West Bank and Gaza

STUDENT ASSESSMENT

Key Policy Areas for Student Assessment

Status

1. Classroom Assessment

Emerging

The *Basis of Success, Completion, and Repetition for Grades 1 to 12* is an informal document authorized by the Ministry of Education in 2010-2011 that provides guidelines for classroom assessment. Some system-level mechanisms, such as pre- and in-service teacher training opportunities, are also in place to ensure that teachers develop skills and expertise in classroom assessment. Despite this, classroom assessment practices are generally considered to be weak because they tend to provide little useful feedback to students, and are used more as administrative tools than as a pedagogical resource. Although classroom assessment information is required to be disseminated to a number of key stakeholders, limited systematic mechanisms are in place to monitor the quality of classroom assessment activities. For example, there is no external moderation system that reviews the difficulty of classroom assessment activities.

2. Examinations

Emerging

The *Tawjihi* is an examination administered to grade 12 students primarily for the purposes of student certification for secondary school completion and student selection to a higher education institution. While the examination has been administered since 1960, most stakeholder groups do not strongly support it, and efforts to improve it are not welcomed by the leadership in charge of the examination. While the examination office—the General Directorate for Measurement, Evaluation, and Examination—employs permanent and full-time staff, staff numbers are insufficient to fully meet the needs of the examination.

3. National Large-Scale Assessment (NLSA)

Emerging

Since 2008, the *National Assessment* has been administered every two years, assessing representative samples of students in grades 4 and 10 in Arabic, Mathematics, and the Sciences. While the Ministry of Education's *Strategic Plan for Educational Development (2008-2012)* refers to the activities of the *National Assessment*, the Ministry of Education has not yet devised a comprehensive strategy for educational assessment. Although the NLSA office, the Assessment and Evaluation Department of the Ministry of Education, is a permanent unit created for running the *National Assessment*, it is inadequately staffed to effectively carry out the NLSA activities.

4. International Large-Scale Assessment (ILSA)

Emerging

The West Bank and Gaza has participated in a number of ILSAs in the last 10 years, including TIMSS (2003, 2007, and 2011) and LAMP (2008). In addition, the Ministry of Education has agreed to participate in TIMSS 2015. Funding for ILSA activities is provided by loans or external donors. A national coordinator and team, who have the relevant experience and training, are responsible for carrying out the ILSA. However, the ILSA office is not sufficiently staffed.

Systems Approach for Better Education Results

WORLD BANK GROUP

Introduction

The West Bank and Gaza has focused on increasing student learning outcomes by improving the quality of education in the country. An effective student assessment system is an important component of efforts to improve 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 assessment system, the West Bank and Gaza 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
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.

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. 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, 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.

the rubric displays four development levels—*Latent*, *Emerging*, *Established*,

Education in the West Bank and Gaza

The West Bank and Gaza are occupied Palestinian territories in the Middle East and North Africa region and are administered by the Palestinian Authority. Due in part to lower levels of donor aid, GDP growth declined from an average of 11 percent in 2010 and 2011 to 6 percent in 2012.

General education in the West Bank and Gaza consists of primary education (grades 1-10) and secondary education (grades 11-12). Since it was first established in 1994, the Ministry of Education and Higher Education (MoEHE) has expanded the education system through a number of activities, including support for school construction, curriculum development, and recruitment and training for teachers. Despite challenges due to persistent conflict, net enrolment rates at the primary and secondary level were 87 percent and 81 percent respectively in 2011. In addition, in 2011, 91 percent of children completed a full course of primary school.

The West Bank and Gaza developed its first five-year Education Development Strategic Plan (EDSP) for the period 2000-2005. However, in October 2000, the second Intifada or "uprising" started, delaying most activities planned under the EDSP. To address the continuing needs of the education sector and the effects of the second Intifada, such as substantial damage to school infrastructure, the government launched the second EDSP for the years 2008-2012. This plan focused on three key goals—increasing access, quality, and management capacity. With respect to access, the plan set out to reduce the number of double-shift classrooms; in order to address quality, it prioritized the revision of teachers' guides and curricula; and with regard to management capacity, the plan targeted improvements in computerized databases and financial management systems at the MoEHE.

Detailed information was collected on the West Bank and Gaza's student assessment system using the SABER-Student Assessment questionnaires and rubrics. It is important to remember that these tools primarily focus on benchmarking a country'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 West Bank and Gaza, 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 West Bank and Gaza's immediate interests and needs. Detailed, completed rubrics for each assessment type are provided in Appendix 5.

Classroom Assessment

Level of development

Emerging

● ● ○ ○

In the West Bank and Gaza, an informal document, the *Basis of Success, Completion, and Repetition for Grades 1 to 12*, authorized by the Ministry of Education, provides guidelines for classroom assessment. In addition, there are some resources available to teachers throughout the system to help them engage in classroom assessment activities. For example, textbooks and workbooks, as well as scoring criteria and rubrics for grading students' work, are available to teachers for their classroom assessment activities.

Some system-level mechanisms are in place to ensure that teachers develop skills and expertise in classroom assessment. Pre-service teacher training on educational evaluation and assessment is available to all students in faculties of education at national universities. In addition, all teachers in primary and secondary schools benefit from in-service teacher training on educational evaluation. Teachers also have opportunities to participate in conferences and workshops, and school inspection and teacher supervision includes a component on classroom assessment.

Nonetheless, classroom assessment practices are considered to be weak in several key ways. For example, classroom assessment activities tend to overly rely on multiple-choice, selection-type questions and to focus on recall of information at the expense of more sophisticated cognitive skills. It is also common for parents to be poorly informed about students' grades and for classroom assessment activities to provide little useful feedback to students.

Limited mechanisms are in place to systematically monitor the quality of classroom assessment activities. For example, classroom assessment is a required component of a teacher's performance evaluation. However, there is no external moderation system that reviews the difficulty of classroom assessment activities, and government funding is not available for research on

the quality of classroom assessment activities and how to improve them.

