

# Arab Republic of Egypt



SABER Country Report  
2013

## STUDENT ASSESSMENT

### Key Policy Areas for Student Assessment

#### 1. Classroom Assessment

In the Arab Republic of Egypt, a formal, publicly available, system-level document provides guidelines for classroom assessment. System-wide resources and materials are also available to teachers for carrying out their assessment activities. These include documents that describe expected learning outcomes for students in different subject areas and grades, as well as student workbooks. At the same time, classroom assessment activities are viewed as relying too much on information recall and are generally considered to be weak. In addition, although classroom assessment information must be disseminated to key stakeholders, required uses of the data to support student learning are limited.

#### Status



#### 2. Examinations

The Certificate of Completion of General Secondary Education has been administered to grade 12 students since 1954. The examination results are officially recognized by certification and selection systems in Egypt and abroad. Regular funding for the examination is provided by the government and covers all core activities. At the same time, limited mechanisms are in place to ensure the quality of the examination. In this regard, a recent OECD/World Bank study (forthcoming) noted that because of the lack of pilot testing of examination questions, ascertaining their difficulty level and comparability across examination cohorts is problematic.



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

The National Standardized Examination has been administered twice in the past five years to a random sample of students in grades 4, 8, and 10 in the subject areas of Arabic language, English language, mathematics, and science. An informal policy document authorizes the NLSA but is not available to the public. In addition, no formal plan is in place for future activities. Funding covers core NLSA activities but is provided on an irregular basis by the government. NLSA results are poorly disseminated, and no mechanisms are in place to monitor the consequences of the assessment in terms of its overall impact on education quality or student learning.



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

In the last 10 years, Egypt has participated in Trends in the International Mathematics and Science Study (TIMSS) 2003 and TIMSS 2007. Currently Egypt has no formal policy document that addresses participation in ILSAs and no official plan to participate in future ILSAs. Funding for TIMSS 2007 was provided by the government. Egypt offers limited opportunities to learn about ILSAs, and these have been made available only to ILSA team members. ILSA results and information have not been systematically disseminated throughout the country.



## Introduction

The Arab Republic of Egypt 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 the quality of education and learning outcomes because it provides the necessary information to meet stakeholders' decision-making needs. To gain a better understanding of the strengths and weaknesses of its existing assessment system, Egypt 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 three to five 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 Egypt

The Arab Republic of Egypt is a lower-middle-income country in North Africa. GDP per capita is \$3,187, with annual growth of approximately 2.2 percent. Egypt has undergone political changes since the 2011 revolution. The political situation remains precarious, with an interim government currently in place.

Egypt made significant progress in educational attainment in the last two decades, particularly in expanding access to basic education and closing the gap between boys' and girls' enrollment. Egypt has attained a primary school net enrollment rate of 96 percent and has managed to close the gender gap at the secondary and higher education levels. However, Egypt faces significant disparities in the area of education quality. Student performance on the 2007 Trends in International Mathematics and Science Study (TIMSS) showed large gaps between the top and bottom achievers, and analysis of the scores showed that performance depends to a large extent on parental education, wealth, and geographic location.

Egypt is committed to investing in education, as illustrated in the *National Strategic Plan* in which education is considered to be a major pillar of economic and social development as well as national security. Government priorities for educational reform include institutionalized decentralization, strengthening partnerships with civil society and the private sector, promoting community participation in the management of schools, and upgrading teachers' socioeconomic and professional status.

Detailed information was collected on Egypt'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 macrolevel. Additional data would need to be collected to determine actual, on-the-ground practices in Egypt, 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 Egypt's immediate interests and needs. Detailed, completed rubrics for each assessment type in Egypt are provided in Appendix 5.

## Classroom Assessment

---

### Level of development



In Egypt, the Ministerial Decision Number 313 document, authorized by the Ministry of Education in 2011, provides guidelines for classroom assessment.

Some system-wide resources are available to help teachers engage in classroom assessment activities, including workbooks that provide support for classroom assessment and a document that outlines what students are expected to learn in different subject areas at different grade levels.

Some system-level mechanisms are in place to ensure that teachers develop skills and expertise in classroom assessment. Such mechanisms include in-service teacher training, online resources on classroom assessment, and a component on classroom assessment in teacher supervision.

Required uses of classroom assessment to support student learning are limited. Although classroom assessment activities are required to be used in diagnosing student learning issues, providing feedback to students on their learning, and planning next steps in instruction, they are not required to be used to inform parents about their child's learning.

In Egypt, classroom assessment practices are known to be weak. Although classroom assessment activities tend to be aligned with the curriculum framework, and grade inflation is not a serious problem, it is common for these assessment activities to be mainly about information recall. It is also common for classroom assessment activities to rely mainly on multiple-choice, selection-type questions and for teachers not to use explicit criteria for scoring students' work.

Some mechanisms are in place to monitor the quality of classroom assessment practices; for example, classroom assessment is a required component of a teacher's performance evaluation and supervision.

### Suggested policy options

1. Provide teachers with a greater variety of system-level resources to engage in classroom assessment, including a document that outlines the levels of performance that students are expected to reach in different subject areas at different grade and age levels, and scoring criteria or rubrics for evaluating students' work.
2. Introduce more system-level mechanisms to ensure that teachers develop skills and expertise in classroom assessment, such as a required course on classroom assessment in all teacher-training programs, opportunities to participate in conferences and workshops, and opportunities to develop or score items for large-scale assessments or exams.
3. Ensure that national reviews of the quality of education include a systematic focus on classroom assessment. Introduce additional, systematic mechanisms, such as an external moderation system and government funding for research on how to improve and ensure the quality of classroom assessment practices.



## Examinations

---

### Level of development



The Certificate of Completion of General Secondary Education was first administered in 1954 and formally authorized in 2012 by the Presidency of the Republic via the Law 20 of 2012 document. The examination is administered to students in grade 12 and covers Arabic language, religious education, civic education, foreign languages, mathematics, sciences, and social sciences.

The government allocates regular funding for the examination. Funding covers examination design and administration, data analysis, and reporting but does not cover long- or medium-term planning of program milestones, staff training, or research and development activities.

The General Directorate for Examinations and the General Administration of Examinations are two departments within the Ministry of Education that have had primary responsibility for running the examination since 1955. They have all of the required facilities for conducting the examination and are adequately staffed with permanent and full-time staff. Reportedly, there are frequent errors in data processing, but no real issues have been identified with other aspects of the examination. However, a study discussed in the report, “Review of National Policies for Education: Schools for Skills—A New Learning Agenda for Egypt,” prepared by the Organisation for Economic Co-operation and Development (OECD) and the World Bank, found that because examinations are scored by readers who have little or no training in essay scoring, inconsistencies may arise in scores within a cohort (OECD/World Bank, forthcoming, 142).

