



# Sindh Province, Pakistan

SABER Province Report

## STUDENT ASSESSMENT

2012

### Key Policy Areas for Student Assessment

#### Status

#### 1. Classroom Assessment

The 2006 National Curriculum provides very general, non-comprehensive guidelines on classroom assessment. Although available online, this document is not offered through other channels (such as teacher training courses or libraries) that are easily accessible to teachers and other stakeholders. There are some system-level mechanisms in place (such as pre- and in-service teacher training opportunities) that are meant to ensure that teachers develop appropriate skills and expertise in classroom assessment. However, on-the-ground classroom assessment practices tend to be mainly about students recalling information, provide little feedback to students and parents, and are generally considered weak.



#### 2. Examinations

The Secondary School Certificate exam is administered to students in Grades 9 and 10. Examination results are used only for making decisions about the promotion of individual students to the next level of schooling. The Sindh Boards of Intermediate and Secondary Education Ordinance of 1972 and the Sindh Boards of Intermediate and Secondary Education (Amendment) Act of 2008 authorize the exam. The Examination Rules and Procedures document (known as the "Calendar") for each Sindh Board covers key aspects of the exam. Officially, the examination is meant to measure student attainment of the knowledge and skills outlined in the 2006 National Curriculum. However, in practice, it assesses textbook knowledge. Moreover, the textbooks are based on an earlier (2002) curriculum, not the more recent (2006) one. There is a permanent staff, but it is insufficient to meet the needs of the examination due to staff members' lack of technical capacity. Additionally, there is no technical documentation available on the examination.



#### 3. National Large-Scale Assessment (NLSA)

The Provincial Education Assessment Center (PEACE) Sindh Assessment was administered for the first time in 2008 as an independent, provincial-level assessment exercise. Previously, PEACE was part of the National Education Assessment System (NEAS). Under NEAS, Sindh administered the assessment to a provincially-representative sample of students and published a provincial report separate from the national report. Although the PEACE office is a permanent agency, it is inadequately staffed to effectively carry out PEACE activities. This also was the case when PEACE was part of NEAS. PEACE results and information have not been adequately used.



#### 4. International Large-Scale Assessment (ILSA)

Sindh has not participated in an ILSA, and it does not have plans to do so in the near future.



## Introduction

The Sindh province in Pakistan has focused on increasing student learning outcomes by improving the quality of education in the province. An effective student assessment system is an important component to improving education quality and learning outcomes as it provides the necessary information to meet stakeholders' decision-making needs. In order to gain a better understanding of the strengths and weaknesses of its existing student assessment system within the context of the province's Public Sector Education system, the Sindh province decided to benchmark this system using standardized tools developed under The World Bank's Systems Approach for Better Education Results (SABER) program. SABER is an evidence-based program to help countries systematically examine and strengthen the performance of different aspects of their education systems.

## What is SABER-Student Assessment?

SABER-Student Assessment is a component of the SABER program that focuses specifically on benchmarking student assessment policies and systems. The goal of SABER-Student Assessment is to promote stronger assessment systems that contribute to improved education quality and learning for all.

National governments and international agencies are increasingly recognizing the key role that assessment of student learning plays in an effective education system. The importance of assessment is linked to its role in:

- (i) providing information on levels of student learning and achievement in the system;
- (ii) monitoring trends in education quality over time;
- (iii) supporting educators and students with real-time information to improve teaching and learning; and
- (iv) holding stakeholders accountable for results.

## SABER-Student Assessment methodology

The SABER-Student Assessment framework is built on the available evidence base for what an effective assessment system looks like. The framework provides guidance on how countries can build more effective student assessment systems. The framework is structured around two main dimensions of assessment systems: the types/purposes of assessment activities and the quality of those activities.

### Assessment types and purposes

Assessment systems tend to be comprised of three main types of assessment activities, each of which serves a different purpose and addresses different information needs. These three main types are: classroom assessment, examinations, and large-scale, system level assessments.

*Classroom assessment* provides real-time information to support ongoing teaching and learning in individual classrooms. Classroom assessments use a variety of formats, including observation, questioning, and paper-and-pencil tests, to evaluate student learning, generally on a daily basis.

*Examinations* provide a basis for selecting or certifying students as they move from one level of the education system to the next (or into the workforce). All eligible students are tested on an annual basis (or more often if the system allows for repeat testing). Examinations cover the main subject areas in the curriculum and usually involve essays and multiple-choice questions.

*Large-scale, system-level assessments* provide feedback on the overall performance of the education system at particular grades or age levels. These assessments typically cover a few subjects on a regular basis (such as every 3 to 5 years), are often sample based, and use multiple-choice and short-answer formats. They may be national or international in scope.

Appendix 1 summarizes the key features of these main types of assessment activities.

## Quality drivers of an assessment system

The key considerations when evaluating a student assessment system are the individual and combined quality of assessment activities in terms of the adequacy of the information generated to support decision making. There are three main drivers of information quality in an assessment system: enabling context, system alignment, and assessment quality.

*Enabling context* refers to the broader context in which the assessment activity takes place and the extent to which that context is conducive to, or supportive of, the assessment. It covers such issues as the legislative or policy framework for assessment activities; institutional and organizational structures for designing, carrying out, or using results from the assessment; the availability of sufficient and stable sources of funding; and the presence of trained assessment staff.

*System alignment* refers to the extent to which the assessment is aligned with the rest of the education system. This includes the degree of congruence between assessment activities and system learning goals, standards, curriculum, and pre- and in-service teacher training.

*Assessment quality* refers to the psychometric quality of the instruments, processes, and procedures for the assessment activity. It covers such issues as design and implementation of assessment activities, analysis and interpretation of student responses to those activities, and the appropriateness of how assessment results are reported and used.

Crossing the quality drivers with the different assessment types/purposes provides the framework and broad indicator areas shown in Table 1. This framework is a starting point for identifying indicators that can be used to review assessment systems and plan for their improvement.

**Table 1: Framework for building an effective assessment system, with indicator areas**

	Assessment types/purposes		
	Classroom assessment	Examinations	Large-scale, system-level assessment
<b>Enabling context</b>	Policies Leadership and public engagement Funding Institutional arrangements Human resources		
<b>System alignment</b>	Learning/quality goals Curriculum Pre- and in-service teacher training opportunities		
<b>Assessment quality</b>	Ensuring quality (design, administration, analysis) Ensuring effective uses		

Source: World Bank.

The indicators are identified based on a combination of criteria, including:

- professional standards for assessment;
- empirical research on the characteristics of effective assessment systems, including analysis of the characteristics that differentiate between the assessment systems of low- versus high-performing nations; and
- theory — that is, general consensus among experts that it contributes to effective assessment.

## Levels of development

The World Bank has developed a set of standardized questionnaires and rubrics for collecting and evaluating data on the three assessment types and related quality drivers.

The questionnaires are used to collect data on the characteristics of the assessment system in a particular country or system. The information from the questionnaires is then applied to the rubrics in order to judge the development level of the country's assessment system in different areas.

The basic structure of the rubrics for evaluating data collected using the standardized questionnaires is summarized in Appendix 2. The goal of the rubrics is to provide a country with some sense of the development level of its assessment activities compared to best or recommended practice in each area. For each indicator, the rubric displays four development

levels—*Latent*, *Emerging*, *Established*, and *Advanced*. These levels are artificially constructed categories chosen to represent key stages on the underlying continuum for each indicator. Each level is accompanied by a description of what performance on the indicator looks like at that level.

- *Latent* is the lowest level of performance; it represents absence of, or deviation from, the desired attribute.
- *Emerging* is the next level; it represents partial presence of the attribute.
- *Established* represents the acceptable minimum standard.
- *Advanced* represents the ideal or current best practice.

A summary of the development levels for each assessment type is presented in Appendix 3.

In reality, assessment systems are likely to be at different levels of development in different areas. For example, a system may be *Established* in the area of examinations, but *Emerging* in the area of large-scale, system-level assessment, and vice versa. While intuition suggests that it is probably better to be further along in as many areas as possible, the evidence is unclear as to whether it is necessary to be functioning at *Advanced* levels in all areas. Therefore, one might view the *Established* level as a desirable minimum outcome to achieve in all areas, but only aspire beyond that in those areas that most contribute to the national vision or priorities for education. In line with these considerations, the ratings generated by the rubrics are not meant to be additive across assessment types (that is, they are not meant to be added to create an overall rating for an assessment system; they are only meant to produce an overall rating for each assessment type). The methodology for assigning development levels is summarized in Appendix 4.

