

Armenia



STUDENT ASSESSMENT

SABER Country Report
2011

Key Policy Areas for Student Assessment

Status

1. Classroom Assessment

In Armenia, an official curriculum/standards document specifies what students are expected to learn in different subject areas at different grade/age levels and to what level of performance. Varied and systematic mechanisms, such as an external moderation system that reviews the difficulty of classroom assessment activities and the appropriateness of scoring criteria, are in place to monitor the quality of classroom assessment practices. There also are some mechanisms to ensure that teachers develop skills and expertise in classroom assessment.

Advanced



2. Examinations

The United Examinations, which are used for high school graduation and university entrance, started to be administered less than five years ago. The examinations receive regular funding from the government and from student fees. Funding covers all core examination activities, as well as research and development. There is a clear understanding of what the United Examinations measure, and comprehensive material to prepare for the examinations is accessible to all students.

Established



3. National Large-Scale Assessment (NLSA)

An NLSA program was introduced in Armenia in the last five years, and one NLSA has been carried out (in 2010). The program receives regular funding from the government; this funding covers all core NLSA activities as well as staff training and research and development. Armenia offers some opportunities to prepare individuals for work on the NLSA. However, the Assessment and Testing Center is inadequately staffed to effectively carry out the NLSA due to the staff's limited experience in conducting NLSAs.

Emerging



4. International Large-Scale Assessment (ILSA)

Armenia has participated in three ILSAs in the last ten years (TIMSS 2003, 2007, and 2011), and has taken concrete steps to participate in two more ILSAs in the next five years (TIMSS 2015 and PIRLS 2016). A formal policy document addresses Armenia's participation in ILSAs, and regular funding for participation in ILSAs is provided by the government. Results from the ILSAs are regularly and widely disseminated, especially through widespread media coverage. However, there are no opportunities available in Armenia to learn about ILSAs.

Established



Introduction

In 2011, Armenia joined the Russia Education Aid for Development (READ) Trust Fund program, the goal of which is to help countries improve their capacity to design, carry out, analyze, and use assessments for improved student learning. As part of the READ Trust Fund program, and in order to gain a better understanding of the strengths and weaknesses of its existing assessment system, Armenia participated in a formal exercise to benchmark its student assessment system 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, the rubric displays four development levels—*Latent*, *Emerging*, *Established*, and *Advanced*.

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

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

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

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

Education in Armenia

Armenia is a lower-middle income country in the Eastern Europe and Central Asia region. GDP per capita (current US\$ 2011) is \$3305, with annual growth of approximately 4 percent. The annual growth reflects a rebound from 2009 levels and can be attributed to improved conditions in mining, manufacturing, and services such as tourism, as well as a rebound in agriculture from pre-financial crisis

levels. Armenia spent 3.2 percent of GDP on education in 2010, compared to an average 6 percent of GDP by OECD countries. Armenia's spending on education is in part reflective of the government's cuts in spending in 2009 in response to the economic crisis. The education sector was one of the most affected sectors, with a 15 percent decrease from the original budget. The Medium Term Expenditure Framework 2012-2014 indicates that spending on education will continue to decline over the next three years. As of 2011, gross enrollment at the primary level is 84 percent and the completion rate is 83 percent. Gross enrollment at the secondary level is 89 percent.

Armenia has committed to reform the education system. Reforms introduced in the past 10 years have included new curriculum for grades 1 through 9 and state standards for secondary education. These changes were accompanied by large-scale teacher training programs to support the new curriculum. The Ministry of Education also invested in improving management efficiency through the development of its Education Management Information System (EMIS). Further, large-scale assessments to evaluate and monitor the quality of learning in schools were introduced. Recently, the Armenian general education system has been extended from 10 to 12 years with the introduction of high schools with specialized streams. All schools in Armenia will have access to the internet by the end of 2011, which will allow students and teachers, especially in rural schools, to be connected to the most up-to-date information.

Detailed information was collected on Armenia's student assessment system using the SABER-Student Assessment questionnaires and rubrics in 2011. 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 Armenia, 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 Armenia's immediate interests and needs. Detailed, completed rubrics for each assessment type in Armenia are provided in Appendix 5.

Classroom Assessment

Level of development

Advanced



In Armenia, formal system-level documents provide guidelines for classroom assessment. These documents, which are available to the public, are the *General Education State order*, and the *Students' Progress Assessment Procedure* document. Additionally, there are official curriculum and standards documents available for all subjects that specify what students are expected to learn in different subject areas at different grade/age levels and to what level of performance.

There are some system-wide resources available to teachers for conducting classroom assessment activities. Along with the official curriculum and standards documents, there are textbooks and workbooks that provide support for classroom assessment activities. Scoring criteria and rubrics for evaluating students' work, as well as item banks or pools with examples of selection/multiple-choice or supply/open-ended questions are also made available to teachers; so too are brochures with sample questions for different subjects.

