Public Disclosure Authorized
Public Disclosure Authorized
Public Disclosure Authorized
Public Disclosure Authorized

Yemen

STUDENT ASSESSMENT

SABER Country Report
2013

Key Policy Areas for Student Assessment	Status
1. Classroom Assessment In Yemen, a formal, system-level document provides guidelines for classroom assessment. In addition, there are some system-wide mechanisms and resources, such as pre- and in-service training and Teachers' Guides, to help teachers develop skills and expertise in classroom assessment. A limited set of mechanisms also are in place to monitor the quality of classroom assessment practices (these include teacher/school supervision and teacher performance evaluations). At the same time, the general perception is that classroom assessment practices are weak, as they seem to provide little useful feedback to students.	Emerging
2. Examinations The General Secondary Certification Examination has been administered to grade 12 students since 1963. The examination results are used to certify student completion of secondary school and determine student selection to higher education institutions. The results are officially recognized by certification and selection systems in Yemen and abroad. Regular funding is provided by the government, but only covers some key activities, specifically, examination design, administration, and staff training. There are no systematic mechanisms in place to ensure the quality of the examination, and inappropriate behavior surrounding the examination process is high.	Emerging
3. National Large-Scale Assessment (NLSA) Yemen has not yet carried out an NLSA. However, the National Assessment System of Student (NASS) is a large-scale assessment program that is currently under development. NASS is intended to assess grades 4 and 7 students in the subjects of Science, Mathematics, and Arabic language.	Latent
4. International Large-Scale Assessment (ILSA) In the last ten years, Yemen has taken part in TIMSS 2003, 2007, and 2011. Currently, Yemen has no formal policy document that addresses participation in ILSAs or an official plan to participate in future ILSAs. The ILSA office is adequately staffed to carry out ILSA activities effectively with minimal issues. However, funding for ILSA participation is sourced entirely from loans and external donors. Yemen-specific ILSA results have not been disseminated in the country, and it is not clear that decisions based on the results have had a positive impact on achievement levels.	Emerging

Systems Approach for Better Education Results

WORLD BANK GROUP

Introduction

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

Yemen is a lower-income country in the Middle East & North Africa region. GDP per capita was $1,494 in 2012, and annual growth was 0.1 percent in 2012 after negative 10.5 percent growth in 2011, resulting from a political crisis. Even before the 2011 political events strained the economy further, Yemen was one of the poorest countries in the Arab region. The poverty rate rose from 42 percent of the nation's population in 2009 to an estimated 54.5 percent in 2012.

Although many challenges exist in providing quality education to all children in Yemen, the country has done much to expand educational opportunities over the past 30 years. Between 1977 and 2001, enrolments in basic education increased from 0.5 million to over 3 million.

The net enrolment rates at the primary and secondary levels were 87 percent and 63 percent respectively in 2011. However, there are still a number of out-of-school children and significant disparities in enrolment by gender. The net enrolment rates at the primary level for males and females in 2011 were 82 percent and 69 percent respectively. In addition, in the same year, only 63 percent of students completed primary school. In order to address these and other challenges, the Government of Yemen has identified several key priorities for the years 2013-2015. These key priorities include improving basic and secondary curricula, improving textbooks and teacher manuals, encouraging families to send their daughters to school, strengthening the capacities of the Ministry of Education, governorate and district offices, and expanding pre-school services.

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

Classroom Assessment

Level of development

In Yemen, a formal, system-level document, *General Regulations for Examinations*, authorized by the General Directorate for Examinations and Educational Assessment at the Ministry of Education in 2001, provides guidance for classroom assessment. However, the document is not available to the general public and is only accessible at the Ministry of Education and the Educational Offices Archives.

There are some system-wide resources for teachers for classroom assessment. These include *Teachers' Guides* for primary and secondary school teachers, which outline what students are expected to learn and the level(s) of performance that students are expected to reach in different subject areas at different grade levels. Although the content covered in the *Teachers' Guides* is also addressed through workshops and training sessions, in practice, the guides are not always available to all school teachers. Other available resources include textbooks and workbooks that provide support for classroom assessment, and scoring criteria and rubrics for grading students' work.

While the *Teachers' Guides* include information on what students are expected to learn, and how well they are expected to learn it, there is no official curriculum or standards document on which this content- and performance-related guidance is based.