## Education in the Sindh province

Pakistan is a lower-middle-income country in South Asia. Current GDP per capita (US, 2012) is \$1,257 with annual growth of 4 percent. Economic growth and

progress in poverty reduction have slowed significantly over the past five years due to a weak macroeconomic environment, political and security conditions, and catastrophic natural shocks.

Sindh is the second-largest province in Pakistan and has the second-highest income among the four provinces in the country. School participation in Sindh, as in the rest of the country, has shown very slow improvement over the past several years, especially for girls in rural areas. In 2010/2011, primary net enrollment rate was 62 percent and secondary net enrollment rate was 35 percent in Sindh. The level of student achievement is also low; a 2009 diagnostic assessment of grade 4 students in mathematics and language showed an average score of less than 50 percent.

The Sindh government recognizes poor sector governance and accountability, and weak administrative systems as principal problems in the education sector. To address these challenges, in 2006/07, the Government of Sindh initiated a multi-faceted Sindh Education Sector Reform Program (SERP) with support from the World Bank. The objectives of this program were to increase school participation and student progression, reduce gender and rural-urban disparities, and improve the measurement of student learning.

Detailed information was collected on Sindh province's student assessment system using the SABER-Student Assessment questionnaires and rubrics in 2012. This SABER-Student Assessment report reflects the status of the student assessment system in the Sindh province as of 2012. It is important to remember that these tools primarily focus on benchmarking a system's policies and arrangements for assessment activities at the system or macro level. Additional data would need to be collected to determine actual, on-the-ground practices in the Sindh province, particularly by teachers and students in schools. The following sections discuss the findings by each assessment type, accompanied by suggested policy options. The suggested policy options were determined in collaboration with key local stakeholders based on Sindh province's immediate interests and needs. Detailed, completed rubrics for each assessment type in the Sindh province are provided in Appendix 5.

## Classroom Assessment

---

### Level of Development: EMERGING

In the Sindh province, the *2006 National Curriculum* document provides general, non-comprehensive guidelines on classroom assessment. The document provides information on assessment purposes, types and questions, and includes some sample test content and scoring criteria. Although available online, this document is not offered through other more easily accessible channels for teachers and other stakeholders (such as teacher training courses or libraries).

The *2006 National Curriculum* document has not been provided to all practicing classroom teachers, and because not all teachers have access to the internet, many teachers are not aware of the guidelines provided in the document. In addition to the *2006 National Curriculum* document, there are scarce system-wide resources available for teachers to engage in classroom assessment activities. Existing resources include textbooks, which contain questions at the end of the chapters, and guidelines for teachers to use when conducting classroom assessment activities. However, the majority of these textbooks are based on a previous version of the national curriculum that was developed in 2002, and all textbooks have not been updated based on the 2006 curriculum.

There are some system-level mechanisms in place (such as pre- and in-service teacher training opportunities) that are meant to ensure that teachers develop appropriate skills and expertise in classroom assessment. For example, a course on assessment is also offered in both undergraduate and graduate education programs.

However, on-the-ground classroom assessment practices are generally considered to be weak. There are no required uses of classroom assessment information to support student learning. Classroom assessment practices tend to focus on students recalling information, and teachers often rely on the questions and the end-of-chapter exercises provided in the textbooks to assess students. Teachers generally do not use explicit criteria for evaluating students' work, and

tend to provide little feedback on student performance to students and parents. Classroom assessment results are often used as an administrative tool rather than a pedagogical resource as classroom assessment information is primarily used internally to promote or fail students.

There are currently no mechanisms in place to systematically monitor the quality of classroom assessment practices.

Suggested policy options:

1. Make the curriculum document more easily accessible to stakeholders through, for example, libraries, teacher training courses, and schools.
2. Develop detailed guidelines for teachers on how to carry out classroom assessment activities to support student learning.
3. Ensure more effective uses of classroom assessment information by: (i) communicating to teachers appropriate uses of the data; and (ii) making it compulsory for classroom assessment information to be disseminated to key stakeholders.

## Examinations

---

### Level of Development: EMERGING

The Secondary School Certificate examination is administered to students in Grades 9 and 10. The Sindh Boards of Intermediate and Secondary Education Ordinance of 1972 and the Sindh Boards of Intermediate and Secondary Education (Amendment) Act of 2008 authorize the examination. The Examination Rules and Procedures document (known as the “Calendar”) for each Sindh Board covers key aspects of the exam including the rules for preparation, the format of the examination, and the procedures to investigate inappropriate behavior.

The Secondary School Certificate examination is intended to measure student attainment of the knowledge and skills outlined in the 2006 National Curriculum. In practice, the examination tests textbook knowledge. At the same time, textbooks are not well aligned with the curricula due to two main issues: (1) majority of the textbooks in circulation are based on a previous version of the national curricula developed in 2002 (new textbooks based on the new curricula developed in 2006 are in various stages of development) and (2) textbooks are not perceived to adequately cover all aspects of the curricula such as higher order thinking skills. Thus, there is no clear link between the examination and the student learning outcomes or the skills identified in the curriculum, resulting in inadequate coverage and the lack of higher order skills tested.

Efforts to improve the examination have been led by the government and by donors. For example, several efforts to improve aspects of the examinations have been made by the Boards along with the Education Department, including the introduction of multiple-choice and short answer questions. The Boards have also tried to ensure that the examiners marking the papers are from the same language background as the language of the paper they are marking.

Funding for the examination is provided by student fees, and covers all core examination activities including

examination design, administration, data processing and data reporting.

There is a permanent staff responsible for running the examination, but it is insufficient to meet the needs of the examination due to staff members’ lack of technical capacity. Issues identified with the performance of the human resources responsible for the examination include item errors, repetition, limited skills tested and inadequate curriculum coverage.

In general, comprehensive material to help prepare for the examination is accessible to most students. Such material includes the textbook (which is considered the main resource) and previous examination papers (as there is little repetition of examination questions from year to year).

There is no technical documentation on the examination and there are no mechanisms in place to ensure the quality of the examination. Inappropriate behavior surrounding the examination process is moderate and includes leakage of the content of the paper prior to the examination, and copying and using unauthorized materials such as prepared answers and notes. As a result, the credibility of the examination and its results are often called into question by many stakeholders.

Suggested policy options:

1. Build the capacity of individuals responsible for examination activities, especially in the areas of test design, curriculum alignment, and report writing. This could be done by: (i) making training opportunities available within the Boards; and (ii) allocating funding for attending international courses or workshops.
2. Develop a technical report on the examination and disseminate it to key stakeholders.
3. Identify and agree upon the official purposes and possible uses of examination results and introduce mechanisms, such as an oversight committee, to monitor actual uses.
4. Institutionalize preventive and reactive mechanisms to address inappropriate behavior during the examination process. Ensure that the mechanisms

are clearly defined, and are communicated through various communication methods (for example, through documents, the radio, etc.) to all key stakeholders in a timely manner.

## **National Large-Scale Assessment (NLSA)**

### **Level of Development: EMERGING**

The Provincial Education Assessment Center (PEACE) Sindh Assessment was administered for the first time in 2008 as an independent, provincial-level assessment exercise. Previously, PEACE was part of the National Education Assessment System (NEAS). Under NEAS, the Sindh province administered the assessment to a provincially representative sample of students and published a provincial report separate from the national report.

Regular funding for PEACE is allocated by the provincial government. Funding covers core PEACE activities, but does not cover research and development activities. Various donors are also providing technical assistance for improving the design of the PEACE assessment and for the training of PEACE administrators.

Although the PEACE office is a permanent agency working under the Bureau of Curriculum, it is inadequately staffed to effectively carry out assessment activities. In general, there is a lack of qualified staff who can ensure that tasks are conducted in an efficient and timely manner. Specifically, there is lack of technical staff, psychometricians, and statisticians for data analysis. There is also lack of clarity on the role that focal persons in the districts are to have in test administration and monitoring. No final review of assessment instruments takes place, which results in errors when the instruments are printed; there is also a need for translation verification between Urdu and Sindhi.