Some mechanisms are in place to ensure that teachers develop the necessary skills and expertise in classroom assessment. For example, teachers are offered the opportunity to participate in conferences and workshops, as well as in item development for, or scoring of, large-scale assessments or examinations. Furthermore, school inspection and teacher supervision includes a component that focuses on classroom assessment. Additionally, the World Bank Education Quality and Relevance Project (EQRP) I and EQRP II introduced and supported teacher training programs, which included training on classroom assessment. However, training on classroom assessment has not been institutionalized in pre-service teacher training institutions (such as pedagogical universities and institutes) or in the in-service teacher training programs.

Generally, classroom assessment practices in Armenia are known to be of moderate quality. Specifically, while classroom assessment activities are mainly about recalling information (as opposed to assessing higher-order thinking skills), classroom assessment activities are nevertheless aligned with pedagogical and curricular frameworks. Teachers use explicit criteria for scoring or grading students' work. Parents are well informed about students' grades and teachers provide substantial useful feedback to students.

Classroom assessment information is required to be disseminated to all key stakeholders, including students, parents, and school district and Ministry of Education officials. Additionally, classroom assessment information is required to be used in various ways in Armenia to support student learning. These required uses include diagnosing student learning issues, providing feedback to students on their learning, informing parents about their child's learning, planning the next steps in instruction, grading students for internal classroom uses, and providing input to an external examination program.

Additionally, varied and systematic mechanisms are in place to monitor the quality of classroom assessment practices, including an external moderation system that reviews the difficulty of classroom assessment activities and the appropriateness of scoring criteria. Classroom assessment is also a required component of a teacher's performance evaluation, as well as of school inspection and teacher supervision.

Suggested policy options:

1. Ensure that classroom assessment activities promote student learning by incorporating assessment activities that evaluate students' higher-order thinking skills.
2. Introduce institutionalized pre- and in-service teacher training courses on classroom assessment.

Examinations

Level of development

Established



The United Examinations, which are used for both high school graduation and university entrance, started to be administered less than five years ago. The formal policy document that authorizes the examination, the *Concept of Assessment System* (authorized by the Republic of Armenia, 2005), describes the official purposes of the examinations and the authorized uses of the results.

The Assessment and Testing Center (ATC) has been responsible for running the examinations since 2006. ATC is adequately staffed to carry out the examinations effectively, with minimal issues. Armenia offers some opportunities that prepare individuals for work on the examinations, including non-university training courses and workshops on educational measurement and evaluation.

Regular funding for the examination is provided by the government and by student fees. Funding covers all core examination activities (design, administration, data processing, and reporting) as well as research and development, staff training, and ATC facilities.

Efforts to improve the examination are generally welcomed by ATC, and the rules of the examinations and methods of administration are reviewed every year on the basis of the previous years' experience.

There is a clear understanding of what the examinations measure, and comprehensive material to prepare for the examinations is made available to all students. Preparatory material includes examples of examination questions, information on how to prepare for the examinations, and the framework document that explains what the examinations measure. Guidelines for taking the examinations, as well as samples of test materials, are annually distributed to all schools.

Some technical documentation on the examinations is developed, but it generally consists of statistical analysis of examination results and is not in a formal report format. There are also limited systematic mechanisms in place to ensure the quality of the examinations, including internal and external review or observers, and pilot or field testing. During the administration of the examinations, the ATC and the Ministry of Education representatives observe the administration procedures in every examination center. Cameras are also placed in examination centers, and parents can observe the examination process on video monitors.

Inappropriate behavior surrounding the examination process is marginal, and the examination results are viewed as credible by all stakeholder groups. All students can take the examinations; there are no language, gender, or other equivalent barriers.

There are very limited options for students who do not perform well on the examinations. For example, students may retake the examination, but only for the graduation certificate and not for university entrance. Instead, students may opt for less selective universities.

There is a committee responsible for reviewing students' appeals and errors in examinations. Currently, there are no mechanisms in place to monitor the consequences of the examinations.

Suggested policy options:

1. Introduce a variety of mechanisms to monitor the consequences of the examinations, including providing funding for independent research on the impact of the examinations, conducting regular focus groups of key stakeholders, and organizing expert review groups.
2. Introduce varied and systematic mechanisms to ensure the quality of the examinations, such as external certification or audit, and making available comprehensive, high-quality technical reports on the examinations to the general public.

National Large-Scale Assessment (NLSA)

Level of development



An NLSA program was introduced in Armenia in the last five years. An NLSA exercise, assessing grade 8 student achievement levels in Armenian language and Armenian history, was carried out in 2010. There is regular funding allocated to the NLSA program, which is provided by the Government of Armenia to the Assessment and Testing Center (ATC). The ATC is responsible for all NLSA activities. The funding covers core NLSA activities (design, administration, analysis, and reporting) as well as research and development, and staff training.