Policy makers, teacher unions, and parents strongly support the examination, along with educators and the media. Although think tanks and nongovernmental organizations (NGOs) oppose the examination, students and employers appear neutral toward it. Some attitudes toward the examination may be informed by the prevalence of private tutoring. The

OECD/World Bank report found that private tutoring can be expensive relative to the average family income and consume students’ time (OECD/World Bank, forthcoming, 140–1).

Egypt offers some opportunities to prepare individuals for work on the examination, including university graduate programs and courses on educational measurement and evaluation in university faculties of education. Nonuniversity training courses and workshops on educational measurement and evaluation are also offered.

At the same time, no courses or workshops on the examination are available to teachers. In general, teachers are involved in very few examination tasks, which include administering and scoring the examination and supervising examination procedures. Teachers are not involved in creating or selecting examination questions or scoring guides, acting as a judge, or resolving inconsistencies between examination scores and school grades.

Inappropriate behavior surrounding the examination process is moderate and has included leakage of the content of an examination paper before the examination and the use of unauthorized materials, such as prepared answers and notes.

Student results are confidential. However, the only mechanism in place to monitor the consequences of the examination for students’ learning and educational opportunities is a set of studies that are regularly updated by the National Center for Examination and Educational Evaluation.

Only limited systematic mechanisms are in place to ensure the quality of the examination, including external and internal reviews conducted by committees responsible for preparing the examination (by the National Center for Examinations, which is external to the Ministry of Education, as well as a committee internal to the Ministry of Education) and external certifications and audits. However, other desirable mechanisms, such as pilots, field testing, or translation verifications, are not in place. The OECD/World Bank report found that because of the lack of pilot testing of the examination questions, the real difficulty level of the questions and their comparability across cohorts is difficult to evaluate (OECD/World Bank, forthcoming, 142).

## Suggested policy options

1. Introduce a document that specifies fair examination practices and test takers' rights, and make it available to all stakeholders to ensure that examination practices are consistent across the country.
2. Provide teachers with opportunities to learn about the examination by developing and making available up-to-date, compulsory courses or workshops on examinations for teachers, and by involving teachers in a variety of examination-related tasks.
3. Introduce varied, systematic mechanisms to ensure the quality of the examination, such as pilots or field testing.
4. Introduce varied and systematic mechanisms to monitor the consequences of the examination, such as funding for independent research on the examination, a permanent oversight committee, regular focus groups or surveys of key stakeholders, or expert review groups.

## National Large-Scale Assessment (NLSA)

---

### Level of development



The National Standardized Examination has been administered twice in the last five years to a representative random sample of students in grades 4, 8, and 10 in Arabic language, English language, mathematics, and science. The main purposes of this NLSA are to monitor education quality at the system level and hold political authorities, educators, and students accountable. The assessment also supports policy design, evaluation, and decision making.

The government allocates irregular funding for the NLSA. The funding covers all core activities, including assessment design and administration, data analysis and reporting, long- and medium-term planning of program milestones, and staff training. However, the funding does not cover research and development activities.

The National Center for Examinations and Educational Evaluation is a permanent unit created for running the NLSA. It is adequately staffed with permanent and full-time staff.

Egypt offers some opportunities to prepare individuals for work on the NLSA, including university graduate programs and courses on educational measurement and evaluation in university faculties of education. Nonuniversity courses and workshops on educational measurement and evaluation are offered during in-service training. However, no internships or short-term employment opportunities are offered in the NLSA office, and no funding is available for attending international programs, courses, or workshops on educational measurement and evaluation.

Mechanisms are in place to ensure the quality of the NLSA, including a standardized manual for assessment administrators and training of scorers to ensure high interrater reliability. However, discrepancies are not required to be recorded on a standard sheet, and no double scoring or processing of data take place.

Although the main reports on the NLSA results contain useful information on achievement levels overall and by subgroups, these results are not disseminated until

more than 12 months after the assessment is administered. In addition, the reports are not made available to all stakeholder groups. The reports do not contain information on trends over time and do not provide standard errors. Furthermore, there is no media briefing organized to discuss the results, and results are not featured in newspapers or magazines or on radio or television.

No mechanisms are in place as well to monitor the consequences of the NLSA in terms of its overall impact on education quality levels in the system.

### Suggested policy options

1. Develop a formal policy document that authorizes the NLSA and provides a timeline for when key NLSA activities are to take place.
2. Ensure that NLSA results are made available to stakeholders and researchers in a timely manner so that they can use them to study the effectiveness and efficiency of the national education system and to inform curriculum, teaching practices, and teacher training programs.
3. Evaluate how NLSA and ILSA results compare to inform curriculum development as well as teaching and learning practices.
4. Introduce mechanisms to monitor the consequences of the NLSA, including expert review groups and funding for independent research on the impact of the NLSA.



## International Large-Scale Assessment (ILSA)

---

### Level of development

Emerging



Egypt has participated in two ILSAs in the past 10 years: the Trends in International Mathematics and Science Study (TIMSS) 2003 and TIMSS 2007. However, the country has not taken concrete steps to participate in another ILSA in the next five years, and no policy document addresses its participation in ILSAs.

Funding for TIMSS 2007 was provided by the Ministry of Education. Funding covered all core activities, including international participation fees, implementation of the assessment exercise in Egypt, processing and analysis of data collected from implementation of the assessment exercise, reporting and dissemination of the assessment results in Egypt, and attendance at international expert meetings for the assessment exercise. However, funding did not cover research and development activities.

Egypt's ILSA office is adequately staffed and trained to carry out the international assessment. The ILSA team has previous experience working on international assessments, and ILSA team members have attended some of the international expert meetings related to the assessment.

However, Egypt offers very limited opportunities for individuals to learn about ILSAs. Although workshops are available on using international assessment databases, no university courses include the topic of international assessments.

Egypt-specific results and information from ILSAs have been disseminated irregularly in the country. Although copies of the international reports were distributed to key stakeholders, copies of the national reports were not distributed, and a national report was not made available online. In addition, media coverage of Egypt's results was limited.