Classroom assessment information is required to be disseminated to a number of key stakeholders. Schools and teachers are required to report on an individual student's performance to school districts, parents, and students. However, classroom assessment information is not reported centrally to the Ministry of Education.

There are adequate required uses of classroom assessment to support student learning. Specifically, classroom assessment activities are supposed to be used to diagnose student learning issues, provide feedback to students on their learning, and inform parents about their child's learning. Although there is no official document covering these issues, teachers' evaluation forms completed by education supervisors and principals emphasize the importance of adequately using classroom assessment to support student learning. Supervisors also ask teachers to provide a detailed analysis of their students' examination results in a student report card.

Suggested policy options

1. Develop a strategy that includes a framework for the design and use of classroom assessment activities in the West Bank and Gaza in order to build momentum around the importance of classroom assessment. Include in the strategy a plan to build teachers' capacity to carry out classroom assessment activities. For example, require that all teachers learn classroom assessment skills as part of their pre-service training and identify opportunities for in-service teachers to participate in developing or scoring items for large-scale assessments or exams.

2. Improve the quality of classroom assessment activities by introducing system-level mechanisms to monitor classroom assessment practices. Such mechanisms could include the use of observers to evaluate whether classroom assessment practices are in line with their intended purposes and uses. Additionally, introduce an external moderation system that reviews the difficulty of classroom assessment activities and the appropriateness of scoring criteria.

Examinations

Level of development

Emerging

●●○○

The *Tawjihi* has been administered regularly since its introduction in 1960, when the West Bank and Gaza was part of Jordan. There is a formal policy document, *Instructions for the General Secondary Education Examination Certificate*, which authorizes the examination.

Regular funding for the examination is provided through a government allocation and student fees. Funding covers all core examination activities: design, administration, data processing, and reporting. Additionally, funding covers stationery for test scoring and transportation for test developers, test administrators, and scoring teams. However, funding does not cover research and development activities.

The General Directorate for Measurement, Evaluation, and Examination, an office within the Ministry of Education, has had primary responsibility for running the examination since the 1960s. Prior to 1967, the Ministry was under the Hashemite Kingdom of Jordan and therefore examination registers before 1967 are kept by the Jordanian Ministry of Education. Since 1968, examination records and all related documents are kept by the Ministry of Education in the West Bank and Gaza.

The General Directorate for Measurement, Evaluation, and Examination has some of the required facilities to carry out the examination. For example, there are computers for all technical staff, secure storage facilities, and adequate communication tools. Although there is a permanent and full-time staff, it is insufficient to meet the needs of the examination.

A wide range of opportunities that prepare individuals for work on the examination are available in the West Bank and Gaza, including university graduate programs and courses specifically focused on educational measurement and evaluation, and internships in the examination office. However, there are no up-to-date courses or workshops on the examination available to teachers, and teachers are involved in very few examination-related tasks.

There are limited systematic mechanisms in place to ensure the quality of the examination. For example, while there are internal reviews or observers, there are no external reviews or observers, and external certifications or audits do not take place.

There is some dissatisfaction with the examination. While policy makers and universities support the examination (policy makers strongly so), educators, students, parents, media, think tanks, and non governmental organizations (NGOs) do not display strong support for it. There is an ongoing discussion about the examination's difficulty level, as there is a belief that the examination questions are too challenging. There are also objections to the paper-and-pencil format. Recommended alternatives include the addition of performance tasks from a student portfolio when calculating the mean scores for students. Despite these concerns, efforts to improve the examination are not welcomed by the leadership in charge of the examination.

Some mechanisms are in place to monitor the consequences of the examination. For example, field follow-ups are undertaken by teachers and researchers in high- or low-performing schools. However, these field follow-ups are not organized or based on a clear methodology.

Suggested policy options

1. Provide teachers with opportunities to learn about the *Tawjihi* by making available up-to-date courses or workshops on the examination. Include teachers in a variety of examination-related tasks, such as selecting or creating the examination scoring guides and administering the examination.

2. Introduce various systematic mechanisms to ensure the quality of the examination, including conducting pilot or field testing, and incorporating internal and external reviews or observers.

3. Ensure that the examination results are used in a fair way by developing guides on the appropriate and inappropriate uses of examination results and making

7

them available to key stakeholder groups, particularly to parents, educators, and the media. The guides should include discussion of the consequences of inappropriate uses of examination results.

4. Introduce a permanent oversight committee, expert review groups, regular focus groups or surveys of key stakeholders to strengthen oversight of the examination and subsequent reviews of its impact.

5. Encourage researchers at higher education institutions to analyze the examination results to inform policy changes in such areas as teacher training. This could be through the provision of funding for research and analysis.

National Large-Scale Assessment (NLSA)

Level of development

Emerging

● ● ○ ○

Since 2008, the *National Assessment* has been administered every two years as a way to collect system-level information on the performance of grade 4 and grade 10 students in Arabic, Mathematics, and Science. The *National Assessment* began in 1998 and initially focused on Arabic and Mathematics performance in grade 6. A representative sample of students from government, United Nations Relief and Works Agency (UNRWA), and private schools in the West Bank and Gaza take part in the assessment exercise. Although the Ministry of Education's plans, such as the *Strategic Plan for Educational Development (2008-2012*), refer to the *National Assessment* activities, the Ministry of Education has not put in place a comprehensive strategy for educational assessment.

The Ministry of Education's development activities are funded through a joint financing program under the Ministry of Finance, with support from donors such as Finland and Ireland. Since 2008, a budget for *National Assessment* activities has been allocated from this joint financing program. Funding covers some core NLSA activities, including assessment administration and data analysis, as well as the printing of test papers, correction of tests, data entry, and the printing of reports that document the results. Although assessment design is not included under this funding, it is carried out free of charge by a specialized team within the Ministry of Education.