The Sindh province offers some opportunities to prepare individuals for work on the PEACE. Graduate and non-graduate university courses on educational measurement and evaluation are made available. The University of Sindh offers a course on educational measurement and evaluation in its B.Ed and M.Ed programs. The Aga Khan University Institute for

Educational Development (AKU-IED) also offers a course on assessment at the M.Ed level.

The PEACE assessment measures competencies identified in the curriculum. However, since assessments are based on the 2006 curricula and the textbooks available are based on the previous curricula (developed in 2002), some skills assessed on PEACE are not available in the textbooks and could cause some confusion amongst teachers who only use the textbook to teach students.

Various mechanisms are in place to ensure the quality of the PEACE, including: (i) all administrators are trained according to a protocol; (ii) there is a standardized manual for administrators; (iii) a pilot is conducted before the main data collection takes place; (iv) all booklets are numbered; and (v) external and internal reviewers and observers are employed. Coding and scoring is a centralized activity, conducted in one place and overseen by persons involved in data analysis. Training of scorers, provision of code sheets and rubrics for open-ended items, and a 10 percent check of scoring by another person also takes place. Computerized data is also compared to the data entered manually to cross-check for errors.

In general, PEACE results are poorly disseminated, and their use has been minimal.

Suggested policy options:

1. Develop the capacity of PEACE staff by (i) introducing workshops and trainings on education measurement and evaluation, including, in particular, on data analysis; (ii) providing funding to PEACE staff to attend international programs on educational measurement and evaluation; and (iii) identifying external human resources for conducting translation verifications between Urdu and Sindhi. Introduce a mandate for capacity development of PEACE staff. Identify which individuals or organizations will be responsible for the workshops and trainings, and provide them with regular funding for and access to the necessary training resources and key experts (at the provincial, national, and international levels).

2. Identify, with the participation of relevant persons at the district level, the intended uses of PEACE results; develop and implement a plan to disseminate the results; introduce mechanisms, such as an oversight committee, to monitor actual uses of PEACE results.

## **International Large-Scale Assessment (ILSA)**

---

### **Level of Development: LATENT**

The Sindh province has not participated in an ILSA, and it does not have plans to do so in the near future.

Suggested policy options:

1. Create an opportunity for high-level discussion among key stakeholders on key education policy questions or problems for which ILSA data could be useful.
2. Determine the need for, and possible next steps in relation to, participation in an ILSA exercise.



### Appendix 1: Assessment Types and Their Key Differences

	Classroom	Large-scale assessment Surveys		Examinations	
		National	International	Exit	Entrance
<b>Purpose</b>	To provide immediate feedback to inform classroom instruction	To provide feedback on overall health of the system at particular grade/age level(s), and to monitor trends in learning	To provide feedback on the comparative performance of the education system at particular grade/age level(s)	To certify students as they move from one level of the education system to the next (or into the workforce)	To select students for further educational opportunities
<b>Frequency</b>	Daily	For individual subjects offered on a regular basis (such as every 3-5 years)	For individual subjects offered on a regular basis (such as every 3-5 years)	Annually and more often where the system allows for repeats	Annually and more often where the system allows for repeats
<b>Who is tested?</b>	All students	Sample or census of students at a particular grade or age level(s)	A sample of students at a particular grade or age level(s)	All eligible students	All eligible students
<b>Format</b>	Varies from observation to questioning to paper-and-pencil tests to student performances	Usually multiple choice and short answer	Usually multiple choice and short answer	Usually essay and multiple choice	Usually essay and multiple choice
<b>Coverage of curriculum</b>	All subject areas	Generally confined to a few subjects	Generally confined to one or two subjects	Covers main subject areas	Covers main subject areas
<b>Additional information collected from students?</b>	Yes, as part of the teaching process	Frequently	Yes	Seldom	Seldom
<b>Scoring</b>	Usually informal and simple	Varies from simple to more statistically sophisticated techniques	Usually involves statistically sophisticated techniques	Varies from simple to more statistically sophisticated techniques	Varies from simple to more statistically sophisticated techniques

**Appendix 2: Basic Structure of Rubrics for Evaluating Data Collected on a Student Assessment System**

Dimension	Development Level				Justification
	LATENT (Absence of, or deviation from, attribute)	EMERGING (On way to meeting minimum standard)	ESTABLISHED (Acceptable minimum standard)	ADVANCED (Best practice)	
<b>EC—ENABLING CONTEXT</b>					
EC1—Policies					
EC2—Leadership, public engagement					
EC3—Funding					
EC4—Institutional arrangements					
EC5—Human resources					
<b>SA—SYSTEM ALIGNMENT</b>					
SA1—Learning/quality goals					
SA2—Curriculum					
SA3—Pre-, in-service teacher training					
<b>AQ—ASSESSMENT QUALITY</b>					
AQ1—Ensuring quality (design, administration, analysis)					
AQ2—Ensuring effective uses					

### Appendix 3: Summary of the Development Levels for Each Assessment Type

Assessment Type	LATENT	EMERGING	ESTABLISHED	ADVANCED
	<i>Absence of, or deviation from, the attribute</i>	<i>On way to meeting minimum standard</i>	<i>Acceptable minimum standard</i>	<i>Best practice</i>
CLASSROOM ASSESSMENT	There is no system-wide institutional capacity to support and ensure the quality of classroom assessment practices.	There is weak system-wide institutional capacity to support and ensure the quality of classroom assessment practices.	There is sufficient system-wide institutional capacity to support and ensure the quality of classroom assessment practices.	There is strong system-wide institutional capacity to support and ensure the quality of classroom assessment practices.
EXAMINATIONS	There is no standardized examination in place for key decisions.	There is a partially stable standardized examination in place, and a need to develop institutional capacity to run the examination. The examination typically is of poor quality and is perceived as unfair or corrupt.	There is a stable standardized examination in place. There is institutional capacity and some limited mechanisms to monitor it. The examination is of acceptable quality and is perceived as fair for most students and free from corruption.	There is a stable standardized examination in place and institutional capacity and strong mechanisms to monitor it. The examination is of high quality and is perceived as fair and free from corruption.
NATIONAL (OR SYSTEM-LEVEL) LARGE-SCALE ASSESSMENT	There is no NLSA in place.	There is an unstable NLSA in place and a need to develop institutional capacity to run the NLSA. Assessment quality and impact are weak.	There is a stable NLSA in place. There is institutional capacity and some limited mechanisms to monitor it. The NLSA is of moderate quality and its information is disseminated, but not always used in effective ways.	There is a stable NLSA in place and institutional capacity and strong mechanisms to monitor it. The NLSA is of high quality and its information is effectively used to improve education.
INTERNATIONAL LARGE-SCALE ASSESSMENT	There is no history of participation in an ILSA nor plans to participate in one.	Participation in an ILSA has been initiated, but there still is need to develop institutional capacity to carry out the ILSA.	There is more or less stable participation in an ILSA. There is institutional capacity to carry out the ILSA. The information from the ILSA is disseminated, but not always used in effective ways.	There is stable participation in an ILSA and institutional capacity to run the ILSA. The information from the ILSA is effectively used to improve education.

#### Appendix 4: Methodology for Assigning Development Levels

1. The country team or consultant collects information about the assessment system in the country or system.

2. Based on the collected information, a level of development and score is assigned to each dimension in the rubrics:

- Latent = 1 score point
- Emerging = 2 score points
- Established = 3 score points
- Advanced = 4 score points

3. The score for each quality driver is computed by aggregating the scores for each of its constituent dimensions. For example:

The quality driver, 'Enabling Context,' in the case of ILSA, has 3 dimensions on which a hypothetical country or system receives the following scores: Dimension A = 2 points; Dimension B = 2 points; Dimension C = 3 points. The hypothetical country's or system's overall score for this quality driver would be:  $(2+2+3)/3 = 2.33$

4. A preliminary level of development is assigned to each quality driver.

5. The preliminary development level is validated using expert judgment in cooperation with the country team and The World Bank Task Team Leader.

For scores that allow a margin of discretion (i.e., to choose between two levels of development), a final decision has to be made based on expert judgment. For example, the aforementioned hypothetical country or system has an 'Enabling Context' score of 2.33, corresponding to a preliminary level of development of 'Emerging or Established.' Based on qualitative information not captured in the rubric, along with expert judgment, the country team chooses 'Emerging' as the most appropriate level.

