Bahrain

STUDENT ASSESSMENT

Key Policy Areas for Student Assessment

Status

1. Classroom Assessment

In Bahrain, a formal, system-level document provides guidelines for classroom assessment. Some resources, such as scoring criteria and rubrics for grading students' work, are available system-wide to support teachers in their classroom assessment activities. Opportunities to build skills in classroom assessment are available to teachers through pre- and in-service training. At the same time, classroom assessment practices are perceived as being of moderate quality with a tendency to overly focus on information recall rather than higher-order skills.

Established
●●●○

2. Examinations

The Examination for the Certificate of the Completion of General Secondary Education is administered to students in grades 10, 11, and 12. It covers the subjects of social studies, mathematics, science, language arts, Islamic studies, sports, and art. Examination results are mainly used for decisions about student promotion to the next grade level in secondary school, and student selection to higher education institutions. Regular funding is allocated by the government, and the Examination Directorate, which has primary responsibility for running the examination, is adequately staffed. However, there are no formal mechanisms in place to monitor the overall impact of the examination on the education system, particularly in terms of its effects on students, teachers, and schools.

Established
●●●○

3. National Large-Scale Assessment (NLSA)

The National Examinations, which monitors education quality at the system level and informs policy design, evaluation, and decision-making, assesses all students in grades 3, 6, 9, and 12. First administered in 2009, it has have since been carried out in 2010 and 2013. Regular funding is allocated by the government. The National Authority for Qualifications and Quality Assurance of Education and Training, which implements the assessment, has an adequate number of staff to carry out activities effectively. However, only some mechanisms are in place to monitor the actual impact of the NLSA on policy dialog, design, and decision-making. While a permanent oversight committee is in place, there are no expert review groups.

Established
●●●○

4. International Large-Scale Assessment (ILSA)

Bahrain has participated in three ILSAs in the past 10 years – TIMSS 2003, 2007, and 2011. Funding for ILSAs is allocated at discretion. A national coordinator and team are responsible for carrying out ILSA activities. In-country opportunities to learn about ILSAs are not very varied, but available to a broad audience. Country-specific ILSA results are widely disseminated in a timely manner. The results are used in a variety of ways to inform decision making in Bahrain, including tracking the impact of reforms on student achievement levels and informing resource allocation.

Established
●●●○

Systems Approach for Better Education Results

WORLD BANK GROUP

Introduction

Bahrain has focused on increasing student learning outcomes by improving the quality of education in the country. An effective student assessment system is an important part of 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, Bahrain 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 Bahrain

Bahrain is a high-income country in the Middle East and North Africa region. In 2010, GDP per capita (current US$, 2010) was $18,334, and annual growth was 4.5 percent.

In Bahrain, basic education lasts for 9 years and is compulsory for children between the ages of six and 14. Secondary education lasts three years. In 2002 (the year for which the latest data are available), net enrolment rates at the primary and secondary levels were 99 percent and 90 percent respectively. In 2008, 11.7 percent of government spending in Bahrain was allocated to education, equivalent to 2.9 percent of GDP.

Bahrain recently implemented King Hamad's Schools of the Future project, with the goal to integrate Information and Communications Technology (ICT) into various aspects of the education system. For example, the project seeks to help teachers individually assess their students' progress through the use of various e-learning activities. In addition, in 2008, the country instituted the National Authority for Qualifications and Quality Assurance of Education and Training (NAQQAET) to monitor the quality of education at all levels. Bahrain is also implementing other reforms to improve its education system, including introducing greater flexibility for students at the secondary stage to pursue vocational training and enter the labor force, as well as improving recruitment, training, and employment conditions for teachers.

Detailed information on Bahrain's student assessment system was collected using the SABER-Student Assessment questionnaires and rubrics. It is important to remember that these tools primarily focus on benchmarking a country's policies and arrangements for assessment activities at the system or macro level. Additional data would need to be collected to determine actual, on-the-ground practices in Bahrain, particularly by teachers and students in schools. The following sections discuss the findings for each assessment type, accompanied by suggested policy options. The suggested policy options were determined in collaboration with key local stakeholders based on Bahrain's immediate interests and needs. Detailed, completed rubrics for each assessment type are provided in Appendix 5.

Classroom Assessment

Level of development

Established
●●●○

In Bahrain, *Educational Assessment System* is a formal, system-level document that provides guidelines for classroom assessment. It identifies methods of classroom assessment, as well as tools that can be used to measure students' performance.

Some resources are available system-wide to support teachers in their classroom assessment activities. For example, curriculum documents for different subject areas outline what students are expected to learn at different grade levels. The *Educational Assessment System* and the *Teacher's Guide for Assessing the Daily Tasks* documents outline the level of performance that students are expected to reach in different subject areas at different grade levels. Textbooks and workbooks, as well as scoring criteria and rubrics for grading students' work, also provide support for teachers' classroom assessment activities.