Although ATC employs full-time staff as well as temporary specialists, it is inadequately staffed to effectively carry out NLSA activities due to a lack of staff with extensive experience in carrying out this particular type of assessment activity. Specifically, the staff responsible for the NLSA program is far more familiar with graduation and university entrance examinations. It has been recognized that ATC staff needs to build capacity specifically in designing NLSAs. Currently, Armenia offers some opportunities to prepare individuals for work on NLSAs, including non-university training courses on educational measurement and evaluation, and funding for attending international programs or workshops on educational measurement and evaluation. However, funding for attending international programs and workshops is limited; therefore, only a few people have been able to participate in such international programs.

The NLSA measures performance against national curriculum or learning standards. Although what the NLSA measures is largely accepted by stakeholder groups, there is sometimes a need to explain that the NLSA is not a high-stakes assessment.

There are ad hoc reviews of the NLSA to ensure that it measures what it is intended to measure. Although there are comparisons of the results of the NLSA with the results of other assessment programs, as well as reviews of the test materials and statistical analysis of results, there are limited internal reviews of the NLSA to ensure that it measures what it is intended to measure.

To better ensure the quality of the NLSA, efforts are made to ensure a representative sample of students. For example, students from all Marzes of Armenia are selected to participate in the NLSA. Additionally, there are various mechanisms in place to ensure the technical quality of the NLSA, including that all proctors or administrators are trained according to a protocol. There is a standardized manual for NLSA administrators, and a pilot is conducted before the main data collection takes place. All booklets are numbered, there is double scoring of data, and scorers are trained to ensure high interrater reliability.

NLSA results are used by some stakeholder groups in a way that is consistent with the purposes and technical characteristics of the assessment. Specifically, the information from the NLSA is used to review the content that is being taught in schools, to provide suggestions to teachers on teaching methods, and to understand the level of students' knowledge in the tested subject areas.

Although conferences to discuss research on the consequences of the NLSA and expert review groups are being planned, there are currently no mechanisms in place to monitor the consequences of the NLSA.

Suggested policy options:

1. Conduct regular internal reviews of the NLSA to ensure that it is aligned with learning goals and measures what it is intended to measure.
2. Build the capacity of the staff of the NLSA office by introducing university graduate programs (at the masters or doctorate level) and courses (graduate and non-graduate) that are focused on educational measurement and evaluation, and specifically on conducting NLSAs.

International Large-Scale Assessment (ILSA)

Level of development

Established



Armenia has participated in three ILSAs in the last ten years – TIMSS 2003, TIMSS 2007, and TIMSS 2011 – and has taken concrete steps to participate in two ILSAs in the next five years, specifically in TIMSS 2015 and PIRLS 2016.

A formal, publically-available policy document, the *Concept of Assessment System* (authorized by the Republic of Armenia, 2005), addresses Armenia's participation in ILSAs. Regular funding for participation in ILSAs is provided by the government. The funding covers international participation fees, implementation of the assessment exercise in Armenia, processing and analyzing data collected from implementation of the assessment exercise, reporting and disseminating the assessment results in the country, and attendance at international expert meetings. The funding does not cover research and development activities.

In Armenia, a National Research Coordinator and a team are responsible for carrying out ILSA activities in the country. The team is composed of staff from the Assessment and Testing Center (ATC) as well as temporary specialists assigned to work on the ILSA. The team is sufficiently staffed, and has previous experience working on international assessments as well as the necessary training to carry out the required assessment activities effectively. While several team members have attended some of the international workshops and meetings, it was primarily the National Research Coordinator who attended the workshops and meetings. There are no opportunities to learn about ILSA in Armenia other than for those individuals who are directly involved in the assessment exercise.

Armenia met all technical standards required to have its data presented in the main displays of the international report. Additionally, Armenia has contributed new knowledge on ILSAs, and country-specific results and information have been regularly and widely disseminated in the country. As part of dissemination activities, copies of the national and international reports were distributed to key stakeholders and the key findings were communicated through a press release. Additionally, ILSA results received wide coverage on the television, radio, and newspapers, and brochures and PowerPoint presentations with Armenia's ILSA results were made available online and distributed to key stakeholders. However, products to provide feedback directly to schools and educators about the ILSA results are only sometimes made available.

Results from ILSAs are used in a variety of ways to inform decision making in Armenia. For example, ILSA results have been used by policy makers at the Ministry of Education as well as by education leaders to improve education quality in Armenia. Uses of ILSA results have included tracking the impact of reforms on student achievement levels, informing curriculum improvement, informing teacher training programs, and informing other assessment activities in the country (e.g., classroom assessment and examinations).

