



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, which teachers and others can easily access. There are some system-level mechanisms in place, including pre- and in-service teacher training opportunities, which are meant to ensure that teachers develop appropriate skills and expertise in classroom assessment. However, on-the-ground classroom assessment practices tend to vary from school to school, provide little feedback to students and parents, and are generally considered weak. There are no required uses of classroom assessment information to support student learning, and there are no mechanisms in place to systematically monitor the quality of classroom assessment practices.



2. Examinations

The Secondary School Certificate exam is administered to students in Grades 9 and 10. The Punjab Boards of Intermediate and Secondary Education Act of 1976 authorizes the exam. The Exam Rules and Procedures document (known as the "Calendar") for each Punjab 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 dedicated to the exam, but it is insufficient to meet the needs of the examination due to staff members' lack of technical capacity. Funding for the examination comes from student fees. Funding covers all core examination activities, but not research and development. There is also no technical documentation on the examination.



3. National Large-Scale Assessment (NLSA)

The Punjab Education Assessment System (PEAS) was administered for the first time in 2011 as an independent, provincial-level assessment exercise. Previously, PEAS was part of the National Education Assessment System (NEAS). Under NEAS, Punjab administered the assessment to a provincially representative sample of students and published a provincial report separate from the national report. Although the PEAS office is a permanent agency, it is inadequately staffed to effectively carry out PEAS activities. This also was the case when PEAS was part of NEAS. PEAS results have been poorly disseminated and inadequately used.



4. International Large-Scale Assessment (ILSA)

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



Introduction

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

Source: World Bank.

assessment system, with indicator areas

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 Punjab 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.

Punjab is Pakistan's largest province (out of four), accounting for approximately three-fifths of the country's population and income. In 2010/2011, primary net enrollment rate was 70 percent, and secondary net enrollment was 37 percent. In particular, there are low enrollment rates among girls in rural areas and children from poor households. In addition, the overall level of student achievement tends to be low; various assessments and surveys find that students perform significantly below curricular standards for key subjects and concepts at their grade level.

In order to increase both school participation and student achievement, Punjab has decided to focus specifically on improving teacher quality and performance. This decision is based on the government's recognition that: 1) the teacher is the main instrument for student learning, and that many initiatives can affect the child only through the teacher; and 2) improved student learning helps retain students in school and attract new children to school. As a result, Punjab plans to promote teacher quality and performance both directly and indirectly through initiatives such as: (i) strengthening the system of field-based advisory support to teachers; (ii) introducing test-based recruitment of teachers; (iii) tying teacher compensation more closely to school performance; (iv) improving the regular collection of credible information on school, teacher, and student performance, and feeding this information back to internal and external stakeholders; and others.

Detailed information was collected on the Punjab 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 Punjab 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 Punjab 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 the Punjab province's immediate interests and needs. Detailed, completed rubrics for each assessment type in the Punjab province are provided in Appendix 5.

Classroom Assessment

Level of development: EMERGING

In Punjab, 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 channels that are more readily accessible by teachers and other stakeholders, such as through teacher training courses or libraries. The Punjab Curriculum Authority Act 2012 established the Curriculum Authority. The Curriculum Authority is responsible for overseeing the supervision of curricula, preparing and approving textbooks, and maintaining standards of education in the Punjab province.

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 some system-wide resources available for teachers to engage in classroom

assessment activities, including textbooks which contain questions at the end of the chapters. The Taleemi Calendar is a document that contains student learning outcomes for each grade and subject. In addition, the Directorate of Staff Development (DSD) has developed Teacher's Guides which contain Student Learning Outcome (SLO)- based lesson plans for each subject and grade in primary school. These guides also include a section on the kinds of activities and homework that teachers can use to assess students on the particular SLO.

There are some system-level mechanisms, such as pre- and in-service teacher training opportunities, in place that are intended to ensure that teachers develop the appropriate skills and expertise in classroom assessment. School inspection and teacher supervision also include a component focused on classroom assessment.