At the same time, mechanisms are in place to systematically ensure that teachers develop skills and expertise in classroom assessment. For example, there is in-service teacher training on how to design achievement tests and use formative classroom assessment methods; pre-service teacher training; and opportunities to participate in developing and scoring questions for large-scale assessments and examinations. In addition, school inspection and teacher supervision include a component focused on classroom assessment.

Classroom assessment practices are perceived to be of moderate quality. Classroom activities tend to be focused on information recall and grade inflation is a serious problem.

Classroom assessment information is required to be disseminated to all key stakeholders, including students, parents, and school district and Ministry of Education officials.

There are adequate required uses of classroom assessment to support student learning. For example,

classroom assessment is used to diagnose student learning issues, provide feedback to students on their learning, inform parents about their child's learning, and plan next steps in instruction.

Limited mechanisms are in place to systematically monitor the quality of classroom assessment practices. For example, classroom assessment is a required component of a teacher's performance evaluation and school inspection. System-wide reviews of the quality of education also include a focus on classroom assessment.

Suggested policy options

1. Introduce additional systematic mechanisms to monitor the quality of classroom assessment practices. For example, expand the NAQQAET school quality assurance reports to include more in-depth reviews of classroom assessment practices. (Responsibility: NAQAAET)

2. Expand the availability of system-wide resources for teachers to conduct classroom assessment activities. For example, establish online assessment resources or computer-based testing with instant reports on students' performance. (Responsibility: MoE)

3. Introduce additional system-level mechanisms to ensure that teachers develop skills and expertise in classroom assessment. For example, mandate that all teacher training programs include a required course on classroom assessment. (Responsibility: MoE)

Examinations

Level of development

Established
●●●○

The Examination for the Certificate of the Completion of General Secondary Education has been administered since 1958. A formal, system-level document, *General Examination Regulations* (2010)*,* authorizes the examination, which is administered to students in grades 10, 11, and 12. The examination covers the subjects of Social Studies, Mathematics, Science, Language Arts, Islamic Studies, Sports, and Art. The main purposes for which examination results are used are decisions about student promotion to the next grade level in secondary school, and student selection to higher education institutions.

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

The Examination Directorate within the MoE is responsible for running the examination. The Examination Directorate has all of the required facilities to carry out the examination and is adequately staffed.

Bahrain offers a wide range of opportunities that prepare individuals to work on the examination. For example, there are university graduate programs in Bahrain University and other private universities in the country, as well as individual university courses (graduate and non-graduate) on educational measurement and evaluation relevant to examinations. There are also opportunities to attend non university training courses or workshops, and funding is available for attending international programs or workshops on educational measurement and evaluation.

Up-to-date, compulsory courses and workshops on the examination are available to teachers. Teachers are also involved in most examination-related tasks, such as creating and selecting examination questions and administering and scoring the examination.

However, there are very limited systematic mechanisms in place to ensure the quality of the examination. Only internal reviews and observers are used in the quality assurance process. There are no external reviews or observers. In addition, there is no piloting or field testing of instruments. Finally, there are no mechanisms, such as a permanent oversight committee or expert review groups, available to monitor the consequences of the examination in terms of its impact on students, teachers, and schools.

Suggested policy options

1. Introduce additional systematic mechanisms to ensure the quality of the examination. For example, conduct a pilot before the examination is administered or commission external certifications or audits of the examination. (Responsibility: MoE)

2. Introduce a variety of system-level mechanisms to monitor the consequences of the examination. For example, establish a permanent oversight committee or expert review groups to review the impact of the examination on students, teachers, and schools. (Responsibility: MoE)

National Large-Scale Assessment (NLSA)

Level of development

Established
●●●○

The National Examinations, which monitor education quality at the system level and inform policy design, evaluation, and decision making, assess all students in grades 3, 6, 9, and 12. For grade 3 students, the assessment covers Arabic language and Mathematics. For grade 6 and grade 9 students, the examination covers Arabic and English languages, Mathematics, and the Sciences. For grade 12 students, the assessment covers Arabic language, English language, and problem solving. Formal, system-level documents, Royal decrees 32 (2008), 6 (2009), and 83 (2012), authorize the National Examinations, and are available to the public online. In addition, there is a written NLSA plan for the coming years, which specifies which grade levels will be tested and in which subject areas.

Regular funding for the National Examinations is allocated by the government. This funding covers all core activities as well as research and development.

The NAQQAET, formerly known as the Quality Assurance Authority for Education and Training (QAAET), is a permanent, independent national authority operating under the guidance of a Board of Directors that is responsible for carrying out the National Examinations. The NAQQAET reports to the Cabinet of Ministers. The NAQQAET has an adequate number of full-time staff to carry out the National Examinations effectively, with no issues.

A variety of mechanisms are in place to ensure the quality of the NLSA. For example, all proctors and administrators are trained according to a protocol and a pilot is conducted before the main data collection takes place. In addition, all booklets are numbered and discrepancies are required to be recorded on a standard sheet. Although a comprehensive technical report is available, its circulation is restricted.

