

Urteilsgenauigkeit von Lehrer(inne)n im emotional-motivationalen Bereich und im Leistungsbereich

Constance Karing und Cordula Artelt

Im beruflichen Alltag wird von Lehrer(inne)n eine Vielzahl von Diagnoseleistungen abverlangt. So fallen beispielweise diagnostische Urteile an, wenn es darum geht, die Leistung von Schüler(inne)n sowohl im Unterricht als auch am Schuljahresende zu beurteilen oder wenn es darum geht, den Unterricht gezielt zu planen und zu gestalten, um eine optimale Lernumgebung für die Schüler(innen) zu schaffen (Helmke, Hosenfeld & Schrader, 2004; Schrader, 2011). In der Forschung wird die Genauigkeit solcher Urteilsleistungen unter dem Stichwort „Diagnostische Kompetenz“ diskutiert (Artelt, 2009). Damit gemeint ist die Kompetenz der Lehrer(innen), Merkmale der Schüler(innen), wie deren Leistung, aber auch deren emotional-motivationalen Merkmale (z. B. Interesse), sowie die Lern- und Aufgabenanforderungen (z. B. Schwierigkeit von Aufgaben) zutreffend einzuschätzen (Artelt, 2009; Helmke et al., 2004).

Um die Genauigkeit solcher Urteile zu ermitteln, werden in der Forschung meistens Lehrerurteile über Schülermerkmale (z. B. Schülerleistung) mit den erfassten Schülermerkmalen (z. B. Leistungen der Schüler(innen) in einem Schulleistungstest) verglichen. Dieser Vergleich ist über verschiedene Maße, wie zum Beispiel die Rang-, Niveau- und Streuungskomponente, möglich. In den meisten wissenschaftlichen Studien wird in der Regel die Rangkomponente herangezogen. Diese Komponente ermöglicht es festzustellen, ob die Lehrer(innen) die Rangordnung zwischen verschiedenen Schüler(inne)n ihrer Klasse im Hinblick auf ein bestimmtes Merkmal (z. B. Schülerleistung) zutreffend wiedergeben können und somit zum Beispiel über Leistungsunterschiede in ihrer Klasse hinreichend informiert sind (Helmke et al., 2004; Schrader, 2008). Um diese Komponente zu berechnen, werden Korrelationen (r) zwischen den Lehrereinschätzungen und den bei den Schüler(inne)n erfassten Merkmalen ermittelt (Helmke et al., 2004; Schrader & Helmke, 1987). Diese Korrelationskoeffizienten können Werte zwischen -1 und $+1$ annehmen. Dabei bedeutet ein Wert von $+1$, dass eine perfekte Übereinstimmung zwischen Lehrerurteil und Schülermerkmal besteht, während bei einem Wert von -1 die von den Lehrer(inne)n angegebene Rangordnung genau entgegengesetzt der tatsächlichen Rangordnung der Schüler(innen) ist. Wohingegen ein Wert von 0 anzeigt, dass gar kein Zusammenhang zwischen Lehrereinschätzung und Schülermerkmal vorhanden ist. Je höher also die Übereinstimmung zwischen Lehrerurteil und den tatsächlichen Schülermerkmalen (je näher bei $+1$) ist, desto genauer fällt die Diagnose der Lehrer(innen) aus (Helmke, Hosenfeld & Schrader, 2003;

Karing & Artelt, 2013). Bisher haben verschiedene wissenschaftliche Studien gezeigt, dass Lehrer(innen) im Allgemeinen recht gut die Leistungen ihrer Schüler(innen) einschätzen können (vgl. Hoge & Coladarci, 1989: mittlere Korrelation: $r = 0,66$; Südkamp, Kaiser & Möller, 2012: mittlere Korrelation: $r = 0,53$), wobei auch diese mittleren Übereinstimmungen durchaus noch Luft nach oben aufweisen.