6. Scores for certain key dimensions under 'Enabling Context' (in the case of EXAM, NLSA, and ILSA) and under 'System Alignment' (in the case of CLASS) were set as ceiling scores, i.e., the overall mean score for the particular assessment type cannot be greater than

the score for these key dimensions. These key variables include formal policy, regular funding, having a permanent assessment unit, and the quality of assessment practices.

**Appendix 5: SABER-Student Assessment Rubrics Sindh Province, Pakistan**





This appendix provides the completed SABER-Student Assessment rubrics for each type of assessment activity in the Sindh province, Pakistan. In each row of the rubric, the relevant selection is indicated by a thick border and an asterisk. The selection may include a superscript number that refers to the justification or explanation for the selection (as indicated by a thick border and an asterisk). The explanation or justification text can be located in the “Development-level rating justifications” section at the end of each rubric. If a row includes a superscript, but not a thick border and an asterisk, this means that insufficient information was available to determine the relevant selection in the row.

## **SINDH PROVINCE, PAKISTAN**

### *Classroom Assessment*





**ENABLING CONTEXT AND SYSTEM ALIGNMENT**

*Overall policy and resource framework within which classroom assessment activity takes place in a country or system, and the degree to which classroom assessment activity is coherent with other components of the education system.*

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
<b>ENABLING CONTEXT AND SYSTEM ALIGNMENT 1: Setting clear guidelines for classroom assessment</b>			
There is no system-level document that provides guidelines for classroom assessment.	There is an informal system-level document that provides guidelines for classroom assessment. <sup>1</sup> *	There is a formal system-level document that provides guidelines for classroom assessment.	This option does not apply to this dimension.
This option does not apply to this dimension.	This option does not apply to this dimension.	The availability of the document is restricted. <sup>2</sup>	The document is widely available.
<b>ENABLING CONTEXT AND SYSTEM ALIGNMENT 2: Aligning classroom assessment with system learning goals</b>			
There are no system-wide resources for teachers for classroom assessment.	There are scarce system-wide resources for teachers for classroom assessment. <sup>3</sup> *	There are some system-wide resources for teachers for classroom assessment.	There are a variety of system-wide resources available for teachers for classroom assessment.
There is no official curriculum or standards document.	There is an official curriculum or standards document, but it is not clear what students are expected to learn or to what level of performance.	There is an official curriculum or standards document that specifies what students are expected to learn and to what level of performance. <sup>4</sup> *	There is an official curriculum or standards document that specifies what students are expected to learn and to what level of performance.
<b>ENABLING CONTEXT AND SYSTEM ALIGNMENT 3: Having effective human resources to carry out classroom assessment activities</b>			
There are no system-level mechanisms to ensure that teachers develop skills and expertise in classroom assessment.	This option does not apply to this dimension.	There are some system-level mechanisms to ensure that teachers develop skills and expertise in classroom assessment. <sup>5</sup> *	There are a variety of system-level mechanisms to ensure that teachers develop skills and expertise in classroom assessment.

## ASSESSMENT QUALITY

Quality of classroom assessment design, administration, analysis, and use.

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
<b>ASSESSMENT QUALITY 1:</b> <i>Ensuring the quality of classroom assessment</i>			
Classroom assessment practices suffer from widespread weaknesses or there is no information available on classroom assessment practices.	Classroom assessment practices are known to be weak. <sup>6</sup> *	Classroom assessment practices are known to be of moderate quality.	Classroom assessment practices are known to be generally of high quality.
There are no mechanisms to monitor the quality of classroom assessment practices. *	There are ad hoc mechanisms to monitor the quality of classroom assessment practices.	There are limited systematic mechanisms to monitor the quality of classroom assessment practices.	There are varied and systematic mechanisms in place to monitor the quality of classroom assessment practices.
<b>ASSESSMENT QUALITY 2:</b> <i>Ensuring effective uses of classroom assessment</i>			
Classroom assessment information is not required to be disseminated to key stakeholders.	This option does not apply to this dimension.	Classroom assessment information is required to be disseminated to some key stakeholders. <sup>7</sup> *	Classroom assessment information is required to be disseminated to all key stakeholders.
There are no required uses of classroom assessment to support student learning. *	There are limited required uses of classroom assessment to support student learning.	There are adequate required uses of classroom assessment to support student learning, excluding its use as an input for external examination results.	There are adequate required uses of classroom assessment to support student learning, including its use as an input for external examination results.

**Classroom Assessment:** [Development-level rating justifications](#)

1. The only system-level document that provides some recommendations on assessment practices is the 2006 National Curriculum, which provides general information on assessment purposes, types, and questions, as well as some sample test content and scoring criteria. These are not, however, comprehensive guidelines. Additionally, teachers in schools have not been provided the curriculum document, for the most part, and are therefore not aware of these guidelines.

2. The document is available online. However, the document is not available through other channels, such as libraries and teacher training courses and colleges, which teachers and other stakeholders can easily access.

3. Textbooks or workbooks that provide support for classroom assessment are typically available on a system-wide basis to teachers. Textbooks contain questions at the end of the chapter; there also are guidelines for teachers on conducting class activities for assessment purposes. For example, it suggests that teachers hold speech competitions, form groups and give them various tasks, etc.

The 2006 National Curriculum document, although available on the Ministry of Education website, has not been provided to all practicing classroom teachers. Not all teachers have access to the internet and can download large documents. The textbook board has yet to develop new textbooks based on the 2006 curriculum.

4. Currently, this is the 2006 National Curriculum. After the 18th constitutional amendment, the authority to develop curriculum has been transferred to the provinces, and Sindh is in the final stages of adopting a new curriculum.

5. Available system-level mechanisms include pre-service and in-service teacher training. A course on assessment is offered in the B.Ed and M.Ed program.

6. It is very common for classroom assessment activities to be mainly about recalling information. Teachers mostly develop questions directly from the textbook and use the exercises at the end of the chapters. Teachers also do not use explicit or a priori criteria for scoring or grading students' work. Parents are poorly informed about students' grades, which, in turn, are very school-dependent. Sharing annual exam results with parents, however, is very common. Classroom assessment activities also provide little useful feedback to students, as students are told only their overall test results. Classroom assessment activities are also mainly used as an administrative or control tool rather than a pedagogical resource, and for the most part are used internally to promote or fail students. (Textbooks are generally not aligned with the curricula because the textbooks in circulation are based on a previous version of the national curricula that was developed in 2002. There are no new textbooks that are based on the new curricula that went into effect in 2006.)

7. Results are reported to the district education department.







# **SINDH PROVINCE, PAKISTAN**

## ***Examinations***

## ENABLING CONTEXT

*Overall framework of policies, leadership, organizational structures, fiscal and human resources in which assessment activity takes place in a country or system and the extent to which that framework is conducive to, or supportive of, the assessment activity.*

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
<b>ENABLING CONTEXT 1:</b> <i>Setting clear policies</i>			
No standardized examination has taken place.	The standardized examination has been operating on an irregular basis.	The examination is a stable program that has been operating regularly. *	This option does not apply to this dimension.
There is no policy document that authorizes the examination.	There is an informal or draft policy document that authorizes the examination.	There is a formal policy document that authorizes the examination. <sup>1</sup> *	This option does not apply to this dimension.
This option does not apply to this dimension.	The policy document is not available to the public. <sup>2</sup> *	The policy document is available to the public.	This option does not apply to this dimension.
This option does not apply to this dimension.	This option does not apply to this dimension.	The policy document addresses some key aspects of the examination. <sup>3</sup> *	The policy document addresses all key aspects of the examination.
<b>ENABLING CONTEXT 2:</b> <i>Having strong leadership</i>			
All stakeholder groups strongly oppose the examination or are indifferent to it. <sup>4</sup>	Most stakeholder groups oppose the examination.	Most stakeholders groups support the examination.	All stakeholder groups support the examination.
There are no attempts to improve the examination by stakeholder groups.	This option does not apply to this dimension.	There are independent attempts to improve the examination by stakeholder groups. <sup>5</sup> *	There are coordinated attempts to improve the examination by stakeholder groups.
Efforts to improve the examination are not welcomed by the leadership in charge of the examination. <sup>6</sup>	This option does not apply to this dimension.	Efforts to improve the examination are generally welcomed by the leadership in charge of the examination.	This option does not apply to this dimension.