Suggested policy options:

1. Provide opportunities to learn about the ILSA to those individuals working directly on the specific ILSA exercise as well as university students studying educational assessment or a related area and professionals and university staff interested in assessment of student learning.
2. Such opportunities might include organizing meetings or workshops on using ILSA databases, university courses on the topic of ILSAs, on-line courses on international assessments, and funding for attending international workshops or training on ILSAs.

Appendix 1: Assessment Types and Their Key Differences

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

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

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

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

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

Appendix 4: Methodology for Assigning Development Levels

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

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.

Appendix 5: SABER-Student Assessment Rubrics for Armenia





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

ARMENIA





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: <i>Setting clear guidelines for classroom assessment</i>			
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.	There is a formal system-level document that provides guidelines for classroom assessment. ¹ *	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: <i>Aligning classroom assessment with system learning goals</i>			
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.	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: <i>Having effective human resources to carry out classroom assessment activities</i>			
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. ⁴ *	There are a variety of system-level mechanisms to ensure that teachers develop skills and expertise in classroom assessment.

ASSESSMENT QUALITY*Quality of classroom assessment design, administration, analysis, and use.*

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
ASSESSMENT QUALITY 1: <i>Ensuring the quality of classroom assessment</i>			
Classroom assessment practices suffer from widespread weaknesses or there is no information available on classroom assessment practices.	Classroom assessment practices are known to be weak.	Classroom assessment practices are known to be of moderate quality. ⁵ *	Classroom assessment practices are known to be generally of high quality.
There are no mechanisms to monitor the quality of classroom assessment practices.	There are ad hoc mechanisms to monitor the quality of classroom assessment practices.	There are limited systematic mechanisms to monitor the quality of classroom assessment practices.	There are varied and systematic mechanisms in place to monitor the quality of classroom assessment practices. ⁶ *
ASSESSMENT QUALITY 2: <i>Ensuring effective uses of classroom assessment</i>			
Classroom assessment information is not required to be disseminated to key stakeholders.	This option does not apply to this dimension.	Classroom assessment information is required to be disseminated to some key stakeholders.	Classroom assessment information is required to be disseminated to all key stakeholders. ⁷ *
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. ⁸ *











Classroom Assessment: Development level rating justifications

1. The two relevant documents are the General Education State order, and the Students Progress Assessment Procedure.
2. The documents are available to the public online, in the public libraries, as well as during in-service courses for teachers. They are also published and distributed to schools.
3. A variety of system-wide resources are available for teachers to carry out classroom assessment activities. For example, there is a document that outlines what students are expected to learn in different subject areas at different grade/age levels and to what performance level. Additionally, textbooks or workbooks are made available that provide support for classroom assessment activities. Furthermore, there are scoring criteria or rubrics for students' work as well as item banks or pools with examples of selection/multiple-choice or supply/open-ended question. There are also many brochures by various authors with examples of questions, but teachers are usually not required to use them. However, the system-wide resources are not available online.
4. The system-level mechanisms include in-service teacher training and on-line resources on classroom assessment. Additionally, all teacher training programs include a required course on classroom assessment. Armenia also offers opportunities for teachers to participate in conferences and workshops, as well as in item development for, or scoring of, large-scale assessments or exams. School inspection or teacher supervision includes a component focused on classroom assessment. The Assessment and Testing Center serves to ensure that teachers develop skills and expertise in classroom assessment. Under EQRPI and EQRPPII (Bank projects), massive training programs were and are implemented for all teachers in Armenia (on-going for more than 5 years now) and all training programs include classroom assessment. However, training on assessment is not institutionalized in teacher pre-service institutions (Pedagogical Universities and Institutes) as well as in the in-service training programs run under the state funding on the regular basis.
5. Although classroom assessment activities are mainly about recalling information (as opposed to assessing higher-order thinking skills), classroom assessment activities are aligned with pedagogical and curricular frameworks. Teachers use explicit criteria for scoring or grading students' work, and it is not common to observe errors in teachers' scoring or grading. Parents are well informed about students' grades and teachers provide substantial useful feedback to students.
6. Specifically, classroom assessment is a required component of a teacher's performance evaluation and of school inspection and teacher supervision. Additionally, there exists an external moderation system that reviews the difficulty of classroom assessment activities, appropriateness of scoring criteria, etc. System-wide reviews of the quality of education include a focus on classroom assessment, and government funding is available for research on the quality of classroom assessment activities and how to improve classroom assessment. There are also annual classroom assessment programs that are organized by the Assessment and Testing Center.
7. The key stakeholders include students, parents, and school district/Ministry of Education officials.
8. The required uses include 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 (e.g., school-based assessment with moderation and quality audit).

