



Key Policy Areas for Student Assessment

Status

1. Classroom Assessment

In the United Arab Emirates, a formal, publicly available system-level document provides guidelines for classroom assessment. In addition, there are some system-wide resources and materials (such as scoring criteria for evaluating students' work) available to teachers for carrying out classroom assessment activities. While there are varied and systematic mechanisms in place to monitor the quality of classroom assessment practices, these practices are considered to be of weak quality due to issues such as grade inflation and the uneven application of standards for grading students' work.



2. Examinations

The Twelfth Grade Examination has been administered to grade 12 students since 1967. Students are assessed in mandatory subjects (Islamic education, Arabic language, English language, mathematics, physics, and biology) along with other subjects depending on their section. Examination results are used to certify student completion of the school cycle and to determine selection to higher-education institutions. Examination results are officially recognized by certification and selection systems in the UAE and abroad. In addition, regular funding for the examination is provided by the government and covers all core examination activities. Although expert review groups are in place to monitor the consequences of the examination, there are very limited options for students who do not perform well on the examination.



3. National Large-Scale Assessment (NLSA)

The United Arab Emirates National Assessment Program (UAENAP) has been operating since 2003 and on a regular basis since 2010, assessing students in grades 3, 5, 7, and 9. A formal, publicly available policy document authorizes the NLSA, and there is a written plan for future NLSA activities. In addition, there is regular government funding covering all core NLSA activities and research and development. Opportunities for teachers to learn about the NLSA are occasionally made available.



4. International Large-Scale Assessment (ILSA)

In the last five years, the UAE has participated in PIRLS (2011), TIMSS (2011), and PISA (2009). At the time of data collection, the UAE had taken concrete steps to participate in PIRLS (2016), TIMSS (2015), and PISA (2012, 2015). Funding for ILSAs is provided through regular government budget and covers all core ILSA activities. Some opportunities to learn about ILSAs are made available in the UAE to a wide audience. While UAE-specific ILSA results are regularly and widely disseminated in the country, it is not known at the time of data collection whether decisions based on ILSA results have had a positive impact on achievement levels.



Introduction

The United Arab Emirates 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 to improving education quality and learning outcomes as it provides the necessary information to meet stakeholders' decision-making needs. In order to gain a better understanding of the strengths and weaknesses of its existing assessment system, United Arab Emirates have decided to benchmark this system using standardized tools developed under The World Bank's Systems Approach for Better Education Results (SABER) program. SABER is an evidence-based program to help countries systematically examine and strengthen the performance of different aspects of their education systems.

What is SABER-Student Assessment?

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

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

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

SABER-Student Assessment methodology

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

Assessment types and purposes

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

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

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

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

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

Quality drivers of an assessment system

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

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

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

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

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

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

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

Source: World Bank.

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

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

Levels of development

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

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

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

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

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

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

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

Education in the United Arab Emirates

The United Arab Emirates (UAE) is a high-income country in the Middle East region. GDP per capita

(current, US\$) is \$39,058, with annual growth of approximately 4.4 percent.

The UAE spends 25 percent of total government expenditure on the education sector, and public education is free through the university level for children who are citizens. The education system is comprised of kindergarten (for children between the ages of 4-5 years); elementary (for students 6-11 years old); intermediate (for students 12-14 years old); and secondary (15-17 years old). The net primary enrollment rate is approximately 90 percent, with a 100 percent completion rate. In addition, the net enrollment rate at the secondary level is 81 percent. The UAE has one of the lowest student-to-teacher ratios (15:1) in the world and has made significant progress to ensure high literacy (which was 91 percent for the overall population in 2005), to increase the use of modern technology in education programs, and to promote and encourage women's education.

Government priorities for education reform are outlined in the Ministry of Education Strategy 2010-2020. In this strategy, the Ministry identified 20 comprehensive initiatives to improve the current education system. Some of these initiatives include: the development of curricula aligned with higher education and job market requirements; supporting the professional development of all teachers and education staff; aligning compulsory school age with international standards; improving national assessments and participating in international examinations; and supporting the technical development of school infrastructure, especially with respect to information technology.

Detailed information was collected on the UAE's student assessment system using the SABER-Student Assessment questionnaires and rubrics in 2011. It is important to remember that these tools primarily focus on benchmarking a country's policies and arrangements for assessment activities at the system or macro level. Additional data would need to be collected to determine actual, on-the-ground practices in the UAE, particularly by teachers and students in schools. While the UAE is a federal system comprised of seven emirates, this report analyzes policies only at the national level, and additional data would need to be

collected to evaluate policies within each emirate (the Ministry of Education in the UAE oversees public schooling in the country, and each emirate oversees its private schools). The following sections discuss the findings by each assessment type, accompanied by suggested policy options. The suggested policy options were determined in collaboration with key local stakeholders based on the UAE's immediate interests and needs. Detailed, completed rubrics for each assessment type in the UAE are provided in Appendix 5.