NLSA results are disseminated within 12 months after the assessment is administered. The results are featured in newspapers and magazines, and on radio and television. The main reports on the NLSA results contain information on achievement levels overall and by subgroups, as well as trends over time, overall and for subgroups.

Some mechanisms are in place to monitor the consequences of the NLSA, including a permanent oversight committee, which is the Education and Training Reform Committee. In addition, funding is made available for internal research on the impact of the NLSA.

Suggested policy options

1. Ensure that results from the National Examinations are used to inform curricula, teacher training, and relevant policies. In addition, ensure that the NLSA is aligned with other assessments carried out in the system, including classroom assessments and high-stakes examinations. Additionally, establish regular focus groups or surveys of key stakeholders to monitor the consequences of the NLSA for policy dialog, design, and decision making. (Responsibility: NAQAAET and MoE)

2. Ensure that the comprehensive technical report on the National Examinations is readily available to the general public. (Responsibility: NAQAAET and MoE)

International Large-Scale Assessment (ILSA)

Level of development

Established
●●●○

Bahrain has participated in three ILSAs in the past 10 years—Trends in International Mathematics and Science Study (TIMSS) 2003, 2007, and 2011—and has taken concrete steps to participate in two more in the near future—Progress in International Reading Literacy Study (PIRLS) 2016 and TIMSS 2015. Although there are formal policy documents addressing Bahrain's participation in ILSAs, they are not available to the public.

Regular funding for ILSAs is allocated at the government's discretion, and covers all core activities, such as international participation fees, implementation of the assessment exercise in Bahrain, and processing and analysis of the data collected. In addition, funding covers the implementation of a remedial action plan to improve students' performance, which is developed after the ILSA results are made available and weaknesses in various competencies have been identified.

A national coordinator and team carry out the ILSA activities in Bahrain. There are a sufficient number of team members with prior experience working on ILSAs who have the training and experience necessary to carry out the assessment activities effectively.

Bahrain offers some opportunities to learn about ILSAs, including workshops and meetings on using international assessment databases, which are available to various audiences, including teachers, principals, and university students studying assessment or a related area. However, there is no funding available to attend international workshops or training on this topic. Further, there are no university or online courses on international assessments.

Bahrain has not only met all technical standards required to have its data presented in the main displays of the international report, but has also contributed to new knowledge on ILSAs through being featured in the *TIMSS 2007 Encyclopedia* and the and *TIMSS 2011 Encyclopedia*, which are publicly available on the TIMSS website.

Country-specific ILSA results are disseminated throughout Bahrain in a timely manner. For example, the national report on the TIMSS 2007 results was published and made available on the MoE website. Copies of the national report were distributed to key stakeholders, and country results were communicated through a press release. Products providing feedback to schools and educators are also systematically made available in Bahrain. However, there has been limited media coverage of ILSA results, including commentary in editorials or columns.

Results from the ILSAs have been used to inform decision making in a variety of ways, including to track the impact of reforms on student achievement levels, as well as to inform curriculum improvement, teacher training programs, and resource allocation. Decisions based on ILSA results have had a positive impact on students' achievement levels as evidenced by the advancement in achievement that occurred between TIMSS 2003 and TIMSS 2007.

Suggested policy options

1. Improve dissemination and reporting of ILSA results through the media. For example, collaborate with various media outlets to ensure extensive coverage of results. (Responsibility: NAQAAET and MoE)

2. Expand local opportunities to learn about ILSAs. For example, develop local university or online courses on international assessments. (Responsibility: NAQAAET and MoE)

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
	Absence of, or deviation from, the attribute	*On way to meeting minimum standard*	*Acceptable minimum standard*	*Best practice*
CLASSROOM ASSESSMENT	There is no system-wide institutional capacity to support and ensure the quality of classroom assessment practices.	There is weak system-wide institutional capacity to support and ensure the quality of classroom assessment practices.	There is sufficient system-wide institutional capacity to support and ensure the quality of classroom assessment practices.	There is strong system-wide institutional capacity to support and ensure the quality of classroom assessment practices.
EXAMINATIONS	There is no standardized examination in place for key decisions.	There is a partially stable standardized examination in place, and a need to develop institutional capacity to run the examination. The examination typically is of poor quality and is perceived as unfair or corrupt.	There is a stable standardized examination in place. There is institutional capacity and some limited mechanisms to monitor it. The examination is of acceptable quality and is perceived as fair for most students and free from corruption.	There is a stable standardized examination in place and institutional capacity and strong mechanisms to monitor it. The examination is of high quality and is perceived as fair and free from corruption.
NATIONAL (OR SYSTEM-LEVEL) LARGE-SCALE ASSESSMENT	There is no NLSA in place.	There is an unstable NLSA in place and a need to develop institutional capacity to run the NLSA. Assessment quality and impact are weak.	There is a stable NLSA in place. There is institutional capacity and some limited mechanisms to monitor it. The NLSA is of moderate quality and its information is disseminated, but not always used in effective ways.	There is a stable NLSA in place and institutional capacity and strong mechanisms to monitor it. The NLSA is of high quality and its information is effectively used to improve education.
INTERNATIONAL LARGE-SCALE ASSESSMENT	There is neither a history of participation in an ILSA nor plans to participate in one.	Participation in an ILSA has been initiated, but there still is need to develop institutional capacity to carry out the ILSA.	There is more or less stable participation in an ILSA. There is institutional capacity to carry out the ILSA. The information from the ILSA is disseminated, but not always used in effective ways.	There is stable participation in an ILSA and institutional capacity to run the ILSA. The information from the ILSA is effectively used to improve education.