There are some system-level mechanisms to ensure that teachers develop skills and expertise in classroom assessment, including pre- and in-service teacher training, a requirement for teacher training programs to include a course on classroom assessment, and school inspections and teacher supervisions which include a component focused on classroom assessment.

In Yemen, classroom assessment practices tend to be weak. It is common for classroom assessment activities to be mainly about recalling information, to be mainly used as an administrative tool rather than as a pedagogical input, and for teachers to not use explicit criteria for grading students' work. Additionally, it is common for classroom assessment activities to rely mainly on multiple-choice, selection-type questions. In general, classroom assessment practices provide little useful feedback to students.

As a way to monitor its quality, classroom assessment is a required component of a teacher's performance evaluation, and of school inspection and teacher supervision. However, there is neither an external moderation system that reviews the difficulty of classroom assessment activities and scoring nor national or other system-wide education reviews that include a focus on classroom assessment.

Classroom assessment information from summative examinations administered by teachers at the classroom level is required to be disseminated to all key stakeholders. Teachers are obligated to report on individual students' performance to the school district and Ministry of Education officials, parents, and students. Students' results are retained in the school, and provided to the Education Directorate, the Provincial Education office, and the Ministry of Education.

Suggested policy options

1. Ensure that the *General Regulations for Examinations* document is widely disseminated in, for example, teacher training programs.

2. Since classroom assessment practices tend to be weak, ensure that a variety of system-wide mechanisms are in place to monitor and support their quality. For example, allocate government funding for research on how to improve the quality of classroom assessment activities.

3. Develop a variety of system-wide resources to help teachers engage in classroom assessment and make these resources widely available so that teachers may use them in their practice. For example, create item banks with examples of multiple-choice and open-ended questions and disseminate them in various ways, including via in-service teacher training programs.

Examinations

Level of development

The General Secondary Certification Examination has been administered to grade 12 students since 1963 and is used for student certification for grade completion and student selection to university or higher education institutions. The Scientific section and the Literature section of the examination cover the subjects of the Holy Koran, Islamic Education, Arabic Language, English Language, and Mathematics. The Scientific section also covers Chemistry, Physics, and Biology, while the Literature section also covers Geography, History, Philosophy, Logic, and Psychology. There is a formal policy document that authorizes the examination, entitled *Examination Rules, Ministerial Decisions*.

Regular funding for the examination is allocated by the government, although the amount tends to be insufficient. Funding covers only examination design, examination administration, and staff training.

The Higher Committee for Examinations works with the General Administration for Examinations within the Ministry of Education to conduct the examination. The General Administration for Examinations has had primary responsibility for running the examinations since 1963 and has state-of-the art facilities to carry out the examination activities. In addition, there is an adequate number of permanent full-time staff to run the examination.

Yemen offers university graduate programs as well as graduate and nongraduate courses specifically focused on educational measurement and evaluation. Yemen also offers non university training courses and workshops on educational measurement and evaluation.

Although courses and workshops on the examination for teachers are compulsory, they are not regularly updated. Teachers are involved in administering the examination, scoring examination papers, and supervising examination procedures. On rare occasions, teachers are asked to participate in the creation and selection of examination questions.

While there are no mechanisms in place to ensure the quality of the examination, there are efforts to improve aspects of examination quality through, for example, internal reviews of the scoring criteria for each question. There is no external review or observers, or external certification or audit.

Students' results are not confidential; once results are officially announced, students' names and results are public and accessible to all.

Inappropriate behaviors that diminish the credibility of the examination typically occur during the examination process and include leakage of the content of an examination paper and impersonation (when an individual other than the registered candidate takes the examination). When either of these behaviors occurs, individuals involved are turned over to the legal authorities.

Suggested policy options

1. Ensure regular funding for all core examination activities, including data analysis and reporting, as well as research and development activities.

2. Introduce more mechanisms to systematically ensure the quality of the examination. For example, introduce external reviews or external observers for quality assurance purposes.

3. Create opportunities for teachers to learn about the examination and ensure that such opportunities are available on a regular basis. For example, conduct courses or workshops that introduce teachers to various aspects of the examination program, including its purpose and content.

4. In order to reduce unintended negative consequences associated with the publication of examination results, develop a protocol to ensure that students' identifying information is not made publically available with their examination results, except to those with a legitimate reason to know.