Darüber hinaus stellten Studien auch fest, dass Lehrer(innen) die Einschätzung von emotionalen und motivationalen Merkmalen der Schüler(innen) (z. B. Interesse, Ängstlichkeit) weniger gut gelingt als die Einschätzung von Leistungen. So finden sich hier nur geringe Zusammenhänge zwischen dem Lehrerurteil und bei den Schüler(inne)n erfassten Merkmalen (vgl. Spinath, 2005: Korrelationen von $r = 0,15$ für Leistungsängstlichkeit, $r = 0,20$ für Lernmotivation, $r = 0,39$ für schulische Fähigkeitsselbstwahrnehmung). Gerade aber das Interesse oder die Leistungsängstlichkeit der Schüler(innen) üben auch einen Einfluss auf die Leistung der Schüler(innen) aus und die korrekte Einschätzung – insbesondere von besonders niedrigen (bei Motivation) oder hohen (bei Leistungsängstlichkeit) Ausprägungen dieser Schülermerkmale – kann eine wichtige Voraussetzung für den adäquaten Umgang mit evtl. Problemen der Schüler(innen) in diesem Bereich darstellen. Insgesamt besteht die Tendenz, dass Schüler(innen) mit hohen leistungsbezogenen Ängsten unterdurchschnittliche Leistungen in der Schule aufweisen (Faber, 2006; Sparfeldt et al. 2005).

Auch in der BiKS-Studie konnten wir (mittlere bzw. niedrige) Zusammenhänge für die beiden Komponenten der Leistungsängstlichkeit, die Aufgeregtheit und Besorgtheit, mit der Schülerleistung feststellen. Dabei bezieht sich die Aufgeregtheit auf die subjektive Wahrnehmung einer Erregung, wie es zum Beispiel in der Aussage "Wenn ich im Mathematikunterricht aufgerufen werde, werde ich nervös" zum Ausdruck kommt. Dagegen bezieht sich die Besorgtheit auf aufgabenirrelevante Überlegungen vor und während der Auseinandersetzung der Schüler(innen) mit Leistungssituationen, wie zum Beispiel vor Mathematikarbeiten: "Vor der Mathematikarbeit habe ich Sorgen, dass ich alles vergesse, was ich geübt habe", die die Schüler(innen) im Rahmen der BiKS-Erhebung beantworteten. Übereinstimmend mit der Literatur stellten wir fest, dass die Besorgtheit der Schüler(innen) insgesamt stärker mit der Leistung im jeweiligen Schulfach zusammenhängt als die Aufgeregtheit.

Ein weiterer Befund, der sich in den meisten Studien zur Urteilsgenauigkeit von Lehrer(inne)n zeigte, sind die großen Unterschiede zwischen Lehrer(inne)n hinsichtlich der Genauigkeit ihrer Urteile im Leistungsbereich sowie im emotional-motivationalen Bereich. Es gibt Lehrer(innen), die sehr gut die Schülerleis-

tung und/oder emotionalen und motivationalen Merkmale ihrer Schüler(innen) einschätzen können, während es anderen Lehrer(inne)n eher schwer fällt, diese Merkmale ihrer Schüler(innen) zu beurteilen.

Befunde zur Urteilsgenauigkeit der Lehrer(innen) im Leistungsbereich und im emotional-motivationalen Bereich im Rahmen der BiKS-Studie

Im Rahmen der BiKS-Studie untersuchten wir, wie akkurat die Lehrer(innen) der Sekundarstufe die Leistungen ihrer Schüler(innen) sowie deren emotionale und motivationale Merkmale einschätzen können. Dazu werden im Folgenden die Befunde zur Urteilsgenauigkeit im Bereich Mathematik und Lesen/Textverstehen sowie im emotional-motivationalen Bereich für das Fachinteresse, die fachspezifische Leistungsängstlichkeit (Aufgeregtheit und Besorgtheit) und die fachspezifische Kompetenzüberzeugung jeweils in Deutsch und Mathematik von der 5. bis zur 7. Jahrgangsstufe berichtet.