Results from ILSAs have been used in a variety of ways to inform decision making in Egypt. Results have been used to track the impact of reforms on student achievement levels and to inform curriculum improvement, teacher training programs, and other assessment activities in the country. For example, results have been used to modify the mathematics and science curricula and train teachers on teaching methods aligned with the modifications.

### Suggested policy options

1. Develop a formal policy document that addresses Egypt's participation in ILSAs.
2. Allocate government funding for ILSA activities to ensure that all ILSA activities are completed correctly and in a timely manner. Provide technical information to decision makers regarding the quality of the outcomes of the Egyptian education system compared with those of other countries.
3. Ensure that targeted reports on ILSA results are disseminated to key stakeholders at the school, district, and governorate levels to support continuous improvement plans.
4. Evaluate how NLSA and ILSA results compare to inform curriculum development as well as teaching and learning practices.

## Appendix 1: Assessment Types and Their Key Differences

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

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

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

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

Assessment type	LATENT	EMERGING	ESTABLISHED	ADVANCED
	<i>Absence of, or deviation from, the attribute</i>	<i>On way to meeting minimum standard</i>	<i>Acceptable minimum standard</i>	<i>Best practice</i>
<b>Classroom assessment</b>	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.
<b>Examinations</b>	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.
<b>National (or system-level) large-scale assessment</b>	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.
<b>International large-scale assessment</b>	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 are 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 three 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 Egypt**





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



# **Egypt, Arab Rep.** *Classroom Assessment*





### ENABLING CONTEXT AND SYSTEM ALIGNMENT

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

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
<b>ENABLING CONTEXT AND SYSTEM ALIGNMENT 1:</b> <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. <sup>1</sup> *	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. <sup>2</sup> *
<b>ENABLING CONTEXT AND SYSTEM ALIGNMENT 2:</b> <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. <sup>3</sup> *	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. <sup>4</sup> *	There is an official curriculum or standards document that specifies what students are expected to learn and to what level of performance.
<b>ENABLING CONTEXT AND SYSTEM ALIGNMENT 3:</b> <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. <sup>5</sup> *	There are a variety of system-level mechanisms to ensure that teachers develop skills and expertise in classroom assessment.

## ASSESSMENT QUALITY

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

<b>LATENT</b> 	<b>EMERGING</b> 	<b>ESTABLISHED</b> 	<b>ADVANCED</b> 
<b>ASSESSMENT QUALITY 1:</b> <i>Ensuring the quality of classroom assessment</i>			
Classroom assessment practices suffer from widespread weaknesses or there is no information available on classroom assessment practices.	Classroom assessment practices are known to be weak. <sup>6</sup> *	Classroom assessment practices are known to be of moderate quality.	Classroom assessment practices are known to be generally of high quality.
There are no mechanisms to monitor the quality of classroom assessment practices.	There are ad hoc mechanisms to monitor the quality of classroom assessment practices. <sup>7</sup> *	There are limited systematic mechanisms to monitor the quality of classroom assessment practices.	There are varied and systematic mechanisms in place to monitor the quality of classroom assessment practices.
<b>ASSESSMENT QUALITY 2:</b> <i>Ensuring effective uses of classroom assessment</i>			
Classroom assessment information is not required to be disseminated to key stakeholders.	This option does not apply to this dimension.	Classroom assessment information is required to be disseminated to some key stakeholders. <sup>8</sup> *	Classroom assessment information is required to be disseminated to all key stakeholders.
There are no required uses of classroom assessment to support student learning.	There are limited required uses of classroom assessment to support student learning. <sup>9</sup> *	There are adequate required uses of classroom assessment to support student learning, excluding its use as an input for external examination results.	There are adequate required uses of classroom assessment to support student learning, including its use as an input for external examination results.

***Classroom Assessment:*** Development-Level Rating Justifications

1. The Ministry of Education authorized the document Ministerial Decision Number 313 in 2011 to provide formal guidelines for classroom assessment.
2. The document Ministerial Decision Number 313 is available to the public online at the Egyptian Ministry of Education's website, in in-service courses for teachers, and in schools, Directorates of Education, and the Provincial Education Administrations.
3. Some system-wide resources are available for teachers to engage in classroom assessment activities. For example, textbooks or workbooks that provide support for classroom assessment and the document that outlines what students are expected to learn in different subject areas at different grade and age levels are available. However, the document that outlines the levels of performance that students are expected to reach in different subject areas at different grade or age levels, scoring criteria or rubrics for students' work, item banks or pools with examples of selection/multiple-choice or supply/open-ended questions, online assessment resources, and computer-based testing with instant reports on students' performance are not available.
4. An official curriculum or standards document that specifies what students at different grade or age levels are expected to learn is available, but it does not specify to what performance level.
5. Some system-level mechanisms, including in-service teacher training, online resources on classroom assessment, and school inspection or teacher supervision, which includes a component focused on classroom assessment, are in place to ensure that teachers develop skills and expertise in classroom assessment. For example, the Professional Academy of Teachers trains teachers on classroom assessment procedures. However, other mechanisms, such as preservice teacher training, a required course on classroom assessment in all teacher training programs, and opportunities to participate in conferences and workshops or item development for, or scoring of, large-scale assessments or exams, are not in place.
6. Classroom assessment practices are known to be weak. Classroom activities are rarely not aligned with a pedagogical or curricular framework, and it is not common for grade inflation to be a serious problem. Additionally, it is not common for parents to be poorly informed about students' grades. It is also not common for classroom activities to provide little useful feedback to students and to mainly be used as administrative or control tools rather than as pedagogical resources. However, it is very common for classroom assessment activities to be mainly about recalling information, and it is common for classroom assessment activities to rely mainly on multiple-choice, selection-type questions. It is also common for teachers to not use explicit or a priori criteria for scoring or grading students' work.
7. Ad hoc mechanisms are in place to monitor the quality of classroom assessment practices. Classroom assessment is a required component of a teacher's performance evaluation and teacher supervision. Although limited in scope, classroom assessment practices are reviewed by external audit committees, which are part of the National Authority for Quality Assurance and Accreditation in Education in Egypt (NAQQAEE). However, no government funding is available for research on the quality of classroom assessment activities or on how to improve classroom assessment, and no external moderation system is set up to review the difficulty of classroom assessment activities and the appropriateness of scoring criteria.





8. Schools or teachers are required to report on individual students' performance to parents and students. Classroom assessment information is not required to be disseminated to school district or Ministry of Education officials.
9. Although classroom assessment activities are required for use in diagnosing student learning issues, providing feedback to students on their learning, and planning next steps in instruction, they are not required to be used in informing parents about their child's learning, grading students for internal classroom uses, or providing input to an external examination program.