The Ministry of Education's Assessment and Evaluation Department is a permanent unit created for running the assessment. While there is a permanent and full-time staff to carry out the NLSA, it is insufficient since staff are responsible for other evaluation activities as well. There are some opportunities to prepare individuals for work on the NLSA. Opportunities include university courses, some of which are considered part of pre-service teacher training; funding for attending international programs, courses, and workshops on educational measurement

and evaluation; as well as internships or short-term employment in the large-scale assessment office.

Most teachers have access to online courses to learn about the NLSA. The Assessment and Evaluation Department also disseminates an introductory leaflet to all directorates of education so that they can distribute them to all schools in the national sample.

There are regular, internal and also ad hoc reviews of the alignment between the assessment instrument and what it is supposed to measure. Experts from the National Curricula Center, education supervision teams, and education assessment teams take part in the working groups that develop the instruments for the NLSA, ensuring that the assessment tools are credible.

General findings and results from the NLSA have been published in reports that are made available to the general public. Additional reports are made available at the school level (but not to the general public). Workshops and presentations on general findings and results from the NLSA are also held for key stakeholders. However, detailed results have only been published for the 2008 NLSA administration, and not for the 2010 and 2012 administrations.

Some mechanisms, such as regular focus groups or surveys of key stakeholders, are in place to monitor the consequences of the NLSA in terms of its impact on policy dialog, teaching, and learning.

Suggested policy options

1. Develop a formal, system-level document that specifies the purposes and uses of the NLSA and make it available to all key stakeholders including parents.

2. Introduce dedicated and predictable funding for key NLSA activities, including for assessment design and data reporting.

3. Make available targeted training opportunities (both local and international in nature) to central NLSA staff on key NLSA topics; identify representatives in the district offices who will collaborate with the central NLSA staff on NLSA activities. Develop and make regularly available

to teachers high quality courses or workshops on the NLSA.

4. Enhance the quality of the NLSA by introducing options to include all student groups in the assessment exercise. Options could include providing accommodations or alternative assessments for students with disabilities and developing plans to ensure that the assessment is administered to students in hard-to-reach areas.

5. Develop a plan to effectively use the results of the NLSA to inform decision making in the West Bank and Gaza. For example, curriculum experts from higher education institutions and specialists from the Ministry of Education could be convened in order to make recommendations for national curriculum guidelines based on NLSA results.

International Large-Scale Assessment (ILSA)

Level of development

Emerging

The West Bank and Gaza has participated in multiple ILSAs in the last 10 years, including the International Association for the Evaluation of Educational Achievement's (IEA) Trends in International Mathematics and Science Study (TIMSS) in 2003, 2007, and 2011. The West Bank and Gaza also participated in the Literacy Assessment and Monitoring Programme (LAMP), developed by the UNESCO Institute for Statistics (UIS), in 2008. The Ministry of Education (MoE) has agreed to participate in TIMSS in 2015, which was included in the five-year education development plan for 2008-2012.

Funding for participation in ILSAs is sourced from loans or external donors and covers all core activities including international participation fees, implementation of the assessment exercise in the country, processing and analysis of data collected from implementation of the assessment exercise, and reporting and disseminating the assessment results in the country.

A national coordinator and team are responsible for carrying out the ILSA activities. The team and the national coordinator have previous experience working on international assessments, have attended some international meetings related to the international assessment, and have the necessary training and experience to carry out the required assessment activities. However, the team is not sufficiently staffed, as it comprises only seven part-time members who are responsible for providing comprehensive follow-up on all activities related to the ILSA.

The West Bank and Gaza has met all technical standards required to have its data presented in the main displays of the international reports. It also has contributed to the global knowledge base on international assessments by generating new knowledge and making it available through publications. For example, research papers were

published in refereed reports, the first being the IEA initiative entitled *Educational Issues in the Middle East and North Africa Region, Outcomes of the IEA Arab Region Training Seminar Series 2006/2007.*

Country-specific results and information from the most recent international assessment were disseminated in the West Bank and Gaza. For example, copies of the international report were distributed to key stakeholders, and results received coverage on the television, radio, and newspapers. Products providing feedback on the results to schools, educators, and Science and Mathematics curricula developers were also made available.

ILSA results have been used in tracking the impact of reforms on student achievement levels and informing curriculum improvement, teacher training programs, and other assessment activities in the country. For example, supplemental material on mathematical topics covered by TIMSS, but not included in the national curriculum, was developed and disseminated to all schools. In addition, decisions based on the ILSA results have had a positive impact on students' achievement levels, as evidenced by the improvement in Mathematics scores on the *National Assessment* between 2010 and 2012.

Suggested policy options

1. Develop and make publicly available a formal policy document that addresses the participation of the West Bank and Gaza in these assessments to institutionalize ILSA activity.

2. Introduce regular and predictable government funding for ILSA activities.

3. Expand the available opportunities to learn about international assessments by, for example, collaborating with higher education institutions in the country to offer courses or workshops on international assessment.

4. Convene key stakeholders in order to develop a plan for using ILSA results to inform resource allocation decision making in the West Bank and Gaza's education system.

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				Justification
	LATENT (Absence of, or deviation from, attribute)	EMERGING (On way to meeting minimum standard)	ESTABLISHED (Acceptable minimum standard)	ADVANCED (Best practice)	
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
	Absence of, or deviation from, the attribute	*On way to meeting minimum standard*	*Acceptable minimum standard*	*Best practice*
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 neither a 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.

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 receives the following scores: Dimension A = 2 points; Dimension B = 2 points; Dimension C = 3 points. The hypothetical country'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 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.