Classroom Assessment

Level of development: EMERGING

In the United Arab Emirates, *The Guidelines to the Implementing of On-going Assessment Tools* document, authorized by the Ministry of Education's Directorate of Evaluation and Examinations in 2010, provides guidelines for classroom assessment.

There are some system-wide resources available for teachers to engage in classroom assessment, including a document that outlines what students are expected to learn in different subject areas at different grade levels and textbooks that provide support for classroom assessment. Scoring criteria or rubrics for grading students' work are also available in an electronic system that is accessible to all schools.

There are some mechanisms available systematically to ensure that teachers develop skills and expertise in classroom assessment. For example, in-service teacher training is available to all teachers, and online resources on classroom assessment include question banks. Some teachers also have the opportunity to participate in conferences and workshops, and in item development for, or scoring of, large-scale assessments or exams. However, not all teacher-training programs include a required course on classroom assessment, and there are no pre-service teacher-training opportunities related to building skills in classroom assessment.

In general, classroom assessment practices are perceived to be weak. Teachers commonly rely on multiple-choice or selection-type questions, and classroom assessment activities tend to be about recalling information. In addition, grade inflation and the uneven application of standards for grading students' work are serious problems.

However, there are varied and systematic mechanisms in place to monitor the quality of classroom assessment practices. For example, government funding is available for research on the quality of classroom assessment activities and how to improve classroom assessment, and national reviews of the quality of education include a focus on classroom assessment. Although the weight of classroom assessment in a teacher's performance

evaluation is low, it is indeed a required component of a teacher's performance evaluation and school inspection.

In addition, there are a number of required uses of classroom assessment to support student learning, including its use as an input for external examination results. Classroom assessment activities are used to diagnose student learning issues, develop students' self-evaluation skills, provide feedback to students on their learning, and inform parents about their child's learning.

Suggested policy options:

1. Introduce various mechanisms to ensure that teachers develop skills and expertise in classroom assessment. For example, make available pre-service teacher-training opportunities that focus on classroom assessment methodologies, and introduce a required course on classroom assessment for all teachers during in-service teacher training. Ensure that teacher-training opportunities address the importance of consistently applying standards for grading across student groups.
2. Introduce a variety of system-wide resources to help teachers improve their classroom assessment practices. For example, develop a document that clearly outlines the levels of performance that students are expected to reach in different subject areas at different grade or age levels, which should be made available to all teachers. Additionally, conduct monitoring of classroom assessment activities to ensure that classroom assessment practices are in line with their intended purposes and uses.

Examinations

Level of development: ESTABLISHED

The Twelfth Grade Examination's main purposes are for student certification of school cycle completion and student selection to higher-education institutions. It was authorized by the Ministry of Education in 2010 with a formal, system-level document, the *Assessment and Examination System for Grades 1-12*. The Twelfth Grade Examination was first administered in 1967, and it continues to be administered to grade 12 students in mandatory subjects, as well as subjects relevant to the student's section. Specifically, all students are assessed in Islamic education, Arabic language, English language, mathematics, physics, and biology. Although students have the option to retake the examination or repeat the grade, they do not have the options to attend remedial or preparatory courses in order to prepare to retake the examination or to opt for less-selective schools, universities, or tracks.

Regular funding is allocated by the government for the examination, which covers all core examination activities, including examination design and administration, data analysis and reporting, and long- or medium-term planning of program milestones. However, funding does not cover staff training, or research and development activities.

The Assessment and Examinations Administration, a unit within the Ministry of Education, has had primary responsibility for running the Twelfth Grade Examination since 1972. While it has state-of-the-art facilities to carry out the examination, the allocated staff is insufficient to meet the needs of the examination. However, this lack of permanent staff is offset by utilizing assistance from technical teams in the Educational Supervision Department within the Ministry of Education and by forming temporary committees to perform the tasks required during examination periods.

The UAE also offers a wide range of opportunities that prepare for work on the examination, including university graduate programs, university courses, and non-university training courses or workshops on educational measurement and evaluation. In addition, internships are offered in the examination office, and

funding is available for attending international programs, courses, or workshops on educational measurement and evaluation.

There is only one systematic mechanism, internal review or observers, in place to ensure the quality of the examination. Other mechanisms, such as pilot or field testing, are not in place.

At the same time, inappropriate behavior surrounding the examination process is low. For example, exam takers' use of unauthorized materials and the provision of external assistance via the supervisor or mobile phone do not typically occur. When they do occur, inappropriate behaviors are dealt with by the legal affairs department on a case by case basis.

Only one systematic mechanism, expert review groups, is in place to monitor the consequences of the examination.

