scheier, M. F. (1998). *On the self-regulation of behavior*. Cambridge: Cambridge University Press.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate behavioural research, 1,* 245-276.

Charbonneau, D., Barling, J. & Kelloway, E. K. (2001). Transformational leadership and sports performance : The mediating role of intrinsic motivation. *Journal of Applied Social Psychology, 31*, 1521-1534.

Chemers, M. M. (1997). *An integrated theory of leadership*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cheung, G. W. (1999). Multifaceted conceptions of self-other ratings disagreement. *Personnel Psychology, 52,* 1-36.

Choi, I. & Nisbett, R. (1998). Situational salience and cultural differences in the correspondence bias and actor-observer bias. *Personality and Social Psychology Bulletin, 24,* 949-960.

Choi, I., Nisbett, R. & Norenzayan, A. (1999). Causal attribution across cultures: Variation and universality. *Psychological Bulletin, 125,* 47-63.

Church, A. H. (1997b). Do you see what I see? An exploration of congruence in ratings from multiple perspectives. *Journal of Applied Social Psychology, 27,* 983-1020.

Church, A. H. (1997a). Managerial self-awareness in high-performing individuals in organizations. *Journal of Applied Psychology, 83,* 281-292.

Church, A. H., Rogelberg, S. G. & Waclawski, J. (2000). Since when is no news good news? The relationship between performance and response rates in multirater feedback. *Personnel Psychology, 53,* 435–451.

Cialdini, R. B. (1997). *Die Psychologie des Überzeugens. Ein Lehrbuch für alle, die ihren Mitmenschen und sich selbst auf die Schliche kommen wollen*. Bern: Hans Huber.

Cillessen, A. H. N. & Bellmore, A. D. (1999). Accuracy of social self-perceptions and peer competence in middle childhood. *Merrill-Palmer Quarterly, 45*, 650-676.

Clements, C. & Washbush, J. B. (1999). The two faces of leadership: considering the dark side of leader-follower dynamics. *Journal of Workplace Learning, 11*, 170-175.

Cleveland, J. N., Murphy, K. R. & Williams, R. E. (1989). Multiple uses of performance appraisal: prevalence and correlates. *Journal of Applied Psychology, 73,* 130-135.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Colvin, C. R. & Block, J. (1994). Do positive illusions foster mental health? An examination of the Taylor and Brown formulation. *Psychological Bulletin, 116,* 3-20.

Colvin, C. R., Block, J. & Funder, D. C. (1995). Overly positive self-evaluations and personality: Negative implications for mental health. *Journal of Personality and Social Psychology, 68,* 1152-1162

Conger, J. A. (1989). *The charismatic leader: Beyond the mystique of exceptional leadership.* San Francisco: Jossey-Bass.

Conger, J. A. (1990). The dark side of leadership. *Organizational Dynamics*, *19,* 44-55.

Conger, J. A. (1999). Charismatic and transformational leadership in organization: An insider's perspective on these developing streams of research. *Leadership Quarterly, 10,*145-170.

Conger, J. A. & Kanungo, R. N. (1987). Toward a behavioural theory of charismatic leadership in organizational settings. *Academy of Management Review, 12,* 637-647.

Conger, J. A. & Kanungo, R. N. (1988). Training charismatic leadership: A risky and critical task. In J.A. Conger & R.N. Kanungo (Eds.), *Charismatic leadership. The elusive factor in organizational effectiveness* (pp. 309-323). San Francisco: Jossey-Bass.

Conger, J. A. & Kanungo, R. N. (1994). Charismatic leadership in organizations: perceived behavioural attributes and their measurement. *Journal of Organizational Behavior, 15,* 439-452

Conway, J. M. (1999). Distinguishing contextual performance from task performance for managerial jobs. *Journal of Applied Psychology, 84*, 3-13.

Conway, J. M. & Huffcutt, A. I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance, 10*, 331-360.

Conway, J. M., Lombardo, K. & Sanders, K. C. (2001). A meta-analysis of Incremental Validity and Nomological Networks for Subordinate and Peer Ratings. *Human Performance, 14,* 267-303.

Cooley, C. H. (1902). *Human nature and the social order.* New York: Scribner.

Crampton, S. M. & Wagner, J. A. (1994). Percept-percept inflation in microorganizational research: an investigation of prevalence and effect. *Journal of Applied Psychology, 79,* 67-76.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Cronbach, L. & Gleser, G. (1953). Assessing similarity between profiles. *Psychological Bulletin, 50,* 456-473.

Cross, P. (1977). Not can but will college teaching be improved. *New Directions for Higher Education, 17,* 1-15.

Dalton, M. A. (1998). Best practices: Five rationals for using 360-degree feedback in organizations. In M. London & W. Tornow, W (Eds.), *Maximizing the value of 360-degree feedback: a process for successful individual and organizational development* (pp. 59-77). San Francisco: Jossey-Bass.

Dauenheimer, D., Stahlberg, D., Frey, D. & Petersen, L.-E. (2002). Die Theorie des Selbstwertschutzes und der Selbstwerterhöhung. In D. Frey & M. Irle (Hrsg.), *Theorien der Sozialpsychologie* (S. 159-190), Bd. 3 (2. Aufl.). Bern: Hans Huber.

Deal, T. E. & Kennedy, A. A. (1982). *Corporate cultures, the rites and rituals of corporate life.* Reading, MA: Addison- Wesley.

DeGroot, T., Kiker, D. S. & Cross, T. C. (2000). A meta-analysis to review organizational outcomes related to charismatic leadership. *Canadian Journal of Administrative Sciences, 17,* 356-371.

Den Hartog, D., House, R. J., Hanges, P. J., Ruiz-Quintanilla, S. A., Dorfman, P. W. & GLOBE Associates (1999). Culture specific and cross-culturally generalizable implicit leadership theories: Are attributes of charismatic/transformational leadership universally endorsed? *Leadership Quarterly, 10*, 219-256.

Den Hartog, D., Muijen, J. J. van & Koopman, P. L. (1997). Transactional versus transformational leadership: An analysis of the MLQ. *Journal of Occupational and Organizational Psychology*, *70,* 19-34.

DeNisi, A. S. & Kluger, A. N. (2000). Feedback effectiveness: Can 360-degree appraisals be improved? *Academy of Management Executive, 14* (1)*,* 129-139.

DePree, M. (1998). What is leadership? In G. R. Hickman (Ed.), *Leading organizations: Perspectives for a new era* (pp. 130-132). Thousand Oaks: Sage.

Diehl, J. M. & Staufenbiel, T. (2001). *Statistik mit SPSS, Version 10.0.* Eschborn: Klotz.

Domsch, M. & Reinecke, P. (1982). Partizipative Personalentwicklung. *Zeitschrift für betriebswirtschaftliche Forschung, 14,* 64-81.

Donat, M. (1991). Selbstbeurteilung. In H. Schuler (Hrsg.), *Beurteilung und Förderung beruflicher Leistung* (S. 135-145). Stuttgart: Verlag für Angewandte Psychologie.

Dörner, D. (1999). *Bauplan für eine Seele*. Reinbek: Rowohlt.

Downton, J.V. (1973). *Rebel leadership. Commitment and charisma in the revolutionary process.* New York: The Free Press.

Drees, H. B. (1994). *Untersuchung zur Validität eines Assessment Centers. Hinweise zur empirischen Überprüfung eines Assessment Centers unter besonderer Berücksichtigung unterschiedlicher Beobachtertrainings.* Aachen: Dissertationsschrift.

Dumdum, U. R., Lowe, K. B. & Avolio, B. J. (2002). A meta-analysis of the transformational and transactional leadership correlates of effectiveness and satisfaction: An update and extension. In B. J. Avolio & F. J. Yammarino (Eds.), *Transformational and charismatic leadership: the road ahead* (pp. 35-66). Amsterdam: JAI Press.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B. & Burke, M. J. (1996). Meta- analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170-177.

Dunnette, M. D. (1966). *Personnel selction and placement.* Belmont, CA: Brooks-Cole.

Dunnette, M. D. (1993). My hammer or your hammer? *Human Resource Management, 32,* 373-384.

Dunning, D., Meyerowitz, J. A. & Holtzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57,* 1082-1090.

Dvir, T., Eden, D., Avolio, B. & Shamir, B. (2002). Impact of transformational leadership on follower development and performance: A field experiment. *Academy of Management Journal, 45,* 735-744.

Eagly, A. H. & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological Bulletin, 108,* 251-273.

Eagly, A. H., Karau, S. J. & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin, 117,* 125-145.

Edwards, J. R. (1994a). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes*, *58,* 51-100.

Edwards, J. R. (1994b). Regression analysis as an alternative to difference scores. *Journal of Management, 20,* 683-689.

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods*, *4*, 264-286.

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. W. Schmitt (Eds.), *Advances in measurement and data analysis* (pp. 350-400). San Francisco: Jossey-Bass.

Edwards, J. R. & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36*, 1577-1613.

Edwards, M. R. & Ewen, A. J. (2000). *360°-Beurteilung: klareres Feedback, höhere Motivation und mehr Erfolg für alle Mitarbeiter.* München: Beck Wirtschaftsverlag.

Elenkov, D. S. (2002). Effects of leadership on organizational performance in Russian companies. *Journal of Business Research, 55* (6), 467-480.

Elke, G. (2001). Sicherheits- und Gesundheitskultur I. Handlungs- und Wertorientierungen im betrieblichen Alltag. In B. Zimolong (Hrsg.), *Management des Arbeits- und Gesundheitsschutzes. Die erfolgreichen Strategien der Unternehmen* (S. 171-200). Wiesbaden: Gabler.

Ellis, R. A. & Taylor, M. S. (1983). Role of self-esteem within the job search process. *Journal of Applied Psychology, 68*, 632-640.

Evans, M. G. (1974). Extensions of a path-goal theory of motivation. *Journal of Applied Psychology, 59,* 172-178.

Ewen, A. J. & Edwards, M. R. (2001). Readiness for multisource feedback. In D. W. Bracken, C. W. Timmreck & A. H. Church (Eds.), *The handbook of multisource feedback* (pp. 33-47). San Francisco: Jossey-Bass.

Farh, J. L. & Dobbins, G. H. (1989). Effects of comparative performance information on the accuracy of self-ratings and agreement between self- and supervisor ratings. *Journal of Applied Psychology, 74*, 606-610.

Farr, J. L. (1991). Leistungsfeedback und Arbeitsverhalten. In H. Schuler (Hrsg.), *Beurteilung und Förderung beruflicher Leistung* (S. 57-80). Stuttgart: Verlag für Angewandte Psychologie.

Fayol, H. (1949). *General and industrial management*. London: Pitman. [zuerst: *Administration industrielle et général*. Paris (1916)].

Felfe, J. & Goihl, K. (2002). Transformational leadership and commitment. In J. Felfe (Eds.), *Organizational development and leadership* (pp. 87-124). Frankfurt: Peter Lang.

Felfe, J. & Goihl, K. (2003). Deutsche überarbeitete und ergänzte Version des Multifactor Leadership Questionnaire (MLQ). In A. Glöckner-Rist (Hrsg.), *ZUMA-Informationssystem. Elektronisches Handbuch sozialwissenschaftlicher Erhebungsinstrumente.* Version 7.00. Mannheim: Zentrum für Umfragen, Methoden und Analysen.

Felfe, J. & Schyns, B. (2004). Der Einfluss von impliziten Führungstheorien und Persönlichkeitsmerkmalen auf die Wahrnehmung und Akzeptanz transformationaler Führung. In W. Bungard, B. Koop & C. Liebig (Hrsg.), *Psychologie und Wirtschaft leben: aktuelle Themen der Wirtschaftspsychologie in Forschung und Praxis* (S. 245-252). München und Mering: Rainer Hampp.

Fengler, J. (1998). *Feedback geben. Strategien und Übungen*. Weinheim: Beltz.

Fennekels, G. P. (2000). *Qualitative Führungsstilanalyse* (2. Aufl.). Göttingen: Hogrefe.

Fennekels, G. P. (2003). *Multidirektionales Feedback-360° (MDF-360°)*. Göttingen: Hogrefe.

Fennekels, G. P. & D'Souza, S. (1999). *Management-Fallstudien.* Göttingen: Hogrefe.

Ferguson, I. R. (1973). *Management by Objectives in Deutschland. Fallstudien*. Frankfurt: Herder & Herder.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117-140.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.

Fiedler, F. E. (1967). *A theory of leadership effectiveness*. New York: McGraw-Hill.

Fiedler, F. E. (1971). Validatoin and extension of the contingency model of leadership effectiveness: a review of empirical findings. *Psychological Bulletin, 76,* 128-148.

Fiedler, F. E. (1973). Personality and situational determinants of leader behavior. In E. A. Fleishman & J. G. Hunt (Eds.), *Current developments in the study of leadership* (pp. 41-61). Carbondale: Southern Illinois University Press.

Fiol, C. M., Harris, D. & House, R. J. (1999). Charismatic leadership: Strategies for effecting social change. *Leadership Quarterly, 10*, 449-482.

Fishbein, M. (1967). Attitude and the prediction of behavior. In M. Fishbein (Ed.), *Readings in attitude theory and measurement* (pp. 477-492). New York, NY: Wiley.

Fittkau-Garthe, H. & Fittkau, B. (1971). *Fragebogen zur Vorgesetzten-Verhaltens-Beschreibung (FVVB).* Göttingen: Hogrefe.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51,* 327-358.

Fleenor, J. W. (1996). Constructs and developmental assessment centers: Further troubling empirical findings. *Journal of Business and Psychology, 10,* 319-333.

Fleenor, J. W., McCauley, C. D. & Brutus, S. (1996). Self-other rating agreement and leader effectiveness. *Leadership Quarterly, 7,* 487-506.

Fleishman, E. A. (1953). The description of supervisory behavior. *Journal of Applied Psychology, 37*, 1-6.

Fleishman, E. A. (1973). Twenty years of consideration and structure. In E. A. Fleishman & J. G. Hunt (Eds.), *Current developments in the study of leadership* (pp. 1-40). Carbondale: Southern Illinois University Press.

Fletcher, C. & Baldry, C. (2001). Multi-source feedback systems: A research perspective. In I. Robertson & C. Cooper (Eds.), *Personnel psychology and human resource management* (pp. 117-161). Chichester: John Wiley & Sons.

Flick, U. (1991). Triangulation. In U. Flick, E. von Kardoff, H. Keupp, L. von Rosenstiel & S. Wolff (Hrsg.), *Handbuch qualitative Sozialforschung* (S. 432-434). München: PVU.

Fox, S., Caspy, T. & Reisler, A. (1994). Variables affecting leniency, halo and validity of self-appraisal. *Journal of Occupational and Organizational Psychology, 67,* 45-56.

Freedman, J. (1978). *Happy people: What happiness is, who has it, and why.* New York: Harcourt Brace Jovanovich.

French, J. R. & Raven, B. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies in social power* (pp. 150-167). Oxford, England: University of Michigan Press.

Frese, M., Beimel, S. & Schoenborn, S. (2003). Action training for charismatic leadership: Two evaluations of studies of a commercial training module on inspirational communication of a vision. *Personnel Psychology, 56,* 671-697.

Fried, Y. & Ferris, G. R. (1987). The validity of the job characteristics model: A review and meta-analysis. *Personnel Psychology, 40,* 287-322.

Fuller, J. B, Morrison,, R., Jones, L., Bridger, D. & Brown, V. (1999). The effects of psychological empowerment on transformational leadership and job satisfaction. *The Journal of Social Psychology, 139*, 389-391.

Fuller, J. B., Patterson, C. P., Hester, K. & Stringer, D. Y. (1996). A quantitative review of research on charismatic leadership. *Psychological Reports, 78*, 271-287.

Furnham, A. & Stringfield, P. (1994). Congruence of self and subordinate ratings of managerial practices as a correlate of supervisor evaluation. *Journal of Occupational and Organizational Psychology, 67*, 57-67.

Furnham, A. & Stringfield, P. (1998). Congruence in job-performance ratings: A study of 360° feedback examining self, manager, peers, and consultant ratings. *Human Relations, 51,* 517–538.

Gabele, E., Liebel, H. & Oechsler, E. (1992). *Führungsgrundsätze und Mitarbeiterführung: Führungsprobleme erkennen und lösen.* Wiesbaden: Gabler.

Gardner, W. L. & Avolio, B. J. (1998). The charismatic relationship: A dramaturgical perspective. *Academy of Management Review, 32,* 32-58.

Gebert, D. (1992). Führungsstilforschung: Ein Vorschlag zur Neuorientierung. *Zeitschrift für Personalwesen, 3,* 245-259.

Gebert, D. (2002). *Führung und Innovation*. Stuttgart: Kohlhammer.

Gebert, D., Steinkamp, T & Wendler, E. (1987). *Führungsstil und Absatzerfolg im Kreditbereich*. Wiesbaden: Gabler.

Geyer, A. & Steyrer, J. (1998). Messung und Erfolgswirksamkeit transformationaler Führung. *Zeitschrift für Personalforschung, 12,* 377-401.

Giacalone, R. A. & Rosenfeld, P. (Eds.) (1989). *Impression management in the organization*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Giacalone, R. A. & Rosenfeld, P. (Eds.) (1991). *Applied impression management: How image-making affects managerial decisions.* Newbury Park, CA: Sage Publications.

Gilbert, D. T. (1995). Attribution and interpersonal perception. In A. Tesser (Ed.), *Advanced social psychology* (pp. 99-147). Boston: McGraw-Hill.

Goodwin, V. L., Wofford, J. C. & Whittington, J. L. (2001). A theoretical and empirical extension to the transformational leadership construct. *Journal of Organizational Behavior, 22,* 759-774.

Gould, S. & Penley, L. E. (1984). Career strategies and salary progression: A study of their relationships in a municipal bureaucracy. *Organizational Behavior and Human Performance, 34,* 244-265.

Graen, G. B. & Scandura, T. A. (1987b). Theorie der Führungsdyaden. In A. Kieser, G. Reber & R. Wunderer (Hrsg.), *Handwörterbuch der Führung* (S. 377-388). Stuttgart: Schäffer-Poeschel.

Graen, G. B. & Scandura, T. A. (1987a). Toward a psychology of dyadic organizing. In L. L. Cummings & B. M. Staw (Eds.), *Research in organiziational behavior* (Vol. 9, pp. 175-208). Greenwich: JAI Press Inc.

Green, S. G. & Mitchell, T. R. (1979). Attributional processes of leaders in leader-member interactions. *Organizational Behavior and Human Performance, 23,* 429-458.

Greguras, G. J., Ford, J. M. & Brutus, S. (2003). Manager attention to multisource feedback. *Journal of Management Development, 22,* 345-361.

Greguras, G. J. & Robie, C. (1998). A new look at within-source interrater reliability of 360-degree feedback ratings. *Journal of Applied Psychology, 83,* 960-968.

Greguras, G. J., Robie, C., Schleicher, D. J. & Goff, M. (2003). A field study of the effects of rating purpose on the quality of multisource ratings. *Personnel Psychology, 56,* 1-21.

Greller, M. M. & Herold, C. K. (1975). Sources of feedback: A preliminary investigation. *Organizational Behavior and Human Performance, 13,* 244-256.

Grieb, C. (1997). *Branchenkultur versus Unternehmenskultur*. Berlin: Institut für Management.

Grün, A. (2003). *Menschen führen. Leben wecken* (5. Aufl.). Münsterschwarzach: Vier-Türme-Verlag.

Guest (1997). Zielsetzungsmethoden (Goal setting). In S. Greif, H. Holling & N. Nicholson (Hrsg.), *Arbeits- und Organisationspsychologie* (S. 467-472).Weinheim: Psychologie Verlags Union.

Guilford, J. P. (1950). Creativity. *American Psychologist, 5*, 444-454.

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.

Guion, R. (1965). *Personnel testing.* New York: McGraw-Hill.

Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, 19, 149-161.

Guzzo, R. A., Jette, R. D. & Katzell, R. A. (1985). The effects of psychologically based intervention programs on worker productivity: A meta-analysis. *Personnel Psychology, 38*, 275-292.

Hackman, J. R. & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance, 16,* 250-279.

Hackman, J. R. & Oldham, G. R. (1980). *Work redesign.* Reading, Mass.: Addison-Wesley.

Harris, M. M. & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*, 43-62.

Harris, M. M., Smith, D. & Champagne, D. (1995). A field study of performance appraisal purpose: research- versus administrative-based ratings. *Personnel Psychology, 48,* 151-160.

Harter, J. K., Schmidt, F. L. & Keyes, C. L. (2003). Well-being in the workplace and its relationship to business outcomes: A review of the Gallup studies. In C.L. Keyes & J. Haidt (Eds.), *Positive psychology and the life well-lived* (pp. 205-224). Washington: APA.

Hater, J. J. & Bass, M. B. (1988). Superiors' evaluations and subordinates' perceptions of transformational and transactional leadership. *Journal of Applied Psychology, 73,* 695-702.

Hatfield, J. & Huseman, R. (1982). Perceptual congruence about communication as related to satisfaction: Some moderating effects of individual characteristics. *Academy of Management Journal, 25,* 349-358.

Hawkins, S. A. & Hastie, R. (1990). Hindsight: Biased judgement of past events ater the outcomes are known. *Psychological Bullettin, 107,* 311-327.

Hazucha, J. F., Hezlett, S. A. & Schneider, R. J. (1993). The impact of 360-degree feeback on management skills developement. *Human Resource Management, 32,* 325-351.

Hedge, J. W., Borman, W. C. & Birkeland, S. A. (2001). History and development of multisource feedback as a methodology. In D. W. Bracken, C. W. Timmreck & A. H. Church (Eds.), *The handbook of multisource feedback* (pp. 15-32). San Francisco: Jossey-Bass.

Heidemeier, H. & Moser, K. (2002). *Self-appraisal of job-performance*. Poster beim 43. Kongress der Deutschen Gesellschaft für Psychologie.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Heilman, M. E., Block, C. J. & Martell, R. F. (1995). Sex stereotypes: Do they influence perceptions of mangers? *Journal of Social Behavior and Personality. Special Issue: Gender in the workplace, 10* (6)*,* 237-252.

Heilman, M. E., Block, C. J., Martell, R. F. & Simon, M. C. (1989). Has anything changed? Current characterizations of men, women, and managers. *Journal of Applied Psychology, 74,* 935-942.

Heller, F., Pusic, E., Strauss, G. & Wilpert, B. (1998). *Organizational participation – myth and reality*. Oxford: Oxford University Press.

Heneman, R. L. (1986). The relationship between supervisory ratings and results-oriented measures of performance: A meta-analysis. *Personnel Psychology, 39,* 811-826.

Hersey, P. & Blanchard, K. H. (1969). Life cycle theory of leadership. *Training and Development Journal, 23,* 26-34.

Hersey, P. & Blanchard, K. H. (1982). Management of organizational behavior (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hersey, P., Blanchard, K. H. & Johnson, D. E. (2001). *Management of Organizational Behavior* (8th ed.). Upper Saddle River, NJ: Prentice Hall.

Herzberg, F., Mausner, B. & Snyderman, B. (1967). *The motivation to work* (2nd ed.). New York: Wiley.

Hickman, G.R. (1998) (Ed.). *Leading organizations: Perspectives for a new era.* Thousand Oaks: Sage.

Hoch, D. J., Roeding, C., Purkert, G. & Lindner, S. K. (1999). *Secrets of software success: Management insights from 100 software firms around the world.* Boston: HBR Press

Hofmann, K., Köhler, F. & Steinhoff, V. (Hrsg.) (1995). *Vorgesetztenbeurteilung in der Praxis. Konzepte, Analysen, Erfahrungen*. Weinheim: PVU.

Hofstede, G. (1983). National cultures in four dimensions: A research-based theory of cultural difference among nations. *International Studies of Management and Organization, 13* (1-2), 46-74.

Hollander, E. P. (1995). Führungstheorien – Idiosynkrasiekreditmodell. In A. Kieser, G. Reber & R. Wunderer (Hrsg.), *Handwörterbuch der Führung* (2. Aufl.). (S. 926-940). Stuttgart: Schäffer-Poeschel.

Homans, G.C. (1972a). *Elementarformen sozialen Verhaltens* (2. Aufl.). Köln, Opladen: Westdeutscher Verlag

Homans, G.C. (1972b). *Grundfragen soziologischer Theorie*. Opladen: Westdeutscher Verlag.

Hoog, M. A. (2000). Social identity and social comparison. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 401-421). New York: Kluwer Academic/Plenum Publishers.

Hooijberg, R. & Choi, J. (2000). From selling peanuts and beer in yankee stadium to creating a theory of transformational leadership: An interview with Bernie Bass. *Leadership Quarterly, 11,* 291-306.

Hossiep, R. & Paschen, M. (1998a). *Das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP). Testmanual*. Göttingen: Hogrefe.

Hossiep, R. & Paschen, M. (1998b). *Selbstbild, Fremdbild und Persönlichkeit. Hinweise für Teilnehmer des Bochumer Inventars zur berufsbezogenen Persönlichkeitsbeschreibung (BIP)*. Göttingen: Hogrefe.

House, R. J. (1971). A path-goal theory of leader effectiveness. *Administrative Science Quarterly, 16,* 321-338.

House, R. J. (1977). A 1976 theory of charismatic leadership. In J. G. Hunt & L. L. Larson (Eds.), *Leadership. The cutting edge* (pp. 189-207). Carbondale: Southern Illinois University Press.

House, R. J. (1996). Path-goal theory of leadership: Lessons, legacy, and a reformulated theory. *Leadership Quarterly, 7*, 323-352.

House, R. J. & Aditya, R. N. (1997). The social scientific study of leadership: Quo vadis? *Journal of Management, 23,* 409-473.

House, R. J. & Chhokar, J. (Eds.). (1999). *Cultures of the world: A GLOBE anthology of in-depth descriptions of the cultures of 14 countries* (Vol. 1).

House, R. J. & Mitchell, T. R. (1974). Path-goal theory of leadership. *Journal of Contemporary Business, 3* (4), 81-97.

House, R. J. & Shamir, B. (1993). Toward the integration of transformational, charismatic, and visionary theories. In M. M. Chemers & R. Ayman (Eds.), *Leadership theory and research. Perspectives and directions* (pp. 81-107). San Diego: Academic Press.

House, R. J. & Shamir, B. (1995). Führungstheorien – Charismatische Führung. . In A. Kieser, G. Reber & R. Wunderer (Hrsg.), *Handwörterbuch der Führung* (2. Aufl.) (S. 878-897). Stuttgart: Schäffer-Poeschel.

House, R. J., Spangler, W. D. & Woycke, J. (1991). Personality and charisma in the U.S. presidency: A psychological theory of leader effectiveness. *Administrative Science Quarterly*, *36*, 364-396.

House, R. J., Wright, N. S. & Aditya, R. N. (1997). Cross-cultural research on organizational leadership: A critical analysis and a proposed theory. In P.C. Earley & M. Erez (Eds.), *New perspectives on international industrial/organizational psychology* (pp. 535-625). San Francisco: New Lexington Press.

Howell, J. M. (1988). Two faces of charisma: Socialized and personalized leadership in organizations. In J. A. Conger & R. N. Kanungo (Eds.), *Charismatic leadership. The elusive factor in organizational effectiveness* (S. 213-236). San Francisco: Jossey-Bass.

Howell, J. M. (1997). *OrganizationcContexts, charismatic and exchange leadership. Transformational leadership, working papers.* Maryland: Academy of Leadership Press.

Howell, J. M., Avolio, B. J. (1993). Transformational leadership, transactional leadership, locus of control, and support for innovation: Key predictors of consolidated-business-unit performance. *Journal of Applied Psychology, 78*, 891-902.

Howell, J. M., Bowen, D. E., Dorfman, P. W., Kerr, S., & Podsakoff, P. M. (1990). Substitutes for leadership: Effective alternatives to ineffective leadership. *Organizational Dynamics, 19*, 21-38.

Howell, J. M. & Frost, P. J. (1989). A laboratory study of charismatic leadership. *Organizational Behavior and Human Decision Processes, 43*, 243-269.

Howell, J. M. & Higgins, C. (1990). Champions of technological innovation. *Administrative Science Quarterly, 35*, 317-341.

Hoyt, W. T. & Kerns, M.- D. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods, 4,* 403-424.

Hunt, J. G. (1999). Transformational/charismatic leadership's transformation of the field: an historical essay. *Leadership Quarterly, 10*, 129-144.

Ilgen, D., Fisher, C. & Taylor, M. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64, 3*49-371.

Jacobs, L., Berscheid, E. & Walster, E. (1971). Self-esteem and attractioin. *Journal of Personality and Social Psychology, 17,* 84-91.

Jago, A. (1995). Führungstheorien – Vroom/Yetton-Modell. In A. Kieser, G. Reber & R. Wunderer (Hrsg.), *Handwörterbuch der Führung* (2. Aufl.) (S. 1958-1975). Stuttgart: Schäffer-Poeschel.

Janis, I. L. (1982). *Groupthink* (2nd ed.). Boston: Houghton Mifflin.

Jannsen, O. (2002). Transformationeel leiderschap en innovatief werkgedrag van medewerkers: Een kwestie van benaderbaarheid van de leider. *Gedrag en Organisatie, 15* (5), 275-293.

Jawahar, I. M. (2001). Attitudes, self-monitoring, and appraisal behaviors. *Journal of Applied Psychology, 86,* 875-883.