ARMENIA *Examinations*

ENABLING CONTEXT

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

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
ENABLING CONTEXT 1: <i>Setting clear policies</i>			
No standardized examination has taken place.	The standardized examination has been operating on an irregular basis. ¹ 	The examination is a stable program that has been operating regularly.	This option does not apply to this dimension
There is no policy document that authorizes the examination.	There is an informal or draft policy document that authorizes the examination.	There is a formal policy document that authorizes the examination. 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: <i>Having strong leadership</i>			
All stakeholder groups strongly oppose the examination or are indifferent to it.	Most stakeholder groups oppose the examination.	Most stakeholders groups support the examination.	All stakeholder groups support the examination. ⁵ 
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.	There are coordinated attempts to improve the examination by stakeholder groups. ⁶ 
Efforts to improve the examination are not welcomed by the leadership in charge of the examination	This option does not apply to this dimension.	Efforts to improve the examination are generally welcomed by the leadership in charge of the examination. ⁷ 	This option does not apply to this dimension.

(CONTINUED)

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





(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 5: <i>Having effective human resources</i>			
There is no staff to carry out the examination.	The examination office is inadequately staffed to effectively carry out the examination, issues are pervasive.	The examination office is adequately staffed to carry out the examination effectively, with minimal issues. ¹⁴ *	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. ¹⁵ *	The country offers a wide range of opportunities that prepare for work on the examination.





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

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

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

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

(CONTINUED)

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
ASSESSMENT QUALITY 3: <i>Using examination information in a fair way</i>			
Examination results are not used in a proper way by all stakeholder groups.	Examination results are used by some stakeholder groups in a proper way.	Examination results are used by most stakeholder groups in a proper way.	Examination results are used by all stakeholder groups in a proper way. *
Student names and results are public. ²³ *	This option does not apply to this dimension.	Students' results are confidential.	This option does not apply to this dimension.
ASSESSMENT QUALITY 4: <i>Ensuring positive consequences of the examination</i>			
There are no options for students who do not perform well on the examination, or students must leave the education system.	There are very limited options for students who do not perform well on the examination. ²⁴ *	There are some options for students who do not perform well on the examination.	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. ²⁵ *	This option does not apply to this dimension.	There are some mechanisms in place to monitor the consequences of the examination.	There is a variety of mechanisms in place to monitor the consequences of the examination.

Examinations: Development level rating justifications










1. The program has been operating for less than five years.
2. The policy document is called "Concept of Assessment System", Government of RA, 2005
3. The policy document is published and distributed.
4. The policy document outlines governance, distribution of power, responsibilities among key entities, describes the purpose of the examination, and it describes authorized uses of results.
5. The Ministry of Education supports the examinations the most.
6. The Government provides funding sources, the Ministry of Education and Science provides support in organizing, administrating and using the examinations. Teachers, educators and parents suggest new ideas.
7. The rules of the examinations, their structures, and methods of administration are revised each year on the basis of previous years' experience.
8. Funding comes from the government as well as from the fees that students pay to take the examinations.
9. Funding also covers long- or medium-term planning of program milestones and staff training, printing of materials, pens, computers, scanners, etc.
10. The Assessment and Testing Center has had the responsibility for running the examinations since 2006.
11. The examination office is accountable to an external board that is comprised of representatives from various stakeholder groups.
12. Students receive certificates with their scores, which they can use to apply to the university of their choice.
13. This includes computers for all technical staff, a secure building, secure storage facilities, access to adequate computer servers, the ability to backup data, and adequate communication tools [phone, email, internet].
14. During the preparation and administration of the examinations there is some need for outsourced human resources. Some errors in the examination questions have been identified as a result of using outsourced human resources. These errors are mostly typos.
15. These include non-university training courses or workshops on educational measurement and evaluation and funding for attending international programs, courses or workshops on educational measurement and evaluation.

16. The examinations measure student knowledge against the national school curriculum guidelines and standards for the purposes of graduating from high school and entering university.
17. Material that is needed to prepare for the examinations is widely accessible by all students (over 90%) in a variety of learning contexts. This material is published and made publicly available. It includes examples of the types of questions that are on the examination, information on how to prepare for the examination and the framework document that explains what the examinations measure. Guidelines and examples of test material are annually distributed to all schools.
18. The Assessment and Testing Center organizes voluntary courses for teachers in selected schools. Additionally, there is compulsory annual training and courses for teachers working on the examinations. The National Institute of Education organizes courses for all teachers as well.
19. Teachers are involved in selecting or creating examination questions, administering the examination, acting as judges (i.e., in orals) and supervising the examination procedures.
20. The documentation usually consists of statistical analysis of examination results.
21. The systematic mechanisms include internal and external review or observers, and pilot or field testing. Additionally, the Assessment and Testing Center and the Ministry of Education have their representatives in every examination center to observe the administration. Parents can also observe the examination process on special monitors.
22. The scoring of the examination answers is done by scanner which compares the student's answers with the correct answer keys. Each student has a copy of his answers and has the opportunity to obtain his score (answer keys are available on the internet).
23. The names of students and their results are posted in the examination center after each examination.
24. Students may retake the examination only for the graduation certificate, but not for university entrance. Students may opt for less selective universities. Students may also repeat a grade.
25. There is a committee the role of which is to consider student's appeals and errors in test items.