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
<b>ENABLING CONTEXT 3:</b> <i>Having regular funding</i>			
There is no funding allocated for the examination.	There is irregular funding allocated for the examination. <sup>7</sup> *	There is regular funding allocated for the examination.	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding covers some core examination activities: design, administration, data processing or reporting.	Funding covers all core examination activities: design, administration, data processing and reporting. <sup>8</sup> *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding does not cover research and development. <sup>9</sup> *	This option does not apply to this dimension.	Funding covers research and development.
<b>ENABLING CONTEXT 4:</b> <i>Having strong organizational structures</i>			
The examination office does not exist or is newly established.	The examination office is newly established.	The examination office is a stable organization. <sup>10</sup> *	This option does not apply to this dimension.
The examination office is not accountable to an external board or agency.	This option does not apply to this dimension.	The examination office is accountable to an external board or agency. <sup>11</sup> *	This option does not apply to this dimension.
Examination results are not recognized by any certification or selection system.	Examination results are recognized by certification or selection system in the country/system. <sup>12</sup> *	Examination results are recognized by one certification or selection system in another country/system.	Examination results are recognized by two or more certification or selection system in another country/system.
The examination office does not have the required facilities to carry out the examination. <sup>13</sup>	The examination office has some of the required facilities to carry out the examination.	The examination office has all of the required facilities to carry out the examination.	The examination office has state of the art facilities to carry out the examination.

(CONTINUED)

<p>LATENT</p> 	<p>EMERGING</p> 	<p>ESTABLISHED</p> 	<p>ADVANCED</p> 
<p><b>ENABLING CONTEXT 5:</b> <i>Having effective human resources</i></p>			
<p>There is no staff to carry out the examination.</p>	<p>The examination office is inadequately staffed to effectively carry out the examination, issues are pervasive.<sup>14</sup> *</p>	<p>The examination office is adequately staffed to carry out the examination effectively, with minimal issues.</p>	<p>The examination office is adequately staffed to carry out the assessment effectively, with no issues.</p>
<p>The country/system does not offer opportunities that prepare for work on the examination.</p>	<p>This option does not apply to this dimension.</p>	<p>The country/system offers some opportunities that prepare for work on the examination.<sup>15</sup></p>	<p>The country/system offers a wide range of opportunities that prepare for work on the examination.</p>

## SYSTEM ALIGNMENT

*Degree to which the assessment is coherent with other components of the education system.*





LATENT ● ○ ○ ○	EMERGING ● ● ○ ○	ESTABLISHED ● ● ● ○	ADVANCED ● ● ● ●
<p><b>SYSTEM ALIGNMENT 1:</b> <i>Aligning examinations with learning goals and opportunities to learn</i></p>			
It is not clear what the examination measures.	This option does not apply to this dimension.	There is a clear understanding of what the examination measures. <sup>16</sup> *	This option does not apply to this dimension.
What the examination measures is questioned by some stakeholder groups. <sup>17</sup> *	This option does not apply to this dimension.	What is measured by the examination is largely accepted by stakeholder groups.	This option does not apply to this dimension.
Material to prepare for the examination is minimal and it is only accessible to very few students.	There is some material to prepare for the examination that is accessible to some students.	There is comprehensive material to prepare for the examination that is accessible to most students. <sup>18</sup> *	There is comprehensive material to prepare for the examination that is accessible to all students.
<p><b>SYSTEM ALIGNMENT 2:</b> <i>Providing teachers with opportunities to learn about the examination</i></p>			
There are no courses or workshops on examinations available to teachers. <sup>19</sup> *	There are no up-to-date courses or workshops on examinations available to teachers.	There are up-to-date voluntary courses or workshops on examinations available to teachers.	There are up-to-date compulsory courses or workshops on examinations for teachers.
Teachers are excluded from all examination-related tasks.	Teachers are involved in very few examination-related tasks.	Teachers are involved in some examination-related tasks.	Teachers are involved in most examination-related tasks. <sup>19</sup> *

## ASSESSMENT QUALITY

*Degree to which the assessment meets quality standards, is fair, and is used in an effective way.*

LATENT ● ○ ○ ○	EMERGING ● ● ○ ○	ESTABLISHED ● ● ● ○	ADVANCED ● ● ● ●
<b>ASSESSMENT QUALITY 1:</b> <i>Ensuring quality</i>			
There is no technical report or other documentation. *	There is some documentation on the examination, but it is not in a formal report format.	There is a comprehensive technical report but with restricted circulation.	There is a comprehensive, high quality technical report available to the general public.
There are no mechanisms in place to ensure the quality of the examination. *	This option does not apply to this dimension.	There are limited systematic mechanisms in place to ensure the quality of the examination.	There are varied and systematic mechanisms in place to ensure the quality of the examination.
<b>ASSESSMENT QUALITY 2:</b> <i>Ensuring fairness</i>			
Inappropriate behavior surrounding the examination process is high.	Inappropriate behavior surrounding the examination process is moderate. <sup>20</sup> *	Inappropriate behavior surrounding the examination process is low.	Inappropriate behavior surrounding the examination process is marginal.
The examination results lack credibility for all stakeholder groups.	The examination results are credible for some stakeholder groups. <sup>21</sup> *	The examination results are credible for all stakeholder groups.	This option does not apply to this dimension.
The majority of the students (over 50%) may not take the examination because of language, gender, or other equivalent barriers.	A significant proportion of students (10%-50%) may not take the examination because of language, gender, or other equivalent barriers.	A small proportion of students (less than 10%) may not take the examination because of language, gender, or other equivalent barriers. <sup>22</sup> *	All students can take the examination; there are no language, gender or other equivalent barriers.

(CONTINUED)

<b>LATENT</b> 	<b>EMERGING</b> 	<b>ESTABLISHED</b> 	<b>ADVANCED</b> 
<b>ASSESSMENT QUALITY 3:</b> <i>Using examination information in a fair way</i>			
Examination results are not used in a proper way by all stakeholder groups.	Examination results are used by some stakeholder groups in a proper way.	Examination results are used by most stakeholder groups in a proper way. <sup>23</sup> *	Examination results are used by all stakeholder groups in a proper way.
Student names and results are public. *	This option does not apply to this dimension.	Students' results are confidential.	This option does not apply to this dimension.
<b>ASSESSMENT QUALITY 4:</b> <i>Ensuring positive consequences of the examination</i>			
There are no options for students who do not perform well on the examination, or students must leave the education system.	There are very limited options for students who do not perform well on the examination.	There are some options for students who do not perform well on the examination. <sup>24</sup> *	There is a variety of options for students who do not perform well on the examination.
There are no mechanisms in place to monitor the consequences of the examination. *	This option does not apply to this dimension.	There are some mechanisms in place to monitor the consequences of the examination.	There is a variety of mechanisms in place to monitor the consequences of the examination.

**Examinations:** *Development-level rating justifications*

1. Sindh Boards of Intermediate & Secondary Education Ordinance 1972 (amended in 1973, 1974, 1978), and most recently the Sindh Boards of Intermediate & Secondary Education (Amendment) Act 2008 (Sindh Act No. XXV of 2011).
2. The acts and ordinances, as well as the Calendar, are technically public documents, but not necessarily easily accessible. Both types of documents should be available on request at the Boards. However, there does not appear to be a clear procedure for making such a request. When a request for these documents was made, the Board did not supply the documents.
3. The act outlines the governance, distribution of power, and responsibilities among key stakeholders. The Calendar outlines the procedures to investigate and address security breaches, cheating, or other forms of inappropriate behavior, and the procedures for special/disadvantaged students, while specifying who can sit for the examination. It also identifies the rules about the preparation and format of the examination.
4. There is little publicly available evidence on stakeholder support or opposition, and there is very little public debate on this issue.
5. Efforts to improve the examination have been led by the government and most recently, by donors. For example, there have been several efforts to improve different aspects of the examinations by the Boards along with the Education Department. The Inter Board Committee of Chairmen (IBCC) has changed the paper pattern in recent years to include multiple choice and short answer questions. The Boards have tried to ensure that the examiners marking the papers are from the same language background as the language of the paper they are marking. There appear to be efforts by the Aga Khan University Examination Board (AKUEB) and USAID-supported EDLINKS project to improve the quality of the exam paper items and curriculum coverage in Boards across Pakistan. However, this is a fairly new initiative and apart from a workshop with paper setters from the Boards, it's not clear what other activities have been implemented. Only Karachi Board respondents mentioned they may be working on developing an item bank in the future.
6. The Boards at the higher levels appear to be aware of the various issues in the examination system and as a result have taken steps to address the issues themselves. However, since efforts to improve the examination from outside the government, such as those by the AKU-EB, have either not occurred or not been implemented as yet, it is difficult to say whether they would be receptive to such changes.
7. The funding is from student fees.
8. Data analysis and data reporting are limited to development and reporting of the exam results for individual candidates.
9. The Karachi Board has a research section; however, this section does not necessarily conduct research. Its functions are usually to develop descriptive analysis of exam results (i.e., by gender or academic groups or over the last 10 years) and handle non-research activities such as scholarships, debates and essay competitions.