Suggested policy options:

1. Introduce varied systematic mechanisms to ensure the quality of the examination. Mechanisms to consider include carrying out pilot or field testing, translation verification, and external review and certification.
2. Introduce more options for students who do not perform well on the examination, such as allowing students to choose to attend less-selective universities, or continue their studies in a vocational track.
3. Ensure that varied and systematic mechanisms are in place to monitor the consequences of the examination. Such mechanisms include holding regular focus groups or surveys of key stakeholders, commissioning studies that are updated regularly, instituting a permanent oversight committee, and providing regular funding for independent research on the impact of the examination.

National Large-Scale Assessment (NLSA)

Level of development: ESTABLISHED

The United Arab Emirates National Assessment Program (UAENAP) was first administered in 2003 to students in grade 5. Since 2010, students in grades 3, 5, 7, and 9 have been assessed in the following subjects: Arabic, English, mathematics, and science. The main purposes of the UAENAP include monitoring education quality at the system level, evaluation, decision making, and supporting schools, teachers, and policy design. The Ministry of Education authorized the UAENAP with the *Implementing the National Assessment Program* document in 2010.

There is regular funding allocated by the government for the UAENAP. Funding covers all core activities of the assessment, including assessment design and administration, data analysis and reporting, long- or medium-term planning of program milestones, and staff training. Funding for the UAENAP also covers research and development activities.

The NLSA office is a permanent unit created for running the assessment within the Administration of Evaluation and Examinations under the supervision of the Ministry of Education. It is adequately staffed with permanent and full-time staff to carry out the assessment effectively.

The UAE also offers a wide range of opportunities to prepare individuals for work on the UAENAP, including university graduate programs, university courses, and non-university courses or workshops on educational measurement and evaluation. In addition, funding for attending international programs, as well as internships or short-term employment in the large-scale assessment office, are available.

Some mechanisms are in place to ensure the quality of the NLSA. For example, all proctors or administrators are trained according to a protocol, all booklets are numbered, and there is double scoring of data.

In addition, NLSA results are disseminated effectively. Reports with the results are made available to all stakeholder groups and contain information on overall

achievement levels and subgroups, as well as trends over time overall and by subgroups. There is also a media briefing organized to discuss results, and results are featured in newspapers, magazines, radio, or television.

However, only two mechanisms are in place to monitor the consequences of the NLSA: expert review groups and themed conferences that provide a forum to discuss research and other data on the consequences of the large-scale assessment.

Suggested policy options:

1. Introduce additional mechanisms to ensure the quality of the NLSA. For example, conduct a pilot before the main data collection takes place, and require the double processing of data.
2. Introduce a variety of system-level mechanisms to monitor the quality of the NLSA. For example, introduce regular focus groups or surveys of key stakeholders, a permanent oversight committee, and funding for independent research on the impact of the large-scale assessment.

International Large-Scale Assessment (ILSA)

Level of development: ESTABLISHED

The United Arab Emirates has participated in three ILSAs in the past five years, the Progress in International Reading Literacy Study (PIRLS) and Trends in International Mathematics and Science Study (TIMSS) in 2011¹, as well as the Programme for International Student Assessment (PISA) in 2009. In addition, at the time of data collection, the UAE has taken concrete steps to participate in several ILSAs in the next 5 years, including PISA 2012 and 2015, TIMSS 2015, and PIRLS 2016. A formal policy document, *Ministerial Council for Services Decision Number (73/65/2)* authorized by the UAE Cabinet in 2010, addresses participation in ILSAs.

Regular funding for international assessment participation is approved by law. Funding covers all core activities of ILSAs, including international participation fees, implementation of the assessment exercise, processing and analyzing data from the implementation of the assessment exercise, and reporting and disseminating assessment results. Research and development activities are also covered by funding for ILSAs.

The ILSA office is adequately staffed and trained to carry out ILSAs effectively, as team members have attended all international meetings related to the assessment and have previous experience working on international assessments. However, there have been some issues with the translation of the assessment instruments. For example, there are some difficulties when translating the original PISA questions from English to Arabic, given that there are strict guidelines for the number of words used in questions.

The UAE also offers some opportunities to learn about ILSAs, including workshops on using international assessment databases, funding for attending international workshops or training on international assessments, and online courses on international assessments. However, online courses are only

periodically available, and there are no stand-alone university courses or workshops on the topic of international assessments.

Country-specific ILSA results and information are regularly and widely disseminated in the UAE, and products providing feedback to schools and educators about ILSA results are systematically made available.

Results from ILSAs are used in a variety of ways to inform decision making in the UAE, including tracking the impact of reforms on student achievement levels, and informing curriculum improvement, teacher training programs, and other assessment activities in the system. However, at the time of data collection, it was not known whether decisions based on ILSA results have had a positive impact on students' achievement levels in the UAE.

Suggested policy options:

1. Make available workshops or presentations about international assessments in the UAE to government officials, and professionals and university staff interested in assessment.
2. Introduce a variety of opportunities to learn about the ILSA that are regularly available, including online courses that are updated, as well as stand-alone university courses or workshops on the topic of international assessments.