15

Appendix 5: SABER-Student Assessment Rubrics for West Bank and Gaza

This appendix provides the completed SABER-Student Assessment rubrics for each type of assessment activity in the West Bank and Gaza. In each row of the rubric, the relevant selection is indicated by a thick border and asterisk. The selection may include a superscript number that refers to the justification or explanation for the selection. The related text can be found 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 asterisk, this means that insufficient information was available to determine the relevant selection in the row.

WEST BANK and GAZA
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.	The document is widely available.[2] *
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:
Ensuring the quality of classroom assessment

LATENT	EMERGING	ESTABLISHED	ADVANCED
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.	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. [7] *	There are varied and systematic mechanisms in place to monitor the quality of classroom assessment practices.

ASSESSMENT QUALITY 2:
Ensuring effective uses of classroom assessment

LATENT	EMERGING	ESTABLISHED	ADVANCED
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.	Classroom assessment information is required to be disseminated to all key stakeholders. [8] *
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. [9] *

Classroom Assessment: Development-level rating justifications

1. An informal system-level document, the *Basis of Success, Completion, and Repetition for Grades 1 to 12*, authorized by the Ministry of Education in 2010-2011, provides guidelines for classroom assessment.

2. *The Basis of Success, Completion, and Repetition for Grades 1 to 12* is widely available. At the beginning of every school year, this document is sent to the directorates of education and then disseminated to all private and public schools, as well as United Nations Relief and Work Agency (UNRWA) schools. UNRWA schools are under the supervision of the UNRWA education sector, serving students from refugee camps in the West Bank and Gaza.

3. There are some system-wide resources available to teachers for classroom assessment. The *Basis for Success, Completion, and Repetition for Grades 1 to 12* document also outlines what students are expected to learn in different subject areas at different grade/age levels. Textbooks or workbooks, as well as scoring criteria or rubrics for students' work are available to teachers for their classroom assessment activities. However, other resources, such as online assessment resources and item banks or pools with examples of selection/multiple-choice or supply/open-ended questions are not available as resources to teachers.

4. 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. The document is the general curricula guidelines developed in 1998 and 1999, which was adopted in principle for the elaboration of textbooks. The document defines teaching objectives and includes a simple description of the content and proposed assessment methods for each subject and grade. However, the document has not been released to the public.

5. There are some system-level mechanisms in place to ensure that teachers develop skills and expertise in classroom assessment. Pre-service teacher training on educational evaluation and assessment is available to all students in faculties of education at national universities. In addition, all teachers in primary and secondary schools benefit from in-service teacher training on educational evaluation. Teachers also have opportunities to participate in conferences and workshops, and school inspection or teacher supervision includes a component on classroom assessment. However, not all teacher training programs include a required course on classroom assessment, and there are no online resources on classroom assessment or opportunities to participate in item development for, or scoring of, large-scale assessments or exams.

6. Classroom assessment practices are known to be weak. While it is uncommon to observe errors in the scoring or grading of students' work, or for classroom assessment activities to not be aligned with the pedagogical or curricular framework, it is common for classroom assessment activities to rely mainly on multiple-choice, selection-type questions and to be mainly about recalling information. At the same time, it is common for teachers to not use explicit or a priori criteria for scoring or grading students' work; grade inflation and the uneven application of standards for grading students' work are serious problems. It is common for parents to be poorly informed about students' grades and for classroom assessment activities to provide little useful feedback to students. Classroom assessment is often used as an administrative or control tool rather than as a pedagogical resource.

7. There are limited systematic mechanisms to monitor the quality of classroom assessment activities. Classroom assessment is a required component of a teacher's performance evaluation and of school inspection or teacher supervision. Additionally, national or other system-wide reviews of the quality of

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

education include a focus on classroom assessment. However, there is no external moderation system that reviews the difficulty of classroom assessment activities, appropriateness of scoring criteria, etc., and there is no government funding available for research on the quality of classroom assessment activities and how to improve classroom assessment.

8. Classroom assessment information is required to be disseminated to a number of key stakeholders. Schools or teachers are required to report on an individual student's performance to school districts, parents, and students. However, classroom assessment information is not reported centrally to the Ministry of Education.

9. There are adequate required uses of classroom assessment to support student learning, including its use as an input for external examination results. Specifically, classroom assessment activities are used for diagnosing student learning issues, providing feedback to students on their learning, informing parents about their child's learning, planning next steps in instruction, grading students for internal classroom uses, and providing input to an external examination program. Although there is no official document governing these issues, teachers' assessment forms completed by education supervisors and principals emphasize the importance of adequately using classroom assessment to support student learning. Supervisors also ask teachers to provide a detailed analysis of their students' exam results in a specific format known as a student report card.

WEST BANK and GAZA
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: Setting clear policies			
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.[1] *	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.[2] *	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.[3] *	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.[4] *	The policy document addresses all key aspects of the examination.
ENABLING CONTEXT 2: Having strong leadership			
All stakeholder groups strongly oppose the examination or are indifferent to it.	Most stakeholder groups oppose the examination.[5] *	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.[6] *	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.[7] *	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.

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 3: _Having regular funding_			
There is no funding allocated for the examination.	There is irregular funding allocated for the examination.	There is regular funding allocated for the examination.[8] *	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.[9] *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding does not cover research and development.[10] *	This option does not apply to this dimension.	Funding covers research and development.
ENABLING CONTEXT 4: _Having strong organizational structures_			
The examination office does not exist or is newly established.	The examination office is newly established.	The examination office is a stable organization.[11] *	This option does not apply to this dimension.
The examination office is not accountable to an external board or agency.[12] *	This option does not apply to this dimension.	The examination office is accountable to an external board or agency.	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.	Examination results are recognized by one certification or selection system in another country.	Examination results are recognized by two or more certification or selection systems in another country.[13] *
The examination office does not have the required facilities to carry out the examination.[14] *	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: *Having effective human resources*		
There is no staff to carry out the examination.	The examination office is inadequately staffed to effectively carry out the examination, issues are pervasive.	The examination office is adequately staffed to carry out the examination effectively, with minimal issues.[15] *	The examination office is adequately staffed to carry out the assessment effectively, with no issues.
The country does not offer opportunities that prepare for work on the examination.	This option does not apply to this dimension.	The country offers some opportunities that prepare for work on the examination.	The country offers a wide range of opportunities that prepare for work on the examination.[16] *