Jawahar, I. M. & Williams, Ch. R. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology, 50*, 905-926.

Jay, M. (2003). Understanding how to leverage executive coaching. *Organization Development Journal, 21* (2), 6-19.

Jeserich, W. (1996). Ohne Feedback keine Optimierung des Führungsprozesses. In Arbeitskreis Assessment Center e.V. (Hrsg.), *Assessment Center als Instrument der Personalentwicklung* (S. 104-118). Hamburg: Windmühle.

Jochum, E. (1991). Gleichgestelltenbeurteilung – ein Instrument der Personalführung und Teaentwicklung. In H. Schuler (Hrsg.), *Beurteilung und Förderung beruflicher Leistung* (S. 107-134). Stuttgart: Verlag für Angewandte Psychologie.

Johns, G. (1981). Difference score measures of organizational behavior variables: A critique. *Organizational Behavior and Human Performance, 27,* 443-463.

Johnson, J. W. & Ferstl, K. L. (1999). The effects of interrater and self-other agreement on performance improvement following upward feedback. *Personnel Psychology, 52,* 271-303.

Jonas, K, Keilhofer, G. & Schaller, J. (Hrsg.). (2004). *Human Resource Management im Automobilbau. Konzepte und Erfahrungen.* Bern: Huber.

Jones, E. E. & Nisbett. R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins & B. Wiener (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79-94). Morristown, NJ: General Learning Press.

Jöns, I. & Mataja, K. (1998). Praxiskonzepte und Erfahrungsberichte. Beteiligung an Vorgesetztenbeurteilungen – Ergebnisse zu Einflußfaktoren auf den Rücklauf bei schriftlichen Befragungen. *Zeitschrift für Arbeits- und Organisationspsychologie, 42,* 33-42.

Judge, T. A., Bono, E. J., Ilies, R. & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quatitative review. *Journal of Applied Psychology 87*, 765-780.

Judge, T. A & Piccolo, R. F. (2004). Transformational and transactional leadership: A meta-analytic test of their relative validity. *Journal of Applied Psychology 89*, 755-769.

Jung, D. I. (2001). Transformational and transactional leadership and their effects on creativity in groups. *Creativity Research Journal, 13*, 185-195.

Jung, D. I. & Avolio, B. J. (2000). Opening the black box: An experimental investigation of the mediating effects of trust and value congruence on transformational and transactional leadership. *Journal of Organizational Behaviour, 21*, 949-964.

Kahneman, D., Slovic, P. & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kaiser, H. F. & Dickman, K. (1959). Analytic determination of common factors. *American Psychologist*, *14,* 425-431.

Kanfer, F., Reinecker, H. & Schmelzer, D. (2000). *Selbstmanagementtherapie* (3. Aufl.). Berlin: Springer.

Kegan, R. & Lahey, L. L. (2001). *How the way we talk can change the way we work: Seven languages for transformation.* San Francisco: Jossey-Bass/Pfeiffer.

Kehr, H.M. (2000). *Die Legitimation von Führung. Ein Kleingruppenexperiment zum Einfluss der Quelle der Autorität auf die Akzeptanz des Führers, den Gruppenprozess und die Effektivität.* Berlin: Duncker & Humblot.

Keller, R.T. (1992). Transformational leadership and the performance of research and development project groups. *Journal of Management, 18*, 489-501.

Kelloway, E. K., Barling, J. & Helleur, J. (2000). Enhancing transformational leadership: The roles of training and feedback. *Leadership and Organization Development, 21* (3), 145-149.

Kenny, D. A. & DePaulo, B. M. (1993). Do people know how others view them? An empirical and theoretical account. *Psychological Bulletin, 114*, 145-161.

Kenrick, D. T., Neuberg, S. L. & Cialdini, R. B. (1998). *Social psychology: Unraveling the mystery.* Boston: Allyn and Bacon.

Kernis, M. H. & Grannemann, B. D. (1990). Excuses in the making: A test and extension of Darley and Goethals' model. *Journal of Experimental Social Psychology, 26,* 333-349.

Kerr, S. (1977). Substitutes for leadership: Some implications for organizational design. *Organization and Administratives Sciences*, *8*, 135-146.

Kerr, S. & Jermier, J. M. (1978). Substitutes for leadership: Their meaning and measurement. *Organizational Behavior and Human Performance, 22,* 375-403.

Kiessling, W. F. & Spannagl, P. (1996). *Corporate Identity: Unternehmensleitbild - Organisationskultur*. Alling: Sandmann.

Kim, K., Park, H.-J. & Suzuki, N. (1990). Reward allocations in the United States, Japan and Korea: A comparison of individualistic and collectivistic cultures. *Academy of Management Journal, 33, 1*88-189.

Kirkpatrick, S. & Locke, E. (1996). Leadership: Do traits matter? In R. Steers, L. Porter & G. Bigley (Eds.), *Motivation and leadership at work* (6th ed.). (pp. 186-199). New York: Mc GrawHill.

Klimoski, R. J. & London, M. (1974). Role of the rater in performance appraisal. *Journal of Applied Psychology, 59,* 445-451.

Kluge, F. (1999). *Etymologisches Wörterbuch der deutschen Sprache* (23. Aufl.). Berlin: de Gruyter.

Kluger, A. N. (2001). Feedback-expectation discrepancy, arousal and locus of cognition. In M. Erez, U. Kleinbeck & H. Thierry (Eds.). *Work motivation in the context of a globalizing economy* (pp. 111-120). Mahwah: Lawrence Erlbaum Associates.

Kluger, A. N. & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254-284.

Koch, S. (2001). *Eigenverantwortliches Handeln von Führungskräften. Schriftenreihe Organisation & Personal (*Bd. 10). München, Mering: Rainer Hampp.

Koene, B., Pennings, J. & Schreuder, H. (1992). Leadership, organizational culture, and organizational outcomes. In K. E. Clark, M. B. Clark & D. P. Campbell (Eds.), *Impact of leadership* (pp. 215-223). Greensboro: The Center for Creative Leadership.

Koh, W. L., Steers, R. M. & Terborg, J. R. (1995). The effects of transformation leadership on teacher attitudes and students performance in Singapore. *Journal of Organizational Behavior 16*, 319-333.

Korman, A. K. (2001). Self-enhancement and self-protection: Toward a theory of work motivation. In M. Erez, U. Kleinbeck & H. Thierry (Eds.). *Work motivation in the context of a globalizing economy* (pp. 121-130). Mahwah: Lawrence Erlbaum Associates.

Kosch, S. (2004, 16./17. Oktober). Ein Michael Moore für Bochum? *Die Tageszeitung*, S. 1.

Kouzes, J. & Posner, B. Z. (1987). *The leadership challenge: How to get extraordinary things done in organizations.* San Francisco: Jossey Bass.

Kristof-Brown, A., Barrick, M. R. & Franke, M. (2002). Applicant impression management: Dispositional influences and consequences for recruiter perceptions of fit and similarity. *Journal of Management, 28,* 27-46.

Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77,* 1121-1134.

Kuhnert, K. W. (1994). Transforming leadership: Developing people through delegation. In B. M. Bass & B. J. Avolio (Eds.), *Improving organizational effectiveness through transformational leadership* (pp. 10-25). Thousand Oaks: Sage.

Kuiper, N. A. & Derry, P. A. (1982). Depressed and nondepressed content self-reference in mild depressives. *Journal of Personality, 50,* 67-80.

Larwood, L. (1978). Swine flu: A field study of self-serving biases. *Journal of Applied Social Psychology, 18,* 283-289.

Larwood, L. & Whittaker, W. (1977). Managerial myopia: Self-serving bias in organizational planning. *Journal of Applied Psychology, 62,* 194-198.

Latham, G. P., Wexley, K. & Pursell, E. D. (1975). Training managers to minimize rating errors in the observation of behavior. *Journal of Applied Psychology, 60,* 550-555.

Laux, L. & Schütz, A. (1996). *„Wir, die wir gut sind" Die Selbstdarstellung von Politikern zwischen Glorifizierung und Glaubwürdigkeit*. München: dtv.

Lebow, R. & Spitzer, R. (2002). *Accountability: freedom and responsiblity without control.* San Francisco: Berrett-Koehler.

Lehner, J. (1995). Führungserfolg – Messung. In A. Kieser, G. Reber & R. Wunderer (Hrsg.), *Handwörterbuch der Führung* (2. Aufl.) (S. 550-562). Stuttgart: Schäffer-Poeschel.

Lewin, K., Lippitt, R. & White, R. (1939). Patterns of aggressive behavior in experimentally created "social climates". *Journal of Social Psychology, 10,* 271-299.

Liebel, H. J. & Oechsler, W. A. (1994). *Handbuch Human-Resource-Management*. Wiesbaden: Gabler.

Lieberson, S. & O'Connor, J. (1972). Leadership and organizational performance: A study of large corporations. *American Sociological Review, 37,* 117-130.

Lievens, F. & Van Geit, P. & Coetsier, P. (1997). Identification of transformational leadership qualities: An examination of potential biases. *European Journal of Work and Organizational Psychology, 6,* 415-430.

Likert, R. (1961). *New patterns of management*. New York: McGraw-Hill.

Limpächer, S. (2003). Vortrag im Rahmen einer Exkursion der Psychologie zu Daimler Chrysler, Sindelfingen.

Locke, E. A., Cartledge, N. & Koeppel, J. (1968). Motivational effects of knowledge of results: A goal-setting phenomenon? *Psychological Bulletin, 70,* 474-485.

Locke, E. A. & Latham, G. P. (1990). *A theory of goal setting & task performance.* Englewood Cliffs: Prentice Hall.

Lombardo, M. M. & McCall, M. W. (1978). Leadership. In M. W. McCall & M. M. Lombardo (Eds.), *Leadership: Where else can we go?* (pp. 3-12). Durham, NC: Duke University.

London, M. (1994). Interpersonal insight in organizations: Cognitive models for human resource development. *Human Resource Management Review, 4*, 311-332.

London, M. (1995a). Giving feedback – source-centered antecedents and consequences of constructive and destructive feedback. *Human Resource Management Review, 5*, 159-188.

London, M. (1995b). *Self and interpersonal insight. How people gain understanding of themselves and others in organizations.* New York: Oxford University Press.

London, M. & Smither, J. W. (1995). Can multi-source feedback change perceptions of goal accomplishment, self-evaluations and performance related outcomes? Theory-based applications and directions for research. *Personnel Psychology, 48*, 803-839.

London, M., Smither, J. W. & Adsit, D. J. (1997). Accountability: The Achilles' heel of multi-source feedback. *Group and Organization Management, 22,* 162-184.

London, M. & Wohlers, A. J. (1991). Agreement between subordinate and self-ratings in upward feedback. *Personnel Psychology, 44,* 375-390.

London, M. & Wohlers, A. J. & Gallagher, P. (1990). 360-degree feedback surveys: A source of feedback to guide management development. *Journal of Management Development, 9*, 17-31.

Longenecker, C. O. & Gioia, D. A. (1992). The executive appraisal paradox. *The Executive, 6* (2), 18-28

Lord, R. G., DeVader, C. L. & Alliger, G. M. (1986). A meta-analysis of the relation between personality traits and leadership perceptions: An application of validity generalization procedures. *Journal of Applied Psychology, 71*, 402-410.

Lord, R. G. & Hall, R. J. (1992). Contemporary views of leadership and individual diffrences. *Leadership Quarterly, 3,* 137-157.

Lord, R. G. & Maher, K.J. (1991). *Leadership and information processing: Linking perceptions to performance.* Boston: Unwin Hyman.

Lowe, K. B., Kroeck, K. G. & Sivasubramaniam, N. (1996). Effectiveness correlates of transformational and transactional leadership: A meta-analytic review of the MLQ literature. *Leadership Quarterly, 7*, 385-425.

Ludorf, S. N. (1998). *Dimensionen des Führungsverhaltens und der Arbeitszufriedenheit und deren Einfluß auf die Leistung, Zielerreichung und Zielvereinbarungsprozesse in drei kulturell unterschiedlichen Settings - Deutschland, Amerika und Südafrika.* Aachen: Shaker.

Luft, J. (1961). The johari window. *Human Relations Training News, 5,* 6-7.

Luft, J. (1963). *Group processes. An introduction to group dynamics*. Palo Alto, CA: National Press.

Luthans, F. & Kreitner, R. (1975). *Organizational behavior modification*. Grenview, IL: Scott, Foresman.

Maaß, D. (in Vorb.). *Der Erfinder und sein Umfeld: Organisations-, gruppen- und persönlichkeitsbezogene Einflussfaktoren auf die Quantität und Qualität von patentfähigen Erfindungen.* Unveröffentlichte Diplomarbeit an der TU Chemnitz.

Mabe, P. A. & West, S. G. (1982). Validity of self evaluations of ability: A review and meta-analysis. *Journal of Applied Psychology, 67,* 280-296.

Malik, F. (2000). *Führen Leisten Leben*. Stuttgart: Deutsche Verlags-Anstalt

Malloy, T. E., Yarlas, A., Montvilo, R. K. & Sugarman, D. B. (1996). Agreement and accuracy in children's interpersonal perceptions: A social relations analysis. *Journal of Personality and Social Psychology, 71,* 692-702.

Marcus, B. & Schuler, H. (2001). Leistungsbeurteilung. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 397-431). Göttingen: Hogrefe.

Markus, H. (1983). Self-knowledge: An expanded view. *Journal of Personality, 51*, 543-565.

Markus, H. & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology, 38,* 299-337.

Martin, R. & Epitropaki, O. (2001). Role of organizational identification on implicit leadership theories (ILTs), transformational leadership and work attitudes. *Group Processes and Intergroup Relations*, *4*, 247-262.

Martineau, J. W. (1998). Using 360-degree surveys to assess change. In M. London & W. Tornow (Eds.), *Maximizing the value of 360-degree feedback: a process for successful individual and organizational development* (pp. 217-248). San Francisco: Jossey-Bass.

Masi, R. J. & Cooke, R. A. (2000). Effects of transformational leadership on subordinate motivation, empowering norms, and organizational productivity. *International Journal of Organizational Analysis, 8*, 16-47.

Mathieu, J. E. & Zajac, D. M. (1990). A review and meta-analysis of the antecedents, correlates and consequences of organizational commitment. *Psychological Bulletin*, *108,* 171-194.

McCall, M. W. (1997). High flyers. *Developing the next generation of leaders.* Boston, MA: Harvard Business Scholl Press.

McCall, M. W. & Lombardo, M. M. (1983). *Off the track: Why and how successful executives get derailed.* Technical Report Number 21. Greensboro: Center for Creative Leadership.

McCall, M. W., Lombardo, M. M. & Morrison, A. M. (1988). *The lessons of experience: How successful executives develop on the job.* Lexington, MA: Lexington.

McClelland, D. C. (1985). *Human motivation*. Glenview, IL: Scott, Foresman.

McClelland, G. H: & Judd, Ch. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114,* 376-390.

McCormick, E. J. & Tiffin, J. (1974). *Industrial psychology* (6th ed.). Englewood Cliffs, New Jersey: Prentice Hall, Inc.

McCormick, I. A., Walkey, F. H. & Green, D. E. (1986). Comparative perceptions of driver ability: A confirmation and expansion. *Accident Analysis & Prevention, 18,* 205-208.

McEvoy, G. M. & Beatty, R. W. (1989). Assessment centers and subordinate appraisals of managers: A seven-year examination of predictive validity. *Personnel Psychology, 42*, 37-52.

Meindl, J. R. (1993). Reinventing leadership: A radical, social psychological approach. In J. K. Murnighan (Ed.), *Social psychology in organizations: advances in theory and research*. Englewood Cliffs: Prentice Hall.

Meindl, J. R. (1998). The romance of leadership as a follower-centric theory: a social constructionist approach. In Dansereau, F. & Yammarino, F. J. (Eds.), *Leadership: The multiple-level approaches.* Stamford: JAI Press.

Meindl, J. R. & Ehrlich, S. B. (1987). The romance of leadership and the evaluation of organizational performance. *Academy of Management Journal, 30*, 90-109.

Meindl, J. R., Ehrlich, S. B. & Dukerich, J. M. (1985). The romance of leadership. *Administrative Science Quarterly, 30,* 78-102.

Messick, D. M., Bloom, S., Boldizar, J. P. & Samuelson, C. D. (1985). Why we are fairer than others. *Journal of Experimental Social Psychology, 21,* 480-500.

Metzger, R. (2004, 16./17. Oktober). Die Reichen am Abgrund. *Die Tageszeitung*, S. 14.

Meyer, H. (1980). Self-appraisal of job performance. *Personnel Psychology, 33*, 291-295.

Meyer, H. (1991). A solution to the performance appraisal feedback enigma. *Academy of Management Executive, 5*, 68-76.

Miller, D.T. & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin, 82*, 213-225.

Miller, K. & Monge, R. (1986). Participation, satisfaction and productivity: a meta-analytic review. *Academy of Management Journal, 29,* 727-753.

Mintzberg, H. (1973). *The nature of managerial work*. New York: Harper & Row.

Mintzberg, H. (1982). If you're not serving Bill and Barbara, then you're not serving leadership. In J. G. Hunt, U. Sekaran & Ch. A. Schriesheim (Eds.), *Leadership. Beyond establishment views* (pp. 239-259). Carbondale and Edwardsville: Southern Illinois University Press.

Morgan, M. (1998). *Traumfänger*. München: Goldmann.

Moser, K. (1999). Selbstbeurteilung beruflicher Leistung: Überblick und offene Fragen. *Psychologische Rundschau, 50,* 14-25.

Moser, K., Donat, M., Schuler, H., Funke, U. & Roloff, K. (1994). Validität der Selbstbeurteilung beruflicher Leistung: Eine Untersuchung im Bereich industrieller Forschung und Entwicklung. *Zeitschrift für experimentelle und angewandte Psychologie, 41,* 473-499.

Moser, K., Zempel, J. & Schultz-Amling, D. (2000). Strategische Elemente in Leistungsbeurteilungen. *Zeitschrift Führung und Organisation, 69,* 218-225.

Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R. & Hezlett, S. A. (1998). Trait, rater, and level effects in 360-degree performance ratings. *Personnel Psychology, 51,* 557-576.

Mount, M. K. & Scullen, S. E. (2001). Multisource feedback ratings: What do they really measure? In M. London (Ed.), *How people evaluate others in organizations* (S. 155-176). Mahwah: Lawrence Erlbaum Associates.

Mummendey, H. D. (1995). *Psychologie der Selbstdarstellung* (2. Aufl.). Göttingen: Hogrefe.

Mummendey, H. D. (2002). Selbstdarstellungstheorie. In D. Frey & M. Irle (Hrsg.) *Theorien der Sozialpsychologie* (2. Aufl). (Bd. 3, S. 212-233). Bern: Hans Huber.

Mummendey, H. D. (2003). *Die Fragebogen-Methode. Grundlagen und Anwendung in Persönlichkeits-, Einstellungs- und Selbstkonzeptforschung* (4. Aufl.). Göttingen: Hogrefe.

Murphy, K. R. & Cleveland, J. N. (1995). *Understanding performance appraisal. Social, organizational, and goal-based perspectives.* Thousand Oaks, CA: Sage.

Murphy, K. R., Cleveland, J. N. & Mohler, C. J. (2001). Reliability, validity, and meaningfulness of multisource ratings. In D. W. Bracken, C. W. Timmreck & A. H. Church (Eds.), *The handbook of multisource feedback* (pp. 130-148). San Francisco: Jossey-Bass.

Nachreiner, F. (1978). *Die Messung des Führungsverhaltens*. Bern: Verlag Hans Huber.

Nagle, B. F. (1953). Criterion development. *Personnel Psychology, 6,* 271-288.

Nanus, B. (1992). *Visionary leadership*. San Francisco: Jossey-Bass.

Nemeth, C. & Kwan, J. (1985). Originality of work assozciations as a function of majority vs. minority influence processes. *Social Psychology Quarterly, 48*, 277-28.

Nemeth, C. & Kwan, J. (1987). Minority influence, divergent thinking and the detection of correct solutions. *Journal of Applied Social Psychology, 9*, 788-799.

Nemeth, C. H. & Wachtler, J. (1983). Creative problem solving as a result of majority vs minority influence. *European Journal of Social Psychology*, *13*, 45-55.

Neuberger, O. (1972). Experimentelle Untersuchungen von Führungsstilen. *Gruppendynamik, 3,* 192-219.

Neuberger, O. (1976). *Führungsverhalten und Führungserfolg*. Berlin: Duncker & Humblot.

Neuberger, O. (2000). *Das 360°-Feedback. Alle fragen? Alles sehen? Alles sagen?* München: Rainer Hampp.

Neuberger, O. (2002). *Führen und führen lassen. Ansätze, Ergebnisse und Kritik der Führungsforschung* (6. Aufl.). Stuttgart: Lucius und Lucius.

Nickel, T. M. & Krems, J. F. (1998). Führungsverhalten und Mitarbeiterkreativität - Eine empirische Untersuchung zum betrieblichen Vorschlagswesen. *Zeitschrift für Arbeits- und Organisationspsychologie, 42*, 27-32.

Nilsen, D. & Campbell, D. P. (1993). Self-observer rating discrepancies: Once an overrater, always an overrater? *Human Resource Management, 32,* 265-281.

Nowack, K. M. (1997). Congruence between self-other ratings and assessment center performance. *Journal of Social Behavior and Personality, 12* (5), 145-166.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

O'Connor, B. P. (1998). SIMPLE: All-in-one programs for exploring interactions in moderated multiple regression. *Educational and Psychological Measurement, 58,* 833-837.

O'Connor-Wilson, P., McCauley, C. D. & Kelly-Radford, L. (1998). 360-degree feedback in the establishment of learning cultures. In M. London & W. Tornow (Eds.), *Maximizing the value of 360-degree feedback: a process for successful individual and organizational development* (pp. 120-146). San Francisco: Jossey-Bass.

Okazaki, S. (2002). Self-other agreement on affective distress scales in Asian Americans and White Americans. *Journal of Counseling Psychology, 49*, 428-437.

Oldham, G. R. (1976). The motivational strategies used by supervisors: Relationships to effectiveness indicators. *Organizational Behavior & Human Decision Making, 15* (1)*,* 66-86.

Ostroff, Ch., Atwater, L. E. & Feinberg, B. J. (2004). Understanding self-Other agreement: A look at rater and ratee characteristics, context and outcomes. *Personnel Psychology, 57*, 333-375.

Otten, S. & Mummendey, A. (2000). Valence-dependent probability of in-group favoritism beeen minimal groups: An integrative view on the positive-negative asymmetry in social discrimination. In D. Capozza & R. Brown (Eds.), *Social identity processes* (pp. 33-48). London: Sage.

Otten, S. & Wentura, D. (2001). Self-anchoring and in-group favoritism: An individual profiles analysis. *Journal of Experimental Social Psychology, 37,* 525-532.

Paglia, A. & Room, R. (1999). Expectancies about the effects of alcohol on the self and on others as determinants of alcohol policy attitudes. *Journal of Applied Social Psychology, 29,* 2632-2651.

Paradise-Tornow, C. A. (1998). The competitive advantage of customer involvement in 360-degree feedback. In M. London & W. Tornow (Eds.), *Maximizing the value of 360-degree feedback: a process for successful individual and organizational development* (pp. 101-119). San Francisco: Jossey-Bass.

Parker, J. W., Taylor, E. K., Barrett, R. S. & Martens, L. (1959). Rating scale content: III. Relationship between supervisory- and self-ratings. *Personnel Psychology, 12,* 49-63.

Pascale, R. T. & Athos, A. G. (1981). *The art of Japanese management*. Harmondsworth: Penguin.

Pascarella, S. V. & Luneburg, F. C. (1988). A field test of Hersey and Blanchard`s situational leadership theory in a school setting. *College Student Journal, 22* (1), 33-37.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598-609.

Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement. A mixed blessing? *Journal of Personality and Social Psychology, 74*, 1197-1208.

Paulhus, D. L. & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66,* 1025-1060.

Paulhus, D. L. & Reid, D. B. (1991). Enhancement an denial in socially desirable responding. *Journal of Personality and Social Psychology, 60,* 307-317.

Pawar, B. & Eastman, K. (1997). The nature and implications of contextual influences on transformational leadership: A conceptual examination. *Academy of Management Review, 22,* 80-109.

Perloff, L. S. & Fetzer, B. K. (1986). Self-other judgments and perceived vulnerability to victimization. *Journal of Personality and Social Psychology, 50,* 502-510.

Peter, L. J. & Hull, R. (2001). *Das Peter-Prinzip oder Die Hierarchie der Unfähigen*. Reinbek: rororo.

Petersen, L.-E. & Stahlberg, D. (1995). Der Integrative Selbstschemaansatz. Die Suche und Verarbeitung selbstkonzeptrelevanter Informationen in Abhängigkeit vom Elaborationsgrad der involvierten Selbstschemata. *Zeitschrift für Experimentelle Psychologie, 42,* 43-62.

Pierce, J. L., Gardner, D. G., Dunham, R. B. & Cummings, L. L. (1993). Moderation by organization-based self esteem of condition-employee response relationship. *Academy of Management Journal, 36*, 271-288.

Pillai, R. (1996). Crisis and the emergence of charismatic leadership in groups: An experimental investigation. *Journal of Applied Social Psychology, 26*, 543-562.

Pillai, R. & Meindl, J. R. (1998). Context and charisma: A `meso' level examination of the relationship of organic structure, collectivism, and crisis to charismatic leadership. *Journal of Management, 24*, 643-671.

Pillai, R., Schriesheim, C. A. & Williams, E. S. (1999). Fairness perceptions and trust as mediators for transformational and transactional leadership: A two-sample study. *Journal of Management, 25*, 897-933

Pillai, R., Williams, E. Y., Lowe, K. B. & Jung, D. I. (2003). Personality, transformational leadership, trust, and the 2000 U.S. presidential vote. *Leadership Quarterly, 14*, 161-193.

Podsakoff, P. M. & Schriesheim, C.A. (1985). Leader reward and punishment behavior: A methodological and substantive review. In B. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (p. 175-196). San Francisco: Jossey-Bass.

Podsakoff, P. M., Todor, W. D. & Skov, R. (1982). Effects of leader contingent and noncontingent reward and punishment behaviors on subordinate performance and satisfaction. *Academy of Management Journal, 25,* 810-821.

Pollack, D. M. & Pollack, L. J. (1996). Using 360-degree feedback in performance appraisal. *Public Personnel Management, 25,* 507-528.

Porter, L. & Lawler, E. (1968). *Managerial attitudes and performance*. Homewwd, Ill.: Irwin-Dorsey.

Pryor, J. B., Gibbons, F. X., Wicklund, R. A., Fazio, R. H. & Hood, R. (1977). Self-focused attention and self-report validity. *Journal of Personality, 45,* 513-527.

Rai, S. & Sinha, A. K. (2000). Transformational leadership, organizational commitment, and facilitating climate. *Psychologcial Studies, 45*, 33-42.

Ram, S. & Jung, H.-S. (1991). "Forced" adoption of innovations in organizations: Consequences and implications. *Journal of Product Innovation Management, 8,* 117-126.

Randall, R., Ferguson, E. & Patterson, F. (2000). Self-assessment accuracy and assessment centre decisions. *Journal of Occupational and Organizational Psychology, 73,* 443–459.

Rapee, R. & Lim, L. (1992). Discrepancy between self- and observer ratings of performance in social phobics. *Journal of Abnormal Psychology, 101,* 728-731.

Rathgeber, K. & Jonas, K. (2002). *Fragebogen zur Führungsbeschreibung – auf der Basis der Führungsgrundsätze von VWS.* Unveröffentlichtes Handbuch.

Rathgeber, K. & Jonas, K. (2003). Transformationale Führung: Mehr Leistung, weniger Stress? In P. Creutzfeldt, *Die gesunde Organisation. Grundlagen, Konzepte, Praxis* (S. 55-75). Düsseldorf: VDM Verlag Dr. Müller.

Raven, B. H. (1992). A power/interaction model of interpersonal influence: French and Raven thirty years later. *Journal of Social Behavior and Personality, 7,* 217-244.

Reddin, W. J. (1977). *Das 3-D-Programm zur Leistungssteigerung des Managements*. München: moderne Industrie.

Reilly, R. R., Smither, J. W. & Vasilopoulos, N. L. (1996). A longitudinal study of upward feedback. *Personnel Psychology, 49,* 599-612.

Reitz, H. J. (1971). Managerial attitudes and perceived contingencies between performance and organizational response. *Proceedings of the Academy of Management,* 227-238.

Rief, W. & Birbaumer, N. (Hrsg.). (2000). *Biofeedback-Therapie*. Stuttgart: Schattauer.

Rieker, J. (1994). Gute Noten. Vorgesetztenbeurteilung. *Manager Magazin, 9,* 180-189.

Riggio, R. E. & Cole, E. (1992). Agreement between subordinate ratings of supervisor performance and effects on self and subordinate job satisfaction. *Journal of Occupational and Organizational Psychology, 65,* 151-158.

Roberts, N. (1985). Transforming leadership: A process of collective action. *Human Relations, 38,* 1023-1046.

Roberts, N. C. & Bradley, R. T. (1988). Limits of charisma. In J. A. Conger & R. N. Kanungo (Eds.), *Charismatic leadership. The elusive factor in organizational effectiveness* (pp. 253-275). San Francisco: Jossey-Bass.

Rosenfeld, P., Giacalone, R. A. & Riordan, C. A (1995). *Impression management in organizations. Theory, measurement, practice.* (2nd ed.). London: Routledge.