Um die Genauigkeit dieser Urteile zu erfassen, wurden die Lehrer(innen) gebeten, die einzelnen Schüler(innen) ihrer Klasse hinsichtlich ihrer Leistungen in Mathematik und im Lesen/Textverstehen sowie hinsichtlich ihres Fachinteresses, ihrer fachspezifischen Leistungsängstlichkeit und ihrer Kompetenzüberzeugung einzuschätzen. Dafür bekamen die Lehrer(innen) Aussagen vorgelegt, wie zum Beispiel für die Einschätzung der Lesekompetenz des jeweiligen Schüler(in)s: „Er/Sie kann Texte gut verstehen.“ oder für die Beurteilung der Kompetenzüberzeugung: „Er/Sie hält sich in Deutsch für kompetent.“. Die Lehrer(innen) schätzten diese Aussagen anhand einer fünfstufigen Skala ein (z. B. „gar nicht“ bis „sehr“).

Zum anderen wurde die Leistung der Schüler(innen) sowie deren fachbezogenes Interesse, ihre Leistungsängstlichkeit und ihre Kompetenzüberzeugung erhoben. So wurde beispielsweise die Lesekompetenz der Schüler(innen) mittels eines Lesekompetenztests erfasst, der verschiedene Textsorten (Sach- und Prosatexte) umfasste. Die Schüler(innen) mussten den jeweiligen Text lesen, bedeutsame Informationen im Text suchen und entsprechende Schlussfolgerungen aus dem Text ziehen, um im Anschluss dazugehörige Multiple-Choice-Aufgaben zu beantworten. Die emotional-motivationalen Merkmale der Schüler(innen) wurden dagegen mittels Fragebögen erhoben. So schätzten zum Beispiel die Schüler(innen) ihr fachspezifisches Interesse anhand von je vier Aussagen ein (z. B. „Wie sehr freust du dich auf eine Stunde im Fach Deutsch?“, „Wie gerne würdest du im Fach Deutsch noch mehr Stunden haben als bisher?“, vgl. Baumert, Gruehn, Heyn, Köller & Schnabel, 1997). Diese Aussagen beantworteten die Schüler(innen) ebenfalls auf einer fünfstufigen Skala (von „gar nicht“ bis „sehr“).

Um nun die Frage zu beantworten, wie genau die Lehrerurteile ausfallen, wurde der Zusammenhang zwischen den Lehrereinschätzungen und den tatsächlichen Merkmalsausprägungen der Schüler(innen) ermittelt. Dafür wurden klassenspezifische Korrelationen (Rangkomponente) berechnet. In der BiKS-Studie konnten wir für die Genauigkeit der Urteile der Sekundarschullehrer(innen) für den Leistungsbereich und den emotional-motivationalen Bereich von der 5. bis zur 7. Jahrgangsstufe folgende Befunde feststellen:

- Lehrer(innen) der Sekundarstufe schätzen die Leistungen ihrer Schüler(innen) in Mathematik akkurater ein als deren Leistungen im Lesen/Textverstehen. So sind die Mathematiklehrer(innen) im Mittel besser über die Leistungsunterschiede in ihrer Klasse informiert als die Deutschlehrer(innen) über die Leistungsunterschiede im Bereich Lesen/Textverstehen.
- Die Einschätzung von emotionalen und motivationalen Merkmalen, wie das fachspezifische Interesse oder die Aufgeregtheit der Schüler(innen) in Deutsch und Mathematik, fällt den Lehrer(inne)n schwer. Wir finden nur geringe Zusammenhänge zwischen Lehrerurteil und den tatsächlichen Ausprägungen dieser Schülermerkmale. In der Regel ist die Urteilsgenauigkeit der Lehrer(innen) in diesem Bereich niedriger als im Leistungsbereich.
- Jedoch zeigen sich auch im emotional-motivationalen Bereich fachspezifische Unterschiede. So schätzen die Mathematiklehrer(innen) sowohl die Kompetenzüberzeugung als auch die Besorgtheit ihrer Schüler(innen) im Fach Mathematik akkurater ein als die Deutschlehrer(innen). Insbesondere die Kompetenzüberzeugung der Schüler(innen) in Mathematik können die Lehrer(innen) genauso gut einschätzen wie deren Leistung in Mathematik.
- Sowohl für den Leistungsbereich als auch für den emotional-motivationalen Bereich stellen wir große Unterschiede zwischen den einzelnen Lehrer(inne)n fest. So gibt es Lehrer(innen), die sehr gute Diagnostiker sind, aber auch Lehrer(innen), deren Urteile in beiden Bereichen nur sehr unzutreffend sind.