Appendix 4: Methodology for Assigning Development Levels

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

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

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

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

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

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

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

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

6. Scores for certain key dimensions under 'Enabling Context' (in the case of EXAM, NLSA, and ILSA) and under 'System Alignment' (in the case of CLASS) were set as ceiling scores, i.e., the overall mean score for the particular assessment type cannot be greater than the score for these key dimensions. These key variables include formal policy, regular funding, having a permanent assessment unit, and the quality of assessment practices.

Appendix 5: SABER-Student Assessment Rubrics for Bahrain

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

BAHRAIN
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.	There is an official curriculum or standards document that specifies what students are expected to learn and to what level of performance.[4]
ENABLING CONTEXT AND SYSTEM ALIGNMENT 3: *Having effective human resources to carry out classroom assessment activities*			✱
There are no system-level mechanisms to ensure that teachers develop skills and expertise in classroom assessment.	This option does not apply to this dimension.	There are some system-level mechanisms to ensure that teachers develop skills and expertise in classroom assessment.[5] ✱	There are a variety of system-level mechanisms to ensure that teachers develop skills and expertise in classroom assessment.

ASSESSMENT QUALITY

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

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

Classroom Assessment: Development-level rating justifications

1. The *Educational Assessment System* document, authorized by the Center of Measurement and Assessment in Bahrain in 2010, provides guidelines for classroom assessment. It identifies the methods of classroom assessment, as well as the tools that may be used to measure student's performance in applying each method. In addition, guidelines for classroom assessment are found in the *Teacher's Guide for Assessing the Daily Tasks of Students in Basic Education*.

2. The document has been made available to all schools and teachers. It is available online, as in-service courses for teachers, and in all schools and educational directorates at the Ministry of Education.

3. There are some system-wide resources available for teachers to engage in classroom assessment. For example, curriculum documents for different subjects outline what students are expected to learn in different subject areas at different grade or age levels, and the *Educational Assessment System* and the *Teacher's Guide for Assessing the Daily Tasks* documents outline the level of performance that students are expected to reach in different subject areas at different grade or age levels. In addition, there are textbooks or workbooks that provide support for classroom assessment and scoring criteria or rubrics for students' work. However, there are no online assessment resources or computer-based testing with instant reports on students' performance.

4. An official curriculum or standards document specifies what students are expected to learn and to what level of performance.

5. There are some system-level mechanisms in place to ensure that teachers develop skills and expertise in classroom assessment. For example, there is in-service teacher training on elaborating achievement tests and using classroom assessment methods, pre-service teacher training, opportunities to participate in item development for or scoring of large-scale assessments or exams, and school inspection or teacher supervision which includes a component focused on classroom assessment. However, not all teacher training programs include a required course on classroom assessment, there are no online resources on classroom assessment, and there are no opportunities to participate in conferences or workshops.

6. Classroom assessment practices are known to be of moderate quality. It is uncommon to observe errors in the scoring or grading of work, and teachers commonly use explicit or a priori criteria for scoring or grading work. Classroom assessments do not rely only on multiple-choice, selection-type questions, and the uneven application of standards for grading students' work is not a serious problem. Classroom assessment activities provide useful feedback to students, and parents are informed about students' grades. However, classroom assessment activities are commonly about recalling information, and grade inflation is a serious problem.

7. There are limited systematic mechanisms in place to monitor the quality of classroom assessment practices. Although classroom assessment is a required component of a teacher's performance evaluation and school inspection or teacher supervision, and there are national or other system-wide reviews of the quality of education which include a focus on classroom assessment, there is no external moderation system which reviews the difficulty of classroom assessment activities and the appropriateness of scoring criteria, and no government funding is available for research on the quality of classroom assessment activities and how to improve classroom assessment.

8. Classroom assessment information is required to be disseminated to all key stakeholders, including students, parents, and school district and Ministry of Education officials.

9. There are adequate required uses of classroom assessment to support student learning, excluding its use as an input for external examination results. Classroom assessment is used to support student learning by diagnosing student learning issues, providing feedback to students on their learning, informing parents about their child's learning, and planning the next steps in instruction.