Rosenstiel, L. von (1999a). Arbeitszufriedenheit. In L. von Rosenstiel, E. Regnet & M. E. Domsch (Hrsg.), *Führung von Mitarbeitern* (4. Aufl.). (S. 193-204). Stuttgart: Schäffer-Poeschel.

Rosenstiel, L. von, Regnet, E. & Domsch, M.E. (Hrsg.). (1999). *Führung von Mitarbeitern* (4. Aufl.). Stuttgart: Schäffer-Poeschel.

Rosenthal, R. L., Rosnow, R. L. & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: a correlational approach.* Cambridge: Cambridge University Press.

Rosnow, R. L. & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods, 1,* 331–340.

Ross, M. & Sicoly, F. (1982). Egocentric biases in availability and attribution. In D. Kahneman, P. Slovic & A. Tversky (Eds.). *Judgment under uncertainty: Heuristics and biases* (pp. 179-189). Cambridge: Cambridge University Press.

Rost, J. C. (1998). Leadership and management. In G. R. Hickman (Ed.), *Leading organizations: Perspectives for a new era* (pp. 97-114). Thousand Oaks: Sage.

Rost, W., Frey, D. & Rost-Schaude, E. (1975). Der Einfluß von Erfolg und Mißerfolg auf Attraktivitätsurteile im Bereich der Schule. In W. Tack (Hrsg.),. *Bericht über den 29. Kongreß der Deutschen Gesellschaft für Psychologie in Salzburg 1974* (S. 321-323). Göttingen: Hogrefe.

Rothstein, H. R. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. *Journal of Applied Psychology, 75,* 322-327.

Roush, P. E. & Atwater L. E. (1992). Using the MBTI to understand transformational leadership and self-perception accuracy. *Military Psychology, 4,* 17-34.

Runde, B., Kirschbaum, D. & Wübbelmann, K. (2001). 360°-Feedback – Hinweise für ein best-practice-Modell. *Zeitschrift für Arbeits- und Organisationspsychologie, 45,* 146-157.

Runge, T. (1996). Studie zur Sozialen Validität. In Arbeitskreis Assessment Center e.V. (Hrsg.), *Assessment Center als Instrument der Personalentwicklung* (S. 286-297). Hamburg: Windmühle.

Sala, F. & Dwight, S. A. (2002). Predicting executive performance with multirater surveys: Whom you ask makes a difference. *Consulting Psychology Journal: Practice and Research, 54,* 166-172.

Salancik, G. R. & Pfeffer, J. (1977). Constraints on administrator discretion: The limited influence of mayors on city budgets. *Urban Affairs Quarterly, 12,* 475-498.

Salgado, J. F., Moscoso, S. & Lado, M. (2003). Test-retest reliability of ratings of job performance dimensions in managers. *International Journal of Selection and Assessment, 11* (1)*,* 98-101.

Sanchez-Burks, J., Lee, F., Choi, I., Nisbett, R., Zhao, S & Koo, J. (2003). Conversing across cultures: East-West communication styles in work and nonwork contexts. *Journal of Personality and Social Psychology, 85,* 363-372.

Sarges, W. (Hrsg.). (1995). *Management-Diagnostik* (2. Aufl.). Göttingen: Hogrefe.

Sashkin, M. (1977). Comments on chapter 12 - the structure of charismatic leadership. In J. G. Hunt & L. L. Larson (Eds.), *Leadership. The cutting edge* (pp. 212-218). Carbondale: Southern Illinois University Press.

Sashkin, M. (1988). The visionary leader. In J. A. Conger & R. N. Kanungo (Eds.), *Charismatic leadership. The elusive factor in organizational effectiveness* (pp. 122-160). San Francisco: Jossey-Bass.

Schaller, J. & Jonas, K. (2004). Einführung: Von Autos und Menschen. In K. Jonas, G. Keilhofer & J. Schaller (Hrsg.), *Human Resource Management im Automobilbau. Konzepte und Erfahrungen.* Bern: Huber.

Schein, V. E., Mueller, R., Lituchy, T. & Liu, J. (1996). Think manger – think male: A global phenomenon? *Journal of Organizational Behavior, 17,* 33-41.

Scherm, M. & Sarges, W. (2002). *360°-Feedback*. Göttingen: Hogrefe.

Schilling, J. (2004). Was Unternehmen unter Führung verstehen. Eine Dokumentenanalyse von Führungsgrundsätzen. In W. Bungard, B. Koop & C. Liebig (Hrsg.), *Psychologie und Wirtschaft leben: aktuelle Themen der Wirtschaftspsychologie in Forschung und Praxis* (S.292-297). München und Mering: Rainer Kampp.

Schlenker, B. R. (1980). *Impression management: The self concept, social identity and interpersonal relations.* Monterey, CA: Brooks.

Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical findings and theoretical implications of 85 years of research findings. *Psycholgical Bulletin, 124*, 262-274

Schmidt, K. H. (1996). Wahrgenommenes Vorgesetztenverhalten, Fehlzeiten und Fluktuation. *Zeitschrift für Arbeits- und Organisationspsychologie, 40,* 54-62.

Scholz, Ch. (2000). *Personalmanagement* (5. Aufl.). München: Verlag Franz Vahlen.

Scholz, G. & Schuler, H. (1993). Das nomologische Netzwerk des Assessment Centers: eine Meta-Analyse. *Zeitschrift für Arbeits- und Organisationspsychologie, 37,* 73-85.

Schriesheim, Ch. A., Tepper, B. J. & Tetrault, L. A. (1994). Least Preferred Co-worker score, situational control, and leadership effectiveness: A meta-analysis of contingence model performance predictions. *Journal of Applied Psychology, 79,* 561-573.

Schuler, H. (1989). Leistungsbeurteilung. In. E. Roth (Hrsg.), *Organisationspsychologie. Enzyklopädie der Psychologie* (Themenbereich D, Ser. 3, Bd. 3, S. 399-430). Göttingen: Hogrefe.

Schulz-Hardt, S. (1997). *Realitätsflucht in Entscheidungsprozessen: Vom Groupthink zum Entscheidungsautismus.* Bern: Huber

Schumacher, B. (1988). *Handbuch: Führungsgrundsätze und Führungshilfen entwickeln und einführen: Bausteine der Organisationskultur.* (2. Aufl.). Landsberg am Lech: Verlag Moderne Industrie.

Schwarz, N. & Hippler, H.-J. (1987). What response scales may tell your respondents: informative functions of response alternatives. In H. J. Hippler, N. Schwarz & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 163-178). New York: Springer.

Scullen, S. E., Mount, M. K. & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology, 85,* 956-970.

Sedikides, C. & Strube, M. J. (1997). *Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and thine own self be better.* In: M. P. Zanna (Ed.), Advances in experimental social psychology (Vol. 29, pp. 209-269). New York: Academic Press

Seligman, L.G. (1975). *Helplessness. On depression, development, and death.* San Francisco: Freeman.

Selznick, P. (1957). *Leadership in administration: A sociological interpretation*. Evanston, IL: Row, Peterson.

Seltzer, J. & Bass, B. M. (1990). Transformational leadership: Beyond initiation and consideration. *Journal of Management, 16*, 693-703.

Seltzer, J. & Numerof, R. E. (1988). Supervisory leadership and subordinate burnout. *Academy of Management Journal, 31*, 439-449.

Seltzer, J., Numerof, R. & Bass, B. (1989). Transformational leadership: Is it a source of more burnout and stress? *Journal of Health and Human Resources, 1,* 174-185.

Senge, P. M. (1999). *Die fünfte Disziplin* (7. Aufl.). Stuttgart: Klett-Cotta.

Senger, N. (in Vorb.). *Entstehungsbedingungen für Diskrepanzen zwischen der Selbst- und Aufwärtsbeurteilung von Führungskräften – eine Interviewstudie.* Unveröffentlichte Diplomarbeit an der TU Chemnitz.

Shamir, B., House, R. & Arthur, M. (1993). The motivational effects of charismatic leadership: A self-concept based theory. *Organization Science, 4*, 1-17.

Shipper, F. & Wilson, C. L. (1992). The impact of managerial behaviors on group performance, stress, and commitment. In K. E. Clark, M. B. Clark & D. P. Campbell (Eds.), *Impact of leadership* (pp. 119-129). Greensboro: The Center for Creative Leadership.

Shore, L. M., Tetrick, L. E. & Shore, T. H. (1998).A comparison of self-, peer, and assessor evaluations of managerial potential. *Journal of Social Behavior and Personality, 13,* 85-101.

Shrauger, J. S. (1975). Responses to evaluation as a function of initial self-perceptions. *Psychological Bulletin, 82,* 581-596.

Shrauger, J. S. & Schoeneman, T. J. (1979). Symbolic interactionist view of self-concept: Through the looking glass darkly. *Psychological Bulletin, 86*, 549-573.

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

Silverthorne, C. & Wang, T. H. (2001). Situational leadership style as a predictor of success and productivity among Taiwanese business organizations. *Journal of Psychology: Interdisciplinary & Applied, 135* (4), 399-412.

Sims, H. P. (1977). The leader as manager of reinforcement contingencies: An empirical example and a model. In J. G. Hunt & C. C. Larson (Eds.), *Leadership: The cutting edge.* Carbondale: Southern Illinois University Press.

Smith, J. E., Carson, K. P. & Alexander, R. A. (1984). Leadership: It can make a difference. *Academy of Management Journal, 27,* 765-776.

Smither, J. W., London, M., Flautt, R., Vargas, Y. & Kucine, I. (2003). Can working with an executive coach improve multisource feedback ratings over time? A quasi-experimental field study, *Personnel Psychology, 56,* 23-44.

Smither, J. W., London, M., Vasilopoulos, N. L., Reilly, R. R., Millsap, R. E. & Salvemini, N. (1995). An examination of the effects of an upward feedback program over time. *Personnel Psychology, 48,* 1-34.

Smither, J. W., Wohlers, A. J. & London, M. (1995). A field study of reactions to normative versus individualized upward feedback. *Group and Organization Management, 20,* 61-89.

Snyder, M. (1976). Attribution and behavior: Social perception and social causation. In J. H. Harvey, W. J. Ickes & R. F. Kidd (Eds.), *New directions in attribution research*. Hillsdale, NJ: Lawrence Erlbaum.

Sohm, R. (1892). *Kirchenrecht*. Leipzig.

Somers, M. J. & Birnbaum, D. (1991). Assessing self-appraisal of job performance as an evaluation device: Are the poor results a function of method or methodology? *Human Relations, 44,* 1081-1091.

Sonntag, K. (1999). (Hrsg.). *Personalentwicklung in Organisationen* (2. Aufl.). Göttingen: Hogrefe.

Sosik, J. J. Avolio, B. J. & Jung, D. I. (2002). Beneath the mask: Examining the relationship of self-presentation attributes and impression management to charismatic leadership. *The Leadership Quarterly, 13,* 217-242.

Sosik, J. J., Avolio, B. J. & Kahai, S. S. (1998). Inspiring group creativity. Comparing anonymous and identified electronic brainstorming. *Small Group Research, 29*, 3-31.

Sosik, J. J., Kahai, S. S. & Avolio, B. J. (1999). Leadership style, anonymity, and creativity in group decision support systems: The mediating role of optimal flow. *The Journal of Creative Behavior, 33* (4), 227-256.

Sosik, J. J. & Megerian, L. E. (1999). Understanding leader emotional intelligence and performance: The role of self-other rating agreement on transformational leadership perceptions. *Group and Organization Management, 24,* 367-390.

Spreitzer, G. M., McCall, M. W. & Mahoney, J. D. (1997). Early identification of international executive potential. *Journal of Applied Psychology,* 82, 6-29.

Sprenger, R. K. (1999). *Das Prinzip Selbstverantwortung* (10. Aufl.). Frankfurt/Main: Campus.

Staehle, W. (1999). *Management. Eine verhaltenswisenschaftliche Perspektive* (8. Aufl.). München: Vahlen.

Staufenbiel, T. (1999). Personalentwicklung. In C. G. Hoyos & D. Frey (Hrsg.), *Arbeits- und Organisationspsychologie. Ein Lehrbuch* (S. 510-525). Weinheim: Beltz.

Steers, R. M. & Mowday, R. T. (1981). Employee turnover and post-decision accommodation processes. *Research in Organizational Behavior*, 3, 235-281.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245-251.

Steiner, I. D. (1972). *Group processes and productivity*. New York: Academic Press.

Steiner, I. D. (1976). Task-performing groups. In J. W. Thibaut, J. T. Spence & R. C. Carson (Eds.), *Contemporary topics in social psychology* (pp. 393-422). Morristown, NJ: General Learning Press.

Steinert, C. (2003). *Feedback als Führungsinstrument*. Verfügbar unter: https://www.personalseite.de/aufsatz/steinert3.htm [24.10.2003].

Steinmann, H. & Schreyögg, G. (1993). *Management: Grundlagen der Unternehmensführung: Konzepte - Funktionen - Fallstudien.* (3. Aufl.). Wiesbaden: Gabler.

Stempfle, J., Hagmayer, Y., Hübner, O., Iwanoff, Ch. & Kaufmann, St. (2004). Weiterentwicklung und Dynamisierung des Assessment Centers auf der Basis eines komplexen computersimulierten Problemlöseszenarios: Das Challenge-AC der Thomas Cook AG. *Wirtschaftspsychologie 2,* 23-36.

Stevens, C. K. & Kristof, A. L. (1995). Making the right impression: A field study of applicant impression management during job interviews. *Journal of Applied Psychology, 80,* 587-606.

Stewart, R. (1983). Managerial behavior: How research has changed the traditional picture. In M. Earl (Ed.), *Perspectives on management – A multidisciplinary analysis.* Oxford: University Press.

Stogdill, R. M. (1948). Personal factors associated with leadership. A survey of the literature. *Journal of Psychology, 25*, 35-71.

Stone-Romero, E. & Anderson, L. (1994). Relative power of moderated multiple regression and the comparison of subgroup correlation coefficients for detecting moderating effects. *Journal of Applied Psychology, 79,* 354-359.

Strack, F. & Deutsch, R. (2002). Urteilsheuristiken. In D. Frey & M. Irle (Hrsg.) *Theorien der Sozialpsychologie* (2. Aufl). (Bd. 3, S. 352-384). Bern: Hans Huber.

Strube, M. J. & Garcia, J. E. (1981). A meta-analytic investigation of Fiedler's contingency model of leadership effectiveness. *Psychological Bulletin, 90,* 307-321.

Sudman, S., Bradburn, N. M. & Schwarz, N. (1996). *Thinking about answers. The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.

Svenson , O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica, 47,* 143-148

Svyantek, D. J. & DeShon, R. P. (1992). Leaders and organizational outcomes in established industries. An analysis of Lee Iacocca and the American Automobile Industry. In K. E. Clark, M. B. Clark & D. P. Campbell (Eds.), *Impact of leadership* (pp. 293-303). Greensboro: The Center for Creative Leadership.

Swann, W. B. (1983). Self-verification: Bringing social reality into harmony with the self. In J. Suls & A. G. Greenwald (Eds.), *Social psychologcial perspectives on the self* (Vol. 2, pp. 33-66). Hillsdale, NJ: Erlbaum.

Tajfel, H. & Turner, J. (1986) The social identity of intergroup behavior. In S. Worchel & W. Austin (Eds.), *Psychology and intergroup relations* (pp. 7-24). Chicago: Nelson-Hall.

Tannenbaum, R. & Schmidt, W. H. (1958). How to choose a leadership pattern. *Harvard Business Review, 36*, 95-101.

Tartler, K., Goihl, K. Kroeger, M. & Felfe, J. (2003). Zum Nutzen zusätzlicher Selbsteinschätzungen bei der Beurteilung des Führungsverhaltens. *Zeitschrift für Personalpsychologie, 2,* 13-21.

Tartler, K., Liepmann, D., Felfe, J. & Nettelnstroth, W. (2004). Single-source single-method. Was bleibt von den Zusammenhängen zwischen transformationaler Führung und Erfolgsindikatoren. In W. Bungard, B. Koop & C. Liebig (Hrsg.), *Psychologie und Wirtschaft leben: aktuelle Themen der Wirtschaftspsychologie in Forschung und Praxis* (S. 259-263). München und Mering: Rainer Kampp.

Tartler, K., Liepmann, D. & Nettelnstroth, W. (2003). *Revisiting the german version of the MLQ – analyzing the effects of lexical overlap in common variance.* Poster präsentiert auf dem elften European Congress on Work and Organizational Psychology, Lissabon.

Taylor, S. E. & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103,* 193–210.

Tedeschi, J. T. (1981). (Ed.). *Impression management theory and social psychological research.* New York: Academic Press.

Tejeda, M. J., Scandura, T. A. & Pillai, R. (2001). The MLQ revisited. Psychometric properties and recommendations. *Leadership Quarterly, 12,* 31-52.

Tesser, A. & Rosen, S. (1975). The reluctance to transmit bad news. In L. Berkowitz (Ed.), *Advances in experimental psychology* (Vol. 8, pp. 193-232). New York: Academic Press.

Thomas, A. B. (1988). Does leadership make a difference to organizational performance? *Administrative Science Quarterly, 33,* 338-400.

Thornton, G. C. (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology, 33,* 263-271.

Tichy, N. M. & Devanna, M. A. (1986a). *The transformational leader*. New York: Wiley.

Tichy, N. M. & Devanna, M. A. (1986b). The transformational leader. *Training & Development Journal, 40* (7), 27-32.

Tichy, N. M., Fombrun, C. J. & Devanna, M. A. (1982). Strategic human resource management. *Sloan Management Review, 23,* 47-61.

Tisak, J. & Smith, C. S. (1994). Defending and extending difference score methods. *Journal of Management, 20,* 675-682.

Tisdale, T. (1998). *Selbstreflexion, Bewusstsein und Handlungsregulation*. Weinheim: PVU.

Titscher, E. & Titscher, S. (1977). Dimensionen des Vorgesetzten-Verhaltens – eine faktorenanalytische Überprüfung des „Fragebogens zur Vorgesetzten-Verhaltens-Beschreibung (FVVB)“. *Psychologie und Praxis, 21* (4)*,* 154-166.

Torjus, N., Tartler, K. & Liepmann, D. (2004). Transformationale Führung und Kommunikation. In W. Bungard, B. Koop & C. Liebig (Hrsg.), *Psychologie und Wirtschaft leben: aktuelle Themen der Wirtschaftspsychologie in Forschung und Praxis* (S. 271- 276). München und Mering: Rainer Kampp.

Tornow, W. (1993). Perceptions or reality: Is multi-perspective measurement a means or an end? *Human Resource Management, 32*, 221-229.

Tornow, W. & London, M. (Eds.). (1998). *Maximizing the value of 360-degree feedback: a process for successful individual and organizational development*. San Francisco: Jossey-Bass.

Tosi, H. L., Mero, N. P. & Rizzo, J. R. (2000). *Managing organizational behavior* (4th ed.). Malden, Oxford: Blackwell.

Towler, A. J. (2003). Effects of charismatic influence training on attitudes, behavior, and performance. *Personnel Psychology 56,* 363-381

Trice, H. M. & Beyer, J. M. (1986). Charisma and its routinization in two social movement organizations. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 8, pp. 113-164). Greenwich, CT: JAI Press.

Trost, A. (1997). Das Antwortverhalten befragter Mitarbeiter – eine kognitionspsycholgische Perspektive. In W. Bungard & I. Jöns (Hrsg.), *Mitarbeiterbefragung. Ein Instrument des Innovations- und Qualitätsmanagements* (S. 130-147). Weinheim: PVU.

Tsui, A. S. & Ohlott, P. (1988). Multiple assessment of managerial effectiveness: Interrater agreement and consensus in effectiveness models. *Personnel Psychology, 41*, 779–803.

Turner, J. (2000). Some current issues in research on social identity and self-categorization theories. In N. Ellemers, R. Spears & B. Doosje (Eds.), *Social identity: context, commitment, content* (pp. 6-34). Oxford: Blackwell.

Turner, N., Barling, J., Epitropaki, O., Butcher, V. & Milner, C. (2002). Transformational leadership and moral reasoning. *Journal of Applied Psychology, 87,* 304-311.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

Twenge, J. M. & Campbell, W. K. (2002). Self-esteem and socioeconomic status: A meta-analytic review. *Personality and Social Psychology Review, 6* (1), 59-71.

Van Velsor, E. & Leslie, J. B. (1995). Why executives derail: perspectives across time and cultures. *Academy of Management Executive, 9* (4)*,* 62-72.

Van Velsor, E., Taylor, S. & Leslie, J. B. (1993). An examination of the relationships among self-perception accuracy, self-awareness, gender, and leader effectiveness. *Human Resource Management, 32*, 249-263.

Vandenberghe, Ch., Stordeur, S. & D'hoore, W. (2002). Transactional and transformational leadership in nursing: Structural validity and substantive relationships. *European Journal of Psychological Assessment, 18*, 16-29.

Vecchio, R. P. (1987). Situational leadership theory: An examination of a prescriptive theory. *Journal of Applied Psychology, 72*, 444-451.

Verband der Automobilindustrie (2004). Verfügbar unter: http://www.vda.de/de/service/jahresbericht/files/VDA_2004.pdf [08.10.2004].

Viswesvaran, C., Ones, D. S. & Schmidt, F. L. (1996). Comparative analysis of the reliability of rob performance ratings. *Journal of Applied Psychology, 81,* 557–574.

Viswesvaran, C., Schmidt, F. L. & Ones, D. S. (2002). The moderating influence of job performance dimensions on convergence of supervisory and peer ratings of job performance: Unconfounding construct-level convergence and rating difficulty. *Journal of Applied Psychology, 87,* 345-354.

Voltz, T. (1998). *Mut zur Kritik. Vorgesetztenbeurteilung einsetzen und durchführen.* Zürich: Orell Füssli Verlag.

Vroom, V.H. (1964). *Work and motivation*. New York: Wiley.

Vroom, V. H. & Jago, A. G. (1991). *Flexible Führungsentscheidungen. Management der Partizipation in Organisationen.* Stuttgart: Poeschel.

Vroom, V. H. & Yetton, P. W. (1973). *Leadership and decision-making*. Pittsburgh: University of Pittsburgh Press.

Waldman, D. A. & Atwater, L. (1998). *The power of 360° feedback: How to leverage performance evaluations for top productivity*. Houston, TX: Gulf Publishing.

Waldman, D. A., Bass, B. M. & Einstein, W. O. (1987). Effort, performance, and transformational leadership in industrial and military settings. *Journal of Occupational Psychology, 60,* 177-186.

Waldman, D. A., Bass, B. M. & Yammarino, F. J. (1990). Adding to contingent-reward behavior: The augmenting effects of charismatic leadership. *Group and Organization Studies, 15*, 381-394.

Waldman, D. A., Ramírez, G. G., House, R. J. & Puranam, P. (2001). Does leadership matter? CEO leadership attributes and profitability under conditions of perceived environmental uncertainty. *Academy of Management Journal, 44,* 134-143.

Walker, A. G. & Smither, J. W. (1999). A five-year study of upward feedback: What managers do with their results matters. *Personnel Psychology, 52*, 393-423.

Walton, D. & Bathurst, J. (1998). An exploration of the perceptions of the average driver's speed compared to perceived driver safety and driving skill. *Accident Analysis & Prevention, 30*, 821-830.

Warech, M. A., Smither, J. W., Reilly, R. R., Millsap, R. E. & Reilly, S. P. (1998). Self-monitoring and 360-degree ratings. *Leadership Quarterly, 9,* 449-473.

Weber, M. (1924). *Wirtschaftsgeschichte* (2. Aufl.). München: Duncker & Humblot.

Weber, M. (1966). *Staatssoziologie* (2. Aufl.). Berlin: Duncker & Humblot.

Weber, M. (1980 [1921]). *Wirtschaft und Gesellschaft* (5. Aufl.). Tübingen: Mohr.

Webster, J. & Starbuck, W. (1988). Theory building in industrial and organizational psychology. In C. Cooper & I. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 3, pp. 93-138). New York: John Wiley.

Weibler, J. (2001). *Personalführung*. München: Vahlen.

Weiner, A. B. (1981). A model of corporate performance as a function of environmental , organizational, and leadership influences. *Academy of Management Journal, 24,* 453-470.

Weinert. A. (1998). *Organisationspsychologie. Ein Lehrbuch* (4. Aufl.). Weinheim: Beltz.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Personality, 39,* 806-820.

Weise, G. (1975). *Psychologische Leistungstests. Göttingen*: Hogrefe.

Weiss, H. M. (1977). Subordinate imitation of supervisor behavior: The role of modeling in organizational socialization. *Organizational Behavior and Human Performance, 19,* 89-105.

Weiss, H. M. (1978). Social learning of work values in organizations. *Journal of Applied Psychology, 63,* 711-718.

Welch, J. (2003). *Was zählt. Die Autobiografie des besten Managers der Welt.* Ulm: Ullstein.

Wendel, T. H. (2003, 22. Juli). Mehr Treibstoff für den Auto-Kanzler. *Berliner Zeitung*, S. 23.

Wexley, K. N., Alexander, R., Greenawalt, J. & Couch, M. (1980). Attitudinal congruence and similarity as related to interpersonal evaluation in manager-subordinate dyads. *Academy of Management Journal, 23,* 320-330.

Wherry, R. J. & Bartlett, C. J. (1982). The control of bias in ratings: A theory of rating. *Personnel Psychology, 35*, 521–551.

Wiendieck, G. (1997). Führungskräfte im Urteil ihrer Mitarbeiter Ein Erfahrungsbericht. In W. Bungard & I. Jöns (Hrsg.), *Mitarbeiterbefragung. Ein Instrument des Innovations- und Qualitätsmanagements* (S. 386-398). Weinheim: PVU.

Wilkens, U. (1998). *Human Resource Management in der europäischen Automobilindustrie. Ein cross-nationaler Vergleich*. Frankfurt/Main: Lang.

Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin, 90,* 245-271.

Wilson, B. (1975). *The noble savage*. Berkeley: University of California Press.

Winckelmann, J. (1952). *Legitimität und Legalität in Max Webers Herrschaftssoziologie*. Tübingen: Mohr.

Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.

Witte, E. (1995). Effizienz der Führung. In A. Kieser, G. Reber & R. Wunderer (Hrsg.), *Handwörterbuch der Führung* (2. Aufl.) (S. 265-276). Stuttgart: Schäffer-Poeschel.

Wohlers, A. J. & London, M. (1989). Ratings of managerial characteristics – Evaluation difficulty, co-worker agreement and self-awareness. *Personnel Psychology, 42,* 235-261.

Wood, J. V., Giordano-Beech, M., Taylor, K. L., Michela, J. L. & Gaus, V. (1994). Strategies of social comparison among people with low self-esteem: Self-protection and self-enhancement. *Journal of Personality and Social Psychology, 67,* 713-731.

Wosinska, W., Dabul, A. J., Whetstone-Dion, R. & Cialdini, R. B. (1996). Self-presentational responses to success in the organization: The costs and benefits of modesty. *Basic and Applied Social Psychology, 18,* 229–242.

Wunderer, R. (1983). (Hrsg.). *Führungsgrundsätze in Wirtschaft und öffentlicher Verwaltung*. Stuttgart: C.E. Poeschel.

Wunderer, R. (1997). *Führung und Zusammenarbeit. Beiträge zu einer unternehmerischen Führungslehre* (2. vollständig überarbeitete und erweiterte Aufl.). Stuttgart: Schäffer-Poeschel.

Wunderer, R. (2001). *Führung und Zusammenarbeit. Eine unternehmerische Führungslehre* (4. vollständig überarbeitete Aufl.). Neuwied: Luchterhand.

Yammarino, F. J. & Atwater L. E. (1993). Understanding self-perception accuracy: Implications for human resources management. *Human Resource Management, 32,* 231-247.

Yammarino, F. J. & Atwater L. E. (1997). Do managers see themselves as others see them? Implications of self-other rating agreement for human resources management. *Organizational Dynamics, 25*, 35-44.

Yammarino, F. J. & Bass, B. M. (1990). Long-term forecasting of transformational leadership and its effects among naval officers: Some preliminary findings. In K. Clark & M. Clark (Eds.), *Measures of leadership* (pp. 151-169). West Orange: Leadership Library of America.

Yammarino, F. J., Dansereau, F. & Kennedy, C. J. (2001). A multiple-level multidimensional approach to leadership: Viewing leadership through an elephants's eye. *Organizational Dynamics, 29,* 149-163.

Yammarino, F. J. & Dubinsky, A. J. (1990). Salesperson performance and managerially controllable factors: An investigation of individual and work group effects. *Journal of Management, 16,* 87-106.

Yukl, G. (1981). *Leadership in organizations*. Englewood Cliffs, New Jersey: Prentice-Hall.

Yukl, G. (1999a). An evaluative essay on current conceptions of effective leadership. *European Journal of Work and Organizational Psychology, 8* (1)*,* 33-48.

Yukl, G. (1999b). *Leadership in organizations* (4th ed.), Englewood Cliffs, NJ: Prentice Hall.

Yukl, G. & Howell, J. (1999). Organizational and contextual influences on the emergence and effectiveness of charismatic leadership. *Leadership Quarterly, 10*, 257-283.

Zacharatos, A., Barling, J. & Kelloway, E. K. (2000). Development and effects of transformational leadership in adolescents. *Leadership Quarterly*, *11,* 211-226.