National Large-Scale Assessment (NLSA)

Level of development

Yemen has not yet carried out an NLSA. However, the National Assessment System of Student (NASS) is a large-scale assessment program that is currently being developed NASS is intended to assess grades 4 and 7 students in the subjects of Science, Mathematics, and Arabic Language.

Suggested policy options

1. Build momentum around the importance of regular system-level assessment of student achievement, and raise awareness about the NASS as a tool to meet this need.

2. Create policy documents that provide a formal policy basis for NASS, ensure a recurrent government budget to fund large-scale assessment activities, and strengthen the organizational structure required to implement NASS.

International Large-Scale Assessment (ILSA)

Level of development

Yemen participated in the Trends in International Mathematics and Science Study (TIMSS) 2003, 2007, and 2011. Although there is no official document or plan for future participation in an ILSA, there is an informational document available to all stakeholders interested in TIMSS called *Introduction to: The Trends in International Mathematics and Science Study (2009)).*

Funding for ILSA participation, sourced from loans and external donors, covers the international participation fees, implementation of the assessment exercise in the country, attendance at international expert meetings for the assessment exercise, external training sessions, and attendance at international education conferences.

In Yemen, a national coordinator and team carry out the ILSA activities.

The ILSA team is sufficiently staffed, has previous experience working on international assessments, and has the necessary training and experience to carry out the required assessment activities effectively. Team members have also attended some of the international meetings related to the assessment.

At the same time, there have been some issues with carrying out these international assessments in the country. These include errors or delays in the printing and layout of the test booklets and the administration of the test, as well as complaints about poor training of test administrators.

Opportunities to learn about the ILSA are available to a wide audience, including individuals working directly on a specific ILSA exercise (including those who participated in program implementation, such as supervisors and researchers from the Center for Educational Development and Research, as well as curricula specialists) and Master of Arts students studying assessment in the faculties of education in five universities (Sanaa University, Ibb University, Taiz University, Hodeidah University, and Hadramout University.

Yemen met all technical standards required to have its data presented in the main displays of the TIMSS international report. Country-specific results from the most recent ILSA were irregularly disseminated in the country, and products to provide feedback to schools and educators were not available.

The results of the TIMSS 2007 assessment were used by policy makers and education leaders to improve education quality in Yemen by elaborating the Basic Classes Initiative program. However, results have not been used to track the impact of reforms on student achievement levels or to inform resource allocation. There is no evidence of a positive impact on student achievement from decisions based on ILSA results.

Suggested policy options

1. Create a document that outlines plans for future ILSA participation and addresses key issue, such as funding sources and organizational structures to be put in place to carry out assessment activities. Ensure that the appropriate bodies authorize this document.

2. Ensure regular government funding for ILSA that covers all core activities, including reporting and disseminating assessment results.

3. Develop and implement a plan for widespread dissemination of ILSA results, which includes, for example, a strategy for collaborating with various media outlets to present results to the public, as well as for providing targeted information to key stakeholders such as schools and educators.

4. Introduce funding for research on ILSA results in Yemen which will encourage researchers to contribute to the global knowledge base on assessment.

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 3Rubrics for Yemen

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

YEMEN
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.[2] *	The document is widely available.
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.[3]	There are some system-wide resources for teachers for classroom assessment. *	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. *	There are varied and systematic mechanisms in place to monitor the quality of classroom assessment practices. [7]
ASSESSMENT QUALITY 2: ***Ensuring effective uses of classroom assessment***			
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. [8]	Classroom assessment information is required to be disseminated to all key stakeholders. *
There are no required uses of classroom assessment to support student learning.	There are limited required uses of classroom assessment to support student learning. *	There are adequate required uses of classroom assessment to support student learning, excluding its use as an input for external examination results.	There are adequate required uses of classroom assessment to support student learning, including its use as an input for external examination results. [9]

Classroom Assessment: Development-level rating justifications

1. *General Regulations for Examinations* (authorized by the General Directorate for Examinations and Educational Assessment Ministry of Education, 2001) provides guidance for classroom assessment. This document refers to examinations administered by teachers in the classroom.

2. The availability of the documents is restricted. *The General Regulations for Examinations* document is not made available to the public and is only available at the Ministry of Education and the Educational Offices Archives.