BAHRAIN
Examinations

BAHRAIN | SABER-STUDENT ASSESSMENT

ENABLING CONTEXT

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

LATENT ○○○○●	EMERGING ○○○●●●	ESTABLISHED ○●●●●	ADVANCED ●●●●
ENABLING CONTEXT 1: *Setting clear policies*			
No standardized examination has taken place.	The standardized examination has been operating on an irregular basis.	The examination is a stable program that has been operating regularly. [1] ✱	This option does not apply to this dimension
There is no policy document that authorizes the examination.	There is an informal or draft policy document that authorizes the examination.	There is a formal policy document that authorizes the examination. [2] ✱	This option does not apply to this dimension.
This option does not apply to this dimension.	The policy document is not available to the public. [3] ✱	The policy document is available to the public.	This option does not apply to this dimension.
This option does not apply to this dimension.	This option does not apply to this dimension.	The policy document addresses some key aspects of the examination.	The policy document addresses all key aspects of the examination. [4] ✱
ENABLING CONTEXT 2: *Having strong leadership*			
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.	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: *Having regular funding*			
There is no funding allocated for the examination.	There is irregular funding allocated for the examination.	There is regular funding allocated for the examination.[8] *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding covers some core examination activities: design, administration, data processing or reporting.	Funding covers all core examination activities: design, administration, data processing and reporting.[9] *	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding does not cover research and development.[10] *	This option does not apply to this dimension.	Funding covers research and development.
ENABLING CONTEXT 4: *Having strong organizational structures*			
The examination office does not exist or is newly established.	The examination office is newly established.	The examination office is a stable organization.[11] *	This option does not apply to this dimension.
The examination office is not accountable to an external board or agency.[12] *	This option does not apply to this dimension.	The examination office is accountable to an external board or agency.	This option does not apply to this dimension.
Examination results are not recognized by any certification or selection system.	Examination results are recognized by certification or selection system in the country.	Examination results are recognized by one certification or selection system in another country.	Examination results are recognized by two or more certification or selection systems in another country.[13] *
The examination office does not have the required facilities to carry out the examination.	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.

(CONTINUED)

ENABLING CONTEXT 5:
Having effective human resources

	LATENT	EMERGING	ESTABLISHED	ADVANCED
	●○○○	●●○○	●●●○	●●●●
	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.	The examination office is adequately staffed to carry out the assessment effectively, with no issues. [15] ✱
	The country does not offer opportunities that prepare for work on the examination.	This option does not apply to this dimension.	The country offers some opportunities that prepare for work on the examination.	The country offers a wide range of opportunities that prepare for work on the examination. [16] ✱

SYSTEM ALIGNMENT

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

LATENT	EMERGING	ESTABLISHED	ADVANCED
●○○○	●●○○	●●●○	●●●●

SYSTEM ALIGNMENT 1:
Aligning examinations with learning goals and opportunities to learn

LATENT	EMERGING	ESTABLISHED	ADVANCED
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.[19] *	There is comprehensive material to prepare for the examination that is accessible to all students.

SYSTEM ALIGNMENT 2:
Providing teachers with opportunities to learn about the examination

LATENT	EMERGING	ESTABLISHED	ADVANCED
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.	Teachers are involved in some examination-related tasks.	Teachers are involved in most examination-related tasks.[21] *

ASSESSMENT QUALITY

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ASSESSMENT QUALITY 1: *Ensuring quality*			
There is no technical report or other documentation.	There is some documentation on the examination, but it is not in a formal report format.	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: *Ensuring fairness*			
Inappropriate behavior surrounding the examination process is high.	Inappropriate behavior surrounding the examination process is moderate.	Inappropriate behavior surrounding the examination process is low.	Inappropriate behavior surrounding the examination process is marginal.[24] ✱
The examination results lack credibility for all stakeholder groups.	The examination results are credible for some stakeholder groups.	The examination results are credible for all stakeholder groups.[25] ✱	This option does not apply to this dimension.
The majority of the students (over 50 percent) may not take the examination because of language, gender, or other equivalent barriers.	A significant proportion of students (10 percent to 50 percent) may not take the examination because of language, gender, or other equivalent barriers.	A small proportion of students (less than 10 percent) may not take the examination because of language, gender, or other equivalent barriers.	All students can take the examination; there are no language, gender or other equivalent barriers.[26] ✱

(CONTINUED)

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

Examinations: Development-level rating justifications

1. The Examination for the Certificate of the Completion of General Secondary Education was first administered in 1958. The examination covers the subjects of Social Studies, Mathematics, Science, Language Arts, Islamic Studies, Sports, and Art in grades 10, 11, and 12. The examination's main purposes are for student selection or promotion to grades, courses, or tracks in secondary school, and student selection to university or other higher education institutions.

2. The Ministry of Education's *General Examination Regulations* from 2010 authorizes the examination.

3. The policy document is available only to parties in charge of the preparation and implementation of the examination.

4. The policy document addresses all key aspects of the examination, including governance, distribution of power, responsibilities among key entities, the purpose of the examination, authorized uses of results, procedures for special or disadvantaged students, and procedures to investigate and address security breaches, cheating, or other forms of inappropriate behavior. Additionally, it states funding sources, specifies who can sit for the examination, identifies rules about preparation, and explains alignment with curricula and standards and the format of the examination questions.