However, on-the-ground classroom assessment practices tend to be mainly about recalling information, provide little feedback to students and parents, and are generally considered weak. Teachers primarily develop the questions directly from the textbook and use the exercises at the end of the chapters; they do not use explicit criteria for scoring or grading students' work. In general, classroom assessment activities tend to vary from school to school and are usually based on the textbooks and not on the curricula. However, textbooks tend to not be aligned with the curriculum because they are based on a previous version of the curriculum (developed in 2002) and not on the most recent curricula that went into effect in 2006.

There are no required uses of classroom assessment information to support student learning, and there are 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 exam is administered to students in Grades 9 and 10. The Secondary School Certificate exam is a major examination for graduation from secondary school, with approximately 1.7 million students taking it every year, and is therefore the focus of this report. The Punjab Boards of Intermediate and Secondary Education Act of 1976 authorizes the exam and outlines the governance, distribution of power, and responsibilities among key stakeholders. Additionally, the Examination Rules and Procedures document (known as the “Calendar”) for each of the Punjab Boards 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) some 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 sought to address the issues of standardization between the exam boards and the variation in the scoring of papers by collectively setting papers,

computerizing the scoring of the multiple-choice questions, and swapping examination papers between boards. The Aga Khan University Examination Board and USAID have also supported efforts to improve the quality of the examination items and curriculum coverage of the Boards across Pakistan.

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 errors in scoring, weakness in test design, omission of curricular topics, and errors in the examination questions and data processing.

In general, comprehensive material to help prepare for the examination is accessible to most students. Such items include the textbook (which is considered the main resource, as the examination is textbook-based) 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, copying, and using unauthorized materials, such as prepared answers and notes. As a result, the credibility of the examination and its results tends to be 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. Introduce a mandate for capacity development of staff responsible for the examination. Identify which individuals or organizations will be responsible for the 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. Develop a technical report on the examination and disseminate it to key stakeholders.
3. Identify and agree on the official purposes and possible uses of examination results, and introduce mechanisms, such as an oversight committee, to monitor actual uses.

National Large-Scale Assessment (NLSA)

Level of development: EMERGING

The Punjab province carries out several national large-scale assessments to students at the primary and secondary levels. The Punjab Examination Commission (PEC) was established in 2006 and awarded a statutory status in 2010 through an act of parliament. PEC is responsible for designing, implementing, and reporting on assessments for all Grade 5 (primary school) and Grade 8 (middle school) students in Punjab. PEC conducts Grade 5 and Grade 8 annual assessments (known as “examinations”) in all subjects for all the public and private schools in Punjab. PEC provides individual student reports which grade the student’s performance on each of the expected student learning outcomes (SLOs) by subject. The main objective of these examinations is to develop a key database about children’s performance and as a result measure the performance of the teachers and schools. The Punjab province also carries out the Punjab Education Assessment System (PEAS). PEAS was administered for the first time in 2011 as an independent, provincial-level assessment exercise administered to students in public sector schools. PEAS is viewed as the most important assessment for use by policymakers. Previously, PEAS was part of the National Education Assessment System (NEAS). Under NEAS, starting in 2005, Punjab administered the assessment to a provincially representative sample of students and published a provincial report separate from the national report.

Regular funding for PEAS is allocated by the government and donors. Funding covers some core PEAS activities, but not research and development.

Although the PEAS office is a permanent agency, it is inadequately staffed to effectively carry out all PEAS activities. In general, there are issues in test design, frequent errors in data processing, and delays due to the dependence on external human resources, including for paper development, data entry, and analysis.

In the past, the University of Education (Lahore) offered short courses and a specialized course was offered by

the University of Melbourne to prepare PEAS staff. However, none of these courses has been sustained, and Punjab province currently offers no opportunities to prepare individuals for work on the PEAS.

The PEAS 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 PEAS 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 PEAS including: (i) all administrators are trained according to a protocol, (ii) there is a standardized manual for the 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. A two-day training is held for scorers, and scorers are also provided with detailed codebooks and rubrics for open-ended questions. 10 percent of the scored assessment papers are also scored by another person. Protocols are also in place for ensuring the quality of data entry and data cleaning.