Zimolong, B. (Hrsg.). (2001). *Management des Arbeits- und Gesundheitsschutzes. Die erfolgreichen Strategien der Unternehmen.* Wiesbaden: Gabler.

Zimolong, B. & Elke, G. (2001). Die erfolgreichen Strategien und Praktiken der Unternehmen. In B. Zimolong (Hrsg.), *Management des Arbeits- und Gesundheitsschutzes. Die erfolgreichen Strategien der Unternehmen* (S. 235-268). Wiesbaden: Gabler.

9 Personenverzeichnis

Aiken 358, 361
Aitkenhead 162
Albach 195
Alicke 123, 210, 331, 435
Allerbeck 96
Allison 122
Amabile 64
Ambrosch 37
Antonakis 100
Antonioni 111
Argyris 26
Ashford 167, 216, 422, 454, 461
Astin 242
Atwater 68, 75, 107, 117, 118, 119, 161, 167, 168, 170, 171, 172, 176, 182, 183, 184, 185, 186, 187, 193, 215, 222, 233, 234, 321, 383, 397, 405, 415, 418, 422, 423, 427, 428, 429, 456, 458
Avolio 55, 56, 58, 66, 68, 71, 72, 73, 75, 76, 80, 81, 82, 99
Backhaus 247
Bailey 121
Barbuto 88
Baril 119, 427, 428
Barling 81, 98
Bass 11, 13, 15, 41, 42, 45, 46, 47, 50, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 68, 70, 71, 72, 73, 75, 76, 77, 78, 80, 81, 82, 83, 84, 85, 86, 88, 89, 90, 91, 95, 98, 99, 100, 106, 107, 108, 112, 117, 120, 130, 170, 171, 175, 201, 210, 214, 215, 235, 268, 276, 277, 355, 383, 399, 419, 427, 428
Bauer 27, 91
Baum 50
Baumeister 159, 177
Becker 128, 140, 149, 229, 463
Beckmann 29
Beer 20
Bennis 49, 50, 72, 75
Bernardin 104, 114, 193, 202
Beyer 128
Binning 252
35, 36, 348
38
Bobko 246, 247, 295, 335, 389, 427
Bodenhausen 147
Böhnisch 38
Bommer 108, 134, 138, 210
Bono 52, 85
Borg 112, 113, 239, 252, 270
Borman 131, 139, 140, 141, 142, 145, 202, 273, 435, 452
Bortz 156, 246, 248, 325, 330
Boski 157
Bozeman 460
Bracken 181, 466
Brandstätter 152
Brehm 16
Brenner 188
Brett 185
Brodbeck 64, 68
Brose 27
Brown 50, 129, 154
Brunner 10
Brutus 10, 111, 113, 114, 128, 141, 169, 189, 218, 302, 321, 363, 383, 385, 387, 400, 456
Bryman 42, 50, 76, 85
Bungard 10, 37, 112, 114, 245, 273
Burns 42, 48, 49, 50, 53, 54, 56, 65, 78
Bycio 56, 69, 72, 73, 98, 106, 108, 459
Cadinu 443
Calder 40
Carver 186, 343, 450
Cattell 257
Charbonneau 69, 72
Chemers 85
Cheung 289
Choi 149
Church 76, 119, 129, 130, 169, 170, 171, 173, 176, 216, 220, 228, 321, 366, 367, 375, 383, 426, 427, 428, 429, 430, 444, 453
Cialdini 76
Cillessen 303, 455, 467
Clements 77
Cleveland 113
Cohen 249
Colvin 169, 179, 462
Conger 51, 77, 78, 80, 82, 85, 87, 88, 101
Conway 10, 13, 105, 108, 109, 114, 116, 118, 119, 124, 126, 127, 130, 131, 138, 139, 142, 143, 144, 145, 146, 147, 155, 161, 166, 167, 185, 196, 210, 213, 278, 279, 289, 296, 297, 418, 419, 421, 422, 425, 426, 435, 448, 450, 452, 455
Cooley 466
Crampton 121, 197, 301
Cronbach 233, 250
Cross 122
Dalton 113
Dauenheimer 122, 150, 156, 157
Deal 465
DeGroot 101
den Hartog 459
Den Hartog 68, 76, 98, 277
DeNisi 185, 186, 189
Diehl 325

Domsch 27
Donat 273
Dörner 369, 454
Drees 120
Dumdum 70, 71, 73
Dunlap 248, 249
Dunnette 143, 181
Dunning 122
Dvir 68, 72, 81
Eagly 128, 434
Edwards 114, 173, 233, 289, 321, 364, 385, 386, 387, 388, 389
Ehrlich 21
Elenkov 72
Elke 177
Ellis 443
Evans 60
Ewen 10
Farh 134, 151
Farr 180
Fayol 28
Felfe 72, 73, 87, 95, 98, 99, 230, 234, 235, 240, 266, 268, 276, 277, 310, 433, 459
Fengler 188
Fennekels 95, 97
Ferguson 25
Festinger 134, 149, 183
Fiedler 36, 37, 39, 84, 179
Fiol 73, 75
Fishbein 20, 106, 447
Fittkau 95, 96
Fittkau-Garthe 211
Flanagan 97
Fleenor 117, 119, 148, 159, 170, 173, 174, 181, 220, 234, 375, 418, 422, 426, 427, 428, 429, 450
Fleishman 35, 37, 84, 96
Fletcher 105, 113
Fox 201
Freedman 122
French 86
Frese 81
Fried 348
Fuller 69, 70, 71, 73, 92, 101
Furnham 106, 117, 119, 171, 172, 375, 427
Gabele 29, 36, 39, 194, 348
Gardner 85
Gebert 19, 27, 93, 193
Geyer 56, 68, 69, 72, 73, 99, 106, 107, 108, 277
Giacalone 159
Gilbert 144
Goodwin 100
Gould 159
Graen 39, 141, 419
Green 40
Greguras 105, 165, 166, 185, 216, 223, 302, 453
Greller 180
Grieb 465
Grün 77, 237
Guion 145
Guttman 257
Guzzo 184, 190
Hackman 27, 181, 347
Harris 116, 117, 118, 119, 126, 127, 130, 132, 140, 143, 144, 145, 146, 147, 155, 156, 157, 160, 204, 243, 296, 297, 421, 422, 425, 435, 450
Harter 74
Hater 56, 68, 72, 73, 75, 98, 107, 108, 113, 210
Hatfield 176
Hawkins 153
Hazucha 169, 185
Hedge 10, 113, 280, 420
Heidemeier 116, 117, 118, 125, 126, 127, 128, 130, 131, 132, 133, 134, 135, 136, 143, 144, 145, 147, 149, 151, 155, 158, 160, 175, 196, 199, 204, 208, 296, 297, 301, 304, 309, 321, 414, 417, 421, 422, 425, 426, 429, 431, 432, 434, 435, 450
Heider 161
Heilman 128
Heller 27
Heneman 134
Hersey 37, 38, 84, 86
Herzberg 269, 270
Hickman 68
Hoch 50, 170
Hofmann 112
Hofstede 75, 128
Hollander 39, 87, 89
Homans 48
Hoog 150
Hooijberg 77, 81, 100
Hossiep 399, 461
House 20, 25, 39, 42, 46, 48, 52, 60, 64, 68, 72, 75, 78, 79, 84, 85, 86, 88, 419
Howell 71, 72, 73, 74, 75, 76, 78, 92
Hoyt 167, 252, 278
Hunt 48, 85
Ilgen 154, 181
Jacobs 155
Jago 38
Janis 182
Janssen 72, 73
Jawahar 129, 135, 141, 160
Jay 112
Jeserich 156, 169, 462
Jochum 109
Johns 120, 234, 289
Johnson 182, 184, 186, 461
Jonas 11, 72
Jones 76, 148
Jöns 229, 464
Judge 33, 34, 69, 70, 73
Jung 72, 73
Kahneman 146
Kaiser 257
Kanfer 26
Katz 17
Kegan 75
Kehr 20

Keller 64, 72, 74, 107, 450
Kelloway 81
Kenny 466
Kenrick 183
Kernis 157
Kerr 40, 86, 92
Kienbaum 49
Kiessling 111, 281
Kim 128
Kirkpatrick 33
Klimoski 120, 144
Kluge 16
Kluger 182, 184, 188, 190
Koch 120
Koene 68, 72, 75, 106, 107
Koh 69, 72
Korman 443
Kosch 11
Kouzes 49
Kristof-Brown 159
Kruger 178, 467
Kuhnert 65
Kuiper 157
Larwood 122, 208
Latham 202
Laux 160
Lebow 112
Lehner 20, 465
Lewin 34, 35, 37, 57, 83, 84, 90
Liebel 24, 25, 281
Lieberson 19
Lievens 99
Likert 35, 84
Limpächer 19
Locke 25, 82, 181, 186, 287, 343
Lombardo 85, 108
London 111, 114, 119, 128, 154, 155, 157, 167, 178, 185, 188, 189, 193, 253, 420, 427, 428
Longenecker 154, 155, 466
Lord 16, 33
Lowe 69, 70, 72, 74, 75, 98, 101, 276, 277
Ludorf 196
Luft 154, 155, 176, 422
Luthans 59
Mabe 111, 115, 116, 118, 120, 121, 126, 127, 132, 133, 134, 136, 145, 147, 149, 151, 155, 158, 160, 205, 212, 256, 300, 315, 417, 421, 431, 435, 450
Malik 23, 25, 26, 28, 30, 33
Malloy 121
Marcus 96, 133
Markus 130, 147
Martin 56, 68, 72, 74
Martineau 113
Masi 68
Maslow 61
Mathieu 73
McCall 174, 181
McClelland 76, 246, 431
McCormick 19, 122
McEvoy 107, 108, 191
Meindl 21, 40, 47, 119, 449
Messick 122
Metzger 11
Meyer 104, 113, 122, 208, 331, 435, 451
Miller 27, 157
Mintzberg 92, 183
Morgan 65
Moser 125, 134, 136, 140, 203, 208, 289, 304, 321, 417, 418, 421, 431
Mount 106, 117, 139, 166, 212, 341, 422
Mummendey 156, 159, 256
Murphy 139, 140, 142, 159, 190, 417, 445
Nachreiner 96
Nagle 143
Nanus 49
Nemeth 64
Neuberger 16, 17, 18, 20, 28, 32, 33, 35, 38, 39, 40, 41, 42, 44, 52, 77, 78, 96, 97, 100, 102, 109, 114, 223, 238, 448
Nickel 107, 196, 211
Nilsen 120, 140, 174, 181
Nowack 168, 169
Nunnally 279
O'Connor 361
O'Connor-Wilson 113
Okazaki 121, 430
Oldham 59
Ostroff 79, 419, 432, 434
Otten 129, 443
Paglia 122
Parker 120
Pascale 29
Pascarella 37
Paulhus 156, 157
Pawar 44, 78
Perloff 122, 331
Peter 94, 159
Petersen 162
Pierce 443
Pillai 68, 72, 78
Podsakoff 59, 60
Pollack 109, 111
Pope 115
Porter 348
Pryor 130
Rai 69, 72, 73, 75
Ram 27
Randall 418
Rapee 121
Rathgeber 72, 74, 469
Raven 86
Reddin 37, 84
Reilly 183, 184, 185
Reitz 59
Rief 181
Rieker 104
Riggio 427, 454
Roberts 49, 69, 72, 75, 76, 78
Rose 77
Rosenfeld 159
Rosenstiel 23, 268
Rosenthal 248
Rosnow 248, 249

Ross 153
Rost 49, 91, 146
Rothstein 143, 204, 397
Roush 130
Runde 109
Runge 239
Sala 106, 420
Salgado 421, 458
Sanchez-Burks 129
Sarges 28
Sashkin 48, 49
Scandura 99
Schaller 428
Schein 128
Scherm 109, 110, 112, 113, 182, 384
Schlenker 135, 155, 156, 159
Schmidt 34, 107, 108, 196, 211
Scholz 50, 158
Schriesheim 37
Schuhmacher 281
Schuler 24, 114, 252, 450
Schultz-Hardt 182
Schwarz 467
Scullen 105, 113, 124, 138, 139, 147, 152, 154, 165, 166, 167, 243, 278, 302, 452
Sedikides 115
Seligman 60
Seltzer 56, 72, 74, 89, 90, 336
Senge 26
Senger 240, 250, 434, 435, 466
Shamir 42, 46, 51, 61, 64, 80, 85, 88
Shipper 106, 107
Shore 148
Shrauger 155, 162
Shrout 246, 278
Silverthorne 37
Sims 59
Smith 19, 21
Smither 112, 120, 182, 184, 186, 187, 189
Snyder 129
Sohm 43
Somers 208
Sonntag 25
Sosik 64, 74, 159, 171, 172, 427, 429
Sprenger 112
Staehle 23
Staufenbiel 25
Steers 448
Steiner 177, 178
Steinert 180
Steinmann 203
Stempfle 120, 168, 169, 421, 423, 467, 469
Stevens 159
Stewart 18
Stogdill 33
Stone-Romero 432
Strack 153
Strube 37
Sudman 145, 255, 256
Svenson 122
Svyantek 19
Swann 161
Tajfel 149, 198, 207
Tannenbaum 37, 84
Tartler 11, 73, 171, 172, 176, 214, 310, 311, 354, 374, 375, 383, 429, 433, 449, 454, 459
Taylor 122, 141, 156, 157, 158, 168, 184
Tejeda 99
Tesser 154
Thomas 19
Thornton 273
Tichy 24, 52, 66, 114
Tisak 289
Tisdale 461
Titscher 96
Torjus 454
Tornow 109, 112, 118, 174, 180, 193, 423
Tosi 67, 84, 86
Towler 81
Trice 69, 72, 75
Trost 255
Tsui 142, 146, 297
Turner 77, 149, 198, 207
Tversky 117
Twenge 158, 443
Van Velsor 118, 128, 129, 161, 171, 172, 174, 380, 405, 423, 430, 469
Vandenberghe 56, 69, 72, 100
VanVelsor 171
Vecchio 37
Viswesvaran 120, 297, 421
Voltz 252
Vroom 25, 26, 37, 38, 58, 84, 88, 91
Waldman 56, 68, 72, 73, 78, 79, 155, 182, 189
Walker 183, 186, 189, 450
Walton 122
Warech 129
Weber 42, 43, 44, 45, 48, 76, 78
Webster 292
Weibler 16, 17, 20, 21, 23, 24, 29, 32, 36, 37, 40, 83, 105, 193, 242
Weinert 268
Weinstein 122
Weise 250
Weiss 281
Welch 63
Wendel 11
Wexley 169, 171, 176, 179
Wherry 131, 138, 143
Wiendieck 104
Wilkens 11, 428
Wills 150
Wilson 47
Wirtz 278, 279
Witte 20
Wohlers 119, 130, 159, 212, 301, 420, 452

Wood 150
Wosinska 159
Wunderer 23, 28, 29, 32, 37, 87, 92, 194, 195, 251
Yammarino 98, 161, 172, 183, 184, 337, 455
Yukl 42, 63, 78, 80, 85, 86
Zacharatos 80
Zimolong 29, 171, 173, 176, 220, 455, 462

Anmerkung. Im Autorenindex sind alle Erstautoren verwendeter Quellen verzeichnet.

10 Abbildungsverzeichnis

Abbildung 1. Aufbau der Arbeit ... 13
Abbildung 2. Human Resource Management des Michigan-Ansatzes (Tichy et al., 1982) ... 24
Abbildung 3. Überblick über führungstheoretische Ansätze (modifiziert nach Wunderer, 2001, S. 274) ... 34
Abbildung 4. Prinzip der Eigenschaftstheorie der Führung (nach Neuberger, 2002) ... 34
Abbildung 5. Prinzip der Führungsstil-Modelle und beispielhafte Übertragung auf das Managerial Grid von Blake und Mouton (1964) ... 36
Abbildung 6. Entwicklungslinien der Führungsstilforschung (aus Weibler, 2001, S. 293) ... 38
Abbildung 7. Prinzip der klassischen Kontingenzansätze (nach Neuberger, 2002) und beispielhafte Übertragung auf das Reifegradmodell nach Hersey und Blanchard (1969) ... 39
Abbildung 8. Prinzip der attributionstheoretischen Betrachtung von Geführten- beziehungsweise Führungsverhalten (nach Neuberger, 2002) ... 42
Abbildung 9. „Full Range of Leadership“ (nach Bass & Avolio, 1994) ... 59
Abbildung 10. Augmentationshypothese transformationaler Führung ... 60
Abbildung 11. Transaktionale Führung und Anstrengung der Geführten (aus Bass, 1985, p. 12) ... 63
Abbildung 12. Transformationale Führung und Extra Effort der Geführten (aus Bass, 1985, p. 23) ... 67
Abbildung 13. Korrelation der transaktionalen und transformationalen Führungsdimensionen mit subjektiven und objektiven Erfolgsmaßen (nach Lowe, Kroeck & Sivasubramaniam, 1996, S. 410) ... 76
Abbildung 14. Skizzenhafte Darstellung: Die Situation moderiert das Auftreten und den Erfolg transformationaler Führung ... 86
Abbildung 15. Transaktionale und transformationale Führung im Führungsstilkontinuum (nach Wunderer, 2001, S. 248) ... 96
Abbildung 16. Multiperspektivische Führungsbeschreibung und ihre theoretischen Grundlagen – Überblick ... 112
Abbildung 18. Prototyp eines 360°-Feedbacks (gesamte Abbildung) und eines Sandwich-Modells (gepunktete Ellipse) (nach Scherm & Sarges, 2002) ... 119
Abbildung 19. Varianzanteile, die durch Ratee-Leistung, Rater-Bias bzw. Fehler aufgeklärt werden (unterschiedliche Perspektiven) (nach Scullen et al., 2000) ... 177
Abbildung 20. Rücklauf: Anzahl eingegangener Fragebogen-Versionen ... 243
Abbildung 21. Schematische Darstellung der personellen Überschneidung bei Beurteilungen nach dem Sandwich-Modell bei einer 270°-Beurteilung ... 256
Abbildung 22. Schematische Darstellung des Verteilungsprinzips der Fragebogen-Versionen (ohne Soll- ... 258
Abbildung 23. Beispielitem zu Arbeitszufriedenheit mit Skalenformat und –verankerung ... 283
Abbildung 24. FzFü-Beurteilung aus der Abwärts-, Selbst- und Aufwärtsperspektive (jeweils $N = 48$) ... 289
Abbildung 25. FzFü-Ideal-Beurteilung aus der Selbst- ($N = 25$) und Aufwärtsperspektive ($N = 17$) ... 290
Abbildung 26. Soll-Beschreibungen auf den sieben FzFü-Dimensionen ... 297
Abbildung 27. Korrelative Zusammenhänge zwischen Beurteilungsperspektiven (Ist), mittlere Korrelation der sieben FzFü-Faktoren ... 310
Abbildung 28. MLQ-Beurteilungen (Aufwärts) in Abhängigkeit von den Selbst-Fremd-Niveau-Unterschieden ... 327
Abbildung 29. Skizzierung der Selbst-Fremd-Diskrepanz ... 339
Abbildung 30. Schematische Darstellung der beiden Betrachtungsrichtungen ... 342

Abbildung 31. Differenz zwischen der Selbstbeurteilung der Führungskräfte und der Beurteilung des Führungsverhaltens anderer Führungskräfte (Vorgesetzte, Mitarbeiter) 343
Abbildung 32. Interaktion von Führung (Aufwärtsbeurteilung) und Wichtigkeit von Führung auf der Dimension B/S als Prädiktor von Leistung 379
Abbildung 33. Einfaktorielle Varianzanalysen für dreifach gestufte Selbst-Aufwärts-Diskrepanz (FzFü-Mittelwert): Starke Effekte ($f \geq .40$) in erwarteter Richtung 389
Abbildung 34. Einfaktorielle Varianzanalysen für dreifach gestufte Selbst-Abwärts-Diskrepanz (FzFü-Mittelwert): Bedeutsame Effekte ($f \geq .10$) in erwarteter Richtung 394
Abbildung 35. Einfaktorielle Varianzanalysen für dreifach gestufte Abwärts-Aufwärts-Diskrepanz (FzFü-Mittelwert): Starke Effekte ($f \geq .40$) in erwarteter Richtung 396
Abbildung 36. Mittlere Selbst-Aufwärts-Diskrepanz für über- und unterdurchschnittlich erfolgreiche Mitarbeiterteams bezüglich verschiedener Kriterien 412
Abbildung 37. Selbst-Aufwärts-Diskrepanz in Abhängigkeit von Erfahrungen des Aufwärtsbeurteilers in der Führungsrolle (eigene Führungsverantwortung bzw. Stellvertreterfunktion) 457

11 Tabellenverzeichnis

Tabelle 1. Charakterisierung von charismatischen Führungskräften (nach House & Shamir, 1995, S. 881ff.) ... 49
Tabelle 2. Charakteristika der verschiedenen Führungsbausteine ... 61
Tabelle 3. Vergleich transaktionaler und transformationaler Führung (modifiziert nach Tichy & Devanna, 1986a sowie Bass & Avolio, 1994) ... 71
Tabelle 4. Transformationale Führung und dadurch angestrebte Wirkung auf Seiten der Mitarbeiter (nach Tosi et al., 2000, S. 474) ... 72
Tabelle 5. Überblick über die Effektivität transformationaler Führung (modifiziert nach Rathgeber & Jonas, 2003) ... 78
Tabelle 6. Ähnlichkeit zwischen den verschiedenen, in dieser Arbeit (Kapitel 2.4 und 2.5) vorgestellten Führungstheorien (modifiziert nach Tosi et al., 2000, S. 472) ... 92
Tabelle 7. Unterschied zwischen Führung und Management (nach Rost, 1998) ... 100
Tabelle 8. Beispiel für eine pragmatische Führungsheuristik (nach Gebert, 1992) ... 102
Tabelle 9. Übersicht über den Zusammenhang zwischen Aufwärtsbeschreibung von Führung und Erfolgsvariablen ... 115
Tabelle 10. Übereinstimmung von Selbst- und Fremdratings (Abwärts, Aufwärts) zu (beruflicher) Leistung: Meta-Analytische Befunde ... 126
Tabelle 11. Meta-analytisch untersuchte Moderatorvariablen für die Übereinstimmung zwischen Selbstbeurteilung von Leistung und Außenkriterien (vgl. Tabelle 10) ... 136
Tabelle 12. Gegenüberstellung von Erklärungsmechanismen und Moderatorvariablen: Plausibilitäts-Passung ... 175
Tabelle 13. Zusammenhänge zwischen Übereinstimmung von Selbst- und Fremdbeschreibung der Führung und Erfolgsmaßen ... 182
Tabelle 14. Überblick über die untersuchten Moderatorvariablen ... 214
Tabelle 15. Untersuchte Variablen zur Charakterisierung erfolgreicher Mitarbeiter ... 235
Tabelle 16. Klassifikation der bearbeiteten Fragestellungen (modifiziert und erweitert nach Neuberger, 2000) ... 238
Tabelle 17. Stichprobencharakteristik ... 241
Tabelle 18. Subskalen des Fragebogens zur Führungsbeschreibung (Hauptteil) ... 244
Tabelle 19. Eindimensionale Skalen des Fragebogens zur Führungsbeschreibung ... 245
Tabelle 20. Subskalen der Transformationalen Führung und Beispielitems ... 249
Tabelle 21. Subskalen der transaktionalen Führung und Beispielitems ... 250
Tabelle 22. Subskala Laissez Faire und Beispielitem ... 250
Tabelle 23. Subskalen zu Erfolgsmaßen (MLQ) und Beispielitems ... 250
Tabelle 24. Erfolgsskalen und Beispielitems ... 251
Tabelle 25. Übersicht der eingesetzten Instrumente und der erfassten Konstrukte unter Angabe der jeweils einbezogenen Beurteilungsperspektiven ... 254
Tabelle 26. Abkürzungen der eingesetzten Skalen zur Messung von Führung und Erfolg ... 255
Tabelle 27: Übersicht über die verwendeten Maße der Effektstärke (nach Bortz & Döring, 2003; Dunlap et al., 1996) ... 263
Tabelle 28. Beispielitem in den drei Perspektiv-Versionen der Ist-Beurteilung ... 268
Tabelle 29. Inhaltliche Verankerung der Skalierung von 1 bis 7 für die vier Modus-Versionen ... 269
Tabelle 30. Überblick über die Versionen des Fragebogens zur Führungsbeschreibung ... 269
Tabelle 31. Faktorkennwerte der rotierten Lösung (Fragebogen zur Führungsbeschreibung) ... 273

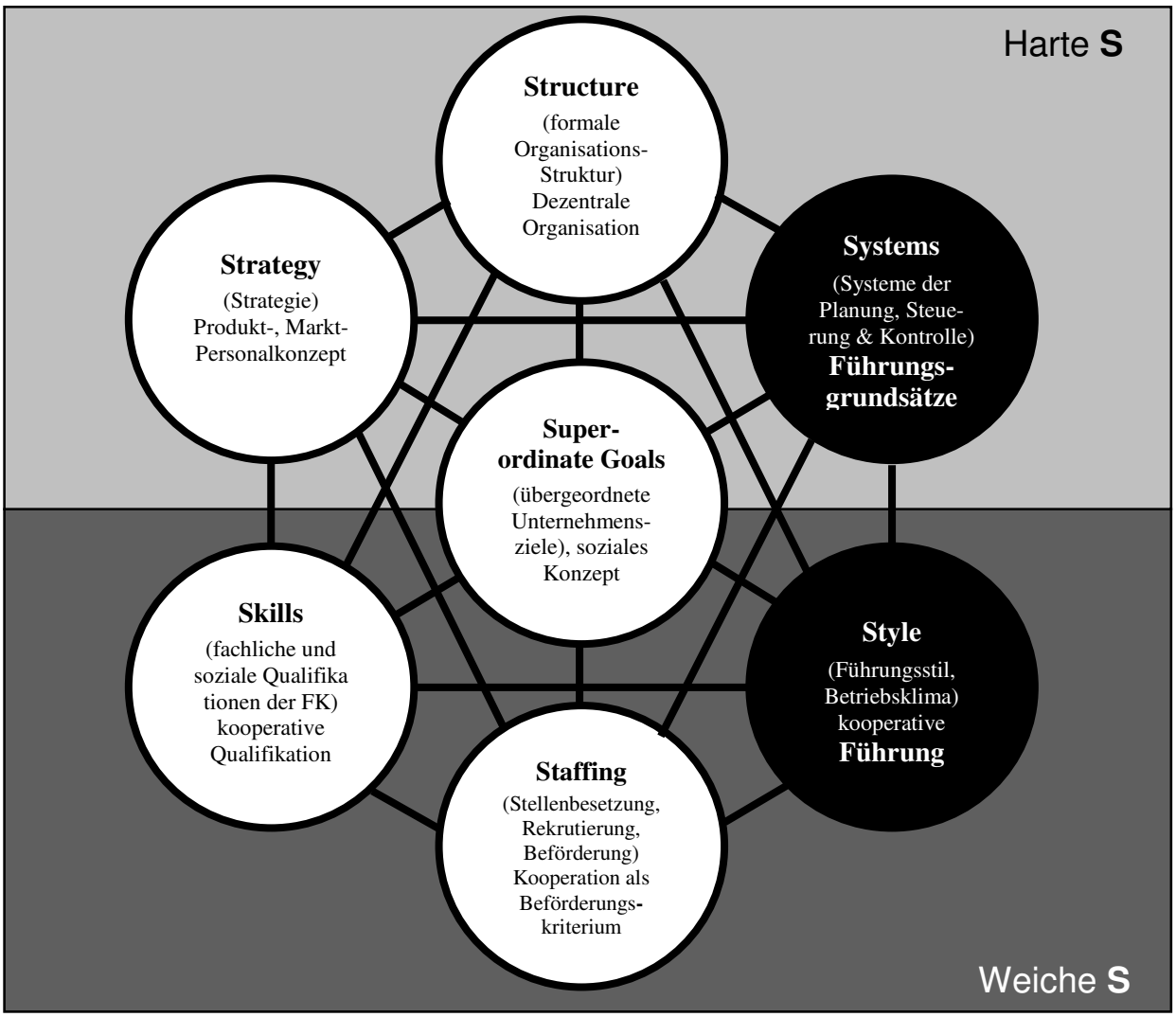
Tabelle 32. Faktorkennwerte der einfaktoriellen Skalen Fördern von Führungsverhalten und Stellvertretung 274
Tabelle 33. Beschreibung der erfassten Führungsdimensionen 275
Tabelle 34. Passung zwischen Führungsgrundsätzen und Führungsfaktoren des FzFü 277
Tabelle 35: Gegenüberstellung der FzFü-Dimensionen mit klassischen Führungsaufgaben 277
Tabelle 36: Korrelationen zwischen den Faktoren des FzFü und denen des MLQ 279
Tabelle 37. Skalen- und Itemkennwerte des MLQ 282
Tabelle 38. Faktor- und Skalenkennwerte der rotierten Lösung (Arbeitszufriedenheit) 284
Tabelle 39. Faktor- und Skalenwerte für die Skalen zu Wohlbefinden und Leistung 285
Tabelle 40. Skaleninterkorrelationen, Mittelwerte und Standardabweichungen der FzFü-Subskalen (Aufwärtsbeurteilung) 287
Tabelle 41. Skaleninterkorrelationen, Mittelwerte und Standardabweichungen der FzFü-Subskalen (Abwärts- und Selbstbeurteilung) 288
Tabelle 42. Skaleninterkorrelationen der MLQ-Subskalen, Mittelwerte, Standardabweichungen und Range 291
Tabelle 43. Skaleninterkorrelationen der MLQ-externen Erfolgsvariablen, Mittelwerte, Standardabweichungen und Range (N = 205) 293
Tabelle 44. Vergleich zwischen aktuellem und erwartetem Soll-Führungsverhalten: Mittelwerte, Standardabweichung und Signifikanzprüfung 297
Tabelle 45. Vergleich von aktuellem und idealem Führungsverhalten (Selbst und Aufwärts): Mittelwerte, Standardabweichung und Signifikanzprüfung 299
Tabelle 46. Vergleich von aktuellem (Abwärts) und idealem (Selbst und Aufwärts) Führungsverhalten: Mittelwerte, Standardabweichung und Unterschiedstestung 300
Tabelle 47. Mittelwertsvergleich zwischen Ideal (Aufwärts, Selbst) und Soll 302
Tabelle 48. Gegenüberstellung der Selbst-Aufwärts- bzw. Abwärts-Aufwärts-Korrelationen für aggregierte vs. individuelle Aufwärtsratings 305
Tabelle 49. Korrelative Zusammenhänge zwischen den Beurteilungsperspektiven der sieben FzFü-Dimensionen 306
Tabelle 50. Korrelationsunterschiede zwischen den Perspektiven 310
Tabelle 51. Niveau-Unterschied zwischen den Beurteilungsperspektiven (Ist-Beurteilung) 315
Tabelle 52. Klassifikation der Übereinstimmung: Selbst-Aufwärts und Selbst-Abwärts 318
Tabelle 53. Ergebnisse der Moderatoranalysen zur Selbst-Aufwärts-Korrelation 322
Tabelle 54. Ergebnisse der Moderatoranalysen zur Selbst-Aufwärts-Diskrepanz 324
Tabelle 55. Ergebnisse der Moderatoranalysen zur Selbst-Abwärts-Korrelation 329
Tabelle 56: Ergebnisse der Moderatoranalysen zur Selbst-Abwärts-Diskrepanz 332
Tabelle 57. Ergebnisse der Moderatoranalysen zur Abwärts-Aufwärts-Korrelation 334
Tabelle 58. Ergebnisse der Moderatoranalysen zur Abwärts-Aufwärts-Diskrepanz 335
Tabelle 59. Zusammenhang zwischen Korrelationen und dem Betrag der Niveau-Unterschiede 339
Tabelle 60. Korrelative Selbst-Aufwärts-Übereinstimmung der Unterschätzer, Übereinstimmer und Überschätzer 340
Tabelle 61. Above-Average-Effekt: Selbstbeurteilung im Vergleich zur Beurteilung der eigenen Vorgesetzten bzw. der unterstellten Mitarbeiter mit Führungsverantwortung (FV) 344
Tabelle 62. Vierfelder-Schema: Vergleich des Above-Average-Effekts nach oben vs. unten 346
Tabelle 63. Überblick über die Erfolgskriterien 350
Tabelle 64. Korrelativer Zusammenhang zwischen der Führungsbeschreibung (FzFü) und den Erfolgskriterien 352