5. Most stakeholder groups support the examination. Policy makers, educators, students, parents, the media, and universities all support the examination. However, it is unclear what level of support teacher unions, think tanks, NGOs, and employers have for the examination.

6. Stakeholder groups have made coordinated attempts to improve the examination.

7. Leadership in charge of the examination generally welcomes efforts to improve the examination.

8. There is regular funding allocated by the government for the examination.

9. Funding covers all core examination activities, including examination design and administration, data analysis and reporting, long- or medium-term planning of program milestones, research and development, and staff training.

10. Funding for the examination does not cover research and development activities.

11. The Examination Directorate, which has had primary responsibility for running the examination since 1958, is an office or branch within the Ministry of Education.

12. The examination office is not accountable to an external board or agency, as it is a branch within the Ministry of Education.

13. Examination results are recognized by certification or selection systems in other Gulf countries.

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

15. The examination office is adequately staffed with full-time and permanent staff to carry out the assessment effectively, with no issues.

16. Bahrain offers a wide range of opportunities that prepare for work on the examination. For example, on the topic of educational measurement and evaluation, there are university graduate programs housed in the Bahrain University and other private universities in the kingdom, as well as university courses (graduate and non-graduate). There are also opportunities to attend non-university training courses or workshops and funding for attendance at 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. There is comprehensive material to prepare for the examinations that is accessible to most students. All previous exam questions and their corresponding answers are available on the Ministry of Education's website and that of the HM King Hamad's Schools of the Future Project, which is accessible at www.moe.gov.bh/khsfp/. However, 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 are unavailable to students preparing for the examination.

20. Up-to-date compulsory courses or workshops on examinations are available for teachers.

21. Teachers are involved in most examination-related tasks. For example, teachers are involved in selecting or creating examination questions and examination scoring guides, administering and scoring the examination, acting as a judge, and supervising examination procedures.

22. A comprehensive technical report is available but with restricted circulation.

23. There are very limited systematic mechanisms in place to ensure the quality of the examination. Internal review or observers are in place; however, there is no external review or observers, external certification or audit, pilot or field testing, or translation verification.

24. Inappropriate behavior surrounding the examination process is marginal. Issues such as copying from other candidates, using unauthorized materials such as prepared answers and notes, and intimidation of examination supervisors do not typically occur during the examination process.

25. All stakeholder groups perceive the examination results as credible.

26. All students can take the examination regardless of location, background, or ability to pay.

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

28. Students' results are confidential as only the student and persons with a legitimate, professional interest in the test taker can know the results.

29. There are some options for students who do not perform well on the examination. For example, students may retake the examination or attend remedial or preparatory courses in order to prepare to retake the examination. However, there are no options for students to repeat the grade, or opt for less selective schools, universities, or tracks.

30. There are no mechanisms, such as a permanent oversight committee or expert review groups, in place to monitor the consequences of the examination.

BAHRAIN

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

BAHRAIN | SABER-STUDENT ASSESSMENT

ENABLING CONTEXT

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

LATENT ◐○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 1: *Setting clear policies for NLSA*			
No NLSA exercise has taken place.	The NLSA has been operating on an irregular basis.	The NLSA is a stable program that has been operating regularly. [1] ✱	This option does not apply to this dimension.
There is no policy document pertaining to NLSA.	There is an informal or draft policy document that authorizes the NLSA.	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: *Having strong public engagement for NLSA*			
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: _Having regular funding for NLSA_			
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: _Having strong organizational structures for NLSA_			
There is no NLSA office, ad hoc unit or team.	The NLSA office is a temporary agency or group of people.	The NLSA office is a permanent agency, institution or unit.[9] *	This option does not apply to this dimension.
This option does not apply to this dimension.	Political considerations regularly hamper technical considerations.	Political considerations sometimes hamper technical considerations.	Political considerations never hamper technical considerations.[10] *
This option does not apply to this dimension.	The NLSA office is not accountable to a clearly recognized body.	The NLSA office is accountable to a clearly recognized body.[11] *	This option does not apply to this dimension.

(CONTINUED)

ENABLING CONTEXT 5:
Having effective human resources for NLSA

LATENT	EMERGING	ESTABLISHED	ADVANCED
○○○	●○○○	●●○○	●●●●
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. [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. [13] *	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:
Aligning the NLSA with learning goals

LATENT	EMERGING	ESTABLISHED	ADVANCED
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:
Providing teachers with opportunities to learn about the NLSA

LATENT	EMERGING	ESTABLISHED	ADVANCED
There are no courses or workshops on the NLSA.	There are occasional courses or workshops on the NLSA.	There are some courses or workshops on the NLSA offered on a regular basis.	There are widely available high quality courses or workshops on the NLSA offered on a regular basis. [17] ✱

ASSESSMENT QUALITY

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

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

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

1. The National Examinations, which monitor education quality at the system level and assist with policy design, evaluation, and decision making, assess all students in grades 3, 6, 9, and 12. These assessments have been in place for less than five years. Initially, the assessments were administered to students in grades 3 and 6 in 2009, with grades 9 and 12 added in 2010 and 2013 respectively. For grade 3 students, the assessment covers Arabic language and Mathematics, while for students in grade 6 and grade 9, the examination covers Arabic and English languages, Mathematics, and Science. For grade 12 students, the assessment covers Arabic language, English language, and problem solving.