Exemplarisch sind die Befunde für die Urteilsgenauigkeit der Lehrer(innen) in Tabelle 1 für die 7. Jahrgangsstufe dargestellt.

Tabelle 1: Urteilsgenauigkeit der Lehrer(innen) im Leistungsbereich und im emotional-motivationalen Bereich für die 7. Jahrgangsstufe

	Mathematiklehrkräfte		Deutschlehrkräfte	
	M	SD	M	SD
Schülerleistung	0,53	0,46	0,37	0,40
<i>Emotional-motivationale Schülermerkmale</i>				
Fachinteresse	0,33	0,42	0,27	0,36
Leistungsängstlichkeit				
Aufgeregtheit	0,31	0,44	0,22	0,45
Besorgtheit	0,41	0,46	0,23	0,41
Kompetenzüberzeugung	0,51	0,49	0,34	0,40
<p><i>Anmerkungen:</i> Die Urteilsgenauigkeit der Lehrer(innen) wurde über die Korrelation zwischen Lehrereinschätzung und Schülermerkmal ermittelt (Rangkomponente). Es handelt sich hierbei um durchschnittliche Korrelationen (M) für die Gesamtstichprobe der Mathematik- bzw. Deutschlehrer(innen). SD = Streuung, diese zeigt die Unterschiede zwischen den einzelnen Lehrer(inne)n hinsichtlich ihrer Urteilsgenauigkeit an. Sehr hohe Werte (SD) zeigen sehr große Unterschiede zwischen Lehrer(inne)n an. Dagegen bedeutet zum Beispiel ein Wert von 0, dass kein Unterschied zwischen Lehrer(inne)n hinsichtlich ihrer Urteilsgenauigkeit besteht, d. h. alle Lehrer(innen) würden denselben Korrelationswert aufweisen und damit dieselbe Urteilsgenauigkeit. Die Daten beziehen sich auf insgesamt 79 Mathematiklehrer(innen) (13 Hauptschul-, 18 Realschul- und 48 Gymnasiallehrer(innen)) und deren 818 Schüler(innen) und zudem auf 74 Deutschlehrer(innen) (12 Hauptschul-, 15 Realschul- und 47 Gymnasiallehrer(innen)) und deren 794 Schüler(innen).</p>				

Interpretation der Befunde und Empfehlungen

Unsere Befunde im Rahmen der BiKS-Studie zeigen, dass Mathematiklehrer(innen) im Großen und Ganzen schon recht gut die Leistungen ihrer Schüler(innen) einschätzen, während es den Deutschlehrer(inne)n nicht so gut gelingt. Wie kann dieser Befund erklärt werden?

Die Genauigkeit von Urteilen ist zum Beispiel davon abhängig, was die Lehrer(innen) unter dem erfragten Schülermerkmal verstehen, welche Informationen sie für ihre Urteilsbildung heranziehen und auch welche Kompetenzen der Leistungstest bei den Schüler(inne)n tatsächlich erfasst. So kann eine größere Übereinstimmung zwischen dem Urteil der Mathematiklehrer(innen) und den erfassten Schülerleistungen in Mathematik schon deshalb vorliegen, weil die Mathematiklehrer(innen) zum einen konkret danach gefragt wurden, ob zum Beispiel der Schüler(in) Rechenaufgaben bzw. Geometrieaufgaben gut lösen konnte. Da schon aufgrund des Lehrplans in Mathematik klar definiert ist, welche mathematischen Kenntnisse und Fähigkeiten die Schüler(innen) bis zum Ende eines Schuljahres erwerben sollen (z. B. Multiplikation und Division na-

türlicher Zahlen in der 5. Jahrgangsstufe, Staatsinstitut für Schulqualität und Bildungsforschung, 2013), wissen die Mathematiklehrer(innen) genau, welche Aufgaben die Schüler(innen) in Mathematik schon bewältigen können. Zum anderen wurden genau solche Aufgaben, entsprechend des Lehrplans der jeweiligen Jahrgangsstufe, in den BiKS-Leistungstests von den Schüler(inne)n abverlangt.