Tabelle 65. Korrelative Zusammenhänge zwischen der Führungsbeschreibung (MLQ) und den Erfolgskriterien ... 355
Tabelle 66. Erklärte Varianz und aufgenommene Variablen in den verschiedenen Beurteilungsperspektiven (schrittweise Regression auf einzelne FzFü-Dimensionen) ... 363
Tabelle 67. Inkrementelle Validität der Beurteilungsperspektiven ... 366
Tabelle 68. Hierarchische Regressionsanalysen zur Vorhersage der Erfolgskriterien durch transaktionale und transformationale Führung (*N* = 90) ... 369
Tabelle 69. Hierarchische Regressionsanalysen zur Vorhersage der Erfolgskriterien durch transformationale, transaktionale und FzFü-Führung ... 373
Tabelle 70. Hierarchische moderierte Regression für verschiedene Erfolgsvariablen (AV) auf E/F-Ist und E/F-Wichtig (Aufwärts), *N* = 106 ... 376
Tabelle 71. Hierarchische moderierte Regression für EF-WOBE auf IB-Ist und IB-Wichtig (Aufwärts), *N* = 106 ... 377
Tabelle 72. Hierarchische moderierte Regression für EF-WOBE und L auf B/S-Ist und B/S-Wichtig (Aufwärts), *N* = 106 ... 378
Tabelle 73. Korrelativer Zusammenhang zwischen Diskrepanzen in der Führungsbeurteilung und den Erfolgskriterien ... 382
Tabelle 74. Mittelwertsunterschiede (Erfolgskriterien) in Abhängigkeit des Ausmaßes (A) bzw. der Richtung (R) der Selbst-Aufwärts-Diskrepanz, *N* = 40 ... 390
Tabelle 75. Übersicht über die mittlere Varianzaufklärung durch die verschiedenen Beurteilungsdiskrepanzen ... 397
Tabelle 76. Mittlere Selbst- und Fremdbeurteilungen des Führungsverhaltens (FzFü) für Überschätzer, Übereinstimmer und Unterschätzer ... 398
Tabelle 77. Mittlere Aufwärtsbeurteilung des Full Ranges of Leadership (MLQ) für Überschätzer, Übereinstimmer und Unterschätzer ... 398
Tabelle 78. Ergebnisse der Regressionsanalysen auf die einfachen Selbst- und Aufwärtsbeurteilungen sowie auf die Selbst-Aufwärts-Differenz, *N* = 40 ... 404
Tabelle 79. Inkrementelle Validität verschiedener Diskrepanzen, *N* = 24 ... 408
Tabelle 80. Erfolgskennwerte (Skalenmittelwert) in Abhängigkeit von der Selbst-Fremd-Diskrepanz *und* der Führungsqualität (transformationale Führung) ... 410
Tabelle 81. Subjektive Erklärungen der befragten Führungskräfte und Mitarbeiter für eine potentielle Selbst-Aufwärts-Diskrepanz ... 422
Tabelle 82. Überblick über die bearbeiteten Hypothesen ... 425
Tabelle 83. Selbst-Aufwärts-Korrelation der vorliegenden Studie (FzFü) im Vergleich zu anderen Studien mit multiperspektivischer Führungsbeurteilung ... 446

12 Anhang

A. Das 7S-Modell ... 526
B. Geltende Führungsgrundsätze im untersuchten Unternehmen ... 527
C. Multifactor Leadership Questionnaire (MLQ) ... 529
D. Faktorladungsmatrix des Fragebogens zur Führungsbeschreibung ... 532
E. Deskriptive Statistik und Skalen-Kennwerte der FzFü-Versionen auf Itemebene ... 535
F. Fragebogen zur Messung von Arbeitszufriedenheit, Leistung und Wohlbefinden sowie offene Fragen ... 548
G. FzFü-Instruktion ... 552
H. Faktorladungsmatrix und Trennschärfen zur Arbeitszufriedenheits-Skala ... 553
I. Range der FzFü-Skalenmittelwerte ... 554
J. Deskriptive Statistik der Perspektiv-Diskrepanzen ... 555
K. Kodierung für Moderatorvariable „Ergebnisbezogene Itemformulierung“ ... 556
L. Korrelative interperspektivische Zusammenhänge in Abhängigkeit von den absoluten interperspektivischen Niveau-Unterschieden ... 557
M. Korrelation zwischen dimensionalen Führungsbeurteilungen (FzFü, verschiedene Richtungen) und Erfolgskriterien ... 558
N. Korrelation zwischen dimensionaler Führungsbeurteilung (MLQ) und Erfolgskriterien ... 560
O. Ergebnisse der blockweisen Regressionen (1. Block: Transformationale Führung, 2. Block: FzFü) ... 561
P. Hierarchische moderierte Regressionsanalysen für Führungserfolg auf Führungsverhalten und Wichtigkeit. ... 562
Q. Hierarchische moderierte Regressionsanalyse für Satisfaction auf Fördern von Führungsverhalten und transformationale Führung ... 563
R. Korrelative Zusammenhänge zwischen den interperspektivischen Diskrepanzen auf den einzelnen FzFü-Dimensionen und den Erfolgsvariablen ... 564
S. Zusammenhang zwischen interperspektivischen Korrelationen und Erfolgskriterien. ... 566
T. Mittlere Selbst- und Fremdbeurteilungen des Führungsverhaltens (FzFü) in Abhängigkeit vom Ausmaß der Abwärts-Aufwärts-Abweichung (Distanzmaß) ... 567
U. Erfolgskennwerte (Skalenmittelwert) in Abhängigkeit von der Selbst-Fremd-Diskrepanz und der Führungsqualität (FzFü) ... 568
V. Selbst-Fremd- und Fremd-Fremd-Korrelationen in Abhängigkeit vom Erfolg ... 569
W. Prädiktive Validität von Führung in Abhängigkeit der Selbst-Fremd- und Fremd-Fremd-Abweichung ... 573
X. Korrelationen zwischen der Selbstbeurteilung und der Beurteilung des Führungsverhaltens anderer Führungskräfte ... 578

A. Das 7S-Modell



Anmerkung. Modell von Pascale und Athos (1981), Abbildung in Anlehnung an Wunderer (2001, S. 387).

B. Geltende Führungsgrundsätze im untersuchten Unternehmen

1. Qualifizieren der Mitarbeiter für ihre Aufgaben (Fähigkeiten entwickeln)

Aufgabe des Vorgesetzten ist es, seine Mitarbeiter

- für die von ihnen wahrzunehmenden Aufgaben anzuleiten sowie ihre berufliche Leistungsfähigkeit zu erhalten und zu verbessern (fachliche Qualifikation),
- zu selbständigem Arbeiten, sachlichem Zusammenarbeiten und – soweit Führungsaufgaben wahrzunehmen sind – kooperativem Führen anzuhalten (überfachliche Qualifikation).

Von allen Mitarbeitern wird Bereitschaft zu weiterer Qualifizierung erwartet.

2. Qualifizieren für höherwertige Aufgaben (Mitarbeiter fördern)

Ziel der Förderung ist es, den Mitarbeiter so einzusetzen, dass seine Fähigkeiten für ihn und für das Unternehmen am besten zur Wirkung kommen können.
Berufliche Möglichkeiten des Mitarbeiters werden zwischen ihm und seinem Vorgesetzten erörtert.

3. Aufgaben und Befugnisse übertragen (Delegieren)

Delegieren ist das Übertragen von Aufgaben und der zu ihrer Lösung erforderlichen Entscheidungsbefugnisse.
Der Vorgesetzte hat die Pflicht zu delegieren, um Führungsaufgaben vorrangig wahrzunehmen.
Die Aufgaben und Befugnisse sind den Mitarbeitern durch den Vorgesetzten zu erläutern und überschaubar zu machen.
Durch die Delegation übernimmt der Mitarbeiter im Rahmen der betrieblichen Möglichkeiten die Verantwortung für die Ausführung der ihm übertragenen Aufgaben.

4. Stellvertretung

Der Vorgesetzte ist für den Fall seiner Abwesenheit verpflichtet, seinen jeweiligen Stellvertreter zu benennen und ihn mit der Wahrnehmung der Stellvertretung zu beauftragen.
Stellvertreter handeln im Rahmen der ihnen übertragenen Aufgaben und Befugnisse eigenverantwortlich.

5. Zusammenarbeiten

Vertrauensvolle und gute Zusammenarbeit zeigt sich in gegenseitiger offener Information und Unterstützung. Vorgesetzte und Mitarbeiter informieren sich gegenseitig über Sachverhalte und betriebliche Zusammenhänge so vollständig, dass sie handeln und entscheiden können.
Der Vorgesetzte unterstützt seine Mitarbeiter bei der Erfüllung ihrer Aufgaben und gibt ihnen Rat und Anregungen. Er soll seine Mitarbeiter zu selbständigem Urteilen und zu Anregungen ermuntern sowie für entsprechenden Informationsfluss sorgen.
Jeder Mitarbeiter soll sich zur Verbesserung seiner Arbeitsergebnisse die Erfahrungen und Kenntnisse anderer nutzbar machen.
Informationen, die andere Mitarbeiter zur Erfüllung ihrer Aufgabe brauchen, sind im Rahmen der gegebenen Befugnisse weiterzugeben, um die allseitige Zusammenarbeit im Unternehmen zu fördern.

6. Ziele erarbeiten und setzen

Der Vorgesetzte setzt seinen Mitarbeitern Ziele, indem er die Aufgaben nach Umfang, Qualität und Termin erläutert.
Dazu gehört auch die Entscheidung über Prioritäten.
Er berücksichtigt bei der Erarbeitung der Ziele übergeordnete Zusammenhänge sowie die Kenntnisse und Erfahrungen der jeweils zuständigen Mitarbeiter. Er beteiligt sie an der Erarbeitung der Ziele im größtmöglichen Umfang.
Den Mitarbeitern ist der Gesamtzusammenhang bewusst zu machen. Dies gilt vor allem für Ziele, bei deren Entwicklung eine Beteiligung der Mitarbeiter nicht möglich war.

7. Arbeitsergebnisse, Arbeits- und Führungsverhalten beurteilen

Arbeitsergebnisse und Arbeitsverhalten der Mitarbeiter werden beurteilt. Hat der Mitarbeiter Führungsaufgaben, so wird auch sein Führungsverhalten beurteilt.
Maßstab für die Beurteilung können nur dem Mitarbeiter bekannte Vorgaben (gesetzte Ziele, geltende Vorschriften und Regelungen) sein. Jede Beurteilung setzt eine Prüfung der Sachverhalte und Einflussfaktoren voraus. Der Mitarbeiter wird dazu gehört. Die Beurteilung ist dem Mitarbeiter bekanntzugeben; ihm die Gelegenheit zur Stellungnahme zu geben.

8. Arbeitsergebnisse, Arbeits- und Führungsverhalten steuern

Die Beurteilung ist Grundlage für die Steuerung des künftigen Arbeits- und Führungsverhaltens.
Die Steuerung besteht darin, dass der Vorgesetzte Arbeitsergebnisse, Arbeits- und Führungsverhalten bestätigt oder notwendige Korrekturen vornimmt.

C. Multifactor Leadership Questionnaire (MLQ)

Instruktion: In diesem Fragebogen geht es um den Führungsstil Ihrer direkten Führungskraft. Bitte beantworten Sie alle Fragen zügig und vertrauen Sie dabei Ihrem spontanen Urteil.

Schätzen Sie ein, wie häufig folgende Aussagen auf Ihre direkte Führungskraft zutreffen. Verwenden Sie dazu bitte die fünf-stufige Skala am rechten Seitenrand.

Item-Nr.	**Itemwortlaut: Meine direkte Führungskraft ...**
IIa – Idealized Influence (attributed)	
5	
f10	Macht mich stolz darauf, mit ihr zu tun zu haben.
f18	stellt die eigenen Interessen zurück, wenn es um das Wohl der Gruppe geht.
f21	handelt in einer Weise, die bei mir Respekt erzeugt.
f25	Strahlt Stärke und Vertrauen aus.
IIb – Idealized Influence (behavior)	
f6	spricht mit anderen über ihre wichtigsten Überzeugungen und Werte.
f14	Macht klar, wie wichtig es ist, sich 100%ig für eine Sache einzusetzen.
f23	berücksichtigt die moralischen und ethischen Konsequenzen von Entscheidungen.
f34	Betont die Wichtigkeit von Teamgeist und einem gemeinsamen Aufgabenverständnis.
IM – Inspirational Motivation	
f9	Äußert sich optimistisch über die Zukunft.
f13	spricht mit Begeisterung über das, was erreicht werden soll.
f26	formuliert eine überzeugende Zukunftsvision.
f36	hat großes Vertrauen, dass die gesteckten Ziele erreicht werden.
IS –Intellectual Stimulation	
f2	überprüft stets aufs neue, ob zentrale/wichtige Annahmen noch angemessen sind.
f8	sucht bei der Lösung von Problemen nach unterschiedlichen Perspektiven.
f30	Bringt mich dazu, Probleme aus verschiedenen Blickwinkeln zu betrachten.
f32	schlägt neue Wege vor, wie Aufgaben/Aufträge bearbeitet werden können.

IC – Individualized Consideration	
f15	verbringt Zeit mir Führung und damit, den Mitarbeitern etwas beizubringen.
f19	berücksichtigt meine Individualität und behandelt mich nicht nur als einen Mitarbeiter unter vielen.
f29	erkennt meine individuellen Bedürfnisse, Fähigkeiten und Ziele.
f31	hilft mir, meine Stärken auszubauen.

CR – Contingent Reward	
f1	bietet im Gegenzug für meine Anstrengungen ihre Hilfe an.
f11	macht deutlich, wer für bestimmte Leistungen verantwortlich ist.
f16	spricht klar aus, was man erwarten kann, wenn die gesteckten Ziele erreicht worden sind.
f35	zeigt Zufriedenheit, wenn andere die Erwartungen erfüllen.

MBA – Management by exception (active)	
f4	konzentriert sich überwiegend auf Unregelmäßigkeiten, Fehler, Ausnahmen und Abweichungen von Vorschriften.
f22	kümmert sich in erster Linie um Fehler und Beschwerden.
f24	erfolgt alle Fehler konsequent.
f27	macht mich auf Fehler aufmerksam, damit die Anforderungen erfüllt werden.

MBP – Management by exception (passive)	
f3	versäumt es, sich um die Probleme zu kümmern, bis sie wirklich ernst geworden sind.
f12	wartet bis etwas schief gegangen ist, bevor sie etwas unternimmt.
f17	ist fest davon überzeugt, dass man ohne Not nichts ändern sollte.
f20	vertritt die Ansicht, dass Probleme erst wiederholt auftreten müssen, bevor man handeln sollte.

LF – Laissez Faire	
f7	ist immer da, wenn sie gebraucht wird. (-)
f28	trifft schnell und ohne Zögern seine Entscheidungen. (-)
f33	klärt wichtige Fragen sofort. (-)

EEF – Extra Effort	
f39	bringt mich dazu, mehr zu schaffen als ich ursprünglich erwartet habe.
f42	spornt mich an, erfolgreich zu sein.
f44	erhöht meine Bereitschaft, mich stärker anzustrengen.

EFF – Effectiveness	
f37	setzt sich effektiv für meine (beruflichen) Bedürfnisse und Interessen ein.
f40	kann meine Interessen gut bei höheren Vorgesetzten vertreten.
f43	setzt sich effektiv für die Belange der Organisation ein.
f45	kann eine Gruppe effektiv führen.

SAT – Satisfaction	
f38	sorgt durch ihr Führungsverhalten für Zufriedenheit.
f41	gestaltet die Zusammenarbeit so, dass ich wirklich zufrieden bin.

AUS – Ausstrahlung	
f46	ist für mich so wichtig, dass ich den Kontakt zu ihr suche/pflege.
f47	vermag mich durch ihre Persönlichkeit zu beeindrucken und zu faszinieren.
f48	versteht es, mich immer wieder zu begeistern.
f49	verfügt über Eigenschaften und Fähigkeiten, die ich bewundere.

D. Faktorladungsmatrix des Fragebogens zur Führungsbeschreibung

Tabelle D-1. Hauptteil des Fragebogens

Item-Nr.	Meine direkte Führungskraft ...	Kom[2]	Ladung auf Faktor[1] f1	f2	f3	f4	f5
21/2	fördert mich auf die Gefahr hin, dass ich irgendwann die Abteilung verlasse.	.69	**.69**				
6/2	plant mit mir meine beruflichen Möglichkeiten und Entwicklungsschritte.	.77	**.69**	.41			
9/2	erkennt recht treffend, zu welcher Höherqualifizierung ich motiviert bin.	.71	**.65**		.41		
1/2	leitet Förderungsmaßnahmen ein, die auf mich persönlich zugeschnitten sind.	.69	**.65**			.40	
19/2	regt mich dazu an, mich über meine derzeitige Tätigkeit hinaus breiter zu informieren und zu qualifizieren.	.72	**.62**	.40			
4/2	fördert mich in einer Art und Weise, die sowohl mir persönlich als auch dem Unternehmen nützt.	.68	**.60**		.42		
25/2	fragt mich immer wieder nach meinen Vorstellungen über meine berufliche Zukunft.	.70	**.59**	.46			
20/1	führt mich Schritt für Schritt an umfangreiche Aufgaben heran.	.70	**.58**	.41			
12/1	motiviert mich dazu, mich selbständig weiterzuqualifizieren.	.66	**.58**				.41
31/1	bietet mir ausreichend Möglichkeiten, mich fachlich weiterzuqualifizieren.	.55	**.56**				.35
32/2	hat ein Gespür für fachliche Trends, so dass sie sich rechtzeitig um meine entsprechende Förderung kümmern kann.	.64	**.54**			.43	
10/2	kann genau einschätzen, wozu ich in Zukunft fähig sein werde.	.64	**.54**		.43		
18/8	erläutert mir, welches Verhalten ich verändern soll.	.66		**.73**			
27/8	erklärt mir, weshalb sie möchte, dass ich ein bestimmtes Verhalten verändere.	.67		**.70**			
22/8	verdeutlicht mir, wie sich die letzte Beurteilung auf unsere künftige Zusammenarbeit auswirkt.	.68		**.69**			
28/3	erinnert mich gleichzeitig an meine Entscheidungsbefugnisse, wenn sie Aufgaben an mich delegiert.	.68		**.64**	.42		
17/7	informiert mich über den Anlass einer Beurteilung.	.56		**.63**			
29/7	will von mir wissen, wie treffend ich die Beurteilung meiner Leistung fand.	.70		**.62**			
24/7	erläutert mir die Abweichungen zwischen den erreichten Ergebnissen und den Zielvorgaben.	.65		**.57**			
38/7	setzt mich über die Kriterien, nach denen sie mich beurteilt, in Kenntnis, bevor ich mit meiner Arbeit beginne.	.65		**.57**			.42
11/1	erläutert mir, welche Fähigkeiten und Fertigkeiten sie zukünftig von mir erwartet.	.65	.43	**.54**			
23/7	beurteilt mich anhand messbarer und nachvollziehbarer Kriterien.	.58		**.48**			

2/3	gibt möglichst alle Aufgaben, die ich selbst erledigen kann, auch an mich weiter.	.70			**.73**		
15/3	bietet mir angemessene Handlungsfreiheit.	.63			**.69**		
5/7	lässt persönliche Zu- oder Abneigung bei meiner Leistungsbeurteilung außen vor.	.57			**.60**		
14/7	kann die fachliche Qualität meiner Leistungen richtig einschätzen.	.60	.37		**.58**		
3/7	beurteilt mich so, dass ich mich im Vergleich zu meinen Kollegen gerecht behandelt fühle.	.64	.42		**.57**		
13/1	setzt mich so ein, dass ich auch langfristig in der Lage bin, optimale Leistung zu zeigen.	.66	.46		**.55**		
16/6	berücksichtigt meine Kenntnisse und Erfahrungen bei der Erarbeitung von Zielen.	.51			**.44**		.41
26/3	achtet darauf, dass ich ausreichend rechtliche Vollmachten habe, um meine Aufgaben gut erledigen zu können.	.53		.38	**.44**		
35/5	sorgt für einen funktionierenden Informationsfluss in unserer Abteilung.	.76				**.70**	
33/8	leitet rechtzeitig – bevor etwas schief läuft – die notwendigen Schritte zur Korrektur ein.	.69	.37			**.65**	
8/5	stellt sicher, dass ich die nötigen Informationen für die Ausführung meiner Aufgaben habe.	.70				**.64**	
34/6	sorgt dafür, dass ich die zwischen uns vereinbarten Ziele bis ins Detail kenne.	.72				**.63**	
39/3	gibt Aufgaben verständlich und inhaltlich korrekt an mich weiter.	.68				**.61**	
42/6	bespricht Zielsetzungen mit mir so lange, bis alle Unklarheiten beseitigt sind.	.72		.39	.36	**.55**	
7/6	erläutert mir auch die Unternehmensziele, so dass ich verstehe, wozu ich etwas tun soll.	.65	.40			**.53**	
41/5	erfragt meine Ideen und Argumente, wenn sie weitreichende Entscheidungen trifft.	.78					**.79**
37/5	gestaltet unsere Zusammenarbeit so, dass ich auch ungewöhnliche Ideen äußern kann.	.70					**.62**
36/6	verbreitet soviel Enthusiasmus, dass ich meine Zielvorgaben wirklich akzeptiere.	.73				.46	**.56**
40/7	will wissen, wie ich meine eigene Leistung einschätze.	.70		.53			**.54**
30/6	beteiligt mich möglichst stark bei der Festlegung der Ziele, die ich im folgenden Zeitraum erreichen soll.	.74				.45	**.53**
	Eigenwerte der rotierten Faktoren		6.93	6.67	5.32	5.11	4.00
	Durch Faktoren aufgeklärter Varianzanteil		16.50	15.87	12.67	12.16	9.53

Anmerkungen. Extraktionsmethode: Hauptkomponentenanalyse, Rotationsmethode: Varimax mit Kaiser-Normalisierung.
[1] Ladungen größer .35, [2] Kommunalitäten, f1: Entwickeln/Fördern, f2: Beurteilen/Steuern, f3: Individuelle Berücksichtigung, f4: Klarheit, f5: Partizipation.

Tabelle D-2. Zusatzdimension „Fördern von Führungsverhalten“ (FvF)

Item-Nr.	Meine direkte Führungskraft ...	Faktor-Ladung
44/1	Sieht es als ihre Aufgabe, auch mein Führungsverhalten zu entwickeln.	**.93**
45/1	hält mich immer wieder zu kooperativem Führen an.	**.88**
46/1	verdeutlicht mir beispielhaft an ihrem eigenen Verhalten, was kooperatives Führen bedeutet.	**.81**
43/7	beurteilt mich auch nach dem Führungsverhalten, das ich meinen Mitarbeitern gegenüber zeige.	**.76**

Anmerkung. Extraktionsmethode: Hauptkomponentenanalyse, 1 Komponente extrahiert.

Tabelle D-3. Zusatzdimension „Stellvertretung“ (S)

Item-Nr.	Meine direkte Führungskraft ...	Faktor-Ladung
48/4	überträgt mir die Verantwortung für Entscheidungen, die ich während meiner Abwesenheit treffe.	**.90**
49/4	steht hinter mir als ihrem Stellvertreter.	**.88**
50/4	bespricht mit mir alle wichtigen Aufgaben, bevor sie in Urlaub geht.	**.85**
47/4	arbeitet mich in Aufgaben ein, die mir nicht geläufig sind.	**.77**

Anmerkung. Extraktionsmethode: Hauptkomponentenanalyse, 1 Komponente extrahiert.

E. Deskriptive Statistik und Skalen-Kennwerte der FzFü-Versionen auf Itemebene

Tabelle E-1. Faktor: Entwickeln und Fördern (E/F), Ist-Version

		Abwärts					Selbst					Aufwärts				
Nr.	Meine direkte Führungskraft …	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*
1/2	leitet Förderungsmaßnahmen ein, die auf mich persönlich zugeschnitten sind.	4.86	.98	.62	-.15	-.85	4.96	1.09	.54	-1.04	.35	3.82	1.56	.76	-.30	-1.02
4/2	fördert mich in einer Art und Weise, die sowohl mir persönlich als auch dem Unternehmen nützt.	5.36	.89	.74	-.64	.01	5.44	.84	.59	-.01	-.54	4.75	1.52	.75	.17	-.17
6/2	plant mit mir meine beruflichen Möglichkeiten und Entwicklungsschritte.	4.48	1.05	.65	.30	-.69	4.42	1.23	.64	.05	.21	3.85	1.75	.82	-.15	-1.08
9/2	erkennt recht treffend, zu welcher Höherqualifizierung ich motiviert bin.	5.07	.93	.70	-.44	-.49	5.29	.90	.55	-.49	3.08	4.47	1.40	.77	-.64	-.00
10/2	kann genau einschätzen, wozu ich in Zukunft fähig sein werde.	5.09	.83	.68	-.13	-.16	5.33	.90	.44	-.20	-.06	4.79	1.31	.70	-.63	.02
12/1	motiviert mich dazu, mich selbständig weiterzuqualifizieren.	4.91	.86	.74	.18	-.52	5.27	.94	.39	-.50	.08	4.31	1.59	.73	-.45	-.50
19/2	regt mich dazu an, mich über meine derzeitige Tätigkeit hinaus breiter zu informieren und zu qualifizieren.	5.09	1.03	.76	-.30	-.90	5.18	.94	.41	-.65	.65	4.05	1.57	.82	-.26	-.71
20/1	führt mich Schritt für Schritt an umfangreiche Aufgaben heran.	5.14	1.00	.72	-.68	.21	5.29	1.10	.63	-1.85	4.62	4.09	1.64	.80	-.39	-.79
21/2	fördert mich auch auf die Gefahr hin, dass ich irgendwann die Abteilung verlasse.	5.18	1.19	.78	-.35	-.64	5.71	1.38	.65	-1.74	3.18	4.05	1.82	.74	-.17	-1.09
25/2	fragt mich immer wieder nach meinen Vorstellungen über meine berufliche Zukunft.	4.55	.90	.53	-.15	-.71	4.64	1.25	.46	-.65	.27	3.20	1.67	.77	.20	-1.02
31/1	bietet mir ausreichend Möglichkeiten, mich fachlich weiterzuqualifizieren.	5.30	.88	.62	.17	-.54	5.36	1.17	.56	-.37	.04	4.53	1.65	.69	-.61	-.55

32/2	hat ein Gespür für fachliche Trends, so dass sie sich rechtzeitig um meine entsprechende Förderung kümmern kann.	5.16	.89	.61	-.30	-.54	5.07	.96	.68	-.34	-.82	4.46	1.48	.72	-.63	-.05
	Cronbachs Alpha	.92					.86					.95				

Anmerkung. Für die Tabellen in Anhang E gilt: *N* vgl. Abbildung 20 (wo nicht anders vermerkt), *M* = Mittelwert, *SD* = Standardabweichung, *TS* = Trennschärfe, *Sch* = Schiefe, *Ex* = Exzess.