3. There are some system-wide resources for teachers for classroom assessment. Resources include *Teachers' Guides* for primary and secondary school teachers, which outline what students are expected to learn and the level(s) of performance that students are expected to reach in different subject areas at different grade/age levels. In practice, the guides are not always available to all school teachers. The content covered in the *Teachers' Guides* is also addressed through workshops and training sessions, which are accessible by all teachers on a yearly basis. Resources also include textbooks or workbooks that provide support for classroom assessment and scoring criteria and rubrics for students' work. These are applied in some schools by individual teacher effort.

However, other resources, such as item banks or pools with examples of selection/multiple-choice or supply/open-ended questions, online assessment resources, or computer-based testing with instant reports on students' performance are unavailable.

4. There is no official curriculum or standards document which specifies what students at different grade/age levels are expected to learn and to what performance level. However, the *Teachers' Guides* do include this type of information.

5. There are some system-level mechanisms to ensure that teachers develop skills and expertise in classroom assessment. These mechanisms include pre-service and in-service teacher training. All teacher training programs include a required course on classroom assessment and school inspections/teacher supervisions include a component focused on classroom assessment.

However, teachers do not have access to online resources on classroom assessment, opportunities to participate in conferences and workshops, or opportunities to participate in item development for, or scoring of, large-scale assessments or exams, in order to develop their skills and expertise in classroom assessment.

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

7. In order to monitor the quality of classroom assessment activities, classroom assessment is a required component of teacher's performance evaluation and of school inspection or teacher supervision.

There is no external moderation system that reviews the difficulty of classroom assessment activities, appropriateness of scoring criteria, etc., there are no national or other system-wide reviews of the quality of education that include a focus on classroom assessment, and government funding is not available for research on the quality of classroom assessment activities and how to improve classroom assessment.

8. Classroom assessment information from examinations administered by teachers at the classroom level is required to be disseminated to all key stakeholders. Teachers are obligated to report on individual students' performance to school district and Ministry of Education officials, parents and students. Students' results are retained in the school, and provided to the Education Directorate, the Provincial Education office, and the Ministry of Education.

9. There are limited required uses of classroom assessment to support student learning and include providing feedback to students on their learning and grading students for internal classroom uses. Classroom assessment is not required to be used for diagnosing student learning issues, planning next steps in instruction, or for providing input to an external examination program (e.g., school-based assessment with moderation and quality audit).

YEMEN
Examinations

ENABLING CONTEXT

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 1: ***Setting clear policies***			
No standardized examination has taken place.	The standardized examination has been operating on an irregular basis.	The examination is a stable program that has been operating regularly. [1] *	This option does not apply to this dimension
There is no policy document that authorizes the examination.	There is an informal or draft policy document that authorizes the examination.	There is a formal policy document that authorizes the examination. [2] *	This option does not apply to this dimension.
This option does not apply to this dimension.	The policy document is not available to the public. [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. [4] *	The policy document addresses all key aspects of the examination.
ENABLING CONTEXT 2: ***Having strong leadership***			
All stakeholder groups strongly oppose the examination or are indifferent to it.	Most stakeholder groups oppose the examination.	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.[9] *	Funding covers all core examination activities: design, administration, data processing and reporting.	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding does not cover research and development.[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.	This option does not apply to this dimension.	The examination office is accountable to an external board or agency.[12] *	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.	The examination office has-state-of-the-art facilities to carry out the examination.[14] *

(CONTINUED)

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

SYSTEM ALIGNMENT

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
SYSTEM ALIGNMENT 1: *Aligning examinations with learning goals and opportunities to learn*			
It is not clear what the examination measures.	This option does not apply to this dimension.	There is a clear understanding of what the examination measures.[17] *	This option does not apply to this dimension.
What the examination measures is questioned by some stakeholder groups.[18] *	This option does not apply to this dimension.	What is measured by the examination is largely accepted by stakeholder groups.	This option does not apply to this dimension.
Material to prepare for the examination is minimal and it is only accessible to very few students.	There is some material to prepare for the examination that is accessible to some students.[19] *	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.
SYSTEM ALIGNMENT 2: *Providing teachers with opportunities to learn about the examination*			
There are no courses or workshops on examinations available to teachers.	There are no up-to-date courses or workshops on examinations available to teachers.[20] *	There are up-to-date voluntary courses or workshops on examinations available to teachers.	There are up-to-date compulsory courses or workshops on examinations for teachers.
Teachers are excluded from all examination-related tasks.	Teachers are involved in very few examination-related tasks.	Teachers are involved in some examination-related tasks.[21] *	Teachers are involved in most examination-related tasks.