10. The oldest Board, the Karachi Board of Secondary & Intermediate Education, was established in 1951, while the most recent Board, the Mirpur Khas Board of Secondary & Intermediate Education, was established in 2004.
11. Accountable to the Governor of Sindh province.
12. The IBCC / Foreign Ministry attests to the use of the Secondary School Certificates (SSC) abroad. Although the SSC exam would not be used for admission into university per se, it would presumably form part of the secondary school education record for admissions.
13. The consultant only visited Karachi and Hyderabad Boards. Given that the Boards are located in very different areas, it is difficult to determine whether these responses in the questionnaire, particularly for technological facilities, are representative of the other Boards as well.
14. While there is permanent full-time staff for running the examination, it is insufficient to meet the needs of the examination. For example, there is a lack of staff who can perform research and data analysis, as well as a lack of technical staff with skills in examination design to guide the paper development process. Issues that have been identified with the performance of the human resources that are responsible for the examination include item errors, repetition, limited skills tested, and inadequate curriculum coverage.
15. Graduate and non-graduate university courses on educational management and evaluation are offered. The University of Sindh, for example, offers a course on educational measurement and evaluation in its B.Ed and M.Ed programs.
16. Officially, the examination seeks to measure the national curriculum. However, there is no clear link to the student learning outcomes or the skills identified in the curriculum, often resulting in inadequate coverage and the lack of higher order skills tested as noted by research studies. In practice, it is essentially the textbook content that is measured by the exam. (Textbooks cannot be said to be well aligned with the curricula: (1) The majority of the textbooks in circulation are based on a previous version of the national curricula developed in 2002 (new textbooks based on the new curricula, i.e. 2006, are in various stages of development), and (2) they are not perceived to adequately cover all aspects of the curricula such as higher order thinking skills.)
17. Different stakeholders have different concerns. Many stakeholders (such as parents, students, and the general public) are more concerned with issues related to cheating, speed, and accuracy of scoring. However, those who recognize that there are inadequacies in the curriculum coverage and skills tested, such as respondents at the Reform Support Unit, do not accept what is measured.
18. Materials required for preparing for the exam include the textbook (the main resource, as the exam is essentially textbook-based) and past papers (as there is apparently not a lot of repetition in exam questions from year to year). Textbooks in government schools are now provided free of cost. Respondents noted that the textbooks are of good quality and that was all that students need to prepare for the exams.
19. Retired teachers select or create examination questions, while active teachers administer and score the examination, supervise the examination procedures, and act as a judge in, for example, orals.

20. Various inappropriate behaviors diminish the credibility of the examinations. Such behaviors include leakage of the content of an examination paper or part of a paper prior to the examination, impersonation, copying from other candidates, using unauthorized materials such as prepared answers and notes, collusion among candidates via mobile phones, passing of paper, or equivalent, and provision of external assistance via the supervisor, mobile phone, etc. To reduce inappropriate behavior, the Boards this year sought to adhere to the scheduled registration date to prevent students from manipulating their center allocation. Also, in addition to the Board's supervisory staff monitoring, there are governor-appointed Special Vigilance Teams, consisting of retired teachers, who monitor and provide independent reports on conduct. To address the leakage of exam papers, there is a secure room for developing papers and processes for preventing leakage at that stage. In Hyderabad, papers are made available at the Banks and center staff must collect and deposit filled papers on the day of the exam. And in Karachi, the papers are delivered to the exam centers by Board representatives themselves. Finally, answer books are coded so that the identity of the candidate is not known to the examiner.
21. The credibility of the exams and its results are called into question with regards to the need for quality improvement at all levels (exam design, administration, scoring, and determining results) and the perceptions in the general public that cheating is still widespread. The views of stakeholders tend to vary in terms of their area of concern. Media reports tend to focus on the conduct (i.e., occurrence of cheating) and accuracy or delays in results, as do parents and students. This year, there has been a substantial amount of criticism over the computerization of the process. Staff within the Boards, government departments, and teachers appears to be concerned with issues in marking. Finally, from reports and interviews, stakeholders such as those in teacher education institutions, NGOs, and selected government officials and educators point out the poor quality of the exam paper and its questions. Universities do not appear to trust the quality of the exams and take their own entrance exams.
22. Respondents have noted that if a student manages to make it to secondary school (i.e., does not drop out at earlier stages or face access-related barriers), then fees or location should not serve as a barrier to take the examinations, as it is the key certification for proving that they have completed secondary school. Exam centers are usually close to the students' schools, and there are no registration barriers based on ethnicity or gender.
23. Because it is unclear what the official purposes and uses of examination results are, it is difficult to identify "improper" uses. Generally, however, exam results are not used beyond determining pass/fail and admission into the next level of education, which are in line with the intended purposes.
24. Students may retake the examination, may opt for less selective schools/universities/tracks, or they may repeat the grade.

**SINDH PROVINCE, PAKISTAN**  
***National (or System-Level) Large-Scale Assessment (NLSA)***

**ENABLING CONTEXT**





*Overall framework of policies, leadership, organizational structures, fiscal and human resources in which NLSA activity takes place in a country or system and the extent to which that framework is conducive to, or supportive of, the NLSA activity.*

<b>LATENT</b> 	<b>EMERGING</b> 	<b>ESTABLISHED</b> 	<b>ADVANCED</b> 
<b>ENABLING CONTEXT 1:</b> <i>Setting clear policies for NLSA</i>			
No NLSA exercise has taken place.	The NLSA has been operating on an irregular basis. <sup>1</sup> *	The NLSA is a stable program that has been operating regularly.	This option does not apply to this dimension.
There is no policy document pertaining to NLSA.	There is an informal or draft policy document that authorizes the NLSA. <sup>2</sup> *	There is a formal policy document that authorizes the NLSA.	This option does not apply to this dimension.
This option does not apply to this dimension.	The policy document is not available to the public. *	The policy document is available to the public.	This option does not apply to this dimension.
There is no plan for NLSA activity.	This option does not apply to this dimension. *	There is a general understanding that the NLSA will take place.	There is a written NLSA plan for the coming years. <sup>3</sup> *
<b>ENABLING CONTEXT 2:</b> <i>Having strong public engagement for NLSA</i>			
All stakeholder groups strongly oppose the NLSA or are indifferent to it. <sup>4</sup>	Some stakeholder groups oppose the NLSA.	Most stakeholders groups support the NLSA.	All stakeholder groups support the NLSA.

(CONTINUED)

LATENT ● ○ ○ ○	EMERGING ● ● ○ ○	ESTABLISHED ● ● ● ○	ADVANCED ● ● ● ●
<b>ENABLING CONTEXT 3:</b> <i>Having regular funding for NLSA</i>			
There is no funding allocated to the NLSA.	There is irregular funding allocated to the NLSA.	There is regular funding allocated to the NLSA. <sup>5</sup> *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding covers some core NLSA activities: design, administration, analysis and reporting.	Funding covers all core NLSA activities: design, administration, analysis and reporting. *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding does not cover research and development activities. *	This option does not apply to this dimension.	Funding covers research and development activities.
<b>ENABLING CONTEXT 4:</b> <i>Having strong organizational structures for NLSA</i>			
There is no NLSA office, ad hoc unit or team.	The NLSA office is a temporary agency or group of people.	The NLSA office is a permanent agency, institution or unit. *	This option does not apply to this dimension.
This option does not apply to this dimension.	Political considerations regularly hamper technical considerations.	Political considerations sometimes hamper technical considerations. *	Political considerations never hamper technical considerations. *
This option does not apply to this dimension.	The NLSA office is not accountable to a clearly recognized body.	The NLSA office is accountable to a clearly recognized body. <sup>6</sup> *	This option does not apply to this dimension.