Dagegen könnte die geringe Übereinstimmung zwischen dem Urteil der Deutschlehrer(innen) und der erfassten Schülerleistung darauf zurückzuführen sein, dass Deutschlehrer(innen) möglicherweise ein weiteres Verständnis vom Textverstehen haben (z. B. Literalitätskonzept zugrundelegen), als es der in der BiKS-Studie erfassten Kompetenz entspricht, was wiederum dazu führen kann, dass deren Urteilsgenauigkeit unterschätzt wird (Artelt, 2009).

Auch die Einschätzung der emotionalen und motivationalen Schülermerkmale ist davon abhängig, was der Lehrer(in) unter diesen Merkmalen versteht und welche Informationen er für sein Urteil verwendet (Schrader, 2009). Da diese Merkmale, wie das Fachinteresse und die Leistungsängstlichkeit der Schüler(innen), nicht direkt beobachtbar sind, muss der Lehrer(in) für die Beurteilung dieser Merkmale beobachtbare Indikatoren heranziehen. Wenn der Lehrer(in) bei der Beurteilung dieser Merkmale ungeeignete Indikatoren verwendet, so kann auch sein Urteil nur unzutreffend ausfallen (Helmke, 2009). Stützt sich zum Beispiel der Lehrer(in) bei der Beurteilung des Fachinteresses des Schüler(in)s auf die Häufigkeit des spontanen Meldens des Schülers(in) im Unterricht als Anzeichen für ein hohes Interesse am Fach, dann verwendet er keinen guten Indikator für dieses Schülermerkmal und lässt geeignetere Indikatoren, wie zum Beispiel die freiwillige Auseinandersetzung mit den Inhalten dieses Faches, außer Acht (Helmke, 2009). Erschwert wird die Genauigkeit solcher Diagnosen noch dadurch, dass häufig für diese Merkmale sehr mehrdeutige und auch für den Lehrer(in) schwer erkennbare Indikatoren vorliegen (Schrader, 2010).

Welche Empfehlungen können wir den Lehrer(inne)n geben, um die Genauigkeit ihrer Urteile zu verbessern? Eine wertvolle Hilfe für akkuratere Urteile sind objektive Informationen über die Merkmale der Schüler(innen). So können zum Beispiel Lehrer(innen) diagnostische Verfahren und Tests verwenden, um die Leistung ihrer Schüler(innen) oder deren emotionalen und motivationalen Merkmale, wie die Leistungsängstlichkeit, einzuschätzen (Karing & Artelt, 2013; Schrader & Helmke, 2001). Voraussetzung dafür ist, dass entsprechende Verfahren bekannt und zugänglich sind sowie auch selbstständig angewendet werden können. In gezielten Weiterbildungsmaßnahmen könnte Lehrer(inne)n

dieses Wissen zu diagnostischen Verfahren sowie deren Anwendbarkeit vermittelt werden. Darüber hinaus wäre es auch erforderlich, Lehrer(innen) für spezifische und valide Indikatoren der Schülerleistung und der emotionalen und motivationalen Schülermerkmale zu sensibilisieren, damit sie wissen, welche Indikatoren überhaupt für ein akkurates Urteil geeignet sind (Helmke et al., 2004).