Tabelle E-2. Faktor: Beurteilen und Steuern (B/S), Ist-Version

		Abwärts					Selbst					Aufwärts				
Nr.	Meine direkte Führungskraft …	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*
17/7	informiert mich über den Anlass einer Beurteilung.	5.07	.77	.52	.21	-.45	5.16	1.22	.56	-1.08	1.24	4.44	1.65	.69	-.51	-.63
18/8	erläutert mir, welches Verhalten ich verändern soll.	5.23	.78	.46	-.50	.57	5.41	.92	.59	-.54	.83	4.51	1.42	.69	-.54	-.42
22/8	verdeutlicht mir, wie sich die letzte Beurteilung auf unsere künftige Zusammenarbeit auswirkt.	4.95	.90	.70	-.49	-.52	4.45	1.34	.72	-.20	-.20	3.97	1.59	.78	-.16	-.76
23/7	beurteilt mich anhand messbarer und nachvollziehbarer Kriterien.	5.02	.99	.66	-.22	-.88	5.09	1.33	.73	-1.09	1.83	4.72	1.56	.71	-.71	-.27
24/7	erläutert mir die Abweichungen zwischen den erreichten Ergebnissen und den Zielvorgaben.	5.40	.82	.55	-.90	2.08	5.27	1.11	.55	-.97	1.10	4.62	1.52	.76	-.64	-.27
27/8	erklärt mir, weshalb sie möchte, dass ich ein bestimmtes Verhalten verändere.	5.14	1.01	.62	-.89	.15	5.50	1.11	.67	-1.18	1.68	4.36	1.56	.75	-.43	-.69
28/3	erinnert mich gleichzeitig an meine Entscheidungsbefugnisse, wenn sie Aufgaben an mich delegiert.	5.16	1.04	.60	-.27	-.78	5.16	1.18	.47	-.83	.36	4.44	1.55	.73	-.56	-.43
29/7	will von mir wissen, wie treffend ich die Beurteilung meiner Leistung fand.	4.70	.83	.63	-.14	-.49	4.34	1.41	.49	.01	-.77	3.85	1.63	.78	.01	-.77
38/7	setzt mich über die Kriterien, nach denen sie mich beurteilt, in Kenntnis, bevor ich mit meiner Arbeit beginne.	4.81	.96	.46	.08	-.52	4.64	1.18	.60	-.35	-.63	4.21	1.51	.71	-.35	-.63
11/1	erläutert mir, welche Fähigkeiten und Fertigkeiten sie von mir zukünftig erwartet.	5.28	.93	.59	-.53	.12	5.39	1.10	.57	-.70	1.09	4.55	1.56	.70	-.31	-.50
	Cronbachs Alpha	.87					.87					.93				

Tabelle E-3. Faktor: Individuelle Berücksichtigung (IB), Ist-Version

		Abwärts					Selbst					Aufwärts				
Nr.	Meine direkte Führungskraft …	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*
2/3	gibt möglichst alle Aufgaben, die ich selbst erledigen kann, auch an mich weiter.	5.60	.99	.70	-.62	-.17	5.74	.98	.36	-.92	1.20	5.69	1.22	.62	-1.49	2.37
3/7	beurteilt mich so, dass ich mich im Vergleich zu meinen Kollegen gerecht behandelt fühle.	5.55	.75	.74	-.36	-.23	5.57	.93	.27	-1.72	3.69	5.22	1.43	.71	-1.05	.80
5/7	lässt persönliche Zu- oder Abneigung bei meiner Leistungsbeurteilung außen vor.	5.40	1.12	.62	-.43	-.53	5.93	.88	.33	-.96	1.55	5.40	1.40	.66	-1.35	1.74
13/1	setzt mich so ein, dass ich auch langfristig in der Lage bin, optimale Leistung zu zeigen.	5.43	.83	.77	-.03	-.45	5.59	.62	.23	-.05	-.19	5.21	1.25	.70	-1.02	.93
14/7	kann die fachliche Qualität meiner Leistungen richtig einschätzen.	5.70	.83	.58	-.57	-.02	5.91	.73	.48	-.63	.96	5.34	1.40	.72	-1.23	1.39
15/3	bietet mir angemessene Handlungsfreiheit	5.60	.95	.68	-.69	.89	6.07	.68	.35	-.97	2.68	5.64	1.26	.68	-1.40	1.82
16/6	berücksichtigt meine Kenntnisse und Erfahrungen bei der Erarbeitung von Zielen.	5.64	.85	.66	-.34	-.33	5.96	.94	.22	-1.96	6.59	5.14	1.36	.64	-.81	.06
26/3	achtet darauf, dass ich ausreichend rechtliche Vollmachten habe, um meine Aufgaben gut erledigen zu können.	5.17	1.01	.38	-.48	.79	5.59	1.15	.39	-1.21	1.62	4.70	1.57	.64	-.76	-.14
	Cronbachs Alpha	.87					.62					.89				

Tabelle E-4. Faktor: Klarheit (K), Ist-Version

		Abwärts					Selbst					Aufwärts				
Nr.	Meine direkte Führungskraft …	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*
7/6	erläutert mir auch die Unternehmensziele, so dass ich verstehe, wozu ich etwas tun soll.	5.63	1.00	.68	-.76	.48	6.09	.75	.53	-.47	-.04	5.24	1.55	.70	-1.26	.93
8/5	stellt sicher, dass ich die nötigen Informationen für die Ausführung meiner Aufgaben habe.	5.56	.77	.77	-1.11	1.77	5.77	.79	.48	-.39	-.06	4.93	1.48	.79	-.88	.16
33/8	leitet rechtzeitig – bevor etwas schief läuft – die notwendigen Schritte zur Korrektur ein.	5.52	.99	.64	-.27	-.35	5.51	.78	.31	-.87	1.36	5.13	1.42	.74	-.99	.74
34/6	sorgt dafür, dass ich die zwischen uns vereinbarten Ziele bis ins Detail kenne.	5.54	.80	.72	-.54	-.23	5.40	.83	.50	-.62	.49	4.83	1.44	.78	-.73	-.13
35/5	sorgt für einen funktionierenden Informationsfluss in unserer Abteilung.	5.67	1.00	.80	-.75	.61	5.60	.90	.62	-.74	1.30	4.78	1.66	.79	-.66	-.42
39/3	gibt Aufgaben verständlich und inhaltlich korrekt an mich weiter.	5.58	.85	.72	-.16	-.47	5.62	.77	.50	-.10	-.31	4.83	1.44	.78	-1.30	1.88
42/6	bespricht Zielsetzungen mit mir so lange, bis alle Unklarheiten beseitigt sind.	5.40	.98	.70	-.47	.02	5.57	.90	.46	-.91	1.48	4.76	1.52	.78	-.63	-.30
	Cronbachs Alpha	.90					.77					.92				

Tabelle E-5. Faktor: Partizipation (P), Ist-Version

		Abwärts					Selbst					Aufwärts				
Nr.	Meine direkte Führungskraft …	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*
30/6	beteiligt mich möglichst stark bei der Festlegung der Ziele, die ich im folgenden Zeitraum erreichen soll.	5.38	.83	.49	-.32	.64	5.33	.81	.12	-1.96	5.94	4.65	1.63	.77	-.62	-.54
36/6	verbreitet soviel Enthusiasmus, dass ich meine Zielvorgaben wirklich akzeptiere.	5.36	.88	.60	-.28	-.05	5.33	.86	.25	-.72	.87	4.68	1.47	.76	-.55	-.12
37/5	gestaltet unsere Zusammenarbeit so, dass ich auch ungewöhnliche Ideen äußern kann.	5.38	1.09	.67	-1.0	.11	5.81	.98	.26	-1.15	3.37	5.16	1.47	.75	-.98	.45
40/7	will wissen, wie ich meine eigene Leistung einschätze.	4.80	.92	.44	-.27	-.71	5.00	.99	.60	-.55	.99	3.88	1.58	.64	-.19	-.70
41/5	erfragt meine Ideen und Argumente, wenn sie weitreichende Entscheidungen trifft.	5.24	.98	.58	-.81	-.05	5.50	1.07	.38	-1.14	1.85	4.79	1.65	.73	-.67	-.51
	Cronbachs Alpha	.78					.56					.89				

Tabelle E-6. Faktor: Fördern von Führungsverhalten (FvF), Ist-Version

		Abwärts					Selbst					Aufwärts				
Nr.	Meine direkte Führungskraft …	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*
43/7	beurteilt mich auch nach dem Führungsverhalten, das ich meinen Mitarbeitern gegenüber zeige.	4.68	1.36	.61	-1.15	1.29	5.16	1.52	.53	-1.50	2.01	5.13	1.27	.59	-.79	.63
44/1	sieht es als ihre Aufgabe, auch mein Führungsverhalten zu entwickeln.	4.74	1.13	.72	-.61	-.59	5.43	1.26	.26	-1.38	2.10	4.82	1.51	.84	-.52	-.48
45/1	hält mich immer wieder zu kooperativem Führen an.	4.68	1.42	.76	-1.01	1.15	5.11	1.15	.40	-.82	.93	4.65	1.48	.76	-.54	-.04
46/1	verdeutlicht mir beispielhaft an ihrem eigenen Verhalten, was kooperatives Führen bedeutet.	5.21	1.04	.23	-.21	-.70	5.64	.92	.11	-.91	1.72	4.82	1.55	.66	-.77	-.13
	Cronbachs Alpha	.77					.53					.86				

Anmerkung. Aufwärts: $N = 134$.

Tabelle E-7. Faktor: Stellvertretung (S), Ist-Version

		Abwärts					Selbst					Aufwärts				
Nr.	Meine direkte Führungskraft …	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*
47/4	arbeitet mich in Aufgaben ein, die mir nicht geläufig sind.	4.98	1.50	.61	-.93	.10	5.19	1.53	.54	-1.03	.55	5.08	1.41	.62	-.60	-.44
48/4	überträgt mir die Verantwortung für Entscheidungen, die ich während ihrer Abwesenheit treffe.	5.53	1.25	.70	-.94	.49	6.00	1.18	.25	-2.48	8.37	5.65	1.38	.79	-1.56	2.85
49/4	steht hinter mir als ihrem Stellvertreter.	6.04	1.02	.70	-1.63	4.20	6.09	1.13	.43	-1.81	3.83	6.04	1.06	.77	-1.62	3.46
50/4	bespricht mit mir alle wichtigen Aufgaben, bevor sie in Urlaub geht.	5.83	1.03	.69	-.41	-.93	6.12	.91	.32	-1.27	2.43	5.47	1.50	.72	-.93	-.01
	Cronbachs Alpha	.82					.59					.86				

Anmerkung. Aufwärts: $N = 54$.

Tabelle E-8. Faktor: Entwickeln und Fördern (E/F), Ideal- und Wichtigkeits-Version

		Ideal-Selbst					Ideal-Aufwärts					Wichtigkeit-Aufwärts				
Nr.	Meine direkte Führungskraft …	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*
1/2	leitet Förderungsmaßnahmen ein, die auf mich persönlich zugeschnitten sind.	6.30	.47	.76	.91	-1.3	5.79	1.02	.56	-1.53	3.51	5.57	1.18	.54	-.75	-.01
4/2	fördert mich in einer Art und Weise, die sowohl mir persönlich als auch dem Unternehmen nützt.	6.46	.59	.36	-.53	-.59	6.10	.77	.46	-1.93	8.41	5.99	1.08	.51	-2.02	6.01
6/2	plant mit mir meine beruflichen Möglichkeiten und Entwicklungsschritte.	6.04	.81	.66	-.62	.30	5.79	1.10	.75	-1.68	3.42	5.65	1.15	.59	-1.24	2.33
9/2	erkennt recht treffend, zu welcher Höherqualifizierung ich motiviert bin.	6.36	.57	.79	-.14	-.68	5.72	.96	.78	-1.73	6.05	5.38	1.20	.58	-1.37	2.23
10/2	kann genau einschätzen, wozu ich in Zukunft fähig sein werde.	6.32	.56	.59	.01	-.57	5.82	.81	.54	-.80	.50	5.64	.99	.43	-.98	1.46
12/1	motiviert mich dazu, mich selbständig weiterzuqualifizieren.	6.25	.79	.73	-1.14	1.36	5.66	1.01	.69	-1.12	1.77	5.27	1.25	.54	-.88	.83
19/2	regt mich dazu an, mich über meine derzeitige Tätigkeit hinaus breiter zu informieren und zu qualifizieren.	6.08	.97	.59	-1.72	3.85	5.53	1.18	.77	-1.23	1.39	5.27	1.21	.67	-1.01	1.23
20/1	führt mich Schritt für Schritt an umfangreiche Aufgaben heran.	6.36	.64	.59	-.47	-.54	5.50	1.22	.75	-1.68	3.16	5.39	1.17	.63	-1.03	1.73
21/2	fördert mich auch auf die Gefahr hin, dass ich irgendwann die Abteilung verlasse.	6.42	.58	.60	-.37	-.68	5.57	1.36	.65	-1.34	1.57	5.06	1.48	.67	-1.12	1.00
25/2	fragt mich immer wieder nach meinen Vorstellungen über meine berufliche Zukunft.	5.83	1.01	.58	-2.42	8.92	5.13	1.45	.54	-1.42	1.60	4.81	1.44	.70	-.62	-.02
31/1	bietet mir ausreichend Möglichkeiten, mich fachlich weiterzuqualifizieren.	6.25	.85	.46	-1.46	2.54	5.99	.94	.66	-1.67	4.32	5.78	1.08	.55	-1.95	6.01
32/2	hat ein Gespür für fachliche Trends, so dass sie sich rechtzeitig um meine entsprechende Förderung kümmern kann.	6.25	.74	.75	-.44	-.96	5.82	.74	.43	-.52	.37	5.54	1.03	.46	-.76	.91
	Cronbachs Alpha	.89					.90					.88				

Tabelle E-9. Faktor: Beurteilen und Steuern (B/S), Ideal- und Wichtigkeits-Version

		Ideal-Selbst					Ideal-Aufwärts					Wichtigkeit-Aufwärts				
Nr.	Meine direkte Führungskraft …	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*
17/7	informiert mich über den Anlass einer Beurteilung.	6.12	1.03	.53	-1.55	2.71	5.82	.92	.58	-1.07	2.24	5.43	1.24	.68	-1.27	2.37
18/8	erläutert mir, welches Verhalten ich verändern soll.	6.36	.57	.68	-.14	-.68	5.67	1.06	.45	-1.29	2.03	5.81	.98	.62	-1.22	2.88
22/8	verdeutlicht mir, wie sich die letzte Beurteilung auf unsere künftige Zusammenarbeit auswirkt.	5.83	1.15	.65	-.99	.29	5.41	1.34	.58	-1.42	2.11	5.16	1.38	.76	-1.06	1.00
23/7	beurteilt mich anhand messbarer und nachvollziehbarer Kriterien.	6.48	.59	.68	-.63	-.47	5.90	.97	.42	-1.85	5.06	5.93	.93	.51	-1.00	.95
24/7	erläutert mir die Abweichungen zwischen den erreichten Ergebnissen und den Zielvorgaben.	6.16	.94	.80	-1.96	5.10	5.74	1.07	.59	-1.56	2.85	5.67	1.15	.65	-1.56	3.59
27/8	erklärt mir, weshalb sie möchte, dass ich ein bestimmtes Verhalten verändere.	6.12	.88	.63	-1.84	5.70	5.77	.96	.70	-1.51	3.92	5.43	1.29	.58	-1.20	1.64
28/3	erinnert mich gleichzeitig an meine Entscheidungsbefugnisse, wenn sie Aufgaben an mich delegiert.	5.92	1.12	.73	-1.99	5.64	5.46	1.26	.38	-1.45	2.35	5.06	1.23	.63	-.49	.07
29/7	will von mir wissen, wie treffend ich die Beurteilung meiner Leistung fand.	5.37	1.61	.64	-1.08	-.11	5.55	1.11	.61	-1.44	3.16	4.93	1.47	.68	-.75	.10
38/7	setzt mich über die Kriterien, nach denen sie mich beurteilt, in Kenntnis, bevor ich mit meiner Arbeit beginne.	5.80	1.29	.67	-1.49	2.34	5.63	1.07	.59	-.93	1.38	5.09	1.23	.65	-.70	.40
11/1	erläutert mir, welche Fähigkeiten und Fertigkeiten sie von mir zukünftig erwartet.	6.40	.58	.68	-.28	-.72	5.82	1.05	.57	-2.00	5.00	5.72	.98	.66	-1.50	4.67
	Cronbachs Alpha	.89					.84					.89				

Tabelle E-10. Faktor: Individuelle Berücksichtigung (IB), Ideal- und Wichtigkeits-Version

		Ideal-Selbst					Ideal-Aufwärts					Wichtigkeit-Aufwärts				
Nr.	Meine direkte Führungskraft …	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*
2/3	gibt möglichst alle Aufgaben, die ich selbst erledigen kann, auch an mich weiter.	6.56	.51	.53	-.56	-2.11	6.28	.76	.35	-1.29	3.08	6.11	.87	.30	-1.62	4.94
3/7	beurteilt mich so, dass ich mich im Vergleich zu meinen Kollegen gerecht behandelt fühle.	6.33	1.05	.46	-3.21	13.09	6.08	1.10	.48	-2.46	9.20	6.07	.88	.38	-1.31	2.64
5/7	lässt persönliche Zu- oder Abneigung bei meiner Leistungsbeurteilung außen vor.	6.56	.51	.52	-.28	-2.11	6.18	.98	.39	-2.09	6.18	6.11	1.06	.29	-1.86	5.00
13/1	setzt mich so ein, dass ich auch langfristig in der Lage bin, optimale Leistung zu zeigen.	6.57	.51	.70	-.28	-2.11	6.14	.83	.55	-2.14	8.09	5.99	.79	.50	-1.41	5.42
14/7	kann die fachliche Qualität meiner Leistungen richtig einschätzen.	6.48	.51	.45	.09	-2.17	6.23	.80	.67	-1.49	3.44	6.15	.69	.38	-.55	.42
15/3	bietet mir angemessene Handlungsfreiheit.	6.48	.51	.62	.09	-2.17	6.30	.55	.37	.04	-.57	6.13	.83	.54	-1.54	5.05
16/6	berücksichtigt meine Kenntnisse und Erfahrungen bei der Erarbeitung von Zielen.	6.40	.50	.65	.44	-1.98	5.97	.90	.44	-1.53	3.48	5.85	.92	.47	-.93	1.17
26/3	achtet darauf, dass ich ausreichend rechtliche Vollmachten habe, um meine Aufgaben gut erledigen zu können.	6.32	.75	.52	-1.27	2.47	5.87	.96	.24	-2.31	8.89	5.79	1.07	.46	-.97	.88
	Cronbachs Alpha	.80					.73					.71				

Tabelle E-11. Faktor: Klarheit (K), Ideal- und Wichtigkeits-Version

		Ideal-Selbst					Ideal-Aufwärts					Wichtigkeit-Aufwärts				
Nr.	Meine direkte Führungskraft …	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*
7/6	erläutert mir auch die Unternehmensziele, so dass ich verstehe, wozu ich etwas tun soll.	6.63	.50	.27	-.55	-1.86	6.05	.98	.10	-2.08	7.45	6.01	.93	.31	-1.26	2.26
8/5	stellt sicher, dass ich die nötigen Informationen für die Ausführung meiner Aufgaben habe.	6.64	.49	.64	-.62	-1.76	6.31	.64	.42	-.64	.66	6.14	.97	.18	-1.91	5.52
33/8	leitet rechtzeitig – bevor etwas schief läuft – die notwendigen Schritte zur Korrektur ein.	6.32	.63	.74	-.35	-.53	6.01	.81	.47	-1.01	1.73	6.15	.80	.58	-.96	1.51
34/6	sorgt dafür, dass ich die zwischen uns vereinbarten Ziele bis ins Detail kenne.	6.20	.65	.72	-.20	-.48	6.10	.63	.55	-1.15	5.69	5.94	.99	.44	-1.17	2.13
35/5	sorgt für einen funktionierenden Informationsfluss in unserer Abteilung.	6.48	.51	.80	.09	-2.17	6.39	.61	.36	-.44	-.63	6.33	.82	.32	-1.93	6.69
39/3	gibt Aufgaben verständlich und inhaltlich korrekt an mich weiter.	6.60	.50	.59	-.44	-1.98	6.30	.66	.38	-1.35	5.62	6.28	.69	.61	-.78	.72
42/6	bespricht Zielsetzungen mit mir so lange, bis alle Unklarheiten beseitigt sind.	6.28	1.02	.41	-3.17	13.23	5.90	.96	.47	-1.55	4.43	5.82	1.00	.47	-1.00	1.53
	Cronbachs Alpha	.81					.66					.69				

Tabelle E-12. Faktor: Partizipation (P), Ideal- und Wichtigkeits-Version

		Ideal-Selbst					Ideal-Aufwärts					Wichtigkeit-Aufwärts				
Nr.	Meine direkte Führungskraft …	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*
30/6	beteiligt mich möglichst stark bei der Festlegung der Ziele, die ich im folgenden Zeitraum erreichen soll.	6.16	.62	.42	-.11	-.27	5.90	.71	.53	-.99	2.82	5.83	.91	.48	-1.09	1.82
36/6	verbreitet soviel Enthusiasmus, dass ich meine Zielvorgaben wirklich akzeptiere.	6.12	.83	.46	-1.18	1.79	5.83	.79	.29	-.77	1.29	5.37	1.03	.61	-.86	.69
37/5	gestaltet unsere Zusammenarbeit so, dass ich auch ungewöhnliche Ideen äußern kann.	6.36	.64	.71	-.47	-.54	5.99	.79	.39	-.94	1.83	5.58	1.09	.73	-.93	.87
40/7	will wissen, wie ich meine eigene Leistung einschätze.	6.04	.73	.77	-.06	-1.0	5.53	.98	.40	-1.51	4.21	5.09	1.29	.57	-.90	.78
41/5	erfragt meine Ideen und Argumente, wenn sie weitreichende Entscheidungen trifft.	6.16	.85	.62	-.77	.06	5.95	.83	.41	-.73	.34	5.91	.96	.40	-1.03	1.89
	Cronbachs Alpha	.80					.65					.78				

Tabelle E-13. Faktor: Fördern von Führungsverhalten (FvF), Ideal- und Wichtigkeits-Version

		Ideal-Selbst					Ideal-Aufwärts					Wichtigkeit-Aufwärts				
Nr.	Meine direkte Führungskraft …	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*
43/7	beurteilt mich auch nach dem Führungsverhalten, das ich meinen Mitarbeitern gegenüber zeige.	6.05	1.33	.51	-2.92	10.28	5.81	1.00	.52	-2.47	7.59	6.07	.88	.62	-1.13	1.62
44/1	sieht es als ihre Aufgabe, auch mein Führungsverhalten zu entwickeln.	5.96	1.17	.67	-1.94	4.69	5.74	1.13	.84	-1.87	3.93	6.03	.93	.79	-.78	.34
45/1	hält mich immer wieder zu kooperativem Führen an.	6.24	.83	.80	-1.07	1.15	5.60	1.36	.70	-1.84	3.36	5.68	1.21	.80	-1.19	2.04
46/1	verdeutlicht mir beispielhaft an ihrem eigenen Verhalten. was kooperatives Führen bedeutet.	6.63	.58	.31	-1.28	.86	6.12	.68	.27	-.93	2.38	5.99	1.00	.72	-.99	.40
	Cronbachs Alpha	.71					.76					.87				

Anmerkungen. Für Ideal-Aufwärts $N = 53$. Wichtigkeit $N = 75$.

Tabelle E-14. Faktor: Stellvertretung (S), Ideal- und Wichtigkeits-Version

		Ideal-Selbst					Ideal-Aufwärts					Wichtigkeit-Aufwärts				
Nr.	Meine direkte Führungskraft …	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*	*M*	*SD*	*TS*	*Sch*	*Ex*
47/4	arbeitet mich in Aufgaben ein, die mir nicht geläufig sind.	6.08	1.47	.60	-2.03	3.61	5.75	1.21	.59	-1.65	3.93	5.97	.91	.67	-1.48	3.42
48/4	überträgt mir die Verantwortung für Entscheidungen, die ich während ihrer Abwesenheit treffe.	6.25	1.22	.76	-3.62	15.72	6.25	.85	.58	-1.10	1.07	6.33	.78	.71	-1.21	1.57
49/4	steht hinter mir als ihrem Stellvertreter.	6.57	.59	.71	-1.00	.16	6.50	.76	.74	-1.99	5.14	6.35	.63	.64	-.41	-.55
50/4	bespricht mit mir alle wichtigen Aufgaben, bevor sie in Urlaub geht.	6.71	.55	.32	-1.80	2.68	6.25	1.25	.80	-2.50	6.93	6.52	.64	.78	-1.01	.07
	Cronbachs Alpha	.73					.84					.85				

Anmerkungen. Für Ideal-Aufwärts $N = 21$, Wichtigkeit $N = 29$.

F. Fragebogen zur Messung von Arbeitszufriedenheit, Leistung und Wohlbefinden sowie offene Fragen

Arbeitszufriedenheit

Im allgemeinen bin ich mit ...	gar nicht zufrieden.	nicht zufrieden.	eher nicht zufrieden.	weder zufrieden noch unzufrieden.	eher zufrieden.	zufrieden.	voll und ganz zufrieden.
meinem Arbeitsplatz (Ort, Status etc.) ...	1	2	3	4	5	6	7
meiner Arbeitstätigkeit selbst (Inhalte, Variation etc.) ...	1	2	3	4	5	6	7
meiner direkten Führungskraft ...	1	2	3	4	5	6	7
der Organisation VW Sachsen ...	1	2	3	4	5	6	7
Den Möglichkeiten meiner Fortbildung, Entwicklung und Förderung ...	1	2	3	4	5	6	7
dem Verhältnis zu meinen Kollegen in unserem Team / unserer Abteilung ...	1	2	3	4	5	6	7
den Arbeitsbedingungen (Raum, Zeit, Material, Temperatur etc.), unter denen ich täglich arbeite, ...	1	2	3	4	5	6	7
Den Maßnahmen zu meinem Gesundheitsschutz ...	1	2	3	4	5	6	7
der finanziellen Entlohnung für meine Leistung ...	1	2	3	4	5	6	7
der Anerkennung für meine Leistung ...	1	2	3	4	5	6	7
meiner Arbeit insgesamt und allem, was dazugehört, ...	1	2	3	4	5	6	7

Wie stark ...	gar nicht	nicht	eher nicht	leicht	eher stark	stark	sehr stark
könnte Ihre direkte Führungskraft Ihre Arbeitszufriedenheit prinzipiell beeinflussen?	1	2	3	4	5	6	7
beeinflusst Ihre direkte Führungskraft derzeit Ihre Arbeitszufriedenheit?	1	2	3	4	5	6	7

Was könnte Ihre direkte Führungskraft tun, um Ihre Arbeitszufriedenheit zu verbessern?

Wohlbefinden

Ich fühle mich ...	trifft überhaupt nicht zu	trifft nicht zu	trifft eher nicht zu	weder noch, unentschieden	trifft eher zu	trifft zu	trifft voll und ganz zu
bei meiner Arbeit wohl.	1	2	3	4	5	6	7
an meinem Arbeitsplatz gestresst.	1	2	3	4	5	6	7

Wie stark ...	gar nicht	nicht	eher nicht	leicht	eher stark	stark	sehr stark
könnte Ihre direkte Führungskraft Ihr persönliches Wohlbefinden prinzipiell beeinflussen?	1	2	3	4	5	6	7
beeinflusst Ihre direkte Führungskraft derzeit Ihr persönliches Wohlbefinden?	1	2	3	4	5	6	7

Was könnte Ihre direkte Führungskraft tun, um Ihr persönliches Wohlbefinden zu verbessern?

Leistung

	trifft überhaupt nicht zu	trifft nicht zu	trifft eher nicht zu	weder noch, unentschieden	trifft eher zu	trifft zu	trifft voll und ganz zu
Ich bringe an diesem Arbeitsplatz meine volle Leistung.	1	2	3	4	5	6	7
Unter einer anderen Führungskraft könnte ich noch mehr leisten.	1	2	3	4	5	6	7

Wie häufig führt Ihre Führungskraft mit Ihnen Beurteilungsgespräche (über Ihre Arbeitsleistung und zukünftige Erwartungen)?

❒ nie ❒ seltener als einmal jährlich ❒ einmal jährlich ❒ zweimal jährlich ❒ häufiger

Wie häufig wünschen Sie sich ein Beurteilungsgespräch?