At the time of data collection for this report, the 2011 PEAS results had not been finalized and thus dissemination had not yet taken place. There are plans, however, to prepare and disseminate three types of reports (technical, thematic for different audiences, and a simplified report for a general audience) as well as to carry out district-based workshops to communicate the results.

Suggested policy options:

1. Develop the capacity of PEAS staff by (i) introducing workshops and trainings on education measurement and evaluation, including, in particular, data analysis and report writing; (ii) providing funding to PEAS staff to attend international programs on educational measurement and evaluation; and (iii) identifying

external human resources for conducting translation verifications between Urdu and Sindhi.

2. Identify the intended uses of PEAS results; develop and implement a plan to disseminate the results to stakeholders; introduce mechanisms, such as an oversight committee, to monitor actual uses of PEAS results.

International Large-Scale Assessment (ILSA)

Level of development: LATENT

The Punjab 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
Purpose	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
Frequency	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
Who is tested?	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
Format	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
Coverage of curriculum	All subject areas	Generally confined to a few subjects	Generally confined to one or two subjects	Covers main subject areas	Covers main subject areas
Additional information collected from students?	Yes, as part of the teaching process	Frequently	Yes	Seldom	Seldom
Scoring	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				
	LATENT (Absence of, or deviation from, attribute)	EMERGING (On way to meeting minimum standard)	ESTABLISHED (Acceptable minimum standard)	ADVANCED (Best practice)	Justification
EC—ENABLING CONTEXT					
EC1—Policies					
EC2—Leadership, public engagement					
EC3—Funding					
EC4—Institutional arrangements					
EC5—Human resources					
SA—SYSTEM ALIGNMENT					
SA1—Learning/quality goals					
SA2—Curriculum					
SA3—Pre-, in-service teacher training					
AQ—ASSESSMENT QUALITY					
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 Punjab Province, Pakistan





This appendix provides the completed SABER-Student Assessment rubrics for each type of assessment activity in the Punjab 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.

PUNJAB 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 
ENABLING CONTEXT AND SYSTEM ALIGNMENT 1: Setting clear guidelines for classroom assessment			
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. ¹ *	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. ²	The document is widely available.
ENABLING CONTEXT AND SYSTEM ALIGNMENT 2: Aligning classroom assessment with system learning goals			
There are no system-wide resources for teachers for classroom assessment.	There are scarce system-wide resources for teachers for classroom assessment.	There are some system-wide resources for teachers for classroom assessment. ³ *	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, but the level of performance required is not clear. ⁴ *	There is an official curriculum or standards document that specifies what students are expected to learn and to what level of performance.
ENABLING CONTEXT AND SYSTEM ALIGNMENT 3: Having effective human resources to carry out classroom assessment activities			
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. ⁵ *	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 
ASSESSMENT QUALITY 1: <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. ⁶ *	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. ⁷ *	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.
ASSESSMENT QUALITY 2: <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. ⁸ *	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 National Curricula 2006. It provides general information on assessment purposes, types, and questions and some sample test content and scoring criteria, but these are not comprehensive guidelines. Additionally, most teachers have not been provided with the curriculum document, and are therefore not aware of these guidelines. The Punjab Curriculum Authority Act 2012 established the Curriculum Authority. The Curriculum Authority is responsible for overseeing the supervision of curriculum, preparing and approving textbooks, and maintaining standards of education in the Punjab province.
2. The National Curricula 2006 is available online but not through other channels, such as libraries and teacher training courses and colleges, that teachers and other stakeholders can easily access.
3. The Taleemi Calendar is a document that contains student learning outcomes for each grade and subject. There are also textbooks and workbooks that provide support for classroom assessment. They contain questions at the end of the chapter, as well as guidelines for teachers on conducting classroom activities for assessment purposes, such as suggesting to teachers to hold speech competitions. Directorate of Staff Development (DSD) has developed Teacher's Guides that contain Student Learning Outcome (SLO)-based lesson plans for each subject and grade in primary school. These include a section on the kinds of activities/homework teachers can use to assess students on that SLO.
4. There is an official national curriculum which went into effect in 2006. After the 18th constitutional amendment, the authority to develop curriculum has been transferred to the provinces and some provinces are developing their own curricula, but it is not clear whether the Punjab province is doing so.
5. System-level mechanisms include pre- and in-service teacher training, as well as school inspection or teacher supervision, which includes a component that is focused on classroom assessment.
6. It is not common for classroom assessment activities to rely mainly on multiple-choice, selection-type questions as short- and long-answer questions. Classroom assessment activities are, however, mainly about recalling information because teachers develop questions directly from the textbook. Teachers also do not use a priori criteria for grading students' work because there are no system-level guidelines or scoring criteria. Additionally, classroom assessment activities vary from school to school and generally provide little useful feedback to students and parents. In general, classroom assessment activities are based on the textbooks not the curricula. (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. There do not appear to be any system-level mechanisms that monitor the quality of assessment activities. The focus is primarily on test results. It appears that formal teacher performance evaluation is only a regular practice at the secondary level and in that case, usually with relation to Board exam results. Generally, district staff visits tend to focus on more administrative aspects, enrollment, or teacher presence. At the key grades and the secondary level where a board exam will be taken, it appears there is an interest in the student results. Finally the District Teacher Educators (DTE) visits also focus on results of their monthly tests, not the teachers' own practices.