# **Egypt, Arab Rep.** *Examinations*



**ENABLING CONTEXT**

*Overall framework of policies, leadership, organizational structures, and 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.*

<b>LATENT</b> 	<b>EMERGING</b> 	<b>ESTABLISHED</b> 	<b>ADVANCED</b> 
<b>ENABLING CONTEXT 1:</b> <i>Setting clear policies</i>			
No standardized examination has taken place.	The standardized examination has been operating on an irregular basis.	The examination is a stable program that has been operating regularly. <sup>1</sup>	This option does not apply to this dimension.
There is no policy document that authorizes the examination.	There is an informal or draft policy document that authorizes the examination.	There is a formal policy document that authorizes the examination. <sup>2</sup>	This option does not apply to this dimension.
This option does not apply to this dimension.	The policy document is not available to the public.	The policy document is available to the public. <sup>3</sup>	This option does not apply to this dimension.
This option does not apply to this dimension.	This option does not apply to this dimension.	The policy document addresses some key aspects of the examination. <sup>4</sup>	The policy document addresses all key aspects of the examination.
<b>ENABLING CONTEXT 2:</b> <i>Having strong leadership</i>			
All stakeholder groups strongly oppose the examination or are indifferent to it.	Most stakeholder groups oppose the examination.	Most stakeholders groups support the examination. <sup>5</sup>	All stakeholder groups support the examination.
There are no attempts to improve the examination by stakeholder groups.	This option does not apply to this dimension.	There are independent attempts to improve the examination by stakeholder groups. <sup>6</sup>	There are coordinated attempts to improve the examination by stakeholder groups.
Efforts to improve the examination are not welcomed by the leadership in charge of the examination	This option does not apply to this dimension.	Efforts to improve the examination are generally welcomed by the leadership in charge of the examination. <sup>7</sup>	This option does not apply to this dimension.

(CONTINUED)





LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
<b>ENABLING CONTEXT 3:</b> <i>Having regular funding</i>			
There is no funding allocated for the examination.	There is irregular funding allocated for the examination.	There is regular funding allocated for the examination. <sup>8</sup> *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding covers some core examination activities: design, administration, data processing or reporting.	Funding covers all core examination activities: design, administration, data processing and reporting. <sup>9</sup> *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding does not cover research and development. <sup>10</sup> *	This option does not apply to this dimension.	Funding covers research and development.
<b>ENABLING CONTEXT 4:</b> <i>Having strong organizational structures</i>			
The examination office does not exist or is newly established.	The examination office is newly established.	The examination office is a stable organization. <sup>11</sup> *	This option does not apply to this dimension.
The examination office is not accountable to an external board or agency.	This option does not apply to this dimension.	The examination office is accountable to an external board or agency. <sup>12</sup> *	This option does not apply to this dimension.
Examination results are not recognized by any certification or selection system.	Examination results are recognized by certification or selection system in the country.	Examination results are recognized by one certification or selection system in another country. <sup>13</sup> *	Examination results are recognized by two or more certification or selection system in another country. <sup>13</sup> *
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. <sup>14</sup> *	The examination office has state of the art facilities to carry out the examination.

(CONTINUED)

<p><b>LATENT</b></p> <p>● ○ ○ ○</p>	<p><b>EMERGING</b></p> <p>● ● ○ ○</p>	<p><b>ESTABLISHED</b></p> <p>● ● ● ○</p>	<p><b>ADVANCED</b></p> <p>● ● ● ●</p>
<p><b>ENABLING CONTEXT 5:</b> <i>Having effective human resources</i></p>			
<p>There is no staff to carry out the examination.</p>	<p>The examination office is inadequately staffed to effectively carry out the examination, issues are pervasive.</p>	<p>The examination office is adequately staffed to carry out the examination effectively, with minimal issues.<sup>15</sup> *</p>	<p>The examination office is adequately staffed to carry out the assessment effectively, with no issues.</p>
<p>The country does not offer opportunities that prepare for work on the examination.</p>	<p>This option does not apply to this dimension.</p>	<p>The country offers some opportunities that prepare for work on the examination.<sup>16</sup> *</p>	<p>The country offers a wide range of opportunities that prepare for work on the examination.</p>





## SYSTEM ALIGNMENT

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





LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
<b>SYSTEM ALIGNMENT 1:</b>			
<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. <sup>17</sup> *	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. <sup>18</sup> *	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. <sup>19</sup> *	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.
<b>SYSTEM ALIGNMENT 2:</b>			
<i>Providing teachers with opportunities to learn about the examination</i>			
There are no courses or workshops on examinations available to teachers. <sup>20</sup> *	There are no up-to-date courses or workshops on examinations available to teachers.	There are up-to-date voluntary courses or workshops on examinations available to teachers.	There are up-to-date compulsory courses or workshops on examinations for teachers.
Teachers are excluded from all examination-related tasks.	Teachers are involved in very few examination-related tasks. <sup>21</sup> *	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 
<b>ASSESSMENT QUALITY 1:</b> <i>Ensuring quality</i>			
There is no technical report or other documentation.	There is some documentation on the examination, but it is not in a formal report format.	There is a comprehensive technical report but with restricted circulation. <sup>22</sup> *	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. <sup>23</sup> *	There are varied and systematic mechanisms in place to ensure the quality of the examination.
<b>ASSESSMENT QUALITY 2:</b> <i>Ensuring fairness</i>			
Inappropriate behavior surrounding the examination process is high.	Inappropriate behavior surrounding the examination process is moderate. <sup>24</sup> *	Inappropriate behavior surrounding the examination process is low.	Inappropriate behavior surrounding the examination process is marginal.
The examination results lack credibility for all stakeholder groups.	The examination results are credible for some stakeholder groups.	The examination results are credible for all stakeholder groups. <sup>25</sup> *	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. <sup>26</sup> *

(CONTINUED)

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
<b>ASSESSMENT QUALITY 3:</b> <i>Using examination information in a fair way</i>			
Examination results are not used in a proper way by all stakeholder groups.	Examination results are used by some stakeholder groups in a proper way.	Examination results are used by most stakeholder groups in a proper way.	Examination results are used by all stakeholder groups in a proper way. <sup>27</sup> *
Student names and results are public.	This option does not apply to this dimension.	Students' results are confidential. <sup>28</sup> *	This option does not apply to this dimension.
<b>ASSESSMENT QUALITY 4:</b> <i>Ensuring positive consequences of the examination</i>			
There are no options for students who do not perform well on the examination, or students must leave the education system.	There are very limited options for students who do not perform well on the examination. <sup>29</sup> *	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.	This option does not apply to this dimension.	There are some mechanisms in place to monitor the consequences of the examination. <sup>30</sup> *	There is a variety of mechanisms in place to monitor the consequences of the examination.