2. HM the King of the Kingdom of Bahrain authorized Royal decree number 32 in 2008 to establish and organize the National Authority for Qualifications and Quality Assurance of Education and Training (NAQQAET). This decree was subsequently amended by decree number 6 in 2009 and decree number 83 in 2012.

3. Royal decrees 32 (from 2008), 6 (from 2009), and 83 (from 2012), which authorize the NLSA, are available online to the public at www.legalaffairs.gov.bh.

4. There is a written NLSA plan for the coming years. As one of the initiatives mentioned in the National Economic Strategy to help achieve Economic Vision 2030, the NAQQAET has been adopted to reform the National Education System. It was given the responsibility "of standard setting and quality control by using a series of national examinations and periodic assessments in public schools and of submitting general reports thereto."(Source: National Economic Strategy 2009-2014, p. 49). The plan specifies who will be tested and in which subject areas. The plan is easily accessible by the public. Information related to the national examinations may be retrieved through the NAQQAET website: http://www.qaa.bh.

5. Most stakeholder groups support the NLSA. While policy makers, educators, the media, universities, and employers support the NLSA, students and parents are neutral to it. It is unclear what level of support teacher unions, think tanks, NGOs, and similar organizations maintain.

6. The government allocates regular funding to the NLSA.

7. Funding covers all core NLSA activities, including design, administration, analysis, reporting, long- or medium- term planning of program milestones, research and development, and staff training.

8. Funding for the NLSA covers research and development activities.

9. The NAQQAET, formerly known as the Quality Assurance Authority for Education and Training (QAAET), is a permanent, independent national authority that operates under the guidance of a Board of Directors established by decree number 7 of 2009, and reports to the Cabinet of Ministers.

10. Political considerations never hamper technical considerations, and assessment results have never been withheld due to political reasons.

11. The NAQQAET is accountable to a clearly recognized body, the Cabinet of Ministers, and is subject to its supervision. NAQQAET's performance is monitored by the Education and Training Reform Committee, as it is one of the national initiatives established to develop education and training in the Kingdom of Bahrain.

12. The NAQQAET has an adequate number of permanent and full-time staff to carry out the NLSA effectively, with no issues.

13. Bahrain offers university courses on educational measurement and evaluation, which prepare individuals for work on NLSAs.

14. The NLSA measures performance against national curriculum guidelines or learning standards for grades 3, 6, and 9 and internationally recognized curriculum guidelines or learning standards for grade 12. Grade 12 National Examinations in Arabic and Problem Solving are set at the Cambridge AS Level (UK), and the grade 12 National Examination in English is set at the Vantage, Level B2 of the Common European Framework (CEF).

15. Stakeholder groups largely accept what is measured by the NLSA.

16. There are regular internal reviews of the alignment between the assessment instrument and what it is supposed to measure.

17. There are widely available high quality courses or workshops on the NLSA offered on a regular basis. Additionally, most courses provide teachers with relevant resources that they can use in their classrooms.

18. Accommodations or alternative assessments are provided for students with disabilities, and the NLSA is offered in the language of instruction for almost all student groups.

19. There are a variety of mechanisms in place to ensure the quality of the NLSA. For example, all proctors or administrators are trained according to a protocol, there is a standardized manual for large-scale assessment administrators, and a pilot is conducted before the main data collection takes place. In addition, all booklets are numbered and discrepancies must be recorded on a standard sheet. Also, there are external and internal reviewers or observers, double scoring and processing of data, and training for scorers to ensure high inter-rater reliability.

20. Although a comprehensive technical report is available, its circulation is restricted.

21. NLSA results are disseminated in an effective way. Results are disseminated within 12 months of the large-scale assessment being administered and are featured by media such as newspapers, magazines, radio, or television. In addition, reports with results are made available for all stakeholder groups, there are workshops or presentations for key stakeholders on the results, and a media briefing is organized to discuss the assessment results. The main reports on the results contain information on overall achievement levels and subgroups, and trends over time overall and for subgroups.

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

23. There are some mechanisms in place to monitor the consequences of the NLSA, including a permanent oversight committee, which is the Education and Training Reform Committee, and funding for internal research on the impact of large-scale assessment. However, there are no expert review groups, themed conferences that provide a forum to discuss research and other data on the consequences of the large-scale assessment, or regular focus groups or surveys of key stakeholders.