SYSTEM ALIGNMENT

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
SYSTEM ALIGNMENT 1: *Aligning examinations with learning goals and opportunities to learn*			
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. [17] ✱	This option does not apply to this dimension.
What the examination measures is questioned by some stakeholder groups. [18] ✱	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. [19] ✱	There is comprehensive material to prepare for the examination that is accessible to all students.
SYSTEM ALIGNMENT 2: *Providing teachers with opportunities to learn about the examination*			
There are no courses or workshops on examinations available to teachers.	There are no up-to-date courses or workshops on examinations available to teachers. [20] ✱	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. [21] ✱	Teachers are involved in some examination-related tasks.	Teachers are involved in most examination-related tasks.

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: *Ensuring quality*			
There is no technical report or other documentation.	There is some documentation on the examination, but it is not in a formal report format. [22] *	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. [23] *	There are varied and systematic mechanisms in place to ensure the quality of the examination.
ASSESSMENT QUALITY 2: *Ensuring fairness*			
Inappropriate behavior surrounding the examination process is high.	Inappropriate behavior surrounding the examination process is moderate. [24] *	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.	The examination results are credible for all stakeholder groups. [25] *	This option does not apply to this dimension.
The majority of the students (over 50 percent) may not take the examination because of language, gender, or other equivalent barriers.	A significant proportion of students (10 percent-50 percent) may not take the examination because of language, gender, or other equivalent barriers.	A small proportion of students (less than 10 percent) may not take the examination because of language, gender, or other equivalent barriers.	All students can take the examination; there are no language, gender or other equivalent barriers. [26] *

(CONTINUED)

LATENT ●○○○	EMERGING ●●●○	ESTABLISHED ●●●○	ADVANCED ●●●●
ASSESSMENT QUALITY 3: *Using examination information in a fair way*			
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.[27] *	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.[28] *	This option does not apply to this dimension.
ASSESSMENT QUALITY 4: *Ensuring positive consequences of the examination*			
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.[29] *	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.[30] *	There is a variety of mechanisms in place to monitor the consequences of the examination.

Examinations: Development-level rating justifications

1. The examination is a stable program that has been operating regularly since it was begun in 1960, when the West Bank and Gaza were part of Jordan.

2. There is a formal policy document, *Instructions for the General Secondary Education Examination Certificate*, authorized by the Ministry of Education via the General Examination Committee yearly.

3. The policy document is available to and easily accessible by the public.

4. The policy document addresses some key aspects of the examination, including a description of the purpose of the examination, authorized uses of results, an outline of procedures for special/disadvantaged students, specifications for who can sit for the examination, and an identification of rules about preparation. However, the policy document does not outline governance, distribution of power and responsibilities among key entities, or state funding sources. Although this policy document does not outline procedures to investigate and address security breaches, cheating, or other forms of inappropriate behavior, or explain alignment with curricula and standards and the format of the examination questions, other policy documents contain content relevant to these issues.

5. Most stakeholder groups do not strongly support the examination. While policy makers and universities support the examination (policy makers strongly so), educators, students, parents, media, think tanks, and NGOs do not support it. Teacher unions and employers are neutral to the examination program. There is an ongoing discussion about the examination's difficulty level, as there is belief that the examination questions are difficult. There are also objections to the paper-and-pencil format, and recommended alternatives include the addition of performance tasks of a student portfolio when calculating the mean scores for students.

6. There are independent attempts to improve the examination by stakeholder groups.

7. Efforts to improve the examination are not welcomed by the leadership in charge of the examination.

8. There is regular funding allocated by the government and obtained from student fees for the examination.

9. Funding covers all core examination activities: design, administration, data processing and reporting. Additionally, funding covers stationery for test scoring, and transportation for test developers, test administrators, and scoring teams.

10. Funding does not cover research and development.

11. The examination office is a stable organization. The General Directorate for Measurement, Evaluation, and Examination, an office or branch within the Ministry of Education, has had primary responsibility for running the examination since the 1960s. Prior to 1967, the Ministry was under the Hashemite Kingdom

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

of Jordan and therefore examination registers before 1967 are kept by the MoE in the West Bank and Gaza. Since 1968, examination records and all related documents are kept by the MoE in the West Bank and Gaza.

12. The examination office is not accountable to an external board or agency.

13. Examination results are recognized by a certification or selection system in the West Bank and Gaza and in other Arab countries.

14. The examination office has some of the required facilities to carry out the examination. There are computers for all technical staff, secure storage facilities, access to adequate computer servers, the ability to backup data, and adequate communication tools. However, the examination office lacks a secure building.

15. There is permanent or full-time staff, but it is insufficient to meet the needs of the examination. However, only a few issues have been identified with the performance of the human resources who are responsible for the examination. These issues are poor training of test administrators or unclear instructions and guidelines for administering the examination, as well as weakness in test design.

16. The country offers a wide range of opportunities that prepare for work on the examination. Opportunities include: university graduate programs (master's degree programs) and university courses specifically focused on educational measurement and evaluation, non-university training courses or workshops on educational measurement and evaluation, funding for attending international programs, courses, or workshops on educational measurement and evaluation, and internships in the examination office.