¹ In addition, two emirates, Dubai and Abu Dhabi, were benchmarking participants for TIMSS 2011. Dubai was also a benchmarking participant for TIMSS 2007.

Appendix 1: Assessment Types and Their Key Differences

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

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

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

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

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

Appendix 4: Methodology for Assigning Development Levels

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

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

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

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

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

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

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

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

6. Scores for certain key dimensions under 'Enabling Context' (in the case of EXAM, NLSA, and ILSA) and under 'System Alignment' (in the case of CLASS) were

set as ceiling scores, i.e., the overall mean score for the particular assessment type cannot be greater than the score for these key dimensions. These key variables include formal policy, regular funding, having a permanent assessment unit, and the quality of assessment practices.

Appendix 5: SABER-Student Assessment Rubrics for the United Arab Emirates

This appendix provides the completed SABER-Student Assessment rubrics for each type of assessment activity in the United Arab Emirates. 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 of the selection (as indicated by a thick border and an asterisk), which is provided 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, such superscript indicates that insufficient information was available to determine the relevant selection in the row.

UNITED ARAB EMIRATES

Classroom Assessment

ENABLING CONTEXT AND SYSTEM ALIGNMENT

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

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

ASSESSMENT QUALITY

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

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

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

1. *The Guidelines to the Implementing of On-going Assessment Tools* document, authorized by the Ministry of Education's Directorate of Evaluation and Examinations in 2010, provides guidelines for classroom assessment.
2. *The Guidelines to the Implementing of On-going Assessment Tools* document is available online and in in-service courses for teachers.
3. Some system-wide resources are available to teachers to engage in classroom assessment, including a document that outlines what students are expected to learn in different subject areas at different grade or age levels and textbooks that provide support for classroom assessment. Scoring criteria or rubrics for students' work are also available in an electronic system that is accessible to all schools. However, teachers do have as a resource a document that outlines the levels of performance students are expected to reach in different subject areas at different grade or age levels, item banks or pools with examples of selection and multiple-choice or supply and open-ended questions, online assessment resources, or computer-based testing with instant reports on students' performance.
4. There is an official curriculum or standards document that specifies what students are expected to learn at different grade or age levels, but it does not specify to what performance level.
5. There are some system-level mechanisms in place to ensure that teachers develop the skills and expertise in classroom assessment. For example, in-service teacher training is available to all teachers, and online resources on classroom assessment include resources such as question banks. Some teachers have the opportunity to participate in conferences and workshops, and in item development for, or scoring of, large-scale assessments or exams. However, not all teacher training programs include a required course on classroom assessment; school inspection or teacher supervision does not include a component focused on classroom assessment; and there are no pre-service teacher training opportunities related to building skills in classroom assessment.
6. Classroom assessment practices are generally considered to be weak. Teachers commonly rely on multiple-choice, selection-type questions, and classroom assessment activities are mainly about recalling information. Classroom assessment information also provides little useful feedback to students. In addition, grade inflation and the uneven application of standards for grading students' work are serious problems. However, it is not common for parents to be poorly informed about students' grades, for teachers to not use explicit or a priori criteria for scoring or grading students' work, to observe errors in the scoring or grading of students' work, or for classroom assessment activities to not be aligned with the pedagogical or curricular framework.
7. There are varied and systematic mechanisms in place to monitor the quality of classroom assessment practices. For example, government funding is available for research on the quality of classroom assessment activities and how to improve classroom assessment, and national reviews of the quality of education include a focus on classroom assessment. Although the weight of classroom assessment in a teacher's performance evaluation is low, it is a required component of a teacher's performance evaluation and school inspection or teacher supervision. However, there is no external moderation system that reviews the difficulty of classroom assessment practices and the appropriateness of scoring criteria.

8. Classroom assessment information is required to be disseminated to all key stakeholders, including school district or Ministry of Education officials, parents, and students.
9. There are adequate required uses of classroom assessment to support student learning, including its use as an input for external examination results. Classroom assessment activities are used to diagnose student learning issues, develop students' self-evaluation skills, provide feedback to students on their learning, inform parents about their child's learning, evaluate teachers' performance, plan next steps in instruction, grade students for internal classroom uses, and provide input to an external examination program.