(CONTINUED)

<p>LATENT</p> 	<p>EMERGING</p> 	<p>ESTABLISHED</p> 	<p>ADVANCED</p> 
<p><b>ENABLING CONTEXT 5:</b> <i>Having effective human resources for NLSA</i></p>			
<p>There is no staff allocated for running an NLSA.</p>	<p>The NLSA office is inadequately staffed to effectively carry out the assessment. <sup>7</sup> *</p>	<p>The NLSA office is adequately staffed to carry out the NLSA effectively, with minimal issues.</p>	<p>The NLSA office is adequately staffed to carry out the NLSA effectively, with no issues.</p>
<p>The country/system does not offer opportunities that prepare individuals for work on NLSA.</p>	<p>This option does not apply to this dimension.</p>	<p>The country/system offers some opportunities to prepare individuals for work on the NLSA. <sup>8</sup></p>	<p>The country/system offers a wide range of opportunities to prepare individuals for work on the NLSA. *</p>





## SYSTEM ALIGNMENT

Degree to which the NLSA is coherent with other components of the education system.

LATENT ● ○ ○ ○	EMERGING ● ● ○ ○	ESTABLISHED ● ● ● ○	ADVANCED ● ● ● ●
<b>SYSTEM ALIGNMENT 1:</b> <i>Aligning the NLSA with learning goals</i>			
It is not clear if the NLSA is based on curriculum or learning standards.	This option does not apply to this dimension.	The NLSA measures performance against curriculum or learning standards. *	This option does not apply to this dimension.
What the NLSA measures is generally questioned by stakeholder groups.	This option does not apply to this dimension.	What the NLSA measures is questioned by some stakeholder groups. <sup>9</sup> *	What the NLSA measures is largely accepted by stakeholder groups.
There are no mechanisms in place to ensure that the NLSA accurately measures what it is supposed to measure.	There are ad hoc reviews of the NLSA to ensure that it measures what it is intended to measure.	There are regular internal reviews of the NLSA to ensure that it measures what it is intended to measure. <sup>10</sup> *	This option does not apply to this dimension.
<b>SYSTEM ALIGNMENT 2:</b> <i>Providing teachers with opportunities to learn about the NLSA</i>			
There are no courses or workshops on the NLSA.	There are occasional courses or workshops on the NLSA. <sup>11</sup> *	There are some courses or workshops on the NLSA offered on a regular basis.	There are widely available high quality courses or workshops on the NLSA offered on a regular basis.

## ASSESSMENT QUALITY

*Degree to which the NLSA meets technical standards, is fair, and is used in an effective way.*

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
<b>ASSESSMENT QUALITY 1:</b> <i>Ensuring the quality of the NLSA</i>			
No options are offered to include all groups of students in the NLSA.	This option does not apply to this dimension.	At least one option is offered to include all groups of students in the NLSA. <sup>12</sup> *	Different options are offered to include all groups of students in the NLSA.
There are no mechanisms in place to ensure the quality of the NLSA.	This option does not apply to this dimension.	There are some mechanisms in place to ensure the quality of the NLSA.	There are a variety of mechanisms in place to ensure the quality of the NLSA. <sup>13</sup> *
There is no technical report or other documentation about the NLSA.	There is some documentation about the technical aspects of the NLSA, but it is not in a formal report format.	There is a comprehensive technical report but with restricted circulation. *	There is a comprehensive, high quality technical report available to the general public.
<b>ASSESSMENT QUALITY 2:</b> <i>Ensuring effective uses of the NLSA</i>			
NLSA results are not disseminated.	NLSA results are poorly disseminated. <sup>14</sup> *	NLSA results are disseminated in an effective way.	This option does not apply to this dimension.
NLSA information is not used or is used in ways inconsistent with the purposes or the technical characteristics of the assessment. <sup>15</sup> *	This option does not apply to this dimension.	NLSA results are used by some stakeholder groups in a way that is consistent with the purposes and technical characteristics of the assessment.	NLSA information is used by all stakeholder groups in a way that is consistent with the purposes and technical characteristics of the assessment.
There are no mechanisms in place to monitor the consequences of the NLSA. <sup>16</sup> *	This option does not apply to this dimension.	There are some mechanisms in place to monitor the consequences of the NLSA.	There are a variety of mechanisms in place to monitor the consequences of the NLSA.



***National (or System-Level) Large Scale Assessment (NLSA): Development-level rating justifications***

1. The first year that the Provincial Education Assessment Center (PEACE) Sindh assessment was administered was in 2008. PEACE Sindh was previously part of NEAS, established to conduct assessments to a provincially representative sample as part of a national study. In 2008, it began conducting assessments independently.
2. There is the original PEACE PC-1 approved by Sindh Education and Literacy Department for the period 2003-2009 and a re-modified PC-1, which includes a three-year plan for 2009-12.
3. The plan for the next three years, 2012-2015, has been drafted and is awaiting approval by the Secretary of Education.
4. There does not appear to be any publicly available evidence for stakeholder perceptions of the large-scale assessments.
5. There is regular funding allocated by the government, as well as technical assistance (for assessment design and training of assessment administrators) from the EU.
6. The NLSA office is accountable to the Sindh Department of Education and Literacy.
7. There is permanent or full-time staff, but it is insufficient to meet the needs of the assessment. Generally, there is lack of qualified staff who can ensure that the tasks are conducted in an efficient and timely manner. Specifically, there is a lack of technical staff, psychometricians, and statisticians for data analysis.  
The issues that have been identified include poor training of test administrators or unclear instructions and guidelines for administering the assessment, weaknesses in test design, and frequent errors in the test questions. It is difficult to find relevant/good-quality persons to develop items, particularly the kind that go beyond the textbook and assess higher order thinking skills. There is also no in-house capacity to develop the sample, and the available in-house persons lack the necessary skills to benefit from technical assistance provided. Finally, data management and storage of scripts is also a key issue.  
Generally, there is a lack of clarity on the part of focal persons in the districts with regards to test administration and monitoring. There is also no final review of tools, which results in some errors in the printed tools and a need for translation verification between Urdu and Sindhi. Finally, there is lack of qualified staff for statistical analysis that caused delays.
8. Graduate and non-graduate university courses on educational measurement and evaluation are made available. The University of Sindh offers a course on educational measurement and evaluation in their B.Ed and M.Ed programs. The Aga Khan University Institute for Educational Development (AKU-IED) also offers a course on assessment at the M.Ed level. In the past, a specialized course was offered by the University of Melbourne in collaboration with the Institute of Educational Research (IER) for NEAS staff.

9. The PEACE assessment measures competencies identified in the curriculum, and it has been conducting these assessments for several years, including under NEAS. Therefore, stakeholders are now familiar with it and are likely to accept it. However, since assessments are based on the 2006 curricula and the textbooks available are based on the previous curricula (developed in 2002), PEACE staff note that some skills are not available in the textbooks and could cause some confusion amongst teachers who are used textbook-based tests.
10. There are several reviews at different stages, during the development of the assessment framework, item development and after the pilot. These are internal reviews that include external experts from the universities, teacher professional development colleges, and schools.
11. PEACE has conducted 13 awareness workshops with teachers in which they shared results of the PEACE assessment and provided guidelines on assessment practices in general, such as how to develop them and how to develop reports, etc.
12. The NLSA is offered in the language of instruction for almost all student groups. Additionally, the sampling frame includes both urban and rural areas. It does not appear that any other plans have been made to ensure inclusion or exclusion of students from hard-to-reach areas.
13. Various mechanisms are in place to ensure the quality of the NLSA. Specifically, all proctors or administrators are trained according to a protocol, there is a standardized manual for large-scale assessment administrators, a pilot is conducted before the main data collection takes place, all booklets are numbered, and external and internal reviewers/observers are employed.