8. Teachers are required to report on individual students' performance to school district and Ministry of Education officials. Specifically, student results are reported to the district education department (usually for annual tests). In some cases for those classes preparing for board exams, particularly at the secondary level (i.e., grades 9 and 10), it appears that term tests are also reported.

PUNJAB 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 
ENABLING CONTEXT 1: <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. ¹ *	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. ² *	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. ³ *	The policy document addresses all key aspects of the examination.
ENABLING CONTEXT 2: <i>Having strong leadership</i>			
All stakeholder groups strongly oppose the examination or are indifferent to it. ⁴	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. ⁵ *	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. ⁶	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 ●●●●
ENABLING CONTEXT 3: <i>Having regular funding</i>			
There is no funding allocated for the examination.	There is irregular funding allocated for the examination. ⁷ *	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. ⁸ *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding does not cover research and development. ⁹ *	This option does not apply to this dimension.	Funding covers research and development.
ENABLING CONTEXT 4: <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. ¹⁰ *	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. ¹¹ *	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. ¹² *	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. ¹³	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)

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

SYSTEM ALIGNMENT

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
SYSTEM ALIGNMENT 1: <i>Aligning examinations with learning goals and opportunities to learn</i>			
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. ¹⁵ *	This option does not apply to this dimension.
What the examination measures is questioned by some stakeholder groups. ¹⁶ *	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. ¹⁷ *	There is comprehensive material to prepare for the examination that is accessible to all students.
SYSTEM ALIGNMENT 2: <i>Providing teachers with opportunities to learn about the examination</i>			
There are no courses or workshops on examinations available to teachers. ¹⁸ *	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. ¹⁸ *

ASSESSMENT QUALITY

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ASSESSMENT QUALITY 1: <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.
ASSESSMENT QUALITY 2: <i>Ensuring fairness</i>			
Inappropriate behavior surrounding the examination process is high.	Inappropriate behavior surrounding the examination process is moderate. ¹⁹ *	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. ²⁰ *	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. ²¹ *	All students can take the examination; there are no language, gender or other equivalent barriers.