**Examinations:** *Development-Level Rating Justifications*

1. The Certificate of Completion of the General Secondary Education was first administered in 1954. The examination is administered to grade 12 students, and it covers the following subjects: Arabic language, religious education, civic education, foreign languages, mathematics, sciences, and social sciences.
2. The Presidency of the Republic authorized the Certificate of Completion of the General Secondary Education with the document Law 20 of 2012 (which amends certain provisions to Education Law 139 of 1981).
3. The Law 20 of 2012 document was published in the widely available Official Journal and is posted on the Ministry of Education's website.
4. The Law 20 of 2012 document addresses some key aspects of the examination. For example, it describes the purpose of the examination and authorized uses of results, specifies who can sit for the examination, and identifies rules about preparation. However, the document does not outline governance, distribution of power, or responsibilities among key entities, state funding sources, outline procedures to investigate and address security breaches, cheating, or other forms of inappropriate behavior, outline procedures for special or disadvantaged students, or explain alignment with curriculum and standards or the format of the examination questions.
5. Policy makers, teacher unions, and parents strongly support the examination, and educators and the media support it as well. Although think tanks and NGOs oppose the examination, students and employers are neutral to it. Some attitudes toward the examination may be informed by the prevalence of private tutoring. A recent OECD/World Bank report found that private tutoring can be expensive for the average family income and consume students' time (OECD/World Bank, forthcoming, 140–1).
6. Attempts have been made to improve the examination by stakeholder groups; however, stakeholder groups have not coordinated their efforts.
7. Leadership in charge of the examination generally welcomes efforts to improve the examination.
8. Regular funding is allocated for the examination by the government.
9. Funding from the Ministry of Education for the examination covers examination design and administration as well as data analysis and reporting. However, funding does not cover long- or medium-term planning of program milestones or staff training.
10. Funding for the examination does not cover research and development activities.
11. The General Directorate for Examinations and the General Administration of Examinations are departments within the Ministry of Education that have been responsible for the examination since 1955.





12. The General Directorate for Examinations has been accountable to the Office of the Minister of Education since 2012.
13. Results of the examination are recognized by certification and selection system in Egypt, as abroad, in countries such as Germany, Romania, and the United States.
14. The examination office has computers for all technical staff, a secure building, secure storage facilities, access to adequate computer servers, the ability to back up data, and adequate communication tools.
15. The number of permanent and full-time staff to carry out the examination effectively is adequate, with minimal issues. Although errors in data processing are frequently made, other issues such as frequent errors in the examination questions, omission of curricular topics, weaknesses in test design, errors in scoring that lead to delays in reporting of results, poor training of test administrators, and delays in administering the examination due to issues with the design of the examination questions have not been identified. However, a study by the OECD/World Bank found that because examinations are scored by readers who have little or no training in essay scoring, inconsistencies may be found in scores within a cohort (OECD/World Bank, forthcoming, 142).
16. Egypt offers university graduate programs and university courses on educational measurement and evaluation at all faculties of education within universities. Although nonuniversity training courses or workshops on educational measurement and evaluation are also available, no internships are open in the examination office. Additionally, funding is not available for attending international programs, courses, or workshops on educational measurement and evaluation.
17. There is a clear understanding that the examination measures national school curriculum guidelines and standards.
18. Stakeholder groups, including teachers, students, and parents, largely accept what is measured by the examination.
19. Although examples of the types of questions on the examination are the only material available to prepare for the examination, they are available to most students through textbooks distributed free of charge, foreign textbooks licensed by the Ministry of Education, and CDs. Other material, such as information on how to prepare for the examination, the framework document explaining what is measured on the examination, and the report on the strengths and weaknesses in student performance, is not available.
20. No courses or workshops on the examination are available to teachers; however, courses are available that provide instructions on the rules related to discipline, monitoring, and scoring of the examination.
21. Teachers are involved in very few examination tasks, which include administering and scoring the examination and supervising examination procedures. For example, teachers are commissioned by the General Directorate for Examination to participate in the discipline, monitoring, and scoring committees. Teachers are not involved in selecting or creating examination questions or scoring guides, acting as a judge, or resolving inconsistencies between examination scores and school grades.

22. Although its circulation is restricted, there is a comprehensive technical report available on the examination.
23. Although external and internal reviews are conducted by committees responsible for preparing the examination (by the National Center for Examinations, which is external to the Ministry of Education, as well as a committee internal to the Ministry of Education), and external certifications and audits take place, mechanisms such as pilots, field testing, or translation verifications are not in place to monitor the quality of the examination. A recent study from the OECD/World Bank reports that because of the lack of pilot testing of the items on the examination, the real difficulty level of the questions and their comparability across cohorts is difficult to understand (OECD/World Bank, forthcoming, 142).
24. Inappropriate behavior surrounding the examination process is moderate, because leakage of the content of an examination paper or part of a paper before the examination, copying from other candidates, use of unauthorized materials such as prepared answers and notes, and intimidation of examination supervisors, markers, or officials does occur. However, impersonation (when an individual other than the registered candidate takes the examination), collusion among candidates via mobile phone or passing of notes, issuance of forged certificates or alteration of information on results, and provision of external assistance via the supervisor or a mobile phone do not typically occur during the examination process. The document Ministerial Decision 319 from 2012 outlines mechanisms to address inappropriate behavior during the examination.
25. All stakeholder groups perceive examination results as credible.
26. All students, regardless of background, location, and ability to pay, can take the examination.
27. All stakeholder groups use examination results in a proper way.
28. Student results are confidential, because only the student and persons with a legitimate, professional interest in the test taker can know the results.
29. Students who do not perform well on the examination can retake the examination or repeat the grade. However, students do not have the option to attend remedial or preparatory courses to prepare to retake the examination or opt for less selective schools, universities, or tracks.
30. Only one mechanism, studies that are updated regularly (by the National Center for Examination and Educational Evaluation), is in place to monitor the consequences of the examination. Other mechanisms, such as funding for independent research on the examination, a permanent oversight committee, regular focus groups or surveys of key stakeholders, and expert review groups, are not in place.