UNITED ARAB EMIRATES

Examinations

ENABLING CONTEXT

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

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

(CONTINUED)

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

(CONTINUED)

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

SYSTEM ALIGNMENT

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
SYSTEM ALIGNMENT 1: <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. ¹⁷ *	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. ¹⁸ *	This option does not apply to this dimension.
Material to prepare for the examination is minimal and it is only accessible to very few students.	There is some material to prepare for the examination that is accessible to some students.	There is comprehensive material to prepare for the examination that is accessible to most students.	There is comprehensive material to prepare for the examination that is accessible to all students. ¹⁹ *
SYSTEM ALIGNMENT 2: <i>Providing teachers with opportunities to learn about the examination</i>			
There are no courses or workshops on examinations available to teachers.	There are no up-to-date courses or workshops on examinations available to teachers.	There are up-to-date voluntary courses or workshops on examinations available to teachers.	There are up-to-date compulsory courses or workshops on examinations for teachers. ²⁰ *
Teachers are excluded from all examination-related tasks.	Teachers are involved in very few examination-related tasks. ²¹ *	Teachers are involved in some examination-related tasks.	Teachers are involved in most examination-related tasks.

ASSESSMENT QUALITY

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ASSESSMENT QUALITY 1: <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. ²² *	There is a comprehensive, high quality technical report available to the general public.
There are no mechanisms in place to ensure the quality of the examination.	This option does not apply to this dimension.	There are limited systematic mechanisms in place to ensure the quality of the examination. ²³ *	There are varied and systematic mechanisms in place to ensure the quality of the examination.
ASSESSMENT QUALITY 2: <i>Ensuring fairness</i>			
Inappropriate behavior surrounding the examination process is high.	Inappropriate behavior surrounding the examination process is moderate.	Inappropriate behavior surrounding the examination process is low. ²⁴ *	Inappropriate behavior surrounding the examination process is marginal.
The examination results lack credibility for all stakeholder groups.	The examination results are credible for some stakeholder groups.	The examination results are credible for all stakeholder groups. ²⁵ *	This option does not apply to this dimension.
The majority of the students (over 50%) 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. ²⁶ *

(CONTINUED)

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

Examinations: Development-level rating justifications

1. The Twelfth Grade Examination's main purposes include student certification for school cycle completion and student selection to higher-education institutions. The Twelfth Grade Examination was first administered in 1967, and it continues to be administered to Grade 12 students in mandatory subjects, as well as subjects relevant to the student's section selection. Specifically, all students are assessed in Islamic education, Arabic language, English language, mathematics, physics, and biology. For students in the scientific section, the examination also covers chemistry and geology, while for students in the literary section, the examination covers geography, history, psychology, and economics.
2. The Ministry of Education's *Assessment and Examination System for Grades 1-12* document from 2010 authorizes the Twelfth Grade Examination.
3. The *Assessment and Examination System for Grades 1-12* document is available to and easily accessible by the public.
4. The *Assessment and Examination System for Grades 1-12* document addresses very few key aspects of the examination. The document outlines governance, distribution of power, and responsibilities among key entities; describes the purpose of the examination; and explains alignment with curricula and standards. However, it does not describe the authorized uses of results; 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; specify who can sit for the examination; identify rules about preparation; or explain the format of the examination questions.
5. Policymakers and universities strongly support the examination, and teacher unions, educators, media, think tanks and NGOs support it as well. Students, parents, and employers are neutral to the examination.
6. Coordinated efforts have been made by stakeholder groups to improve the examination.
7. Leadership in charge of the examination generally welcomes efforts to improve the examination.
8. Regular funding is allocated by the government for the examination.
9. Funding covers all core examination activities, including examination design and administration, data analysis and reporting, and long- or medium-term planning of program milestones. However, staff training is not covered by the funding.
10. Research and development activities are not covered by the funding for the examination. Training and research budgets are allocated by specialized executive entities and are unrelated to the funding of examinations.
11. The Assessment and Examinations Administration, a unit within the Ministry of Education, has been in charge of the Twelfth Grade Examination since 1972.
12. The Assessment and Examinations Administration is not accountable to an external board or agency.

13. Examination results are recognized by certification or selection systems in the UAE and abroad, including in the USA, UK, Germany, and other Arab countries.
14. The Assessment and Examinations Administration has state-of-the-art facilities to carry out the examination, which include computers for all technical staff, a secure building, secure storage facilities, access to adequate computer servers, the ability to backup data, and adequate communication tools.
15. The Assessment and Examinations Administration has permanent and full-time staff, but it is insufficient to meet the needs of the examination. However, this lack of permanent staff is offset by utilizing assistance from technical teams in the Educational Supervision Department within the Ministry of Education and by forming temporary committees to perform the necessary tasks required during examination periods. Issues, such as frequent errors in data processing or in the examination questions, have not been identified with the performance of the human resources responsible for the examination.
16. The UAE offers a wide range of opportunities to prepare for work on the examination, including university graduate programs, university courses, and non-university training courses or workshops on educational measurement and evaluation. In addition, internships are offered in the examination office, and funding is 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 or standards.
18. Stakeholder groups largely accept what is measured by the examination.
19. Comprehensive material to prepare for the examination is available and accessible to all students. For example, information on how to prepare for the examination, examples of the types of questions that are on the examination, and the report on the strengths and weaknesses in student performance are available. However, the framework document explaining what is measured on the examination is not available.
20. Up-to-date compulsory courses or workshops on the examination are available to teachers.
21. Although teachers are involved in scoring the examination and supervising examination procedures, they are not involved in selecting or creating examination questions or scoring guides, administering the examination, acting as a judge, or resolving inconsistencies between examination scores and school grades.
22. Although a comprehensive technical report is available, its circulation is restricted.
23. Only one systematic mechanism, internal review or observers, is in place to ensure the quality of the examination. Other systematic mechanisms, such as pilot or field testing, external certification or audits, and external review or observers, are not in place.
24. Inappropriate behavior surrounding the examination process is low. Although copying from other candidates and collusion among candidates via mobile phones or passing of paper occurs, leakage of the content of an examination paper or part of a paper prior to the examination, impersonation (when an