Coding and scoring is a centralized activity, conducted in one place and overseen by persons involved in analysis. There is training of scorers, provision of code sheets and rubrics for open-ended items, and a 10 percent check of scoring by another person. In terms of processing of data, the computerized data is compared to the manual data to cross-check for errors. The persons involved in monitoring the administration of the assessment consisted of district focal persons and PEACE staff, as well as the Bureau of Curriculum and Provincial Institute of Teacher Education staff.
14. The NLSA results are disseminated in a variety of ways to different stakeholders. Reports containing results include a technical report, a general report for a general audience, a teacher feedback report, and a leaflet or flyer that provides brief information. For dissemination, they have conducted a provincial workshop in which they called the Executive District Education Officer from the districts and people from other provinces. In the future, they would like to hold district-based workshops as well. For NEAS assessment, there were also yearly reports and newsletters, stakeholder conferences, and coverage in the media. Dissemination of both the current PEACE and previous NEAS assessments is noted as one of the weaker aspects of the program by a current and former staff member. For the current PEACE assessment, one district officer noted that the relevant persons from the districts are not involved in dissemination, which has caused a lack of awareness in the assessment and its uses at the district level.
15. There has been little use of results by teachers, districts, or policymakers. Reasons cited include that there are no mechanisms for ensuring that recommendations are implemented or that the assessment results inform teacher practices. There has only recently been some interest by apex body Sindh Teacher Education Development Authority (STEDA) to use the results for teacher training.

16. Working groups have been established under PEACE to look at the gaps in curricula in relation to results. While PEACE recognizes the need to conduct research, and there is now two years' worth of data available, no initiatives have been made to conduct research.





**SINDH PROVINCE, PAKISTAN**  
***International Large-Scale Assessment (ILSA)***

## ENABLING CONTEXT

*Overall framework of policies, leadership, organizational structures, fiscal and human resources in which ILSA takes place in a country or system and the extent to which that framework is conducive to, or supportive of, ILSA activity.*





LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
<b>ENABLING CONTEXT 1:</b> <i>Setting clear policies for ILSA</i>			
The country/system has not participated in an ILSA in the last 10 years. <sup>1</sup> *	This option does not apply to this dimension.	The country/system has participated in at least one ILSA in the last 10 years.	The country/system has participated in two or more ILSA in the last 10 years.
The country/system has not taken concrete steps to participate in an ILSA in the next 5 years. *	This option does not apply to this dimension.	The country/system has taken concrete steps to participate in at least one ILSA in the next 5 years.	This option does not apply to this dimension.
There is no policy document that addresses participation in ILSA.	There is an informal or draft policy document that addresses participation in ILSA.	There is a formal policy document that addresses participation in ILSA.	This option does not apply to this dimension.
This option does not apply to this dimension.	The policy document is not available to the public.	The policy document is available to the public.	This option does not apply to this dimension.
<b>ENABLING CONTEXT 2:</b> <i>Having regular funding for ILSA</i>			
There is no funding for participation in ILSA.	There is funding from loans or external donors.	There is regular funding allocated at discretion.	There is regular funding approved by law, decree or norm.
This option does not apply to this dimension.	Funding covers some core activities of the ILSA.	Funding covers all core activities of the ILSA.	This option does not apply to this dimension.
Funding does not cover research and development activities.	This option does not apply to this dimension.	This option does not apply to this dimension.	Funding covers research and development activities.

(CONTINUED)

<b>LATENT</b> 	<b>EMERGING</b> 	<b>ESTABLISHED</b> 	<b>ADVANCED</b> 
<b>ENABLING CONTEXT 3:</b> <i>Having effective human resources for ILSA</i>			
There is no team or national/system coordinator to carry out the ILSA activities.	There is a team or national/system coordinator to carry out the ILSA activities.	There is a team and national/system coordinator to carry out the ILSA activities.	This option does not apply to this dimension.
This option does not apply to this dimension.	The national/system coordinator or other designated team member may not be fluent in the language of the assessment.	The national/system coordinator is fluent in the language of the assessment.	This option does not apply to this dimension.
This option does not apply to this dimension.	The ILSA office is inadequately staffed or trained to carry out the assessment effectively.	The ILSA office is adequately staffed or trained to carry out the ILSA effectively, with minimal issues.	The ILSA office is adequately staffed and trained to carry out the ILSA effectively, with no issues.





**SYSTEM ALIGNMENT**

*Degree to which the ILSA meets technical quality standards, is fair, and is used in an effective way.*

<b>LATENT</b> 	<b>EMERGING</b> 	<b>ESTABLISHED</b> 	<b>ADVANCED</b> 
<b>SYSTEM ALIGNMENT 1:</b> <i>Providing opportunities to learn about ILSA</i>			
The ILSA team has not attended international workshops or meetings.	The ILSA team attended some international workshops or meetings.	The ILSA team attended all international workshops or meetings.	This option does not apply to this dimension.
The country/system offers no opportunities to learn about ILSA.	This option does not apply to this dimension.	The country/system offers some opportunities to learn about ILSA.	The country/system offers a wide range of opportunities to learn about ILSA.
This option does not apply to this dimension.	This option does not apply to this dimension.	Opportunities to learn about ILSA are available to the country's/system's ILSA team members only.	Opportunities to learn about ILSA are available to a wide audience, in addition to the country's/system's ILSA team members.

## ASSESSMENT QUALITY

*Degree to which the ILSA meets technical quality standards, is fair, and is used in an effective way.*

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
<b>ASSESSMENT QUALITY 1:</b> <i>Ensuring the quality of ILSA</i>			
Data from the ILSA has not been published.	The country/system met sufficient standards to have its data presented beneath the main display of the international report or in an annex.	The country/system met all technical standards required to have its data presented in the main displays of the international report.	This option does not apply to this dimension.
The country/system has not contributed new knowledge on ILSA.	This option does not apply to this dimension.	This option does not apply to this dimension.	The country/system has contributed new knowledge on ILSA.
<b>ASSESSMENT QUALITY 2:</b> <i>Ensuring effective uses of ILSA</i>			
If any, country/system-specific results and information are not disseminated in the country/system.	Country/system-specific results and information are disseminated irregularly in the country/system.	Country/system-specific results and information are regularly disseminated in the country/system.	Country/system-specific results and information are regularly and widely disseminated in the country/system.
Products to provide feedback to schools and educators about the ILSA results are not made available.	This option does not apply to this dimension.	Products to provide feedback to schools and educators about the ILSA results are sometimes made available.	Products to provide feedback to schools and educators about ILSA results are systematically made available.
There is no media coverage of the ILSA results.	There is limited media coverage of the ILSA results.	There is some media coverage of the ILSA results.	There is wide media coverage of the ILSA results.
If any, country/system-specific results and information from the ILSA are not used to inform decision making in the country/system.	Results from the ILSA are used in a limited way to inform decision making in the country/system.	Results from the ILSA are used in some ways to inform decision making in the country/system.	Results from the ILSA are used in a variety of ways to inform decision making in the country/system.
It is not clear that decisions based on ILSA results have had a positive impact on students' achievement levels.	This option does not apply to this dimension.	This option does not apply to this dimension.	Decisions based on the ILSA results have had a positive impact on students' achievement levels.



***International Large Scale Assessment (ILSA):*** Development-level rating justifications

1. The Sindh province has not participated in an ILSA, and it does not have plans to do so in the near future.

## Acknowledgements

This report was prepared by The World Bank SABER-Student Assessment team in collaboration with Umbreen Arif, World Bank Senior Education Specialist and Task Team Leader for education projects in Pakistan.

## References

Clarke, M. 2012. "What Matters Most for Student Assessment Systems: A Framework Paper." READ/SABER Working Paper Series. Washington, D.C.: World Bank.

World Bank. 2012. Pakistan Country Indicator Data. Washington, D.C.: World Bank. Data retrieved from <http://data.worldbank.org/> on March 10, 2014.

World Bank. 2009. Project Appraisal Document on a Proposed Credit in the Amount of SDR 200.6 Million (US \$300 Million Equivalent) to the Islamic Republic of Pakistan for a Sindh Education Sector Project (SEP). Report No. 47642-PK. Washington D.C.: The World Bank.

The **Systems Approach for Better Education Results (SABER)** initiative produces comparative data and knowledge on education policies and institutions, with the aim of helping countries systematically strengthen their education systems. SABER evaluates the quality of education policies against evidence-based global standards, using new diagnostic tools and detailed policy data. The SABER country reports give all parties with a stake in educational results—from administrators, teachers, and parents to policymakers and business people—an accessible, objective snapshot showing how well the policies of their country's education system are oriented toward ensuring that all children and youth learn.

This report focuses specifically on policies in the area of student assessment.

This work is a product of the staff of The World Bank with external contributions. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