**Egypt, Arab Rep.**  
***National (or System-Level) Large-Scale Assessment (NLSA)***

**ENABLING CONTEXT**





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

<b>LATENT</b> 	<b>EMERGING</b> 	<b>ESTABLISHED</b> 	<b>ADVANCED</b> 
<b>ENABLING CONTEXT 1:</b> <i>Setting clear policies for NLSA</i>			
No NLSA exercise has taken place.	The NLSA has been operating on an irregular basis. <sup>1</sup>	The NLSA is a stable program that has been operating regularly.	This option does not apply to this dimension.
There is no policy document pertaining to NLSA.	There is an informal or draft policy document that authorizes the NLSA. <sup>2</sup> *	There is a formal policy document that authorizes the NLSA.	This option does not apply to this dimension.
This option does not apply to this dimension.	The policy document is not available to the public. <sup>3</sup> *	The policy document is available to the public.	This option does not apply to this dimension.
There is no plan for NLSA activity. <sup>4</sup> *	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.
<b>ENABLING CONTEXT 2:</b> <i>Having strong public engagement for NLSA</i>			
All stakeholder groups strongly oppose the NLSA or are indifferent to it.	Some stakeholder groups oppose the NLSA.	Most stakeholders groups support the NLSA. <sup>5</sup> *	All stakeholder groups support the NLSA.

(CONTINUED)

LATENT ● ○ ○ ○	EMERGING ● ● ○ ○	ESTABLISHED ● ● ● ○	ADVANCED ● ● ● ●
<b>ENABLING CONTEXT 3:</b> <i>Having regular funding for NLSA</i>			
There is no funding allocated to the NLSA.	There is irregular funding allocated to the NLSA. <sup>6</sup> *	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. <sup>7</sup> *	Funding covers all core NLSA activities: design, administration, analysis and reporting. <sup>7</sup> *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding does not cover research and development activities. <sup>8</sup> *	This option does not apply to this dimension.	Funding covers research and development activities.
<b>ENABLING CONTEXT 4:</b> <i>Having strong organizational structures for NLSA</i>			
There is no NLSA office, ad hoc unit or team.	The NLSA office is a temporary agency or group of people.	The NLSA office is a permanent agency, institution or unit. <sup>9</sup> *	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. <sup>10</sup> *	Political considerations never hamper technical considerations.
This option does not apply to this dimension.	The NLSA office is not accountable to a clearly recognized body.	The NLSA office is accountable to a clearly recognized body. <sup>11</sup> *	This option does not apply to this dimension.





(CONTINUED)

<b>LATENT</b> 	<b>EMERGING</b> 	<b>ESTABLISHED</b> 	<b>ADVANCED</b> 
<b>ENABLING CONTEXT 5:</b> <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.	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. <sup>12</sup> *
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. <sup>13</sup> *	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.*

<b>LATENT</b> 	<b>EMERGING</b> 	<b>ESTABLISHED</b> 	<b>ADVANCED</b> 
<b>SYSTEM ALIGNMENT 1:</b> <i>Aligning the NLSA with learning goals</i>			
It is not clear if the NLSA is based on curriculum or learning standards.	This option does not apply to this dimension.	The NLSA measures performance against curriculum or learning standards. <sup>14</sup> *	This option does not apply to this dimension.
What the NLSA measures is generally questioned by stakeholder groups.	This option does not apply to this dimension.	What the NLSA measures is questioned by some stakeholder groups. <sup>15</sup> *	What the NLSA measures is largely accepted by stakeholder groups.
There are no mechanisms in place to ensure that the NLSA accurately measures what it is supposed to measure.	There are ad hoc reviews of the NLSA to ensure that it measures what it is intended to measure.	There are regular internal reviews of the NLSA to ensure that it measures what it is intended to measure. <sup>16</sup> *	This option does not apply to this dimension.
<b>SYSTEM ALIGNMENT 2:</b> <i>Providing teachers with opportunities to learn about the NLSA</i>			
There are no courses or workshops on the NLSA. <sup>17</sup> *	There are occasional courses or workshops on the NLSA.	There are some courses or workshops on the NLSA offered on a regular basis.	There are widely available high quality courses or workshops on the NLSA offered on a regular basis.



## ASSESSMENT QUALITY

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

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
<b>ASSESSMENT QUALITY 1:</b> <i>Ensuring the quality of the NLSA</i>			
No options are offered to include all groups of students in the NLSA.	This option does not apply to this dimension.	At least one option is offered to include all groups of students in the NLSA. <sup>18</sup> *	Different options are offered to include all groups of students in the NLSA.
There are no mechanisms in place to ensure the quality of the NLSA.	This option does not apply to this dimension.	There are some mechanisms in place to ensure the quality of the NLSA. <sup>19</sup> *	There are a variety of mechanisms in place to ensure the quality of the NLSA.
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. <sup>20</sup> *	There is a comprehensive, high quality technical report available to the general public.
<b>ASSESSMENT QUALITY 2:</b> <i>Ensuring effective uses of the NLSA</i>			
NLSA results are not disseminated.	NLSA results are poorly disseminated. <sup>21</sup> *	NLSA results are disseminated in an effective way.	This option does not apply to this dimension.
NLSA information is not used or is used in ways inconsistent with the purposes or the technical characteristics of the assessment.	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. <sup>22</sup> *	NLSA information is used by all stakeholder groups in a way that is consistent with the purposes and technical characteristics of the assessment.
There are no mechanisms in place to monitor the consequences of the NLSA. <sup>23</sup> *	This option does not apply to this dimension.	There are some mechanisms in place to monitor the consequences of the NLSA.	There are a variety of mechanisms in place to monitor the consequences of the NLSA.