ARMENIA
National (or System-Level) Large-Scale Assessment (NLSA)

ENABLING CONTEXT

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

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

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 3: <i>Having regular funding for NLSA</i>			
There is no funding allocated to the NLSA.	There is irregular funding allocated to the NLSA.	There is regular funding allocated to the NLSA. ⁵ *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding covers some core NLSA activities: design, administration, analysis and reporting.	Funding covers all core NLSA activities: design, administration, analysis and reporting. ⁶ *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding does not cover research and development activities.	This option does not apply to this dimension.	Funding covers research and development activities. *
ENABLING CONTEXT 4: <i>Having strong organizational structures for NLSA</i>			
There is no NLSA office, ad hoc unit or team.	The NLSA office is a temporary agency or group of people.	The NLSA office is a permanent agency, institution or unit. ⁷ *	This option does not apply to this dimension.
This option does not apply to this dimension.	Political considerations regularly hamper technical considerations.	Political considerations sometimes hamper technical considerations.	Political considerations never hamper technical considerations. *
This option does not apply to this dimension.	The NLSA office is not accountable to a clearly recognized body.	The NLSA office is accountable to a clearly recognized body. ⁸ *	This option does not apply to this dimension.





(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 5: <i>Having effective human resources for NLSA</i>			
There is no staff allocated for running an NLSA.	The NLSA office is inadequately staffed to effectively carry out the assessment. ⁹ *	The NLSA office is adequately staffed to carry out the NLSA effectively, with minimal issues.	The NLSA office is adequately staffed to carry out the NLSA effectively, with no issues.
The country 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. ¹⁰ *	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: <i>Aligning the NLSA with learning goals</i>			
It is not clear if the NLSA is based on curriculum or learning standards.	This option does not apply to this dimension.	The NLSA measures performance against curriculum or learning standards. *	This option does not apply to this dimension.
What the NLSA measures is generally questioned by stakeholder groups.	This option does not apply to this dimension.	What the NLSA measures is questioned by some stakeholder groups.	What the NLSA measures is largely accepted by stakeholder groups. ¹¹ *
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. ¹² *	There are regular internal reviews of the NLSA to ensure that it measures what it is intended to measure.	This option does not apply to this dimension.
SYSTEM ALIGNMENT 2: <i>Providing teachers with opportunities to learn about the NLSA</i>			
There are no courses or workshops on the NLSA.	There are occasional courses or workshops on the NLSA.	There are some courses or workshops on the NLSA offered on a regular basis. ¹³ *	There are widely available high quality courses or workshops on the NLSA offered on a regular basis.

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

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
ASSESSMENT QUALITY 1: <i>Ensuring the quality of the NLSA</i>			
No options are offered to include all groups of students in the NLSA.	This option does not apply to this dimension.	At least one option is offered to include all groups of students in the NLSA. ¹⁴ *	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. ¹⁵ *
There is no technical report or other documentation about the NLSA.	There is some documentation about the technical aspects of the NLSA, but it is not in a formal report format.	There is a comprehensive technical report but with restricted circulation. ¹⁶ *	There is a comprehensive, high quality technical report available to the general public.
ASSESSMENT QUALITY 2: <i>Ensuring effective uses of the NLSA</i>			
NLSA results are not disseminated.	NLSA results are poorly disseminated.	NLSA results are disseminated in an effective way. ¹⁷ *	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 technical characteristics of the assessment. ¹⁸ *	NLSA information is used by all stakeholder groups in a way that is consistent with the purposes and technical characteristics of the assessment.
There are no mechanisms in place to monitor the consequences of the NLSA. ¹⁹ *	This option does not apply to this dimension.	There are some mechanisms in place to monitor the consequences of the NLSA.	There are a variety of mechanisms in place to monitor the consequences of the NLSA.