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ASSESSMENT QUALITY 3: <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. ²² *	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.
ASSESSMENT QUALITY 4: <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. ²³ *	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. The formal policy document is the Punjab Boards of Intermediate and Secondary Education Act of 1976. Additionally, there are "Examination Rules and Procedures" documents that are known as the "Calendar" for each of the Punjab Boards, which covers key areas.
2. All acts, ordinances, and the Calendar are technically public documents, but they are not necessarily easily accessible. The full Calendar is available in hard copy at the Board on request, and it must be purchased. Parts of the Calendar are easily accessible on the website, but internet is not easily accessible to all stakeholders.
3. The Punjab Boards of Intermediate and Secondary Education Act outline 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, the procedures for special/disadvantaged students, and specifies 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 questions and short answer questions. Under the guidance of the Chief Minister last year, they created a talent pool of best paper setters and provided capacity development to them. The Boards have sought to address issues of standardization between the exam boards by setting papers collectively. They have also sought to deal with variations in the scoring of the papers through processes such as computerizing the scoring of the multiple-choice question sections, swapping papers between Boards and syndicated marking of open-ended questions (in which one person marks one question and the other the next to minimize effects of the marker on the overall score). The Boards have also sought to computerize the system from registration to scoring to result production; although there are issues with this system currently, this may in the future reduce errors in the result and instances of unfair means. The Punjab Education Department and the Boards have instituted several measures to curb cheating and other inappropriate behaviors. Finally, there appear to be efforts by the Aga Khan University Examination Board (AKUEB) and a USAID-supported EDLINKS project to improve the quality of the exam paper items and curriculum coverage of the Boards across Pakistan. However this is a fairly new initiative and apart from a workshop with paper setters—individuals who set the examination question papers—from the Boards, it is not clear what other activities have been implemented.
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 not occurred, it is difficult to say whether they would be receptive to such changes.
7. The funding for the examination are from student fees collected in previous years and other fees for issuing or validating certificates, conducting a re-evaluation of papers, etc.





8. Data analysis and reporting are limited to development and reporting of exam results for individual candidates.
9. Only the Lahore 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 location) and handle non-research activities such as scholarships, debates, and essay competitions. There is no staff development.
10. The oldest Board, the Lahore Board of Secondary & Intermediate Education, was established in 1954, while the most recent Board, the Dera Ghazi Khan Board of Secondary & Intermediate Education, was established in 1989.
11. Accountable to the Department of Higher Education, Punjab.
12. The IBCC/Foreign Ministry attests to the use of the SSC exam certificates 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 one Lahore BISE physically, and the selections for this question are based on this Board, either through observation or respondent comments. 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 or full-time staff, it is insufficient to meet the needs of the examination. For example, there is lack of staff who can perform research and data analysis, and 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 responsible for the examination include errors in scoring that have led to delays in results being reported, weakness in test design, omission of curricular topics, and frequent errors in the examination questions and data processing.
15. 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 (in Punjab, there is one set of provincially approved textbooks in schools). Textbooks cannot be said to be well aligned with the curricula: (1) Some 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.
16. Different stakeholders have different concerns. Many stakeholders (such as parents, students, general public) are more concerned with issues related to cheating and speed, and accuracy of scoring. However, those who recognize that there are inadequacies in the curriculum coverage and skills tested do not accept what is measured. Such respondents for this study have included the Chairman of a Board, staff at NGOs and professional development institutions, and some educators.

17. 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 feel that both these resources should be readily available in the market even in rural areas, and cost is not too high even for those who have to pay. Thus only a very small percentage may have difficulty accessing these items. The material on the examination includes examples of the types of questions that are on the examination, past examination papers, and examples of examination papers developed by the Boards.
18. 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 and practical.
19. Various inappropriate behaviors diminish the credibility of the examinations. Such behaviors include leakage of the examination paper content 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, an Examination Monitoring Cell was established within the Education Department, which, in addition to the monitoring done by the Boards themselves and the monitoring by retired government officials, has strengthened the monitoring of exams. Press conferences are also used to raise awareness about the consequences of inappropriate behavior during the examination.
20. 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 perception of 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. In addition to these issues, 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 have their own entrance exams.
21. 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 they've completed secondary school. Exam centers are usually close to the students' schools and there no registration barriers based on ethnicity or gender. With regards to fees, as of 2010, the fees have been waived by the Punjab government, stating that fees were a cause of dropout at secondary level. For the last two years, the fees have not served as a barrier. When looking at the numbers of students who appeared for the exams in 2010, news reports estimate about 1.1 million for all Boards and this is similar to the enrollment figures for Grade 10 from 2008 of about 1 million (PSLM). One could assume that even with a population increase, the number is quite similar.
22. 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.
23. Students may retake the examination, may opt for less selective schools/universities/tracks, or they may repeat the grade.