ASSESSMENT QUALITY

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ASSESSMENT QUALITY 1: ***Ensuring quality***			
There is no technical report or other documentation.	There is some documentation on the examination, but it is not in a formal report format. [22] *	There is a comprehensive technical report but with restricted circulation.	There is a comprehensive, high quality technical report available to the general public.
There are no mechanisms in place to ensure the quality of the examination. [23] *	This option does not apply to this dimension.	There are limited systematic mechanisms in place to ensure the quality of the examination.	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. [24] *	Inappropriate behavior surrounding the examination process is moderate.	Inappropriate behavior surrounding the examination process is low.	Inappropriate behavior surrounding the examination process is marginal.
The examination results lack credibility for all stakeholder groups.	The examination results are credible for some stakeholder groups. [25] *	The examination results are credible for all stakeholder groups.	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. [28] *	This option does not apply to this dimension.	Students' results are confidential.	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.	There is a variety of options for students who do not perform well on the examination. [29] *
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 General Secondary Certification Examination has been administered to grade 12 students since 1963. It is used for student certification for grade completion and selection to university or higher education institutions.

The Scientific section and the Literature section of the examination cover the subjects of the Holy Koran, Islamic Education, Arabic Language, English Language, and Mathematics. The Scientific section also covers Chemistry, Physics, and Biology, while the Literature section also covers Geography, History, Philosophy, Logic, and Psychology. The examination is offered in the paper-and-pencil format and includes multiple-choice, supply/open-ended and true/false questions, as well as essays.

2. The formal policy document that authorizes the examination is *Examination Rules, Ministerial Decisions,* authorized by the General Administration for Examinations and Educational Assessment-Ministry of Education, 1965 and 2001 (New edition).

3. The *Examination Rules, Ministerial Decisions* policy document is not available to the public.

4. The *Examination Rules, Ministerial Decisions* policy document covers many aspects of the examination: it outlines governance, distribution of power, responsibilities among key entities; describes the purpose of the examination; describes authorized uses of results; outlines procedures to investigate and address security breaches, cheating or other forms of inappropriate behavior; outlines procedures for special/disadvantaged students; specifies who can sit for the examination; identifies rules about preparation; explains alignment with curricula and standards; explains the format of the examination questions; and determines penalties and sanctions. However, the document does not state funding sources.

5. Policy makers strongly support the General Secondary Certification Examination. Teacher unions, educators, and media also support the General Secondary Certification Examination. In general, students, parents, think tanks, NGOs, universities and employers are neutral to the General Secondary Certification Examination.

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

7. Efforts to improve the examination are generally welcomed by the leadership in charge of the examination.

8. There is regular funding allocated for the examination, allocated by the government. In practice, the amount tends to be scarce and insufficient.

9. Funding covers examination design, examination administration and staff training. However, funding does not cover data analysis, data reporting, or long- or medium-term planning of program milestones.

10. Funding does not cover research and development activities.

11. The Higher Committee for Examinations works with the General Administration for Examinations within the Ministry of Education to conduct the examination. The General Administration for Examinations has had primary responsibility for running the examinations since 1963.

12. The General Administration for Examinations and the Higher Committee for Examinations are accountable to the Ministry of Education.

13. The examination results are officially recognized in Yemen and by more than one certification and selection system abroad, including all Arab states and the USA.

14. The examination office has state-of-the-art facilities to carry out the examination, including computers for all technical staff, a secure building, secure storage facilities, access to adequate computer servers, the ability to backup data, and adequate communication (telephone, email, and internet) tools.

15. There is an adequate number of permanent full time staff to run the examination. A sufficient number of full time employees are always available to manage and organize the examinations and all related issues and tasks. The Higher Committee for Examinations and its various sub-committees manage the examinations (from the preparation phase to the announcement of results) but can outsource work to teachers, supervisors and others, who are hired on a temporary or part-time basis.

There has been some evidence of poor training of test administrators or unclear instructions and guidelines in administering the examination, errors in scoring that have led to delays in reporting results, and frequent errors in data processing. However, there have been no issues reported with regard to delays in administering the examination due to issues with the design of the examination questions, weaknesses in test design, omission of curricular topics, or frequent errors in the examination questions.