17. There is a clear understanding that the examination measures the national school curriculum guidelines or standards.

18. What the examination measures is questioned by some stakeholder groups.

19. There is comprehensive material to prepare for the examination that is accessible to most students. The competencies tested by the examination are included in textbooks provided at a low cost by the Ministry. Examples of the types of questions that are on the examination are also available. Additionally, examination instructions, which describe how to calculate mean scores (since subjects carry different weights) are distributed to all schools hosting students that are taking the examination, registered students, and Education offices. These examination instructions are available nine months before the examination. However, a framework document explaining what is measured on the examination and information on how to prepare for the examination are not publicly available.

20. There are no up-to-date courses or workshops on the examination available to teachers.

21. Teachers are involved in very few examination-related tasks. Teachers are involved in selecting or creating examination questions and supervising examination procedures; however, they are not involved in tasks such as administering the examination, scoring the examination, or selecting or creating examination scoring guides. They are also not involved in acting as a judge or in resolving inconsistencies between examination scores and school grades.

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

22. There is some documentation on the examination but it is not in a formal report format.

23. There are limited systematic mechanisms in place to ensure the quality of the examination. For example, while there is internal review or observers, there is no external review or observers, and external certification or audit does not take place. Additionally, pilot or field testing or translation verification do not take place.

24. Inappropriate behavior surrounding the examination process is moderate. Issuing forged certificates or altering results information, collusion among candidates, copying from other candidates, and impersonation (when an individual other than the registered candidate takes the examination) are behaviors that occur, although rarely. When these behaviors are detected, students are expelled from the examination. There is a document which clarifies the measures corresponding to each case of reported inappropriate behavior.

25. The examination results are credible for all stakeholder groups.

26. All students can take the examination; there are no language, gender or other equivalent barriers.

27. Most stakeholder groups, including policy makers, teacher unions, students, think tanks and NGOs, universities, and employers, use the examination results in a proper way. However, educators, parents, and the media have used the examination results in an improper way. It is believed that the media uses students' results out of the context of the examinations. Also, it is believed that educators and parents improperly use examination results when ranking schools, evaluating the quality of the education system, or using results as predictors of students' success in university.

28. Students' results are confidential. Although schools publish students' results, students' names are not disclosed, only their seating number.

29. There are some options for students who do not perform well on the examination. For example, students may retake the examination or repeat a grade; however, options to attend remedial or preparatory courses to prepare to retake the examination or options for students to choose less selective schools, universities, or tracks, are not available.

30. There are some mechanisms in place to monitor the consequences of the examination. For example, field follow-ups are undertaken by teachers and researchers to high or low-performing schools; however, these are not organized or based on a clear methodology. Mechanisms such as a permanent oversight committee, expert review groups, regular focus groups or surveys of key stakeholders, are not in place.

WEST BANK and GAZA

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: *Setting clear policies for NLSA*			
No NLSA exercise has taken place.	The NLSA has been operating on an irregular basis.	The NLSA is a stable program that has been operating regularly. [1] *	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. [3] *	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. [4] *	There is a written NLSA plan for the coming years.
ENABLING CONTEXT 2: *Having strong public engagement for NLSA*			
All stakeholder groups strongly oppose the NLSA or are indifferent to it.	Some stakeholder groups oppose the NLSA. [5] *	Most stakeholders groups support the NLSA.	All stakeholder groups support the NLSA.

(CONTINUED)

WEST BANK AND GAZA | SABER-STUDENT ASSESSMENT

LATENT ⬤〇〇〇	EMERGING ⬤⬤〇〇	ESTABLISHED ⬤⬤⬤〇	ADVANCED ⬤⬤⬤⬤
	ENABLING CONTEXT 3: _Having regular funding for NLSA_		
There is no funding allocated to the NLSA.	There is irregular funding allocated to the NLSA.[6] *	There is regular funding allocated to the NLSA.	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.[7] *	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.[8] *	This option does not apply to this dimension.	Funding covers research and development activities.
	ENABLING CONTEXT 4: _Having strong organizational structures for NLSA_		
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.[9] *	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.[10] *
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.[11] *	This option does not apply to this dimension.

(CONTINUED)

LATENT ○○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 5: *Having effective human resources for NLSA*			
There is no staff allocated for running an NLSA.	The NLSA office is inadequately staffed to effectively carry out the assessment. [12] *	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 does not offer opportunities that prepare individuals for work on NLSA.	This option does not apply to this dimension.	The country offers some opportunities to prepare individuals for work on the NLSA. [13] *	The country 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: *Aligning the NLSA with learning goals*			
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. [14] *	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. [15] *	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. [16] *	This option does not apply to this dimension.
SYSTEM ALIGNMENT 2: *Providing teachers with opportunities to learn about the NLSA*			
There are no courses or workshops on the NLSA.	There are occasional courses or workshops on the NLSA. [17] *	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: *Ensuring the quality of the NLSA*			
No options are offered to include all groups of students in the NLSA. [18] *	This option does not apply to this dimension.	At least one option is offered to include all groups of students in the NLSA.	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. [19] *
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. [20] *	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: *Ensuring effective uses of the NLSA*			
NLSA results are not disseminated.	NLSA results are poorly disseminated.	NLSA results are disseminated in an effective way. [21] *	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.	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 the technical characteristics of the assessment. [22] *	NLSA information is used by all stakeholder groups in a way that is consistent with the purposes and the technical characteristics of the assessment.
There are no mechanisms in place to monitor the consequences of the NLSA.	This option does not apply to this dimension.	There are some mechanisms in place to monitor the consequences of the NLSA. [23] *	There are a variety of mechanisms in place to monitor the consequences of the NLSA.

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

1. Since 2008, the national assessment has been administered every two years, covering Arabic, Mathematics, and the Sciences for grade 4 and grade 10 students. The national assessment was first administered in 1998, at first covering Arabic and Mathematics for grade 6. When it was administered in 2000, the national assessment covered Arabic, Mathematics, and Science for grades 4 and 6, and in 2005, it covered Arabic and Mathematics for grade 4 students. A representative sample of students from the West Bank and Gaza in government, UNRWA, and private schools take part in the national assessment exercise.