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





1. The program has been operating for less than five years. One NLSA has been carried out (in 2010).
2. The policy document is called "Concept of Assessment System", Government of RA, 2005.
3. Yes, in the form of a booklet.
4. Policymakers, teachers unions, students, and the media are supportive of the NLSA, while the educators, parents, employers, think-tanks and universities are neutral to it. The NLSA is a focus of the Ministry of Education and Science, as well as teachers, parents, and students, and is highlighted by the media.
5. The funding is regular and provided by the government and is part of the annual budget of the Assessment and Testing Center.
6. Funding also includes staff training, printing of booklets, and trips to regional schools.
7. It is the Assessment and Testing Center with permanent staff as well as temporary specialists.
8. The Assessment and Testing Center is accountable to its board and to the government.
9. There is the full-time staff of the Assessment and Testing Center and some temporarily assigned specialists. However, the office is inadequately staffed because there are some issues due to the lack of extensive experience of the staff in carrying out the NLSA. Specifically, the staff responsible for the NLSA is more familiar with graduation exams and initially designed items using a similar approach as the one used for exams. The staff needs to build capacity in specifically designing NLSAs.
10. Opportunities include non-university training courses or workshops on educational measurement and evaluation, and funding for attending international programs or courses or workshops on educational measurement and evaluation. Training courses are considered very productive because they are conducted on a regular basis and are available to more people. Funding for attending international programs and workshops is rare, and only a few people have participated in them.
11. Sometimes there is a need to explain that the NLSA is not high-stakes.
12. Although there are comparisons of the results of the NLSA with the results of other assessment programs, there are limited internal reviews of the NLSA to ensure that it measures what it is intended to measure. Specifically, there are reviews of the test materials, and statistical analysis of results is conducted.
13. There are regularly available and occasional courses or workshops, as well as occasional presentations on the NLSA. Most teachers have access to the live courses or workshops. Trainings and seminars to present the results of the NLSA also have been planned.
14. Schools from all Marzes of Armenia are selected.

15. Such mechanisms include: all proctors or administrators are trained according to a protocol; there is a standardized manual for NLSA administrators; a pilot is conducted before the main data collection takes place; all booklets are numbered; there is double data scoring; scorers are trained to ensure high interrater reliability; and the Coordinator of the NLSA program has considered and approved all disputable situations with scores. The results of the statistical analyses is used to improve the quality of the test items and to make decisions regarding testing subjects.
16. While the current version of the report has restricted circulation, a comprehensive, high-quality technical report is being planned.
17. Specifically, results are disseminated within twelve months after the NLSA is administered; reports with results are made available for all stakeholder groups; the main reports on the results contain information on overall achievement levels and subgroups; the main reports on the results contain information on trends over time, overall and for subgroups; the main reports on the results contain standard errors (measure of uncertainty); there is a media briefing organized to discuss results; there are workshops or presentations for key stakeholders on the results; and results are featured in newspapers, magazines, radio and television. Currently, a report is being prepared to be published.
18. Assessment information is used to review the content of the subjects being tested, to provide suggestions to teachers on teaching methods, and to understand the level of students' knowledge in the tested subject areas.
19. Themed conferences that provide a forum to discuss research and other data on the consequences of the NLSA and expert review groups are being planned.

ARMENIA
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: <i>Setting clear policies for ILSA</i>			
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.	There is a formal policy document that addresses participation in ILSA. ³ *	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: <i>Having regular funding for ILSA</i>			
There is no funding for participation in ILSA.	There is funding from loans or external donors.	There is regular funding allocated at discretion.	There is regular funding approved by law, decree or norm. *
This option does not apply to this dimension.	Funding covers some core activities of the ILSA.	Funding covers all core activities of the ILSA. ⁵ *	This option does not apply to this dimension.
Funding does not cover research and development activities. *	This option does not apply to this dimension.	This option does not apply to this dimension.	Funding covers research and development activities.





(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 3: <i>Having effective human resources for ILSA</i>			
There is no team or national/system coordinator to carry out the ILSA activities.	There is a team or national/system coordinator to carry out the ILSA activities.	There is a team and national/system coordinator to carry out the ILSA activities. ⁶ *	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. ⁷ *	This option does not apply to this dimension.
This option does not apply to this dimension.	The ILSA office is inadequately staffed or trained to carry out the assessment effectively.	The ILSA office is adequately staffed or trained to carry out the ILSA effectively, with minimal issues.	The ILSA office is adequately staffed and trained to carry out the ILSA effectively, with no issues. ⁸ *

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

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
SYSTEM ALIGNMENT 1: <i>Providing opportunities to learn about ILSA</i>			
The ILSA team has not attended international workshops or meetings.	The ILSA team attended some international workshops or meetings. ⁹ *	The ILSA team attended all international workshops or meetings.	This option does not apply to this dimension.
The country/system offers no opportunities to learn about ILSA. *	This option does not apply to this dimension.	The country/system offers some opportunities to learn about ILSA.	The country/system offers a wide range of opportunities to learn about ILSA.
This option does not apply to this dimension.	This option does not apply to this dimension.	Opportunities to learn about ILSA are available to the country's/system's ILSA team members only.	Opportunities to learn about ILSA are available to a wide audience, in addition to the country's/system's ILSA team members.