PUNJAB 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.

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
ENABLING CONTEXT 1: <i>Setting clear policies for NLSA</i>			
No NLSA exercise has taken place.	The NLSA has been operating on an irregular basis. ¹ *	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. ² *	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. ³ *	There is a written NLSA plan for the coming years.
ENABLING CONTEXT 2: <i>Having strong public engagement for NLSA</i>			
All stakeholder groups strongly oppose the NLSA or are indifferent to it. ⁴	Some stakeholder groups oppose the NLSA.	Most stakeholders groups support the NLSA.	All stakeholder groups support the NLSA.

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 3: <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. ⁵	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. ⁶	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.
ENABLING CONTEXT 4: <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. ⁷	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. ⁸	This option does not apply to this dimension.

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 5: <i>Having effective human resources for NLSA</i>			
There is no staff allocated for running an NLSA.	The NLSA office is inadequately staffed to effectively carry out the assessment. ⁹	The NLSA office is adequately staffed to carry out the NLSA effectively, with minimal issues.	The NLSA office is adequately staffed to carry out the NLSA effectively, with no issues.
The country/system does not offer opportunities that prepare individuals for work on NLSA. ¹⁰ *	This option does not apply to this dimension.	The country/system offers some opportunities to prepare individuals for work on the NLSA.	The country/system offers a wide range of opportunities to prepare individuals for work on the NLSA.

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
SYSTEM ALIGNMENT 1: <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. ¹¹ *	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. ¹² *	This option does not apply to this dimension.
SYSTEM ALIGNMENT 2: <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.	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 ●●●●
ASSESSMENT QUALITY 1: <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. ¹³ *	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. ¹⁴ *
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. ¹⁵ *	There is a comprehensive technical report but with restricted circulation.	There is a comprehensive, high quality technical report available to the general public.
ASSESSMENT QUALITY 2: <i>Ensuring effective uses of the NLSA</i>			
NLSA results are not disseminated.	NLSA results are poorly disseminated. ¹⁶ *	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. ¹⁷ *	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. ¹⁸ *	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 Punjab Education Assessment System (PEAS) was administered in 2011 independently. Prior to that, PEAS was part of the National Assessment (National Education Assessment System). Under NEAS, the province implemented the assessment to a provincially representative sample of students and published a provincial report separate from the national report. The Punjab Examination Commission (PEC) was established in 2006 and awarded a statutory status in 2010 through an act of parliament. PEC is responsible for designing, implementing and reporting on assessments for all Grade 5 (primary school) and Grade 8 (middle school) students in Punjab. PEC conducts Grade 5 and Grade 8 annual examinations in all subjects for all the public and private schools in Punjab. PEC provides individual student reports that grade the student's performance on each of the expected student learning outcomes (SLOs) by subject. The main objective of these examinations is to develop a key database about children's performance and, as a result, measure the performance of the teachers and schools.
2. Notification of Punjab Education Assessment System (PEAS) covers the period 2005-2008, issued in 2005 by Government of Punjab, Education Department (while PEAS was still a part of NEAS). This notification has not been updated to include the current period. This is an issue pointed out by staff members themselves, who recognize that it needs to be raised with the Secretary, but no timeline was provided for when this will happen.
3. The plan for the next two years has been developed. The written plan, unofficially approved by the Secretary of Education, is currently under development, and they expect to receive official approval in the next few months.
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 from the World Bank and DFID.
6. PEAS receives single-line funding that is not allocated to any one activity, then it seeks approval for planned activities. Planning does not appear as a separate activity in the work plan, however planning occurs twice a year as per government rules, in which funding can be reallocated to existing or new activities.
7. The NLSA office is a permanent institution created for running the NLSA. It was first established in 2002 as the Provincial Education Assessment Centre (PEACE), Punjab, and in 2005 it was renamed to the Punjab Education Assessment System (PEAS).
8. It is accountable to the Punjab Department of Education.
9. There is permanent or full-time staff, but it is insufficient to meet the needs of the assessment. Issues that have been identified with the performance of the human resources that are responsible for the NLSA include weaknesses in test design, frequent errors in data processing, and delays due to dependence on external human resources for, for example, paper development, data entry, and analysis. With regard to issues with item development, it has been difficult to find relevant/good quality specialists to develop items, especially the kind of items that go beyond the textbook and that assess higher order thinking

skills. There is also no in-house capacity to develop the sample. Additionally, in-house staff lack the required skills to benefit from the technical assistance that was provided. Effective data management and storage is also an issue.