individual other than the registered candidate takes the examination), using unauthorized materials such as prepared answers and notes, intimidation of examination supervisors, issuing forged certificates or altering results information, or provision of external assistance via the supervisor or mobile phone do not occur. Inappropriate behaviors are dealt with by the legal affairs department on a case-by-case basis.

25. All stakeholder groups perceive examination results as credible.

26. Regardless of background, location, or the ability to pay, all students may take the examination.

27. All stakeholder groups use examination results in a proper way.

28. Student results are not confidential, as student names and results are public. Students are first informed of the results via SMS from the Ministry of Education's e-learning system, after which the results are made available to the media for dissemination to the public.

29. Although students have the option to retake the examination or repeat the grade, they do not have the options to attend remedial or preparatory courses in order to prepare to retake the examination or to opt for less-selective schools, universities, or tracks.

30. Only one systematic mechanism, expert review groups, is in place to monitor the consequences of the examination. Other mechanisms, such as regular focus groups or surveys of key stakeholders, studies that are updated regularly, a permanent oversight committee, and funding for independent research on the impact of the examination, are not in place.

UNITED ARAB EMIRATES
National (or System-Level) Large-Scale Assessment (NLSA)

ENABLING CONTEXT

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

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

(CONTINUED)

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

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 5: <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. ¹²	The NLSA office is adequately staffed to carry out the NLSA effectively, with no issues. ¹² *
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.	The country offers a wide range of opportunities to prepare individuals for work on the NLSA. ¹³ *

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

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

ASSESSMENT QUALITY

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

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

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

1. The United Arab Emirates National Assessment Program (UAENAP) was first administered in 2003 as the National Assessment of Student Achievement and Progress (NASAP) to students in Grade 5, and subsequently administered in 2005 to students in Grades 5 and 7, to students in Grades 3-12 in 2009 as the External Measurement of Student Achievement (EMSA), and to students in Grades 3, 5, 7, and 9 in 2010 as the UAENAP. (The NASAP, EMSA, AND UAENAP all refer to the same national assessment.) The main purposes of the assessment program include monitoring the education quality at the system level, supporting schools, teachers, and policy design, evaluation, and decision making. All students at specified grade levels are assessed in the following subjects: Arabic language, English language, mathematics, and science.
2. The Ministry of Education authorized the UAENAP with the *Implementing the National Assessment Program* document in 2010.
3. The *Implementing the National Assessment Program* document is available to the public.
4. There is a publicly available written plan specifying who will be tested and in which subject areas. The plan is available to, and easily accessible, by the public.
5. Most stakeholder groups support the UAENAP. While policymakers, teacher unions, and educators strongly support the UAENAP, and the media and universities support it as well, students, parents, think tanks and NGOs, and employers are neutral to it.
6. There is regular funding allocated by the government for the UAENAP.
7. Funding for the UAENAP covers all core activities of the assessment, including assessment design and administration, data analysis and reporting, long- or medium-term planning of program milestones, and staff training.
8. Funding for the UAENAP covers research and development activities.
9. The office in charge of the UAENAP is a permanent unit created for running the assessment. It is within the Administration of Evaluation and Examinations under the supervision of the Ministry of Education.
10. Political considerations never hamper technical considerations; large-scale assessment results have never been withheld from publication because of political reasons.
11. The office in charge of the UAENAP is held accountable to the Policy Council, a higher office in the Ministry of Education.
12. The office in charge of the UAENAP is adequately staffed with permanent and full-time staff to carry out the assessment effectively, with no issues identified with the performance of the human resources that are responsible for the large-scale assessment.