2. The Ministry of Education's plans, such as the *Strategic Plan for Educational Development (2008-2012)*, refer to the national assessment activities; however, the Ministry of Education has not devised a clear strategy for educational assessment.

3. The policy document, the *Strategic Plan for Educational Development (2008-2012)*, is available to the public at the Ministry of Education's website. There is a common understanding that the assessment will take place, but there is no formally written plan.

4. There is a common understanding that the assessment will take place, but there is no formally written plan.

5. Some stakeholder groups oppose the NLSA. Educators, think tanks, NGOs, and universities support the NLSA, and students and parents oppose it due to issues related to test anxiety. Teacher unions and employers are neutral to it. While policy makers and the media generally support the NLSA, some policy makers have expressed concern over the lack of transparency regarding assessment information, and some media sources, such as newspaper and magazine editorials or columns, have criticized the assessment and its uses.

6. There is irregular funding allocated to the NLSA. The Ministry of Education's education development activities are financed through a joint financing program under the Ministry of Finance, with support from donors such as Finland, Ireland, and KFW. Since 2008, budgets have been allocated to national assessment activities within this joint financing program.

7. Funding covers some core NLSA activities, including assessment administration and data analysis, as well as the printing of examination papers, correction of tests, data entry, and the printing of reports that document the examination results. Although assessment design is not included under this funding, it is done by a specialized team within the Ministry of Education free of charge. However, funding does not cover data reporting, long- or medium- term planning of program milestones, research and development, or staff training.

8. Funding does not cover research and development activities.

9. The Ministry of Education Assessment and Evaluation Department is a permanent unit created for running the assessment.

10. Political considerations never hamper technical considerations.

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

11. The Ministry of Education Assessment and Evaluation Department is accountable to a higher office in the MoE. It is under the direct supervision of the Ministry of Education Assistant Deputy for Planning and Development Affairs.

12. The NLSA office is inadequately staffed to effectively carry out the assessment. While there is permanent or full-time staff to carry out the NLSA, it is insufficient since staff are responsible for other evaluation activities as well. Unlike other departments within the MoE, which have staff in district offices as well as in the headquarters, the Assessment and Evaluation Department has staff only in the headquarters. As a result, it is required to rely on part-time members to manage and follow-up on field activities. Recently, the NLSA faced some difficulties in obtaining the latest data on schools, teachers, and students, which has delayed sampling.

13. The West Bank and Gaza offer some opportunities to prepare individuals for work on the NLSA. University courses, some of which are considered part of pre-service teacher training, funding for attending international programs, courses, or workshops on educational measurement and evaluation, as well as internships or short-term employment in the large-scale assessment office, are all available. However, there are no university graduate programs or non-university training courses or workshops on educational measurement and evaluation. In addition, the quality of courses offered by universities differs by institution.

14. The NLSA measures performance against curriculum or learning standards.

15. Some stakeholder groups question what the NLSA measures. .

16. There are regular internal and ad hoc reviews of the alignment between the assessment instrument and what it is supposed to measure. To ensure that the assessment tools are credible, experts from the National Curricula Center, education supervision teams, and education assessment teams take part in the working groups to develop achievement examinations for the NLSA.

17. Occasional courses, workshops and presentations on the NLSA are offered to teachers. In addition, most teachers have access to courses online and the Assessment and Evaluation Department disseminates an introductory leaflet to all the directorates of education so that they may distribute them to all of the schools in the national sample. However, courses or workshops are not offered on a regular basis, and when courses are offered, most teachers do not have access to live courses or workshops.

18. No options are offered to include all groups of students in the NLSA. Students with special needs are exempted from the national assessment; instead, their evaluation is undertaken by their own schools according to special arrangements.

19. There is a variety of mechanisms in place to ensure the quality of the NLSA. For example, 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, there is double data scoring, and scorers are trained to ensure high inter-rater reliability. Additionally, staff from the Assessment and Evaluation Department, as well as a team of education supervisors from the Directorates of Education, conduct visits to some schools participating in the assessment.

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

20. There is some documentation about the technical aspects of the assessment but it is not in a formal report format. This documentation exists in a guide on the application of the assessment, which includes the model of the NLSA administration.

21. Reports with general NLSA indicators have been published in the yearly follow-up and assessment reports (which have been made available to the public) and at the school level (which have not been made available to the public). Specifically, these results are disseminated within 12 months of the large-scale assessment being administered and reports with results are made available for all stakeholder groups. The main reports on the results contain information on overall achievement levels and sub-groups, as well as information on trends over time overall and for subgroups. These reports also contain standard errors. Workshops or presentations on the general results are also held for key stakeholders. However, detailed results have been only published for the 2008 NLSA administration. The 2010 and 2012 detailed results have not been published yet as they are currently being used to develop central policies.

22. NLSA results are used by some stakeholder groups in a way that is consistent with the purposes and technical characteristics of the assessment.

23. There are some mechanisms, such as regular focus groups or surveys of key stakeholders, in place to monitor the consequences of the NLSA. However, other mechanisms, such as a permanent oversight committee, expert review groups, themed conferences that provide a forum to discuss research and other data on the consequences of the large-scale assessment, and funding for independent research on the impact of the large-scale assessment are not available.

WEST BANK and GAZA
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.	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. [1] *
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. [2] *	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. [3] *	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. [4] *	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. [5] *	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. [6] *	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. [7] *

(CONTINUED)

ENABLING CONTEXT 3:
Having effective human resources for ILSA

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
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.[8] *	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.[9] *	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.[10] *	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: *Providing opportunities to learn about ILSA*			
The ILSA team has not attended international workshops or meetings.	The ILSA team attended some international workshops or meetings. [11] *	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. [12] *	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. [13] *

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: Ensuring the quality of ILSA				
	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. [14] *	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. [15] *
ASSESSMENT QUALITY 2: Ensuring effective uses of ILSA				
	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. [16] *
	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. [17] *	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. [18] *	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. [19] *	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. [20] *

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

1. The West Bank and Gaza has participated in multiple ILSAs in the last 10 years, including the International Association for the Evaluation of Educational Achievement's (IEA) Trends in Mathematics and Science Study (TIMSS) in 2003, 2007, and 2011. Additionally, the West Bank and Gaza participated in the Literacy Assessment and Monitoring Programme (LAMP), developed by the UNESCO Institute for Statistics (UIS), in 2008.