16. Yemen offers university graduate programs (for masters or doctorate level) as well as graduate and non graduate courses specifically focused on educational measurement and evaluation. Yemen also offers non university training courses or workshops on educational measurement and evaluation. However, Yemen does not provide funding for attending international programs, courses, or workshops on educational measurement and evaluation, or internships in the examination office.

17. The examination measures objectives presented in the *Teachers' Guides.* .

18. While what is measured by the examination is largely accepted by stakeholders, some stakeholder groups question what the examination measures.

19. Materials needed to prepare for the examination are accessible by most students, but certain student sub-groups may have greater access than others, particularly those with access to electronically-available resources. Samples from previous examinations and examples of the types of questions that are on the examination are publicly available; however this material is mostly available to students in urban areas. Information on how to prepare for the examination, the

framework document explaining what is measured on the examination, and reports on the strengths and weaknesses in student performance are not publically available.

20. The only courses or workshops on the examinations available to teachers are compulsory, but are not regularly updated.

21. Teachers are involved in administering the examination, scoring the examination, and supervising examination procedures. On rare occasions, teachers are asked to participate in the selection and creation of examination questions. Teachers are not involved in selecting or creating examination scoring guides, acting as a judge (i.e. in orals), or resolving inconsistencies between examination scores and school grades (i.e. moderation.)

22. There is some documentation about the technical aspects of the examination but it is not in a formal report format.

23. While there are no mechanisms in place to ensure the quality of the examination, there are efforts to improve aspects of examination quality through: a) internal reviews of the scoring for each question and b) a new pilot mechanism to review the implementation potential. There is no external review or observers, external certification or audit.

24. Inappropriate behavior surrounding the examination process is high. Inappropriate behaviors that diminish the credibility of the examination which typically occur during the examination process include leakage of the content of an examination and impersonation (when an individual other than the registered candidate takes the examination). When either of these behaviors occurs, all results are canceled and the individuals involved are turned over to the legal authorities.

Other inappropriate behaviors include copying from other candidates, using unauthorized materials such as prepared answers and notes, and collusion among candidates via mobile phones, passing of papers or the equivalent. When these students are caught in such activities, their results in the given subject are canceled.

Further inappropriate behaviors include: intimidation of examination supervisors, markers or officials which results in the student being deprived of the results of all subjects for the entire academic year; issuing of forged certificates or altering results information which results in expulsion and the prohibition of the candidate to receive an examination score; and provision of external assistance via the supervisor, mobile phone, etc. which results in student results in the given subject being canceled and expulsion of the supervisor.

25. The results are perceived as credible by some stakeholder groups as examination results tend to be more credible in urban areas than in rural areas.

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

27. There is no systematic evidence on improper use of the results.

28. Once results are officially announced, students' names and results are public and accessible to all.

29. There are a variety of options for students who do not perform well on the examination. Students may retake the examination the following year, attend remedial or preparatory courses in order to prepare to retake the examination, opt for less selective schools/universities/tracks, or repeat the grade. If the student does not pass the general examination for two consecutive years, he/she may not retake the examination. If a student fails in the science section of the examination for two consecutive years, he/she is no longer permitted to pursue scientific studies. At no point must the student leave the education system.

30. There are some mechanisms in place to monitor the consequences of the examination. There is a permanent oversight committee within the Ministry's examination departments to monitor the consequences of the examination. However, there is no funding for independent research on the impact of the examination, a permanent oversight committee, studies (e.g., predictive validity) that are updated regularly, regular focus groups or surveys of key stakeholders, or expert review groups.

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

ENABLING CONTEXT

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 1: *Setting clear policies for NLSA*			
No NLSA exercise has taken place.[1] *	The NLSA has been operating on an irregular basis.	The NLSA is a stable program that has been operating regularly.	This option does not apply to this dimension.
There is no policy document pertaining to NLSA. *	There is an informal or draft policy document that authorizes the NLSA.	There is a formal policy document that authorizes the NLSA.	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.	This option does not apply to this dimension.
There is no plan for NLSA activity.	This option does not apply to this dimension.	There is a general understanding that the NLSA will take place. *	There is a written NLSA plan for the coming years.
ENABLING CONTEXT 2: *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.	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.	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.	This option does not apply to this dimension.
This option does not apply to this dimension.	Funding does not cover research and development activities.	This option does not apply to this dimension.	Funding covers research and development activities.
ENABLING CONTEXT 4: *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.	This option does not apply to this dimension.
This option does not apply to this dimension.	Political considerations regularly hamper technical considerations.	Political considerations sometimes hamper technical considerations.	Political considerations never hamper technical considerations.
This option does not apply to this dimension.	The NLSA office is not accountable to a clearly recognized body.	The NLSA office is accountable to a clearly recognized body.	This option does not apply to this dimension.