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

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
ASSESSMENT QUALITY 1: <i>Ensuring the quality of ILSA</i>			
Data from the ILSA has not been published.	The country/system met sufficient standards to have its data presented beneath the main display of the international report or in an annex.	The country/system met all technical standards required to have its data presented in the main displays of the international report.	The country/system met all technical standards required to have its data presented in the main displays of the international report. *
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. ¹⁰ *
ASSESSMENT QUALITY 2: <i>Ensuring effective uses of ILSA</i>			
If any, country/system-specific results and information are not disseminated in the country/system.	Country/system-specific results and information are disseminated irregularly in the country/system.	Country/system-specific results and information are regularly disseminated in the country/system.	Country/system-specific results and information are regularly and widely disseminated in the country/system. ¹¹ *
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. ¹² *	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. ¹⁷	There is some media coverage of the ILSA results.	There is wide media coverage of the ILSA results. ¹³ *
If any, country/system-specific results and information from the ILSA are not used to inform decision making in the country/system.	Results from the ILSA are used in a limited way to inform decision making in the country/system.	Results from the ILSA are used in some ways to inform decision making in the country/system.	Results from the ILSA are used in a variety of ways to inform decision making in the country/system. ¹⁴ *
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. ¹⁵ *

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

1. Armenia has participated in TIMSS 2003, 2007, and 2011.
2. The country has taken concrete steps to participate in PIRLS 2016 and TIMSS 2015. Specifically, the Assessment and Testing Center has requested funding from the government for participation in these ILSAs.
3. The policy document is called "Concept of Assessment System", Government of RA, 2005.
4. The policy document is available in the form of a booklet.
5. Funding also covers attendance at international expert meetings for the assessment exercise.
6. By the order of the Minister of Education (in 1999), there is a national coordinator responsible for the ILSA. The staff of the Assessment and Testing Center is responsible for carrying out the ILSA in Armenia, along with some temporary specialists assigned to work on the ILSA.
7. The team has participated in four rounds of TIMSS.
8. The team is sufficiently staffed, and has previous experience working on international assessments as well as the necessary training to carry out the required assessment activities effectively.
9. While team members have attended some of the international workshops/meetings, it was mostly the National Research Coordinator who attended the workshops/meetings.
10. There are printed and distributed brochures on the research of TIMSS.
11. Dissemination activities included: copies of the national report were distributed to key stakeholders; copies of the international report were distributed to key stakeholders and the key findings communicated through a press release; results received coverage on the television, radio or newspapers, and brochures and PowerPoint presentations with the country's/system's results were made available online or distributed to key stakeholders.
12. However, not all educators receive the results.
13. There are editorials or columns commenting on the international assessment results, and the assessment results have been reported on in newspapers, radio and on television.
14. ILSA results have been used by policy makers in the Ministry of Education and education leaders to improve education quality in Armenia by tracking the impact of reforms on student achievement levels, informing curriculum improvement, informing teacher training programs and informing other assessment activities in the country (e.g., classroom assessment, examinations).
15. The student achievement levels for TIMSS 2007 were significantly better than for TIMSS 2003. The student achievement levels for TIMSS 2011 will be released in 2012.

Acknowledgements

This report was prepared by The World Bank SABER-Student Assessment team in collaboration with Cristian Aedo, World Bank Senior Education Specialist and Task Team Leader for education projects in Armenia, and Arsen Baghdasaryan, Deputy Director of the Assessment and Testing Center in Armenia.

References

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

Organisation for Economic Co-operation and Development (OECD). 2011. "Education at a Glance 2011: OECD Indicators." Paris: OECD.

United Nations Children's Fund (UNICEF). 2013. Education in Armenia. Data retrieved from <http://www.unicef.org/armenia/education.html> on March 12, 2013.

———. 2010. "Country profile: Education in Armenia 2010." Data retrieved from http://www.unicef.org/ceecis/Armenia_2010.pdf on March 12, 2013.

United Nations Educational, Scientific, and Cultural Organization (UNESCO) - Institute for Statistics. Armenia Country Indicator Data. Montreal, QC: UNESCO. Data retrieved from <http://stats.uis.unesco.org/unesco> on March 12, 2013.

World Bank. 2013. World Bank Development Indicators: Armenia Country Indicator Data. Washington, DC: World Bank. Data retrieved from <http://databank.worldbank.org/data> on March 12, 2013.

———. 2009. Country Partnership Strategy for the Republic of Armenia for the Period FY09-FY12. Report No. 48222-AM. Washington, DC: World Bank.

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

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

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



THE WORLD BANK