10. In the past, the University of Education (Lahore) offered short courses, and a specialized course was offered by the University of Melbourne in collaboration with the Punjab University Institute of Educational Research (IER) for NEAS staff. However none of these have been sustained.

11. PEAS staff note that there has been some concern, particularly from teachers, over what the assessment covers because they are used to textbook-based exams and an assessment that links itself to the curricula and asks questions in a different manner is unfamiliar to them (textbooks cannot be said to be well aligned with the current curricula as they are not only based on an older version of the curricula, developed in 2002, but they also do not adequately cover all aspects of the curricula such as higher order thinking skills). However, this sentiment has decreased over the years with teachers' increased familiarity of the assessment and their realizing that it is not a high stakes test.

12. For the PEAS 2011 assessment, there have been 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.

13. The NLSA is offered in the language of instruction for almost all student groups and in various geographical regions of Punjab.

14. A variety of mechanisms are in place to ensure the quality of the NLSA instrument. Specifically, all proctors or administrators are trained according to a protocol, and there is a standardized manual for large-scale assessment administrators. Additionally, a pilot is conducted before the main data collection takes place and all booklets are numbered. There is also internal and external review by observers. With regard to scoring, there is a two-day training of scorers, provision of detailed codebooks and rubrics for open-ended items, and 10-15 percent check of scoring by another person. In terms of processing of data, there are protocols for ensuring the quality of data entry, 10 percent data check and cleaning. There does not appear to be double processing of data.

15. The technical report for PEAS has been developed and is currently being printed. For NEAS there were regular technical reports but with limited circulation.

16. The PEAS 2011 Assessment results and reports have not been finalized and thus there has been no dissemination at this stage. Plans include three types of reports (technical, thematic for different audiences, and simplified for a general audience), as well as district-based workshops. Under NEAS there were yearly reports and newsletters, stakeholder conferences and coverage in the media. However, dissemination was noted as one of the weakest areas in the entire program by a former staff member. Due to the lack of capacity and dedicated staff throughout most of the project, developing reports often took a great deal of time and caused delays in dissemination.

17. Given that the current PEAS assessment results have yet to be determined it is not possible to say how the information from this particular assessment will be used. In previous experiences, there has been some, although sporadic, use of results. For example the National Education Policy 2009 mentions NEAS results, and the Directorate of Staff Development (DSD) in Punjab claims to have used results in training. However, for the most part very little use is found, and implementation of the assessment recommendations has been almost non-existent. This is due to lack of a mechanism for ensuring the use of results

and to some extent lack of motivation to ensure implementation on the part of policymakers. It also has partly to do with the lack of a research-based decision making practices and understanding of the uses of assessment results.

18. The need is recognized by PEAS and this may happen in the future provided there is funding.

PUNJAB 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 ●●●●
ENABLING CONTEXT 1: Setting clear policies for ILSA			
The country/system has not participated in an ILSA in the last 10 years. ¹ *	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.
ENABLING CONTEXT 2: Having regular funding for ILSA			
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)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 3: <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.

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
SYSTEM ALIGNMENT 1: <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 ●●●●
ASSESSMENT QUALITY 1: <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.
ASSESSMENT QUALITY 2: <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 Punjab 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.

The Government of the Punjab Province. 2012. *The Punjab Curriculum Authority Act 2012*. Punjab province: The Government of the Punjab Province, Pakistan.

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

World Bank. 2012. *Pakistan - Second Punjab Education Sector Project*. Report No. 66291-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.



THE WORLD BANK