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 5: *Having effective human resources for NLSA*			
There is no staff allocated for running an NLSA.	The NLSA office is inadequately staffed to effectively carry out the assessment.[12]	The NLSA office is adequately staffed to carry out the NLSA effectively, with minimal issues.	The NLSA office is adequately staffed to carry out the NLSA effectively, with no issues.
The country does not offer opportunities that prepare individuals for work on NLSA.	This option does not apply to this dimension.	The country offers some opportunities to prepare individuals for work on the NLSA.	The country offers a wide range of opportunities to prepare individuals for work on the NLSA.

SYSTEM ALIGNMENT

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
SYSTEM ALIGNMENT 1: ***Aligning the NLSA with learning goals***			
It is not clear if the NLSA is based on curriculum or learning standards.	This option does not apply to this dimension.	The NLSA measures performance against curriculum or learning standards.	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.
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.	This option does not apply to this dimension.
SYSTEM ALIGNMENT 2: ***Providing teachers with opportunities to learn about the NLSA***			
There are no courses or workshops on the NLSA.	There are occasional courses or workshops on the NLSA.	There are some courses or workshops on the NLSA offered on a regular basis.	There are widely available high quality courses or workshops on the NLSA offered on a regular basis.

ASSESSMENT QUALITY

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ASSESSMENT QUALITY 1: ***Ensuring the quality of the NLSA***			
No options are offered to include all groups of students in the NLSA.	This option does not apply to this dimension.	At least one option is offered to include all groups of students in the NLSA.	Different options are offered to include all groups of students in the NLSA.
There are no mechanisms in place to ensure the quality of the NLSA.	This option does not apply to this dimension.	There are some mechanisms in place to ensure the quality of the NLSA.	There are a variety of mechanisms in place to ensure the quality of the NLSA.
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.
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.	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.
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.	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. National Assessment System of Student (NASS) is a large-scale assessment program that is currently being developed. NASS is intended to assess grade 4 and 7 students in the subjects of Science, Mathematics, and Arabic Language.

Yemen has never carried out a national large-scale assessment and has no policies or plan for carrying out one in the future. This set of facts provides us with sufficient information to determine the development level of NLSA in Yemen, and the remainder of the NLSA rubric is intentionally left blank as a result.

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

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 3: *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. [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: ***Providing opportunities to learn about ILSA***			
The ILSA team has not attended international workshops or meetings.	The ILSA team attended some international workshops or meetings. [11] *	The ILSA team attended all international workshops or meetings.	This option does not apply to this dimension.
The country/system offers no opportunities to learn about ILSA.	This option does not apply to this dimension.	The country/system offers some opportunities to learn about ILSA. [12] *	The country/system offers a wide range of opportunities to learn about ILSA.
This option does not apply to this dimension.	This option does not apply to this dimension.	Opportunities to learn about ILSA are available to the country's/system's ILSA team members only.	Opportunities to learn about ILSA are available to a wide audience, in addition to the country's/system's ILSA team members. [13] *

ASSESSMENT QUALITY

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ASSESSMENT QUALITY 1: ***Ensuring the quality of ILSA***			
Data from the ILSA has not been published.	The country/system met sufficient standards to have its data presented beneath the main display of the international report or in an annex.	The country/system met all technical standards required to have its data presented in the main displays of the international report. [14] *	This option does not apply to this dimension.
The country/system has not contributed new knowledge on ILSA. [15] *	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.
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. [16] *	Country/system-specific results and information are regularly disseminated in the country/system.	Country/system-specific results and information are regularly and widely disseminated in the country/system.
Products to provide feedback to schools and educators about the ILSA results are not made available. [17] *	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.
There is no media coverage of the ILSA results. [18] *	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.
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. [19] *	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.
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. Yemen has participated in TIMSS in 2003, 2007 and 2011. The rubric is completed with regard to Yemen's participation in TIMSS 2007 (results from TIMSS 2011 have not yet been released.)