*National (of System-Level) Large Scale Assessment (NLSA): Development-Level Rating Justifications*

1. The National Standardized Examination's main purposes are monitoring education quality at the system level, and holding government or political authorities, schools or educators, and students accountable. In addition, the assessment supports policy design, evaluation, or decision making. A representative random sample of students in grades 4, 8, and 10 are assessed in Arabic language, mathematics, science, and English language. The assessment was administered in all provinces in Egypt two times in the past five years.
2. An informal or draft policy document authorized the NLSA, which was carried out by the National Center for Examinations and Educational Evaluation in cooperation with the Ministry of Education.
3. The informal policy document authorizing the NLSA is not available to the public.
4. The government does not have a large-scale assessment plan for the coming years or future assessment rounds.
5. Although policy makers strongly support the NLSA, and educators and think tanks and NGOs support it as well, teacher unions, students, parents, and media are neutral to it. It is unclear whether universities and employers support or oppose the assessment.
6. Irregular funding from the government is allocated for the NLSA.
7. Funding for the NLSA covers all core NLSA activities, including assessment design and administration, data analysis and reporting, long- or medium-term planning of program milestones, and staff training.
8. Funding for the NLSA does not cover research and development activities.
9. The National Center for Examinations and Educational Evaluation is a permanent unit created for running the NLSA.
10. Political considerations sometimes hamper technical considerations with respect to issues in sampling, logistics, and administration. However, large-scale assessment results have never been withheld from publication because of political reasons.
11. The National Center for Examinations and Educational Evaluation is accountable to the Minister of Education, who is also the chairman of the Center's Board of Directors.
12. The National Center for Examinations and Educational Evaluations is adequately staffed with permanent and full-time staff to carry out the NLSA effectively, with no issues identified with the performance of the human resources that are responsible for the assessment.





13. Egypt offers some opportunities to prepare individuals for work on the NLSA, including university graduate programs and university courses on educational measurement and evaluation in faculties of education at Egyptian universities. Although nonuniversity courses or workshops on educational measurement and evaluation also are offered during in-service training, no internships or short-term employment opportunities are offered in the large-scale assessment office, and no funding is available for attending international program, courses, or workshops on educational measurement and evaluation.
14. The NLSA measures performance against national and internationally recognized curriculum or learning standards.
15. Some stakeholder groups, including students and teachers, question what the assessment measures.
16. Regular internal and regular independent reviews are performed by qualified experts on the alignment between the assessment instrument and what it is supposed to measure.
17. No courses or workshops for teachers are offered on the NLSA.
18. Special plans are made to ensure that the large-scale assessment is administered to students in hard-to-reach and remote areas and that the large-scale assessment is offered in the Arabic language, which is the language of instruction for almost all student groups. However, accommodations or alternative assessments are not provided for students with disabilities.
19. Some mechanisms, such as training for all proctors or administrators according to a protocol, a standardized manual for large-scale assessment administrators, a requirement for all booklets to be numbered, and training for scorers to ensure high interrater reliability, are in place to ensure the quality of the NLSA. However, discrepancies are not required to be recorded on a standard sheet, no double scoring or processing of data is done, no external or internal reviewers or observers are in place, and no external certification or audits are performed.
20. Although a comprehensive technical report is available, its circulation is restricted.
21. Although the main reports on the NLSA results contain information on overall achievement levels and by subgroups, and although workshops or presentations are held for key stakeholders on the results, results are not disseminated within 12 months after the large-scale assessment is administered and reports with results are not made available for all stakeholder groups. In addition, the main reports on the results do not contain information on trends over time overall and for subgroups and do not provide standard errors. No media briefing is organized to discuss results, and results are not featured in newspapers, magazines, radio, or television.
22. Some stakeholder groups use NLSA results in a way that is consistent with the stated purposes or technical characteristics of the assessment.

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

# **Egypt, Arab Rep.** *International Large-Scale Assessment (ILSA)*

**ENABLING CONTEXT**

*Overall framework of policies, leadership, organizational structures, and 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.*

<b>LATENT</b> 	<b>EMERGING</b> 	<b>ESTABLISHED</b> 	<b>ADVANCED</b> 
<b>ENABLING CONTEXT 1:</b> <i>Setting clear policies for ILSA</i>			
The country/system has not participated in an ILSA in the last 10 years.	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. <sup>1</sup>
The country/system has not taken concrete steps to participate in an ILSA in the next 5 years. <sup>2</sup> *	This option does not apply to this dimension.	The country/system has taken concrete steps to participate in at least one ILSA in the next 5 years.	This option does not apply to this dimension.
There is no policy document that addresses participation in ILSA. <sup>3</sup> *	There is an informal or draft policy document that addresses participation in ILSA.	There is a formal policy document that addresses participation in ILSA.	This option does not apply to this dimension.
This option does not apply to this dimension. <sup>4</sup> *	The policy document is not available to the public.	The policy document is available to the public.	This option does not apply to this dimension.
<b>ENABLING CONTEXT 2:</b> <i>Having regular funding for ILSA</i>			
There is no funding for participation in ILSA.	There is funding from loans or external donors.	There is regular funding allocated at discretion. <sup>5</sup> *	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. <sup>6</sup> *	This option does not apply to this dimension.
Funding does not cover research and development activities. <sup>7</sup> *	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 ● ● ● ●
<b>ENABLING CONTEXT 3:</b> <i>Having effective human resources for ILSA</i>			
There is no team or national/system coordinator to carry out the ILSA activities.	There is a team or national/system coordinator to carry out the ILSA activities.	There is a team and national/system coordinator to carry out the ILSA activities. <sup>8</sup> *	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. <sup>9</sup> *	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. <sup>10</sup> *	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.*

<p><b>LATENT</b> ● ○ ○ ○</p>	<p><b>EMERGING</b> ● ● ○ ○</p>	<p><b>ESTABLISHED</b> ● ● ● ○</p>	<p><b>ADVANCED</b> ● ● ● ●</p>
<p><b>SYSTEM ALIGNMENT 1:</b> <i>Providing opportunities to learn about ILSA</i></p>			
<p>The ILSA team has not attended international workshops or meetings.</p>	<p>The ILSA team attended some international workshops or meetings.<sup>11</sup> *</p>	<p>The ILSA team attended all international workshops or meetings.</p>	<p>This option does not apply to this dimension.</p>
<p>The country/system offers no opportunities to learn about ILSA.</p>	<p>This option does not apply to this dimension.</p>	<p>The country/system offers some opportunities to learn about ILSA.<sup>12</sup> *</p>	<p>The country/system offers a wide range of opportunities to learn about ILSA.</p>
<p>This option does not apply to this dimension.</p>	<p>This option does not apply to this dimension.</p>	<p>Opportunities to learn about ILSA are available to the country's/system's ILSA team members only.<sup>13</sup> *</p>	<p>Opportunities to learn about ILSA are available to a wide audience, in addition to the country's/system's ILSA team members.</p>