13. The UAE offers a wide range of opportunities to prepare individuals for work on the NLSA, including university graduate programs, university courses, and non-university courses or workshops on educational measurement and evaluation. In addition, funding for attending international programs, courses, or workshops on educational measurement and evaluation, and internships or short-term employment in the large-scale assessment office are available as well.
14. The UAENAP measures performance against national curriculum guidelines or learning standards.
15. Stakeholder groups largely accept what is measured by the UAENAP.
16. There are ad hoc reviews, regular internal reviews, and regular independent reviews of the alignment between the assessment instrument and what it is supposed to measure.
17. Live courses or workshops on the NLSA are offered occasionally, and they are accessible by most teachers. These courses tend to be of high quality and provide teachers with relevant resources that they can use in their classrooms. However, teachers do not have access to courses online.
18. The large-scale assessment is offered in the language of instruction for almost all student groups. However, special plans or accommodations are not made to ensure that the large-scale assessment is administered to students with disabilities or students in hard-to-reach areas.
19. There are some mechanisms in place to ensure the quality of the NLSA. For example, all proctors or administrators are trained according to a protocol, there is a standardized manual for large-scale assessment administrators, all booklets are numbered, there is double scoring of data, scorers are trained to ensure high inter-rater reliability, there are external and internal reviewers or observers, as well as external certification or audits. However, discrepancies are not required to be recorded on a standard sheet, there is no double processing of data, and a pilot is not conducted before the main data collection takes place.
20. A comprehensive, high-quality technical report on the general results is available to the public. Although detailed reports on the results are also produced, their circulation is restricted.
21. Results from the UAENAP are disseminated in an effective way. Reports with the results are made available to all stakeholder groups. Reports contain information on overall achievement levels and subgroups, as well as trends over time overall and by subgroups. Although results are not disseminated within 12 months after the large-scale assessment is administered and the main reports on the results do not contain standard errors, there is a media briefing organized to discuss results, workshops or presentations are organized for key stakeholders on the results, and the results are featured in newspapers, magazines, radio, or television.
22. Information on the UAENAP is used by all stakeholder groups in a way that is consistent with the stated purposes or technical characteristics of the assessment.

23. There are only two mechanisms in place to monitor the consequences of the NLSA: Expert review groups are in place and themed conferences that provide a forum to discuss research and other data on the consequences of the large-scale assessment are held. Regular focus groups or surveys of key stakeholders, a permanent oversight committee, and funding for independent research on the impact of the large-scale assessment are not in place.

UNITED ARAB EMIRATES

International Large-Scale Assessment (ILSA)

ENABLING CONTEXT

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

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
ENABLING CONTEXT 1: <i>Setting clear policies for ILSA</i>			
The country/system has not participated in an ILSA in the last 10 years.	This option does not apply to this dimension.	The country/system has participated in at least one ILSA in the last 10 years.	The country/system has participated in two or more ILSA in the last 10 years. ¹ *
The country/system has not taken concrete steps to participate in an ILSA in the next 5 years.	This option does not apply to this dimension.	The country/system has taken concrete steps to participate in at least one ILSA in the next 5 years. ² *	This option does not apply to this dimension.
There is no policy document that addresses participation in ILSA.	There is an informal or draft policy document that addresses participation in ILSA.	There is a formal policy document that addresses participation in ILSA. ³ *	This option does not apply to this dimension.
This option does not apply to this dimension.	The policy document is not available to the public. ⁴ *	The policy document is available to the public.	This option does not apply to this dimension.
ENABLING CONTEXT 2: <i>Having regular funding for ILSA</i>			
There is no funding for participation in ILSA.	There is funding from loans or external donors.	There is regular funding allocated at discretion. ⁵	There is regular funding approved by law, decree or norm. *
This option does not apply to this dimension.	Funding covers some core activities of the ILSA.	Funding covers all core activities of the ILSA. ⁶ *	This option does not apply to this dimension.
Funding does not cover research and development activities.	This option does not apply to this dimension.	This option does not apply to this dimension.	Funding covers research and development activities. ⁷ *

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 3: <i>Having effective human resources for ILSA</i>			
There is no team or national/system coordinator to carry out the ILSA activities.	There is a team or national/system coordinator to carry out the ILSA activities.	There is a team and national/system coordinator to carry out the ILSA activities. ⁸ *	This option does not apply to this dimension.
This option does not apply to this dimension.	The national/system coordinator or other designated team member may not be fluent in the language of the assessment.	The national/system coordinator is fluent in the language of the assessment. ⁹ *	This option does not apply to this dimension.
This option does not apply to this dimension.	The ILSA office is inadequately staffed or trained to carry out the assessment effectively.	The ILSA office is adequately staffed or trained to carry out the ILSA effectively, with minimal issues. ¹⁰ *	The ILSA office is adequately staffed and trained to carry out the ILSA effectively, with no issues.