❒ nie ❒ seltener als einmal jährlich ❒ einmal jährlich ❒ zweimal jährlich ❒ häufiger

Führungsgrundsätze und Fragebogen

Wenn Ja: Was ist für Sie der Kerngedanke der Führungsgrundsätze? Versuchen Sie bitte, diesen in einem Satz aufzuschreiben:

❒ ja ❒ teilweise ❒ nein

	trifft überhaupt nicht zu	trifft nicht zu	trifft eher nicht zu	weder noch, unentschieden	trifft eher zu	trifft zu	trifft voll und ganz zu
Meine direkte Führungskraft verhält sich bereits nach diesem Kerngedanken.	1	2	3	4	5	6	7

Ich habe bereits an einem Workshop zu den Führungsgrundsätzen bei VWS teilgenommen.

❒ Ja, am (ungefähre Angabe): _____ ❒ Nein

Wie Sie wissen, bearbeitet auch Ihre direkte Führungskraft diesen Fragebogen. Sie beschreibt ihr eigenes Führungsverhalten (Selbstbild), während Sie es aus Ihrer Sicht beschrieben haben (Fremdbild).

1. Wenn sich Unterschiede zwischen Selbst- und Fremdbild ergeben, woran könnte das Ihrer Meinung nach liegen?

- ..
- ..
- ..

Was sollte man dann tun (möglichst konkrete Angaben)?

- ..
- ..
- ..

2. Wenn sich keine beachtenswerten Unterschiede zwischen Selbst- und Fremdbild ergeben, woran könnte das liegen?

- ..
- ..
- ..

Was sollte man dann tun (möglichst konkrete Angaben)?

- ..
- ..
- ..

Der Fragebogen, der in Teil 1 abgedruckt ist, kann zukünftig in der vorliegenden Version oder in entsprechenden Auszügen hilfreich sein für ...

- ❒ die Beschreibung von Führungskräften.
- ❒ die Beratung von Führungskräften.
- ❒ die Beratung von Abteilungen/Teams.
- ❒ folgendes: _____
- ❒ folgendes: _____
- ❒ folgendes: _____

Welche Alternativen zum Fragebogen würden Sie vorziehen, um das Führungsverhalten von Führungskräften zu beschreiben?

G. FzFü-Instruktion

Bitte schätzen Sie in der rechten Antwortspalte ein, inwiefern die einzelnen Aussagen auf *Ihre direkte* Führungskraft *momentan* zutreffen. Es geht darum, wie *Sie persönlich* diese Verhaltensweisen wahrnehmen.

Verwenden Sie bei der Beantwortung bitte folgende Abstufung, indem Sie die entsprechende Nummer ankreuzen. Sie finden diese Skala rechts neben den Aussagen.

trifft überhaupt nicht zu	trifft nicht zu	trifft eher nicht zu	weder noch, bin unentschieden	trifft eher zu	trifft zu	trifft voll und ganz zu
1	2	3	4	5	6	7

Um welche Person geht es?

Bitte beschreiben Sie im Fragebogen die Person, die auf dem oben angehefteten Zettel namentlich genannt ist. Werfen Sie diesen Zettel weg, nachdem Sie den Fragebogen ausgefüllt haben.

H. Faktorladungsmatrix und Trennschärfen zur Arbeitszufriedenheits-Skala

Item-Nr.	Zufriedenheitsbereich	Ladung auf Faktor[1]			
		Kom[2]	f1	f2	TS[3]
10	Anerkennung	.76	**.86**		.74
3	Führungskraft	.60	**.77**		.53
6	Team/Abteilung	.46	**.67**		.53
5	Förderung, Entwicklung	.57	**.66**	.38	.62
4	Organisation	.43	**.62**		.53
2	Tätigkeit	.35	**.49**		.44
7	Arbeitsbedingungen	.74		**.85**	.63
8	Gesundheitsschutz	.68		**.83**	.48
1	Arbeitsplatz (Ort, Status)	.49	.48	**.52**	.46
9	Entlohnung	.30		**.43**	.36
	Eigenwerte der rotierten Faktoren		3.17	2.20	
	Durch Faktoren aufgeklärter Varianzanteil		31.68	21.96	

Anmerkungen. [1] Ladungen größer .35, [2] Kommunalitäten, [3] Trennschärfe.

I. Range der FzFü-Skalenmittelwerte

		E/F	B/S	IB	K	P	FvF	S
Ist-Abwärts	*Min*	3.67	3.80	3.75	3.29	3.40	2.50	3.50
(*N* = 48)	*Max*	6.25	6.40	6.75	6.86	6.40	6.33	7.00
Ist-Selbst	*Min*	3.00	2.50	4.75	4.43	3.80	2.00	3.75
(*N* = 48)	*Max*	6.50	6.70	6.50	6.71	6.40	6.75	7.00
Ist-Aufwärts	*Min*	1.00	1.11	1.00	1.00	1.00	1.00	1.75
(*N* = 207/134/54)	*Max*	6.75	6.90	7.00	7.00	7.00	7.00	7.00
Ideal-Selbst	*Min*	5.42	4.20	5.75	5.57	5.20	2.00	3.50
(*N* = 25)	*Max*	7.00	7.00	7.00	7.00	7.00	7.00	7.00
Ideal-Aufwärts	*Min*	2.42	2.40	4.38	4.86	4.00	2.67	4.00
(*N* = 93/53/21)	*Max*	7.00	7.00	7.00	7.00	7.00	7.00	7.00
Wichtigkeit	*Min*	3.17	2.40	3.75	3.00	3.40	3.50	3.00
(*N* = 109/75/29)	*Max*	6.83	7.00	7.00	7.00	7.00	7.00	7.00

J. Deskriptive Statistik der Diskrepanzmaße

Tabelle J-1. D-Statistik: Selbst-Aufwärts

	N	*Min*	*Max*	*M*	*SD*
E/F	40	.17	2.87	1.38	.77
B/S	40	.00	3.35	1.40	.82
IB	40	.25	3.02	.91	.62
K	40	.15	2.69	1.09	.66
P	40	.20	3.25	1.36	.88
FvF	30	.25	2.80	1.16	.68
S	26	.00	3.00	1.01	.90

Anmerkungen. für Tabellen J-1 bis J-5 gilt: Min: Minimum, Max: Maximum.

Tabelle J-2. Differenz: Selbst minus Aufwärts (aggregiert)

	N	*Min*	*Max*	*M*	*SD*
E/F	40 (142)	-1.67 (-1.75)	3.05 (3.67)	.83 (1.05)	1.12 (1.30)
B/S	40 (142)	-1.90 (-2.40)	2.90 (4.69)	.68 (.85)	1.21 (1.45)
IB	40 (142)	-.63 (-1.13)	2.56 (4.75)	.40 (.55)	.79 (1.01)
K	40 (142)	-1.57 (-1.57)	2.50 (4.14)	.52 (.67)	.87 (1.18)
P	40 (142)	-1.60 (-2.00)	3.10 (5.00)	.61 (.96)	1.24 (1.52)
FvF	30 (79)	-1.50 (-1.75)	2.75 (4.00)	.52 (.69)	1.05 (1.33)
S	26 (36)	-2.25 (-2.25)	3.00 (4.00)	.20 (.37)	1.24 (1.31)

Anmerkung. In Klammern: Angaben für die Differenz *Selbst minus individuelle Aufwärtsratings*.

Tabelle J-3. Differenz: Selbst minus Abwärts

	N	*Min*	*Max*	*M*	*SD*
E/F	32	-2.17	1.17	-.03	.82
B/S	32	-1.50	1.30	-.12	.81
IB	32	-1.50	1.38	.13	.75
K	32	-1.71	1.43	-.02	.77
P	32	-1.20	1.80	.08	.69
FvF	31	-3.75	2.75	.46	1.25
S	31	-2.50	2.25	.13	1.06

Tabelle J-4. Differenz: Abwärts minus Aufwärts (aggregiert)

	N	*Min*	*Max*	*M*	*SD*
E/F	32	-1.04	3.45	1.10	1.01
B/S	32	-1.35	3.83	1.11	1.06
IB	32	-.75	2.43	.35	.74
K	32	-.86	3.36	.73	1.19
P	32	-1.00	2.47	.70	1.11
FvF	24	-1.25	2.50	.27	1.21
S	19	-1.50	1.75	.29	.98

Tabelle J-5. Als Differenzmaße operationalisierte Erfolgskriterien, *N* = 205

	EF-WOBE	ZB	*Min*	*Max*	*M*	*SD*
EF-AZ	.43**	.32**	-3.00	6.00	.55	1.24
EF-WOBE		.08	-1.00	6.00	.61	1.13
ZB			-4.00	3.00	.90	1.11

Anmerkung. Spalte 2-3: Interkorrelationen.

K. Kodierung für Moderatorvariable „Ergebnisbezogene Itemformulierung“

Folgende Items wurden als ergebnisbezogen (vs. nicht ergebnisbezogen) kodiert. Es ist jeweils angegeben, auf welchem Faktor sie laden.

	Meine direkte Führungskraft ...	Faktor				
		E/F	B/S	IB	K	P
4/2	fördert mich in einer Art und Weise, die sowohl mir persönlich als auch dem Unternehmen nützt.	x				
20/1	führt mich Schritt für Schritt an umfangreiche Aufgaben heran.	x				
32/2	hat ein Gespür für fachliche Trends, so dass sie sich rechtzeitig um meine entsprechende Förderung kümmern kann.	x				
29/7	will von mir wissen, wie treffend ich die Beurteilung meiner Leistung fand.		x			
3/7	beurteilt mich so, dass ich mich im Vergleich zu meinen Kollegen gerecht behandelt fühle.			x		
26/3	achtet darauf, dass ich ausreichend rechtliche Vollmachten habe, um meine Aufgaben gut erledigen zu können.			x		
35/5	sorgt für einen funktionierenden Informationsfluss in unserer Abteilung.				x	
33/8	leitet rechtzeitig – bevor etwas schief läuft – die notwendigen Schritte zur Korrektur ein.				x	
8/5	stellt sicher, dass ich die nötigen Informationen für die Ausführung meiner Aufgaben habe.				x	
34/6	sorgt dafür, dass ich die zwischen uns vereinbarten Ziele bis ins Detail kenne.				x	
42/6	bespricht Zielsetzungen mit mir so lange, bis alle Unklarheiten beseitigt sind.				x	
7/6	erläutert mir auch die Unternehmensziele, so dass ich verstehe, wozu ich etwas tun soll.				x	
37/5	gestaltet unsere Zusammenarbeit so, dass ich auch ungewöhnliche Ideen äußern kann.					x
36/6	verbreitet soviel Enthusiasmus, dass ich meine Zielvorgaben wirklich akzeptiere.					x

L. Korrelative interperspektivische Zusammenhänge in Abhängigkeit von den absoluten interperspektivischen Niveau-Unterschieden

	Korrelation zwischen Selbst und Aufwärts$_a$		Korrelation zwischen Selbst und Abwärts		Korrelation zwischen Aufwärts$_a$ und Abwärts	
	NU gering	NU hoch	NU gering	NU hoch	NU gering	NU hoch
F/E	.59**	.68**	.64**	.82**	.61*	.72**
B/S	.53*	.55	.66**	.70**	.22	.58*
IB	.65**	.50*	.26	.78**	.77**	.72**
K	.46*	.80**	.39	.75**	.74**	.42
P	.36	.07	.60*	.46	.68**	.50*
FvF	.67**	.42	.36	.76**	.62*	.57
S	.34	.38	.73**	.57*	.68*	.88**

Anmerkungen. NU (gering/hoch) = Niveau-Unterschied (absolut) unter/über Median. Für Selbst-Aufwärts: Differenz: Selbst minus Auf$_a$ (aggregiert), *N* vgl. Tabelle 49.

M. Korrelation zwischen dimensionalen Führungsbeurteilungen (FzFü, verschiedene Richtungen) und Erfolgskriterien

Tabelle M-1. Entwickeln/Fördern

	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB	EEF[a]	EFF	SAT
Auf_i	.72**	.24**	-.36**	.21**	-.26**	.43**	-.48**	.69**	.76**	.75**
Auf_a	.70**	.10	-.40*	-.06	-.49**	.45**	-.50**	.74**	.69**	.79**
Auf_c	.25**	.09	-.10	.06	-.22**	.15*	-.11	.22**	.23**	.27**
Ab	.31+	-.05	.07	.03	.02	.28	-.44*	.47**	.30	.32
Selbst	-.13	.03	.13	.28+	-.08	-.36*	.16	.16	.08	.21
Fremd_a	.70**	.01	-.47**	-.09	-.50**	.52**	-.51**	.74**	.69**	.81**

Anmerkungen. Für Tabellen M-1 bis M-7 gilt: *N* vgl. Tabelle 64, $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$.

Tabelle M-2. Beurteilen/Steuern

	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB	EEF[a]	EFF	SAT
Auf_i	.64**	.19**	-.35**	.25 **	-.28**	.37**	-.44**	.68**	.72**	.68**
Auf_a	.58**	-.06	-.31+	.13	-.37*	.31+	-.51**	.60**	.61**	.73**
Auf_c	.20**	-.02	-.07	.01	-.13+	.10	-.08	..22**	.26**	.31**
Ab	.23	-.07	.02	-.01	.04	.19	-.28	.40*	.23	.21
Selbst	-.05	-.01	.20	.22	.16	-.03	.17	.12	-.07	.03
Fremd_a	.56**	-.08	-.42**	.11	-.32+	.41**	-.55**	.63**	.66**	.78**

Tabelle M-3. Individuelle Berücksichtigung

	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB	EEF[a]	EFF	SAT
Auf_i	.69**	.16*	-.40**	.20**	-.30**	.55 **	-.35**	.62**	.74**	.74**
Auf_a	.50**	-.18	-.44**	-.14	-.46**	.53 **	-.38*	.65**	.67**	.76**
Auf_c	.19**	.01	-.12+	-.05	-.17*	.13 +	-.08	.26**	.24**	.30**
Ab	.07	-.30+	-.28	-.07	-.04	.36 *	-.33+	.44*	.42*	.44*
Selbst	-.11	.09	.20	.25	.11	-.21	.19	.15	.08	.10
Fremd_a	.46**	-.18	.-45**	-.14	-.47**	.50 **	-.40**	.59**	.63**	.72**

Tabelle M-4. Klarheit

	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB	EEF[a]	EFF	SAT
Auf_i	.66**	.16*	-.40**	.25**	-.29**	.51**	-.36**	.69**	.81**	.81**
Auf_a	.61**	-.02	-.56**	.07	-.47**	.50**	-.34*	.69**	.73**	.82**
Auf_c	.19**	-.04	-.12+	.03	-.23**	.16*	-.06**	.22**	.26**	.30**
Ab	-.10	-.12	-.11	-.05	.04	.23	-.32+	.28	.14	.24
Selbst	-.07	-.07	-.00	.33*	-.02	-.28+	.03	.05	.09	.24
Fremd_a	.60**	-.07	-.58**	.06	-.46**	.56**	-.40	.72**	.75**	.83**

Tabelle M-5. Partizipation

	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB	EEF[a]	EFF	SAT
Auf_i	.60**	.11	-.40**	.18 **	-.31*	.52**	-.35**	.67**	.75**	.78**
Auf_a	.54**	.02	-.46**	.05	-.38*	.46**	-.47**	.62**	.62**	.76**
Auf_c	.21**	-.002	-.07	.004	-.16*	.14*	-.13+	.18**	.23**	.28**
Ab	.17	-.24	-.08	.01	-.11	.07	-.29	.30+	.22	.35+
Selbst	-.20	-.001	.31+	.23	.43**	-.28+	.11	-.25	-.22	-.15
Fremd_a	.51**	.06	-.46**	.03	-.42*	.44**	-.51**	.58**	.56**	.70**

Tabelle M-6. Fördern von Führungsverhalten

	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB	EEF[a]	EFF	SAT
Auf_i	.42**	-.06	-.41**	.09	-.34**	.51**	-.39**	.64**	.68**	.68**
Auf_a	.12	-.22	-.09	-.27	-.12	.16	-.18	.28	.34+	.55**
Auf_c	.17+	-.08	.02	.02	-.16+	.14	-.02	.11	.20*	.23**
Ab	.23	-.28	.25	-.01	.07	-.06	-.18	.29	.14	.12
Selbst	.01	.07	.27+	.07	-.10	-.12	.23	.13	-.07	.07
Fremd_a	.37*	-.24	.08	-.11	-.33+	.25	-.38**	.47**	.31+	.54**

Anmerkungen. Auf_a: $N = 30$, Auf_i und Auf_c: $N = 133$, Fremd_a: $N = 48$.

Tabelle M-7. Stellvertretung

	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB	EEF[a]	EFF	SAT
Auf_i	.53**	.20	-.51**	-.04	-.34*	.32*	-.39**	.60**	.60**	.62**
Auf_a	.22	-.15	-.74**	-.31	-.31	.18	-.27	.48*	.39*	.52**
Auf_c	.20	.10	-.07	-.09	-.14	.04	.10	.06	.14	.26+
Ab	.28	-.03	.08	.01	.05	.11	-.36*	.46**	.34+	.29
Selbst	.01	.27+	.16	.07	-.09	-.01	.04	.38*	.23	.21
Fremd_a	.31+	-.03	-.36*	-.10	-.23	.18	-.30*	.53**	.44**	.42*

Anmerkungen. Auf_a: $N = 30$, Auf_i und Auf_c: $N = 52$, Fremd_a: $N = 44$.

N. Korrelation zwischen dimensionaler Führungsbeurteilung (MLQ) und Erfolgskriterien

Tabelle N-1. Individuelle Aufwärtsratings

	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB	EEF	EFF	SAT
IIa	.64**	.15	-.40**	.19+	-.24*	.44**	-.48**	.84**	.89**	.86**
IIb	.65**	.24*	-.48**	.28*	-.25*	.43**	-.54**	.83**	.84**	.85**
IM	.66**	.28**	-.36**	.29**	-.27*	.32**	-.45**	.78**	.76**	.79**
IS	.64**	.31**	-.45**	.27*	-.24*	.46**	-.47**	.83**	.83**	.80**
IC	.67**	.29**	-.48**	.20+	-.27*	.49**	-.48**	.82**	.85**	.86**
AUS	.58**	.09	-.37**	.22**	-.30**	.46**	-.36**	.78**	.83**	.82**
TF	.63	.21	-.42	.24	-.27	.44	-.45	.81	.84	.83

Anmerkungen. $N = 91$, AUS: $N = 201$, TF = mittlere Korrelation aus den sechs Korrelationen, die die sechs transformationalen Skalen mit den Erfolgskriterien aufweisen.

Tabelle N-2. Aggregierte Aufwärtsratings

	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB	EEF	EFF	SAT
IIa	.43*	-.04	-.38*	.08	-.43*	.60**	-.47**	.71**	.78**	.79**
IIb	.63**	.13	-.41*	.14	-.48**	.54**	-.45*	.61**	.63**	.67**
IM	.44*	.12	-.28	.21	-.47**	.51**	-.42*	.70**	.63**	.68**
IS	.47**	.15	-.38*	.10	-.49**	.57**	-.41*	.68**	.59**	.59**
IC	.41*	.01	-.54**	.04	-.32	.56**	-.41*	.59**	.67**	.68**
AUS	.62**	.01	-.37*	.02	-.45**	.56**	-.42**	.78**	.76**	.86**
TF	.51	.06	-.39	.09	-.44	.56	-.43	.69	.69	.73

Anmerkungen. $N = 31$, AUS: $N = 40$.

Tabelle N-3. Gekreuzte Korrelationen

	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB	EEF	EFF	SAT
IIa	.20+	-.09	-.05	.08	-.10	.22*	-.17	.18+	.24*	.25*
IIb	.23*	-.02	.02	.05	-.11	.12	-.16	.07	.13	.17
IM	.19+	.00	.02	.04	-.10	.12	-.15	.17	.16	.18+
IS	.13	-.05	.06	.00	-.10	.14	-.04	.07	.03	.09
IC	.27**	.09	-.13	.17	-.14	.14	-.15	.13	.23*	.26*
AUS	.24**	.00	-.12+	.04	-.20**	.21**	-.12	.26**	.29**	.31**
TF	.22	-.01	-.05	.06	-.14	.17	-.13	.17	.20	.23

Anmerkungen. nach dem Criss-Cross-Design, $N = 91$, AUS: $N = 201$.

O. Ergebnisse der blockweisen Regressionen auf transformationale Führung und FzFü-Dimensionen

Kriterium	Prädiktoren		B°	$SE\ B^{\circ}$	β°	p°	*Korr.* R^2	$p\ (\Delta F)$	f^2
AZ1	TF		.06	.27	.05	(.83)	.24**	(< .01)	.06
	E/F	(aufwärts)	.53	.20	.61	(.01)	.38*	(.01)	.21
AZ2	TF		-.24	.23	-.30	(.30)	.03	(.68)	.01
	E/F	(aufwärts)	.28	.16	.48	(.10)	.04+	(< .10)	.11
EF-AZ	TF		.02	.25	.02	(.93)	.16*	(.02)	.23
	K	(aufwärts)	-.48	.20	-.59	(.03)	.28*	(.03)	.48
WOBE	TF		.05	.12	.07	(.71)	.02	(.50)	.02
	K	(selbst)	.42	.17	.42*	(.02)	.13*	(.02)	.23
EF-WOBE	TF		-.38	.29	-.33	(.20)	.19*	(.01)	.29
	E/F	(aufwärts)	.01	.22	.02	(.95)			
	P	(selbst)	.87	.34	.60	(.02)			
	FvF	(selbst)	-.35	.26	-.29	(.18)	.31+	(.08)	.70
L	TF		.40	.23	.42	(.09)	.34**	(< .01)	.57
	E/F	(selbst)	-.43	.15	-.40	(.01)			
	IB	(aufwärts)	.16	.21	.18	(.47)	.47*	(.02)	1.09
ZB	TF		-.32	.26	-.28	(.25)	.33**	(< .01)	.58
	E/F	(abwärts)	-.26	.17	-.28	(.13)			
	B/S	(aufwärts)	-.22	.16	-.34	(.19)	.45+	(.07)	1.15
EEF	TF		.35	.16	.34	(.05)	.48**	(< .01)	1.04
	E/F	(abwärts)	.20	.10	.24	(.06)			
	E/F	(aufwärts)	.15	.11	.25	(.20)			
	B/S	(selbst)	.33	.12	.39	(.02)			
	P	(selbst)	-.65	.23	-.52	(.01)			
	S	(selbst)	.26	.10	.36	(.02)	.80**	(< .01)	4.03
EFF	TF		.27	.27	.24	(.33)	.52**	(< .01)	1.17
	IB	(abwärts)	.00	.18	.00	(.99)			
	K	(aufwärts)	.41	.18	.60	(.04)	.60+	(.07)	1.98
SAT	TF		.33	.26	.26	(.22)	.62**	(< .01)	1.76
	IB	(abwärts)	.17	.16	.14	(.31)			
	K	(aufwärts)	.43	.18	.55	(.03)			
	E/F	(aufwärts)	.04	.14	.06	(.76)	.75*	(.02)	4.08

Anmerkungen. erster Block: TF (= transformationale Führung), zweiter Block: FzFü-Modell (angegebene Prädiktoren = „bestes Modells“), $N = 31$, °Kennwerte des Gesamtmodells (letzter Block), **/*/+*Änderung* (ΔR^2) in Signifikanz von F: $p < .01/.05/.10$, Determinationskoeffizient für die einzelnen Blöcke angegeben.

P. Hierarchische moderierte Regressionsanalysen für Führungserfolg auf Führungsverhalten und Wichtigkeit.

Tabelle P-1. Hierarchische moderierte Regression für AZ2 auf IB-Ist (Selbst) und IB-Wichtig (Aufwärts), $N = 32$

	Modell 1				Modell 2					
	B	*SE B*	*β*	*p*	*B*	*SE B*	*β*	*p*	*korr.* R^2	*p* (*ΔF*)
AV: AZ2										
1. Block									.03	.26
Selbst	.12	.10	.21	.26	.01	.11	.02	.91		
Wi	.12	.10	.20	.26	.09	.10	.15	.37		
2. Block									.15	.03
Selbst x Wi					-.24	.11	-.42	.03		

Anmerkungen. *p*(Δ*F*): Änderung in Signifikanz von *F,* AV: Abhängige Variable, Selbst: Selbstbeurteilung des Führungsverhaltens (Individuelle Berücksichtigung), Wi: Wichtigkeitsbeurteilung (Aufwärts, Individuelle Berücksichtigung).

Tabelle P-2. Hierarchische moderierte Regression für ZB auf S-Ist (Selbst) und S-Wichtig (Aufwärts), $N = 19$

	Modell 1				Modell 2					
	B	*SE B*	*β*	*p*	*B*	*SE B*	*β*	*p*	*korr.* R^2	*p* (*ΔF*)
AV: ZB										
1. Block									-.09	.75
Selbst	.01	.20	.01	.97	.17	.18	.21	.36		
Wi	-.11	.15	-.19	.46	-.01	.13	-.02	.94		
2. Block									.25	.01
Selbst x Wi					-.81	.29	-.64	.01		

Anmerkungen. *p*(Δ*F*): Änderung in Signifikanz von *F,* AV: Abhängige Variable, Selbst: Selbstbeurteilung des Führungsverhaltens (Stellvertretung), Wi: Wichtigkeitsbeurteilung (Aufwärts, Stellvertretung).

Q. Hierarchische moderierte Regressionsanalyse für Satisfaction auf Fördern von Führungsverhalten und transformationale Führung

Tabelle Q-1. Hierarchische moderierte Regression für SAT auf FvF-Ist (Aufwärts) und transformationale Führung (Aufwärts), $N = 49$

	Modell 1				Modell 2					
	B	*SE B*	*β*	*p*	*B*	*SE B*	*β*	*p*	*korr.* R^2	*p* (*ΔF*)
AV: SAT										
1. Block									.84	<.01
Aufwärts-FvF (Auf)	.22	.07	.24	<.01	.23	.07	.25	<.01		
Transformationale Führung (Tf)	.87	.09	.74	<.01	.87	.09	.74	<.01		
2. Block									.86	<.01
Auf x Tf					-.17	.06	-.16	<.01		

Anmerkungen. $p(\Delta F)$: Änderung in Signifikanz von *F,* AV: Abhängige Variable.

R. Korrelative Zusammenhänge zwischen den interperspektivischen Diskrepanzen auf den FzFü-Dimensionen und den Erfolgsvariablen

Tabelle R-1. Korrelative Zusammenhänge zwischen den Führungsdiskrepanzen und E/F

	EEF	EFF	SAT	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB
S-Auf_a	-.54**	-.55**	-.50**	-.64**	.09	.30+	.18	.36*	-.50**	.34*
S-Auf_i	-.59**	-.66**	-.67**	-.66**	-.20*	.31**	-.08	.36**	-.44**	.48**
S-Ab	-.29	-.10	-.18	-.38+	.18	.12	.15	.60	-.57**	.55**
Ab-Auf_a	-.46**	-.44*	-.57**	-.59**	-.32+	.49**	-.29	.50**	-.33+	.38*

Anmerkungen. Für Tabellen R-1 bis R-7 gilt: Anmerkungen vgl. Tabelle 73.