2. There is no official document or plan for future participation in an ILSA.

3. Yemen does not have a policy document that addresses participation in international assessments, however, there is an introductory document available to all stakeholders—*Introduction to: The Trends in International Mathematics and Science (2009).*

4. The policy document does not exist therefore this option does not apply to this dimension.

5. There is funding for international assessment participation sourced from loans and external donors. Funding for the participation fees and international travel was provided for by the World Bank. Funding for the domestic implementation fees was provided by the Basic Education Development Project (BEDP) which is supported by donors, including the World Bank.

6. Funding covers international participation fees, implementation of the assessment exercise in the country, attendance at international expert meetings for the assessment exercise, external training sessions and attendance at international education conferences. However, funding does not cover processing and analyzing data collected from implementation of the assessment exercise or reporting and disseminating the assessment results in Yemen.

7. Funding does not cover research and development activities.

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

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

10. The national coordinator is fluent in the language in which the international-level meetings are conducted and related documentation is available. Additionally, the ILSA team is sufficiently staffed, has previous experience working on international assessments, and has the necessary training and experience to carry out the required assessment activities effectively. Team members have also attended some of the international meetings related to the assessment. The national coordinator is fluent in the language in which the international-level meetings are conducted and related documentation is available.

However, there have been some issues with the carrying out of the international assessment in the country. There were errors or delays in the printing and layout of the test booklets and in the administration of the test, as well as complaints about poor training of test administrators. There have been no issues reported with translation of the assessment instruments (e.g., test booklets, background questionnaires) or errors or delays in scoring student responses to questions.

11. ILSA team members have attended some of the international meetings and workshops related to the assessment.

12. Opportunities to learn about international assessments have been funded by the World Bank and included workshops on using international assessment databases and international workshops and trainings on international assessments. There are no university or online courses on the topic of international assessments.

13. Opportunities to learn about the ILSA are available to a wide audience. Individuals working directly on the specific international assessment exercise including those who participated in the program implementation such as supervisors and researchers from the Center for Educational Development and Research as well as curricula specialists are able to attend learning opportunities. In addition, MA students studying assessment from the faculties of education in five universities, Sana'a Univ., Ibb Univ., Taiz Univ., Hodeidah Univ., and Hadramout Univ, were trained. However, professionals or university staff interested in assessment have not benefited from such learning opportunities.

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

15. Yemen has not contributed new knowledge on ILSA.

16. ILSA country specific results and information from the most recent international assessment are disseminated irregularly in the country.

17. Products to provide feedback to schools and educators about the ILSA are not available.

18. International assessment results have not been covered by the media.

19. The results of the international assessment exercise (TIMSS 2007) have been used by policy makers or education leaders to improve education quality in Yemen by elaborating the Basic Classes Initiative program. Results have not been used to track the impact of reforms on student achievement levels, inform curriculum improvement, inform teacher training programs, inform other assessment activities in the system (e.g., classroom assessment, examinations) or inform resource allocation.

20. There is no evidence of a positive impact on student achievement from decisions based on ILSA results.

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 Tomomi Miyajima, Education Specialist, and Lianqin Wang, Senior Education Specialist and Task Team Leader for education projects in Yemen in the World Bank's Education Global Practice, as well as comments received during a national validation workshop held in Yemen.

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. 2013. *Yemen Education Sector Plan: Midterm Results Framework 2013-2015.* Sana'a: Republic of Yemen.

Republic of Yemen and the World Bank. 2010. *Republic of Yemen Education Status Report: Challenges and Opportunities.* Washington, D.C.: World Bank.

United Nations Educational, Scientific, and Cultural Organization (UNESCO)- Institute for Statistics. Education Profile- Yemen. Montreal, QC: UNESCO. Data retrieved from http://stats.uis.unesco.org/unesco on October 10, 2013.

World Bank. 2013. Yemen Country Indicator Data. Washington, DC: World Bank. Data retrieved from http://databank.worldbank.org/data on October 9, 2013.

----------. -----. "Yemen Overview." Washington, DC: World Bank. Data retrieved from http://www.worldbank.org/en/country/yemen/overview on October 9, 2013.

www.worldbank.org/education/saber

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.