Abgesehen davon, können Lehrer(innen) bei der Beurteilung ihrer Schüler(innen) aber auch Urteilsfehler und -tendenzen unterliegen, die die Genauigkeit ihrer Urteile beeinträchtigen (Karing & Artelt, 2013; Helmke, 2009). Typische Urteilsfehler und -tendenzen sind zum Beispiel der Strenge- und Mildeeffekt, der Halo-Effekt oder der logische Fehler (vgl. Helmke, 2009). Beim Strengeeffekt und Mildeeffekt beurteilt der Lehrer(in) die Leistung der Schüler(innen) durchweg strenger bzw. günstiger, als dies tatsächlich der Fall ist. Dagegen wird beim Halo-Effekt aufgrund weniger unpassender Indikatoren (z. B. Aussehen, Kleidung des Schüler(in)s) auf globale Merkmale des Schüler(in)s (z. B. Einschätzung der Gesamtleistung in mündlicher Prüfung, Persönlichkeit) geschlossen. Unterliegt hingegen der Lehrer(in) einem logischen Fehler, so schließt er von der Ausprägung eines Schülermerkmals (z. B. Rechtschreibprobleme) auf ein anderes Merkmal des Schülers(in)s (z. B. Indiz für geringe Intelligenz), ohne dass dies empirisch begründet wäre (Karing & Artelt, 2013; Helmke, 2009). Auch hier sollte im Rahmen von Lehrerfortbildungen entsprechendes Wissen über diese Beurteilungsfehler vermittelt werden und Lehrer(innen) für solche Urteilsfehler und -tendenzen sensibilisiert werden.

Literatur

- Artelt, C. (2009). Diagnostische Urteile von Lehrkräften im Bereich der Lesekompetenz. In A. Bertschi-Kaufmann & C. Rosebrock (Hrsg.), *Literalität. Bildungsaufgabe und Forschungsfeld* (S. 125-136). Weinheim: Juventa Verlag.
- Artelt, C. & Gräsel, C. (2009). Gasteditorial. Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23(3-4), 157-160.
- Baumert, J., Gruehn, S., Heyn, S. Köller, O. & Schnabel. K.-U. (1997). *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU). Dokumentation Bd 1. Skalen Längsschnitt I, Wellen 1-4*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Faber, G. (2006). Die Erfassung rechtschreibängstlicher Besorgtheit und Aufgeregtheit. Zur Bedeutung ausgewählter Forschungsergebnisse für lerntherapeutische Diagnose- und Interventionskonzepte. *Sprachrohr Lerntherapie. Zeitschrift für integrative Lerntherapie*, 2, 5-14.
- Helmke, A. (2009). *Unterrichtsqualität: Erfassen, Bewerten, Verbessern*. Seelze: Kallmeyer.

- Helmke, A., Hosenfeld, I. & Schrader, F. W. (2003). Diagnosekompetenz in Ausbildung und Beruf entwickeln. *Karlsruher pädagogische Beiträge*, 55, 15-34.
- Helmke, A., Hosenfeld, I. & Schrader, F. W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griesse (Hrsg.), *Schulleitung und Schulentwicklung* (S. 119-144). Hohengehren: Schneider Verlag.
- Hoge, R. D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Psychology*, 59(3), 297-313.
- Karing, C. & Artelt, C. (2013). Genauigkeit von Lehrerurteilen und deren Förderung in der Aus- und Weiterbildung von Lehrkräften. *Beiträge zur Lehrerbildung*, 31(2), 166-173.
- Schrader, F.-W. (2008). Diagnoseleistungen und diagnostische Kompetenzen von Lehrkräften. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 168-177). Göttingen: Hogrefe.
- Schrader, F.-W. (2010). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 102-108). Weinheim: Beltz.
- Schrader, F.-W. (2011). Lehrer als Diagnostiker. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch zur Forschung zum Lehrerberuf* (S. 683-698). Münster: Waxmann.
- Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1(1), 27-52.
- Schrader, F.-W. & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 45-58). Weinheim: Beltz.
- Sparfeldt, J.R., Schilling, S.R., Rost, D.H., Stelzl, I. & Peipert, D. (2005). Leistungsängstlichkeit: Facetten, Fächer, Fachfacetten? Zur Trennbarkeit nach Angstfacette und Inhaltsbereich. *Zeitschrift für Pädagogische Psychologie*, 19(4), 225-236.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19 (1/2), 85-95.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A Meta-Analysis. *Journal of Educational Psychology*, 104 (3), 743-762.
- Staatsinstitut für Schulqualität und Bildungsforschung (2013). Lehrplan für das Gymnasium in Bayern. Verfügbar unter: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26350>. Geprüft am 01.05.2014.