SYSTEM ALIGNMENT

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

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

ASSESSMENT QUALITY

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

LATENT 	EMERGING 	ESTABLISHED 	ADVANCED 
ASSESSMENT QUALITY 1: <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. ¹⁴ *	This option does not apply to this dimension.
The country/system has not contributed new knowledge on ILSA.	This option does not apply to this dimension.	This option does not apply to this dimension.	The country/system has contributed new knowledge on ILSA. ¹⁵ *
ASSESSMENT QUALITY 2: <i>Ensuring effective uses of ILSA</i>			
If any, country/system-specific results and information are not disseminated in the country/system.	Country/system-specific results and information are disseminated irregularly in the country/system.	Country/system-specific results and information are regularly disseminated in the country/system.	Country/system-specific results and information are regularly and widely disseminated in the country/system. ¹⁶ *
Products to provide feedback to schools and educators about the ILSA results are not made available.	This option does not apply to this dimension.	Products to provide feedback to schools and educators about the ILSA results are sometimes made available.	Products to provide feedback to schools and educators about ILSA results are systematically made available. ¹⁷ *
There is no media coverage of the ILSA results.	There is limited media coverage of the ILSA results.	There is some media coverage of the ILSA results.	There is wide media coverage of the ILSA results. ¹⁸ *
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. ¹⁹ *
It is not clear that decisions based on ILSA results have had a positive impact on students' achievement levels. ²⁰ *	This option does not apply to this dimension.	This option does not apply to this dimension.	Decisions based on the ILSA results have had a positive impact on students' achievement levels.

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

1. The United Arab Emirates (UAE) has participated in three ILSAs in the past five years, including PIRLS and TIMSS in 2011, and PISA in 2009.
2. At the time of data collection, the UAE had taken concrete steps to participate in several ILSAs in the next five years, including PISA 2012 and 2015, TIMSS 2015, and PIRLS 2016.
3. The UAE Cabinet authorized the *Ministerial Council for Services Decision Number (73/6S/2)* document in 2010 to address participation in ILSAs.
4. The *Ministerial Council for Services Decision Number (73/6S/2)* document is not available to the public; however, it is available to strategic partners who participate in the implementation of the assessment exercises.
5. Regular funding for ILSAs in the UAE is approved by law, decree, or norm.
6. Funding covers all core activities of ILSAs, including international participation fees, implementation of the assessment exercise, processing and analyzing data from the implementation of the assessment exercise, reporting and disseminating assessment results, and attendance at international expert meetings for the assessment exercise.
7. Research and development activities are covered by funding for ILSAs.
8. The UAE has a team and national coordinator responsible for ILSA activities.
9. The national coordinator for ILSA in the UAE is fluent in the language of the assessment.
10. The ILSA office is adequately staffed and trained to carry out ILSAs effectively, with minimal issues. In addition, team members have attended all international meetings related to the assessment and have previous experience working on international assessments. However, there have been some issues with the translation of the assessment instruments; for example, regarding the translation of the PISA questions, given that there are strict guidelines for the number of words used in questions, there are some difficulties encountered when translating the original questions from English to Arabic. Other issues, such as errors or delays in the printing or layout of the test booklets or in the administration of the assessment, have not been identified.
11. The ILSA team has attended all international meetings related to the assessment.
12. The UAE offers some opportunities to learn about ILSAs, including workshops or meetings on using international assessment databases, funding for attending international workshops or training on international assessments, and online courses on international assessments. However, online courses are only periodically available, and there are no stand-alone university courses or workshops on the topic of international assessments.

13. University students studying assessment or a related area and ILSA team members benefit from opportunities to learn about international assessments in the UAE. However, professionals and university staff interested in assessment are not able to benefit from such opportunities.
14. The UAE met all technical standards required to have its data presented in the main displays of the international report.
15. The UAE has contributed to the global knowledge base on international assessments by generating new knowledge and making it available through publications and presentations.
16. Country-specific ILSA results and information are regularly and widely disseminated in the UAE.
17. In the UAE, products providing feedback to schools and educators about ILSA results are systematically made available.
18. In addition to ILSA results appearing on the front page of newspapers or as the main story on television news, there are editorials and columns commenting on international assessment results.
19. Results from ILSAs are used to inform decision making by tracking the impact of reforms on student achievement levels, as well as by informing curriculum improvement, teacher-training programs, and other assessment activities in the system. However, ILSA results have not been used to inform resource allocation.
20. At the time of data collection, it was not known whether decisions based on ILSA results have had a positive impact on students' achievement levels in the UAE.

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 from Ernesto Cuadra, Lead Education Specialist, and Simon Thacker, Education Specialist, in the World Bank's Middle East & North Africa region, as well as comments received during a national validation workshop held in UAE.

References

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

Ministry of Education, 2010. "The Ministry of Education Strategy 2010-2020." Dubai: Government of the United Arab Emirates.

Ministry of Education, 2013. "Education in the UAE." Dubai: Government of the United Arab Emirates. Data retrieved from www.moe.gov.ae/English/Pages/UAE/UaeEdu.aspx on October 4, 2013.

———. The United Arab Emirates Country Indicator Data. Washington, DC: World Bank. Data retrieved from <http://data.worldbank.org/> on October 3, 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.