## ASSESSMENT QUALITY

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

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
<b>ASSESSMENT QUALITY 1:</b> <i>Ensuring the quality of ILSA</i>			
Data from the ILSA has not been published.	The country/system met sufficient standards to have its data presented beneath the main display of the international report or in an annex.	The country/system met all technical standards required to have its data presented in the main displays of the international report. <sup>14</sup> *	This option does not apply to this dimension.
The country/system has not contributed new knowledge on ILSA. <sup>15</sup> *	This option does not apply to this dimension.	This option does not apply to this dimension.	The country/system has contributed new knowledge on ILSA.
<b>ASSESSMENT QUALITY 2:</b> <i>Ensuring effective uses of ILSA</i>			
If any, country/system-specific results and information are not disseminated in the country/system.	Country/system-specific results and information are disseminated irregularly in the country/system. <sup>16</sup> *	Country/system-specific results and information are regularly disseminated in the country/system.	Country/system-specific results and information are regularly and widely disseminated in the country/system.
Products to provide feedback to schools and educators about the ILSA results are not made available.	This option does not apply to this dimension.	Products to provide feedback to schools and educators about the ILSA results are sometimes made available. <sup>17</sup> *	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. <sup>18</sup> *	There is some media coverage of the ILSA results.	There is wide media coverage of the ILSA results.
If any, country/system-specific results and information from the ILSA are not used to inform decision making in the country/system.	Results from the ILSA are used in a limited way to inform decision making in the country/system.	Results from the ILSA are used in some ways to inform decision making in the country/system.	Results from the ILSA are used in a variety of ways to inform decision making in the country/system. <sup>19</sup> *
It is not clear that decisions based on ILSA results have had a positive impact on students' achievement levels. <sup>20</sup> *	This option does not apply to this dimension.	This option does not apply to this dimension.	Decisions based on the ILSA results have had a positive impact on students' achievement levels.

***International Large Scale Assessment (ILSA):*** Development-Level Rating Justifications

1. In the last 10 years, Egypt participated in two ILSAs, in TIMSS 2003 and TIMSS 2007.
2. Egypt has not taken concrete steps to participate in an ILSA in the next five years.
3. No policy document addresses Egypt's participation in ILSAs.
4. This option does not apply to this dimension.
5. Funding for the administration of TIMSS in 2007 was provided by Ministry of Education.
6. Funding covers all core activities of the ILSA, including international participation fees, implementation of the assessment exercise in Egypt, processing and analyzing data collected from implementation of the assessment exercise, reporting and disseminating assessment results in Egypt, and attendance at international expert meetings for the assessment exercise.
7. Research and development is not among the activities covered under ILSA funding.
8. A team and national coordinator are responsible for carrying out the ILSA activities.
9. The national coordinator is fluent in the language in which the international-level meetings are conducted and related documentation is available.
10. The ILSA office is adequately staffed and trained to carry out the international assessment effectively, and the ILSA team has previous experience working on international assessments. Minimal issues have been identified, including errors or delays in scoring student responses to questions and issues with translation of assessment instruments. However, no errors or delays have been seen in, for example, the printing or layout of the test booklets or in the administration of the assessment, and complaints about poor training of test administrators have been received.
11. ILSA team members have attended some meetings related to the assessment.
12. Egypt offers very limited opportunities to learn about ILSAs. Although workshops or meetings are available on using international assessment databases, no university courses are offered on the topic of international assessments, funding for attending international workshops or training on international assessments, or online courses on international assessment.

13. Only individuals working directly on the international assessment exercise benefit from opportunities to learn about ILSAs. University students studying assessment or a related area and professionals or university staff interested in assessment also do not benefit from such opportunities.
14. Egypt met all technical standards required to have its data presented in the main displays of the international report.
15. Egypt has not contributed to the global knowledge base on international assessments by generating new knowledge and making it available through publications or presentations.
16. Although copies of the international report were distributed to key stakeholders, copies of the national report were not distributed to key stakeholders, and a national report was not made available online. In addition, Egypt's results were not communicated through a press release or covered in television, radio, or newspapers. Brochures and PowerPoint presentations with Egypt's results were also not made available online or distributed to key stakeholders.
17. Products providing feedback to schools and educators about ILSA results are sometimes made available through, for example, a summary of results that is prepared and disseminated to specific directorates.
18. Media coverage of ILSA results is limited to a few small articles.
19. Results from ILSAs are used in a variety of ways to inform decision making in Egypt. Results are used in tracking the impact of reforms on student achievement levels and in informing curriculum improvement, teacher training programs, and other assessment activities in the country. The Center for Developing Curricula and Educational Materials of the Ministry of Education has studied the results of TIMSS 2007 and subsequently developed the science and mathematics curricula in light of the standards on which said study was built. The Central Training Agency followed by the Technical Academy for Teachers offered training courses to teachers based on TIMSS standards. The Technical Academy for Teachers trained teachers on the curricula developed based on the TIMSS results and questions in sciences and mathematics. However, TIMSS results have not been used to inform resource allocation.
20. It is not clear that decisions based on ILSA results have had a positive impact on students' achievement levels.

## Acknowledgements

This report, part of a 16-country benchmarking exercise in the Middle East & North Africa and Africa regions, was prepared by the World Bank SABER-Student Assessment team, in partnership with the Arab League Educational, Cultural, and Scientific Organization (ALECSO) and the United Nations Educational, Scientific, and Cultural Organization (UNESCO), which led data collection efforts. This effort is part of the Arab Regional Agenda for Improving Education Quality (ARAIEQ), led by ALECSO in partnership with the World Bank. It benefited from feedback and review by Ernesto Cuadra, Lead Education Specialist, and Dina Abu-Ghaida, Task Team Leader for education projects in Egypt at the World Bank, as well as comments received during a national validation workshop held in Egypt.

## References

- Clarke, M. 2012. "What Matters Most for Student Assessment Systems: A Framework Paper." READ/SABER Working Paper Series. World Bank, Washington, DC.
- Ministry of Education. 2006. "National Strategic Plan—2008–2012." Government of Egypt, Cairo.
- OECD and World Bank. Forthcoming. "Review of National Policies for Education: Schools for Skills—A New Learning Agenda for Egypt." OECD and World Bank, Washington, DC.
- World Bank. 2012. "Arab Republic of Egypt: Inequality of Opportunity in Educational Achievement." Report No. 70300-EG. World Bank, Washington, DC.
- . "Egypt, Arab Rep. Country Indicator Data." World Bank, Washington, DC. Data retrieved from <http://data.worldbank.org/> on October 15, 2013.
- . 2013. "Egypt Overview." World Bank, Washington, DC. Data retrieved from <http://www.worldbank.org/en/country/egypt/overview> 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.

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.