Tabelle R-2. Korrelative Zusammenhänge zwischen den Führungsdiskrepanzen und B/S

	EEF	EFF	SAT	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB
S-Auf_a	-.43**	-.57**	.54**	-.32*	.18	.37*	.13	.42**	-.39*	.35*
S-Auf_i	-.52**	-.59**	-.58**	-.58**	-.17*	.24**	-.18*	.38**	-.29**	.40**
S-Ab	-.16	-.13	-.31	-.33	.19	.30	.23	-.01	-.36+	.51*
Ab-Auf_a	-.45*	-.43*	-.59**	-.58**	-.20	.37*	-.25	.34+	-.32+	.48**

Tabelle R-3. Korrelative Zusammenhänge zwischen den Führungsdiskrepanzen und IB

	EEF	EFF	SAT	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB
S-Auf_a	-.56**	-.67**	-.75**	-.51**	.21	.53**	.02	.38+	-.57**	.25
S-Auf_i	-.50**	-.68**	-.70**	-.63**	-.04	.33**	-.09	.36**	-.51**	.37**
S-Ab	-.23	-.13	-.30	-.10	.29	.33	.19	.01	-.43*	.33
Ab-Auf_a	-.39*	-.32+	-.48**	-.51**	-.22	.34+	-.24	.47**	-.31+	.26

Tabelle R-4. Korrelative Zusammenhänge zwischen den Führungsdiskrepanzen und K

	EEF	EFF	SAT	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB
S-Auf_a	-.55**	-.69**	-.71**	-.49**	.12	.44**	.03	.32+	-.49**	.28+
S-Auf_i	-.65**	-.76**	-.76**	-.69**	-.25**	.37**	-.24**	.37**	-.54**	.36**
S-Ab	-.21	.00	-.04	-.01	.22	.14	.36	.05	-.48*	.30
Ab-Auf_a	-.50**	-.61**	-.65**	-.65**	-.19	.49**	-.35+	.42*	-.35+	-.31+

Tabelle R-5. Korrelative Zusammenhänge zwischen den Führungsdiskrepanzen und P

	EEF	EFF	SAT	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB
S-Auf_a	-.53 **	-.61 **	-.62 **	-.42 **	-.02	.46 **	.09	.41 *	-.47 **	.41 **
S-Auf_i	-.60 **	-.67 **	-.73 **	-.62 **	-.18 *	.39 **	-.16 +	.46 **	-.52 **	.33 **
S-Ab	-.33	-.22	-.46 *	-.33	.28	.40 +	.18	.24	-.32	.40 +
Ab-Auf_a	-.54 **	-.52 **	-.65 **	-.57 **	-.20	.57 **	-.20	.33 +	-.55 **	.55 **

Tabelle R-6. Korrelative Zusammenhänge zwischen den Führungsdiskrepanzen und FvF

	EEF	EFF	SAT	AZ1	AZ2	EF-AZ	WOBE	EF-WOBE	L	ZB
S-Auf_a	-.27	-.34 +	-.35 +	-.12	.23	.37 *	.44 *	.32	-.26	.20
S-Auf_i	-.52 **	-.54 **	-.55 **	-.41 **	.09	.37 **	.10	.44 **	-.35 **	.41 **
S-Ab	-.14	-.14	-.32	-.28	.44 *	.20	-.17	-.14	-.11	.36 +
Ab-Auf_a	-.35 +	-.39 +	-.56 **	-.15	-.15	.33	-.03	.38 +	-.32	.35 +

Tabelle R-7. Korrelative Zusammenhänge zwischen den Führungsdiskrepanzen und S

	EEF	EFF	SAT	AZ1	AZ2	EF-AZ	WO-BE	EF-WOBE	L	ZB
S-Auf_a	-.07	.09	-.06	.23	.26	.45 *	.29	.28	.09	-.15
S-Auf_i	-.29 +	-.20	- .30 +	-.18	-.02	.40 *	.24	.39 *	-.05	.09
S-Ab	-.33	-.23	-.29	-.43 *	.24	.17	-.07	-.05	-.18	.46 *
Ab-Auf_a	-.10	.04	-.24	-.06	-.26	.58 *	-.03	.46 *	-.14	.16

S. Zusammenhang zwischen interperspektivischen Korrelationen und Erfolgskriterien.

	τ -Korrelationen		
	Selbst-Abwärts (N = 22)	Selbst-Aufwärts (N = 136)	Abwärts-Aufwärts (N = 117)
AZ1	-.62**	.16	.05
AZ2	-.14	-.07	.15
EF-AZ	.32	-.04	-.05
WOBE	-.05	-.07	.05
EF-WOBE	.30	-.12	-.08
L	-.36	.26**	.08
ZB	.41	.00	.21*
EEF	-.30	.20*	-.09
EFF	-.22	.26**	-.06
SAT	-.38	.28**	.00

Anmerkung. Die τ-Korrelationen wurden für jeden Beurteiler berechnet. Sie geben die interperspektivische Korrelation über die beurteilten Führungsdimensionen hinweg an.

T. Mittlere Selbst- und Fremdbeurteilungen des Führungsverhaltens (FzFü) in Abhängigkeit vom Ausmaß der Abwärts-Aufwärts-Abweichung (Distanzmaß)

Abweichung	Selbst-Rating	Aufwärts-Rating	Abwärts-Rating
gering	5.33 (8)	5.14 (10)	5.36 (10)
mittel	5.39 (9)	4.70 (11)	5.31 (11)
hoch	5.19 (7)	4.00 (11)	5.53 (11)

Anmerkungen. Terzilsplit nach dem Ausmaß der Abweichung, *N* in Klammern.

U. Erfolgskennwerte (Skalenmittelwert) in Abhängigkeit von der Selbst-Fremd-Diskrepanz und der Führungsqualität (FzFü)

	S-Auf-Übereinstimmer		S-Auf-Überschätzer		S-Ab-Übereinstimmer		S-Ab-Überschätzer		S-Ab-Unterschätzer	
	A (N = 4)	B (N = 4)	A (N = 7)	B (N = 7)	C (N = 3)	D (N = 7)	C (N = 4)	D (N = 2)	C (N = 2)	D (N = 5)
AZ1	5.94	5.68	4.65	5.08	4.50	5.74	4.96	4.84	4.78	5.81
AZ2	5.81	5.56	5.44	5.36	4.88	5.70	5.71	5.75	4.83	5.38
EF-AZ	.25	-.25	.42	.29	-.50	.14	.67	-.50	.67	-.25
WOBE	5.13	4.75	4.79	4.71	4.83	4.89	5.00	4.25	4.33	4.75
EF-WOBE	.00	.25	.64	.43	1.00	.00	.58	.00	.33	.25
L	5.44	5.75	5.43	5.29	5.83	5.18	5.13	5.75	5.33	6.38
ZB	.50	.38	1.50	.64	1.00	.57	1.42	1.00	.67	.50
EEF	3.33	3.63	2.79	3.19	2.44	3.41	2.95	4.00	3.11	3.88
EFF	3.69	3.75	3.04	3.45	3.00	3.64	3.17	4.38	3.08	3.88
SAT	3.88	3.88	2.82	3.57	3.17	3.86	2.88	4.00	3.33	4.00

Anmerkungen. A: Abwärtsbeurteilung gering, B: Abwärtsbeurteilung hoch, C: Aufwärtsbeurteilung gering, D: Aufwärtsbeurteilung gering.

V. Selbst-Fremd- und Fremd-Fremd-Korrelationen in Abhängigkeit vom Erfolg

Tabelle V-1. Selbst-Aufwärts-Korrelationen der Führungsbeurteilungen (FzFü-Dimensionen) für gering und hoch erfolgreiche Teilstichprobe

		N	E/F	B/S	IB	K	P	FvF	S	$g_{FzFü}$
AZ1	gering[2]	19	.30	-.03	-.09	.61**	-.32	-.09	.24	.13
	hoch	20	-.10	-.01	.25	-.02	-.15	.16	-.47+	.04
q			.41	.02	.35	.73	.18	.25	.76	.09
AZ2	gering	20	.13	-.02	.07	.49*	-.53*	.05	-.16	.08
	hoch	19	-.11	-.01	.14	.04	-.13	.09	.00	.03
q			.24	.01	.07	.50	.46	.04	.16	.05
EF-AZ	gering	23	-.06	.00	.31	.17	-.12	.23	-.25	.12
	hoch	15	.30	-.02	-.01	.50+	-.34	-.09	-.11	.07
q			.37	.02	.33	-.38	.23	.32	.15	.05
WOBE	gering	19	.31	.09	-.02	.46*	-.29	.28	.05	.16
	hoch	20	-.22	-.22	.30	-.22	-.44+	.00	-.27	-.19
q			.55	.31	.33	.72	.17	.29	.33	.35
EF-WOBE	gering	24	.20	.15	.15	.27	-.04	.30	-.04	.25
	hoch	13	.26	.19	-.13	.65*	-.15	-.71*	-.09	.27
q			.06	.04	.28	.50	.11	1.20	.05	.02
L	gering	17	.13	-.19	.11	.57*	-.37	.13	-.16	-.02
	hoch	22	.06	.24	.33	.18	-.11	.32	-.14	.26
q			.07	.44	.23	.47	.28	.20	.02	.29
ZB	gering	18	-.08	-.04	.21	.08	-.04	.17	-.61*	.13
	hoch	22	.36	.22	.03	.51*	-.47*	-.10	.07	.26
q			.46	.26	.18	.48	.47	.27	.78	.14
EEF	gering	21	.25	.06	-.04	.67*	-.41+	-.17	-.27	.19
	hoch	19	-.06	.08	-.06	-.01	.09	.31	-.26	.11
q			.32	.02	.02	.82	.53	.49	.01	.08
EFF	gering	21	.29	-.00	.05	.61**	-.46*	-.06	-.11	.18
	hoch	19	-.11	.28	.11	.05	.08	.30	-.27	.17
q			.41	.29	.06	.66	.58	.37	.17	.01
SAT	gering	19	.14	-.01	-.16	.39	-.31	.00	-.28	.07
	hoch	21	-.01	.13	.25	.13	-.25	.10	-.22	.09
q			.15	.14	.42	.28	.07	.10	.06	.02

Anmerkungen. [1] $g_{FzFü}$: Summenscore aus E/F, B/S, IB, K und P, [2]Einteilung nach Median-Split.

Tabelle V-2. Selbst-Abwärts-Korrelationen der Führungsbeurteilungen (FzFü-Dimensionen) für gering und hoch erfolgreiche Teilstichprobe

		N	E/F	B/S	IB	K	P	FvF	S	$g_{FzFü}$
AZ1	gering	13	.44	.34	-.30	-.21	.00	.63*	.62*	.15
	hoch	10	-.23	.18	.17	.08	.53	-.44	.24	-.09
q			.71	.17	.48	.29	.59	1.21	.48	.24
AZ2	gering	12	-.12	.26	-.54+	-.16	-.08	.01	-.06	-.30
	hoch	11	.21	.18	.27	.09	.38	.55+	.85**	.25
q			.33	.08	.88	.25	.48	.61	1.32	.57
EF-AZ	gering	15	-.13	.17	-.01	.24	.25	.08	.27	-.08
	hoch	7	.50	.36	-.11	-.21	.30	.50	.68+	.38
q			.68	.21	.10	.46	.06	.47	.55	.48
WOBE	gering	11	.11	.12	-.44	-.28	-.15	.26	.02	-.13
	hoch	12	.10	.35	.24	.24	.50	.06	.67*	.25
q			.01	.24	.72	.53	.70	.21	.77	.39
EF-WOBE	gering	15	.11	.25	.04	.06	.39	.27	.34	.17
	hoch	8	.38	.12	-.63+	-.07	-.27	.76*	.12	.01
q			.29	.13	.78	.13	.69	.72	.23	.16
L	gering	10	.64*	.41	.52	.36	.25	-.15	.67*	.59+
	hoch	13	-.34	.14	-.58*	-.10	.03	.56*	-.03	-.46
q			1.11	.30	1.25	.48	.23	.78	.84	1.18
ZB	gering	9	.21	.21	.23	.17	.42	-.39	.87**	.24
	hoch	15	.33	.20	-.43	-.13	-.01	.47+	.06	.01
q			.13	.01	.69	.30	.44	.92	1.27	.24
EEF	gering	13	.54+	.33	.02	-.04	.14	.20	.46	.31
	hoch	11	-.29	-.01	-.51	.06	.13	.22	-.21	-.38
q			.90	.35	.57	.10	.01	.02	.71	.72
EFF	gering	12	.55+	.28	.06	-.06	.13	.12	.47	.31
	hoch	12	-.30	.18	-.55+	.07	.14	.50+	-.05	-.38
q			.93	.11	.68	.13	.01	.43	.56	.72
SAT	gering	10	.59+	.27	.13	-.06	.12	.47	.52	.35
	hoch	14	-.20	.25	-.60*	.02	.20	.03	.20	-.33
q			.93	.02	.82	.08	.08	.48	.37	.71

Tabelle V-3. Abwärts-Aufwärts-Korrelationen der Führungsbeurteilungen (FzFü-Dimensionen) für gering und hoch erfolgreiche Teilstichprobe

		N	E/F	B/S	IB	K	P	FvF	S	$g_{FzFü}$
AZ1	gering	13	.50^{+}	.72**	.64*	.08	.50^{+}	.14	.67^{+}	.57*
	hoch	10	.15	.09	.54	.73*	-11	.02	.17	.16
q			.40	.82	.15	.85	.66	.12	.64	.49
AZ2	gering	12	.32	.39	.39	-15	.14	.37	.60	.26
	hoch	11	.50	.67*	.64*	.46	.47	-20	.51	.63*
q			.22	.40	.35	.65	.37	.59	.13	.48
EF-AZ	gering	15	.36	.34	.39	.47^{+}	.41	.07	.12	.40
	hoch	7	.70^{+}	.83*	.78*	.25	.59	-09	.79	.75^{+}
q			.49	.83	.63	.26	.24	.16	.95	.54
WOBE	gering	11	.57^{+}	.81**	.78**	.25	.58^{+}	.46	.70^{+}	.66*
	hoch	12	.11	.05	.06	43	-25	-50	-17	-11
q			.54	1.08	.99	.21	.92	1.05	1.04	.90
EF-WOBE	gering	15	.34	.57*	.57*	.08	.34	.11	.51	.39
	hoch	8	.72*	.57	.66^{+}	.20	.20	-05	.995^{+}	.70^{+}
q			.55	.00	.15	.12	.15	.16	2.43	.46
L	gering	10	.41	.40	.61^{+}	.27	.49	-27	.57	.50
	hoch	13	.26	.52^{+}	-08	-29	.07	.29	.22	.14
q			.17	.15	.79	.58	.47	.58	.42	.41
ZB	gering	9	-79*	-.53	.16	.31	-18	-21	-66	-61^{+}
	hoch	15	.68**	.62*	.70**	.05	.44	.24	.71*	.62*
q			1.90	1.32	.71	.27	.65	.46	1.68	1.43
EEF	gering	13	.32	.29	.57*	.09	.31	-05	49	.35
	hoch	11	.63*	.60^{+}	.38	.36	.41	.17	.50	.62*
q			.41	.39	.25	.29	.12	.22	.01	.36
EFF	gering	12	33	.31	.58*	.06	.31	-.05	.49	.34
	hoch	12	.56^{+}	.63*	.13	.35	.41	.17	.50	.59*
q			.29	.42	.53	.31	.12	.22	.01	.32
SAT	gering	10	.62^{+}	.70*	.67*	-01	.44	-31	.41	.56
	hoch	14	.13	.25	-04	.07	24	.19	.50	.13
q			.59	.61	.85	.08	.23	.51	.11	.50

Zu Anhang V: Hervorzuheben sind für einzelne Führungsdimensionen und einzelne Erfolgsmaße: Die *Selbst-Abwärts*-Korrelationen auf der Dimension *Klarheit* liegen für neun der zehn Kriterien in der erfolgreichen Teilstichprobe höher als in der weniger erfolgreichen Teilstichprobe – allerdings mit überwiegend geringen Unterschieden ($q < .30$). Auch die *Abwärts-Aufwärts*-Korrelationen können auf dieser Dimension für acht Kriterien erfolgreiche von weniger erfolgreichen Führungskräften differenzieren, zum Teil mit starken Effekten. Im Unterschied dazu ist es für die *Selbst-Aufwärts*-Korrelation ausgerechnet die Dimension Klarheit, auf der alle Effekte in die unerwartete Richtung weisen. Klarheit scheint also insbesondere in der Gegenüberstellung mit den Abwärtsbeurteilungen eine wichtige Rolle für verschiedene Erfolgsvariablen zu spielen. Dies kann als weiterer Hinweis dafür gewertet werden, dass nicht nur die Selbstüberschätzung, sondern auch transparente, klare Führung, die von außen beobachtbar ist, im Wirkzusammenhang mit Erfolg steht. Die vorliegenden Korrelationsmuster scheinen im Übrigen unabhängig von der generellen Selbst-Fremd-Übereinstimmung zu sein, denn in der Gesamtstichprobe fällt die Selbst-Aufwärts-Korrelation gerade auf der Dimension Klarheit erstaunlich hoch aus. Die Selbst-Aufwärts-Korrelation differenziert erfolgreiche von weniger erfolgreichen Führungskräften (für nahezu alle Erfolgskriterien) auf den Dimensionen IB ($\bar{q} = .25$), P ($\bar{q} = .30$) und FvF ($\bar{q} = .36$).[230] Zumindest IB und P umfassen Führungsverhaltensweisen, die relativ zu den anderen Dimensionen stark auf die Aufwärtsbeurteiler ausgerichtet sind.

Richtet man den Fokus nicht auf die Führungsdimensionen, sondern auf die Erfolgskriterien, dann zeigt sich, dass sich überdurchschnittliche *Zufriedenheit mit Bedingungen (*AZ2) und überdurchschnittliches Wohlbefinden insbesondere in erhöhten Selbst-Abwärts-Korrelationen widerspiegeln ($\bar{q} = .65$ bzw $\bar{q} = .59$). Auf den ersten Blick scheint dies den oben berichteten positiven Korrelationen zwischen der Selbst-Abwärts-Diskrepanz und AZ2 sowie WOBE zu widersprechen. Das Besondere ist jedoch, dass die Selbst-Abwärts-Korrelation nur für die Teilstichprobe derjenigen, die überdurchschnittlich hohe Werte auf AZ2 und WOBE haben, positiv ausfällt. Umgekehrt bedeutet dies, dass die Führungskräfte im Vergleich zu ihren Vorgesetzten sogar gegenläufige Beurteilungen abgeben, wenn ihre Mitarbeiter – also Dritte – geringe Werte auf AZ2 und WOBE haben. Da die Führungsbeurteilungen und Erfolgskennwerte nicht von den gleichen Personen stammen, ist es wahrscheinlich, eine kausale Richtung von der Führungsdiskrepanz zu AZ2 und WOBE anzunehmen. Eine geringe Zufriedenheit mit den Kontextbedingungen (AZ2) kann beispielsweise dadurch weiter gefestigt werden, dass Führungskräfte und ihre Vorgesetzten nicht die gleiche Sprache sprechen.

Die *Abwärts-Aufwärts*-Übereinstimmung fällt für überdurchschnittliches *EEF* und überdurchschnittliche *EFF* auf allen Führungsdimensionen außer IB höher aus als für unterdurchschnittliche Ausprägungen von EEF ($\bar{q} = .24$) bzw. EFF ($\bar{q} = .23$). Demgegenüber kristalliert sich kein Kriterium heraus, für das sich der Selbst-Aufwärts-Zusammenhang besonders konsistent differenzieren ließe. Dies deutet erneut auf die verhältnismäßig geringe Validität der *Selbst-Aufwärts*-Gegenüberstellung hin.

[230] Lediglich für WOBE oder EEF ist die Selbst-Aufwärts-Korrelation bei überdurchschnittlichen Werten nicht auf allen drei genannten Dimensionen höher. Für die Berechnung der mittleren Effektstärken wurden diese Ausnahmen nicht berücksichtigt.

W. Prädiktive Validität von Führung in Abhängigkeit von der Selbst-Fremd- und Fremd-Fremd-Abweichung

Tabellen W-1. Korrelation zwischen Führung und Erfolg in Abhängigkeit von der Selbst-Aufwärts-Abweichung

	Korrelation zwischen **AZ1** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
Unterschätzer	.19	1.00	.13	.95*	.87	.79	-.80	-.87
Übereinstimmer	-.38	-.17	-.07	.18	.52	.10	-.24	-.49
Überschätzer	.46*	.45	.53*	.44*	.22	.51*	.05	-.35

Anmerkungen. Für die Tabellen W-1 und W-2 gilt: Unterschätzer (N = 2-4), Übereinstimmer (N = 8-14), Überschätzer (N = 13-22). N variiert in Abhängigkeit der Beurteilungsrichtung und Fragebogenversion.

	Korrelation zwischen **AZ2** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
Unterschätzer	-.36	-1.00**	-.21	-.52	-.64	-.75	.74	.64
Übereinstimmer	-.42	-.11	-.35	.25	.11	.15	-.30	-.32
Überschätzer	.51*	-.24	.18	.06	.03	.25	-.09	-.04

	Korrelation zwischen **WOBE** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
Unterschätzer	-.53		-.55	-.85	-.28	-.13	.14	.28
Übereinstimmer	-.04	-.27	.07	.55 *	.48	.21	-.13	-.50
Überschätzer	.54*	.02	.20	.13	.28	.34	.09	-.05

	Korrelation zwischen **L** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
Unterschätzer	-.00	1.00**	-.16	.48	.85	.92	-.91	-.85
Übereinstimmer	-.32	.33	-.04	.53+	.32	.70	-.83**	-.60+
Überschätzer	-.29	.09	.32	.49*	.52*	.11	-.34	-.48*

	Korrelation zwischen **EEF** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
Unterschätzer	.71		.64	-.01	-.28	-.13	.14	.28
Übereinstimmer	-.08	.43	-.04	.57 *	.09	.09	-.48	-.40
Überschätzer	.47 *	.51 +	.80 **	.78 **	.64 **	-.03	-.51 *	-.68 **

	Korrelation zwischen **EFF** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
Unterschätzer	.32		.36	.80	.28	.13	-.14	-.28
Übereinstimmer	-.09	.20	.08	.39	.27	.07	-.54	-.58
Überschätzer	.34	.26	.71 **	.70 **	.68 **	-.02	-.50 *	-.72 **

	Korrelation zwischen **SAT** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
Unterschätzer	.53		.55	.85	.28	.13	-.14	-.28
Übereinstimmer	.31	.15	.41	.65 *	.67 *	-.06	.06	-.47
Überschätzer	.41	.51	.86 **	.81 **	.74 **	-.05	-.49 *	-.68 **

Tabellen W-2. Korrelation zwischen Führung und Erfolg in Abhängigkeit von der Selbst-Abwärts-Abweichung

	Korrelation zwischen **AZ1** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
Unterschätzer	-.52	.11	.66	.80 *	.82 *	.32	-.44	-.67
Übereinstimmer	-.42	-.38	.54	.67 *	.54	.66 +	-.68 +	-.60
Überschätzer	.95 **	.81 *	.65	.55	.53	.20	.48	-.54

	Korrelation zwischen **AZ2** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
Unterschätzer	-.60	-.47	.19	.15	.06	-.42	.11	-.05
Übereinstimmer	-.32	-.23	.17	.44	.34	.57	-.76 *	-.50
Überschätzer	.49	.65	.16	-.12	-.12	.04	.59	.25

	Korrelation zwischen **WOBE** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
Unterschätzer	-.55	-.08	.31	.59	.56	.05	-.24	-.41
Übereinstimmer	-.36	-.60^{+}	.07	.21	.44	.60	-.33	-.52
Überschätzer	.80*	.55	.08	.06	.23	.56	.78^{+}	.04

	Korrelation zwischen **L** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
Unterschätzer	-.74^{+}	-.04	.47	.90**	.94**	.43	-.79*	-.90**
Übereinstimmer	-.56^{+}	-.45	-.51	-.27	.04	.44	-.07	-.49
Überschätzer	.14	.29	.73^{+}	.85*	.80^{+}	-.65	-.60	-.42

	Korrelation zwischen **EEF** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
Unterschätzer	-.22	.30	.85*	.64	.52	-.08	-.27	-.38
Übereinstimmer	.43	.19	.68*	.85**	.17	.51	-.58	-.66^{+}
Überschätzer	.56	.71^{+}	.98**	.93**	.83*	-.55	-.47	-.85*

	Korrelation zwischen **EFF** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
Unterschätzer	-.28	-.15	.78*	.88**	.61	.24	-.55	-.62
Übereinstimmer	.19	.09	.45	.59^{+}	.25	.26	-.73*	-.30
Überschätzer	.42	.48	.92**	.95**	.83*	-.55	-.47	-.85*

	Korrelation zwischen **SAT** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
Unterschätzer	-.14	.07	.70^{+}	.73^{+}	.48	.06	-.53	-.48
Übereinstimmer	.60^{+}	.25	.78**	.81**	.16	.12	-.67^{+}	-.15
Überschätzer	.47	.55	.93**	.96**	.88*	-.58	-.40	-.74^{+}

Tabellen W-3. Korrelation zwischen Führung und Erfolg in Abhängigkeit von der Abwärts-Aufwärts-Abweichung

	Korrelation zwischen **AZ1** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
keine/geringe D[1]	-.63 +	-.09	-.10	-.09	.27	.73	.30	.18
mittlere D	-.42	.46	.35	.77 *	.64 +	-.01	-.48	-.65 +
hohe D	.67	.67 *	.36	.42	.31	-.24	.28	-.36

Anmerkung. Für die Tabellen W-3 gilt: N = 8/9/7, [1]Ausmaß der Abweichung (Distanz) unabhängig von der Richtung der Abweichung.

	Korrelation zwischen **AZ2** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
keine/geringe D	.02	-.44	-.18	-.60	-.31	.17	.54	.08
mittlere D	-.36	.34	-.05	.38	.16	-.09	-.42	-.55
hohe D	.46	-.17	.18	-.70	-.74 +	.29	.58	.61

	Korrelation zwischen **WOBE** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
keine/geringe D	-.02	-.12	-.18	-.49	-.04	.72	.54	.11
mittlere D	.05	-.07	-.18	.32	.49	.56	-.02	-.63 +
hohe D	.35	.20	.09	-.16	-.38	.51	.77 +	.18

	Korrelation zwischen **L** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
keine/geringe D	-.35	.40	.47	.70	.50	-.00	-.51	-.41
mittlere D	-.59 +	.45	.17	.55	.49	-.25	-.68 *	-.50
hohe D	-.75 +	1.00	-.09	-.02	.73 +	-.67	-.60	-.72

	Korrelation zwischen **EEF** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
keine/geringe D	.48	.24	.53	.33	.23	-.41	-.03	-.41
mittlere D	-.28	.75 **	.65 *	.74 *	.51	-.59 +	-.38	-.15
hohe D	.63	.86 **	.38	.74 +	.59	-.57	-.09	-.62

	Korrelation zwischen **EFF** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
keine/geringe D	.48	.02	.32	.26	-.01	-.52	-.20	-.46
mittlere D	-.43	.39	.43	.66	.55	-.42	-.68 *	-.25
hohe D	.20	.86 **	.37	.59	.53	-.56	-.18	-.52

	Korrelation zwischen **SAT** und							
	Selbstbeurteilung	Abwärtsbeurteilung	Aufwärtsbeurteilung					
	FzFü	FzFü	FzFü	Tf	CR	MBA	MBP	LF
keine/geringe D	.61	.56	.66 *	.46	-.23	-.46	-.54	-.08
mittlere D	-.36	.51	.73 *	.80 **	.71 *	-.48	-.59 *	-.07
hohe D	.27	.85 **	.50	.93 **	.84 *	-.64	-.36	-.83 *

X. Korrelationen zwischen der Selbstbeurteilung und der Beurteilung des Führungsverhaltens anderer Führungskräfte

		Beschreibung des Führungsverhaltens unterstellter Führungskräfte						
		E/F	B/S	IB	K	P	FvF	S
Selbst-beschreibung	E/F	.31	.39	.12	-.02	.13	.04	.28
	B/S	.18	.50	.07	.11	.05	.27	.47
	IB	.36	.52	.42	.45	.35	.16	.42
	K	.19	.15	.24	.21	.21	.20	.09
	P	.14	.30	-.04	-.09	.02	.36	.37
	FvF	-.00	.18	.01	.06	.01	.16	-.02
	S	.27	.39	.38	.37	.20	-.06	.15

Anmerkung. gleicher Rater, verschiedene Beurteilte, $N = 41$.

		Beschreibung des Führungsverhaltens vorgesetzter Führungskräfte						
		E/F	B/S	IB	K	P	FvF	S
Selbst-beschreibung	E/F	.21	.35	.25	.13	.34	.33	.40
	B/S	.16	.46	.28	.23	.51	.31	.35
	IB	.07	-.06	.04	.18	.15	.05	.11
	K	.30	.67	.57	.31	.51	.49	.42
	P	.23	.49	.25	.20	.42	.37	.27
	FvF	.27	.57	.40	.12	.44	.34	.44
	S	.01	.11	-.15	-.11	.21	.02	.12

Anmerkung. gleicher Rater, verschiedene Beurteilte, $N = 61$, für S: $N = 24$.

Lebenslauf

Katrin Rathgeber
Schmähinger Str. 23
86756 Reimlingen
* 01.07.2005 in Nördlingen

Chemnitz, 02.02.2005

Abitur

1993	allgemeine Hochschulreife
1984 – 1993	Theodor-Heuss-Gymnasium in Nördlingen

Studium

1993 – 1999	Studium der Psychologie an der Otto-Friedrich-Universität Bamberg Abschluss: Diplom-Psychologin (Abschlussnote: sehr gut)
1997 – 1999	Diplomarbeit: „Eine ganze Welt zu zweit: Neuropsychologische Therapie des visuellen Neglects im Dyadensetting – Eine Einzelfallanalyse“

Ausbildung

1996 – 2000	Ausbildung zur „verhaltenstherapeutisch qualifizierten Diplompsychologin (AVM)“
derzeit	Ausbildung zur Psychologischen Psychotherapeutin am Zentrum für Psychotherapie Chemnitz e.V.

Psychologische Praktika

04/95	Sozial-psychiatrischer Dienst Donauwörth
02/96 – 04/96	Bezirkskrankenhaus Erlangen, Abteilung für Neurologische Rehabilitation
08/96 – 10/96	Baar Klinik Donaueschingen, Fachklinik für Psychosomatik & Verhaltensmedizin
04/97 – 10/97	Fachklinik Herzogenaurach, neurologische, orthopädische und kardiologische Rehabilitation
03/98 – 07/98	Forschungs- und Beratungsstelle Universität Bamberg

Tätigkeiten als Psychologin

12/99 – 03/00	Fachklinik Herzogenaurach; Neurologische, orthopädische und kardiologische Rehabilitation
01/02 – 12/02	Klinikum Chemnitz gGmbH; Klinik für Psychiatrie, Verhaltensmedizin und Psychosomatik (50%)
seit 04/00	TU Chemnitz; Wissenschaftliche Mitarbeiterin an der Professur für Wirtschafts-, Organisations- und Sozialpsychologie, seit 01/05 Wissenschaftliche Mitarbeiterin an der Professur für Klinische Psychologie
seit 2000	freiberufliche Tätigkeiten als Dozentin Aufbau des Zentrums für Psychotherapie Chemnitz e.V.

Erklärung

Hiermit erkläre ich, dass ich diese Dissertation selbständig verfasst habe und keine anderen als die angegebenen Hilfsmittel verwendet habe.

Dipl.-Psych. Katrin Rathgeber

Chemnitz, 02.02.2005