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

BAHRAIN
International Large-Scale Assessment (ILSA)

ENABLING CONTEXT

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 1: *Setting clear policies for ILSA*			
The country/system has not participated in an ILSA in the last 10 years.	This option does not apply to this dimension.	The country/system has participated in at least one ILSA in the last 10 years.	The country/system has participated in two or more ILSA in the last 10 years.[1] ✶
The country/system has not taken concrete steps to participate in an ILSA in the next 5 years.	This option does not apply to this dimension.	The country/system has taken concrete steps to participate in at least one ILSA in the next 5 years.[2] ✶	This option does not apply to this dimension.
There is no policy document that addresses participation in ILSA.	There is an informal or draft policy document that addresses participation in ILSA.	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: *Having regular funding for ILSA*			
There is no funding for participation in ILSA.	There is funding from loans or external donors.	There is regular funding allocated at discretion.[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.		Funding covers research and development activities.[7] ✶

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
	ENABLING CONTEXT 3: *Having effective human resources for ILSA*		
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.	The ILSA office is adequately staffed and trained to carry out the ILSA effectively, with no issues. [10] ✱

SYSTEM ALIGNMENT

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

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

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

1. Bahrain has participated in three ILSAs in the past 10 years, TIMSS 2003, 2007, and 2011.

2. Bahrain has taken concrete steps to participate in two ILSAs in the next five years, including PIRLS 2016 and TIMSS 2015.

3. There are formal policy documents addressing Bahrain's participation in ILSAs.

4. The policy documents addressing ILSA participation are not available to the public.

5. There is regular funding allocated at the government's discretion.

6. Funding covers all core activities of the ILSA, including international participation fees, implementation of the assessment exercise in Bahrain, processing and analyzing data collected from the implementation of the assessment exercise, reporting and disseminating the assessment results in Bahrain, attendance at international expert meetings for the assessment exercise, and research and development. In addition, funding covers the implementation of a remedial action plan to improve students' performance, which is developed after results of the assessment are available and the weaknesses in various competencies are identified.

7. Funding covers research and development activities.

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

9. The ILSA coordinator in Bahrain is fluent in the language of the assessment.

10. The ILSA office is adequately staffed and trained to carry out the ILSA effectively, with no issues. In addition to the national coordinator being fluent in the language of the assessment, the team is sufficiently staffed, has previous experience working on international assessments, has the necessary training or experience to carry out the required assessment activities effectively, and has attended all international meetings related to the assessment.

11. The ILSA team has attended all international workshops or meetings.

12. Bahrain offers some opportunities to learn about ILSA, including workshops or meetings on using international assessment databases. However, there are no university or online courses on international assessments, or funding to attend international workshops, or training on international assessments.

13. Opportunities to learn about ILSA are available to a wide audience, in addition to Bahrain's ILSA team members. For instance, opportunities are available to teachers, principals, and university students studying assessment or a related area.

14. Bahrain met all technical standards required to have its data presented in the main displays of the international report.

15. Bahrain has contributed new knowledge on ILSA. For example, it is featured in the *TIMSS 2007 Encyclopedia* and *TIMSS 2011 Encyclopedia*, which are publicly available at the TIMSS website.

16. Country-specific results and information are regularly and widely disseminated in Bahrain. For example, the Kingdom of Bahrain's report on the results of TIMSS 2007 was published and previously available on the Ministry of Education's website. In addition, copies of the national report were distributed to key stakeholders, and country results were communicated through a press release. Results received coverage on the television, radio, and newspapers, and brochures and PowerPoint presentations with Bahrain's results were made available online or distributed to key stakeholders. In addition, products providing feedback to the schools or educators about the results were made available.

17. Products providing feedback to schools and educators about ILSA results are systematically made available in Bahrain.

18. There is limited media coverage, including commentary in editorials or columns, of the ILSA results.

19. Results from the ILSA are used in a variety of ways to inform decision making in Bahrain, including tracking the impact of reforms on student achievement levels, as well as informing curriculum improvement, teacher training programs, other assessment activities in the system, and resource allocation.

20. Decisions based on the ILSA results have had a positive impact on students' achievement levels, evidenced by the advancement in achievement between TIMSS 2003 and TIMSS 2007.

Acknowledgements

References

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

The Education Project. 2009. "Bahrain: Education and Training." Data retrieved from http://www.educationprojectbahrain.org/download/EDB%20-%20Education%20and%20Training%20September%2009.pdf on October 3, 2013.

Embassy of the Kingdom of Bahrain. 2013. "Bahrain's Education System." Data retrieved from http://www.mofa.gov.bh/Default.aspx?tabid=7741 on October 3, 2013.

Kingdom of Bahrain Ministry of Education.2003. "King Hamad's Schools of the Future Project." Data retrieved from http://www.moe.gov.bh/khsfp/khsfpdoc/future-eng.pdf on October 3, 2013.

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

World Bank. 2012. Bahrain 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.

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

WORLD BANK GROUP