2. The MoE in the West Bank and Gaza has agreed to participate in TIMSS in 2015.

3. The West Bank and Gaza has an informal policy document that addresses participation in international assessments. TIMSS was included in the five-year education development plan for 2008-2012.

4. *The Strategic Plan for Education Development (2008-2012)*, authorized in 2008 by the MoE, is published on the MoE website, along with the yearly follow-up and assessment report in Arabic and English. The Plan is also disseminated to private and public institutions, as well as directorates of education.

5. Funding for participation in the ILSA is sourced from loans or external donors. In the past three years, a special funding method was adopted by the MoE—the allocation of a special funding basket from a number of donor countries.

6. Funding covers all core activities of the ILSA, including international participation fees, implementation of the assessment exercise in the country, processing and analyzing data collected from implementation of the assessment exercise, reporting and disseminating the assessment results in the country, attendance at international expert meetings for the assessment exercise, and research and development.

7. Funding covers research and development activities. There are development projects within the MoE funded by donors which often cover some activities related to international assessment studies, especially those pertaining to quality assessment or impact of such programs on pilot schools. These are funded independently of joint funding programs, such as those sponsored by the World Bank and USAID.

8. There is a team and a national coordinator to carry out the ILSA activities.

9. The national coordinator is fluent in the language of the assessment.

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

10. The team responsible for carrying out the ILSA has previous experience working on international assessments, has attended some international meetings related to the international assessment, and has the necessary training or experience to carry out the required assessment activities. However, the team is not sufficiently staffed, as it is composed of seven part-time members who are responsible for providing comprehensive follow-up of all activities. These team members work under the Assessment and Evaluation Department, which carries out many of the educational assessment activities within the MoE. Additionally, there have been some issues with translation of the assessment instruments.

11. The ILSA team has attended some international workshops or meetings. However, the West Bank and Gaza is represented in many of these meetings by no more than two members due to budget constraints. In some instances, no one can attend such meetings because of security measures or visa delays.

12. The country offers some opportunities to learn about ILSA, including online courses on international assessments. However, other opportunities, including workshops or meetings on using international assessment databases, university courses on the topic of international assessments, and funding for attending international workshops or training on international assessments are not available.

13. Opportunities to learn about ILSA are available to a wide audience, in addition to the country's ILSA team members. For instance, opportunities are also available to university students studying assessment or a related area, and professionals or university staff interested in assessment. Additionally, information related to international assessments is often made available under the supervision of international and local organizations targeting quality improvement.

14. The West Bank and Gaza has met all technical standards required to have its data presented in the main displays of the international report.

15. The West Bank and Gaza has contributed to the global knowledge base on international assessments by generating new knowledge and making it available through publications. For example, research papers were published in refereed reports, the first being the IEA initiative entitled *Educational Issues in the Middle East and North Africa Region, Outcomes of the IEA Arab Region Training Seminar Series 2006/2007*. Additionally, a special chapter on the West Bank and Gaza was published in the TIMSS report, which can be accessed at www.iea.nl.

16. Country specific results and information from the most recent international assessment were disseminated in the West Bank and Gaza. For example, copies of the international report were distributed to key stakeholders, results were communicated through a press release, results received coverage on the television, radio, or newspapers, and brochures and PowerPoint presentations with the results were made available online or distributed to key stakeholders. Products providing feedback to schools and educators about the results were also made available. Additionally, Science and Mathematics curricula developers were informed about the results.

17. Products to provide feedback to schools and educators about the ILSA results are sometimes made available.

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

18. Media coverage of the ILSA results is limited to a few short articles and brief television/radio reports.

19. Results from the ILSA have been used in tracking the impact of reforms on student achievement levels, informing curriculum improvement, teacher training programs, and other assessment activities in the system. However, results have not been used in informing resource allocation.

20. Decisions based on the ILSA results have had a positive impact on students' achievement levels. For instance, supportive material covering mathematical skills encompassed in TIMSS and not included in national curricula was developed and disseminated to all schools. Additionally, translated samples of paragraphs allowed by the IEA were sent to all schools, providing a good example for teachers to use in developing their classroom exams. Also, an improvement in the educational achievement level in Mathematics was noted in the national assessment study between 2010 and 2012, which may be due to steps taken on the basis of TIMSS 2007 results at the national level.

Acknowledgements

References

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

Ministry of Education and Higher Education. 2008. "Education Development Strategic Plan 2008-2012." Ramallah: Palestinian National Authority.

United Nations Children's Fund. 2011. "Child Friendly Schools in the Occupied Palestinian Territory: A Case Study." Occupied Palestinian Territory: UNICEF.

United Nations Educational, Scientific, and Cultural Organization – Institute for Statistics. 2011. UIS Statistics in Brief. "Education Profile: Palestine." Montreal: UNESCO-UIS.

World Bank. 2013. "Country Brief: West Bank and Gaza." Washington, DC: World Bank. Data retrieved from http://web.worldbank.org/WBSITE/EXTERNAL/COUNTRIES/MENAEXT/WESTBANKGAZAEXTN/0,,menuPK:294374~pagePK:141132~piPK:141107~theSitePK:294365,00.html on October 15, 2013.

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**.

READ — Russia Education Aid for Development — When children learn, nations prosper

WORLD BANK GROUP