# Oman

## STUDENT ASSESSMENT

SABER Country Report

2013

## Key Policy Areas for Student Assessment

**Status**

### 1. Classroom Assessment

Emerging ●●○○

In Oman, formal, system-level documents provide guidelines for classroom assessment. A variety of resources, such as guides describing expected learning outcomes for different subject areas and grades, are available to teachers system-wide to support them in carrying out their assessment activities. Classroom assessment activities tend to be aligned with the curriculum framework. However, they are generally considered to be weak due, for example, to the fact that the assessments tend to provide little useful feedback to students and are mainly used as administrative tools rather than pedagogical inputs.

### 2. Examinations

Established ●●●○

The General Education Diploma Examination has been administered to grade 12 students since 1980. The examination results are used to certify student completion of secondary school and determine student selection to higher education institutions. Examination results are officially recognized by certification and selection systems in Oman and abroad. Regular funding is provided by the government and covers all core examination activities, including research and development. However, there are limited systematic mechanisms in place to ensure the quality of the examination and no up-to-date courses or workshops on examinations are available to teachers.

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

Established ●●●○

The Cognitive Development Test was administered for the first time in 2007-2008. Since then, it has been administered every year, assessing all students in grade 7 in Mathematics, Science, and Environmental Geographic concepts. A written plan for future NLSA activities, specifying who will be tested and in which subject areas, is in place. Regular funding for the NLSA is allocated by the government and covers core activities. While there are some opportunities to prepare individuals to work on the NLSA, there are limited opportunities for teachers to learn about the assessment.

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

Established ●●●○

In the last 10 years, Oman has participated in TIMSS 2007 and 2011, and PIRLS 2011. It also has taken concrete steps to participate in TIMSS 2015 and PIRLS 2016. Regular funding for ILSAs is provided by the government and covers all core activities. Oman offers a range of opportunities to learn about ILSAs that are available to ILSA team members, teachers, and specialists. These opportunities include workshops, online courses, and funding for attending international training. Oman-specific ILSA results are widely disseminated. However, it is not clear that decisions based on the results have had a positive impact on student achievement levels.

Systems Approach for Better Education Results

WORLD BANK GROUP

## Introduction

Oman 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, Oman 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 stakehol ders 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 Oman

Oman is a high-income country in the Middle East & North Africa region. GDP per capita is $23,133 and GDP growth was 5.6 percent in 2010 and 0.3 percent in 2011.

Oman has made major progress in primary and secondary enrolments, expanding considerably from 1970, when there were only 3 schools in the country and a total of 909 students, to 600,000 students enrolled in 1,300 schools in the country in 2008. In 2011, net enrollment rates were 97 percent and 94 percent at the primary and secondary levels respectively.

The Government of Oman has focused on improving the quality, efficiency, and relevance of education. In addition, due to concerns that the Omani economy is largely reliant on oil and gas revenues, the government has prioritized education reform in order to facilitate the development of other sectors. Oman's 'Vision 2020', the country's long-term strategy for growth, called for several reforms in the education sector including strengthening the teaching and learning of English in basic education, greater attention to science subjects, and greater alignment of secondary education with labor market needs. Other reforms already introduced in the country include the implementation of continuous assessment, a reduction in class size, and a school-based management process.

Recent figures on spending in the education sector reflect the country's commitment to its growth. In 2005 and 2007, total public expenditure on education represented 4.6 percent of GDP, which further increased to 5.3 percent of GDP in 2009.

Detailed information was collected on Oman'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 Oman, 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 Oman's immediate interests and needs. Detailed, completed rubrics for each assessment type are provided in Appendix 5.

## Classroom Assessment

### Level of development

**Emerging**

In Oman, formal, system-level documents, such as *The General Document for Learning Assessment of Students in Grades 1 to 12*, authorized in 2012 by the General Directorate of Educational Assessment in the Ministry of Education, provide guidelines for classroom assessment. Various resources are also available to teachers throughout the system to help them engage in classroom assessment activities. These include documents that outline what students are expected to learn (and how well) in different subject areas at different grade levels. In addition, the *Scope and Sequence Matrix* document specifies the objectives that the students must achieve at a particular grade level in order to pass to the next grade. There are also textbooks or workbooks that provide support for classroom assessment, scoring criteria for grading students' work, item banks with examples of selection/multiple-choice and supply/open-ended questions, and online assessment resources.

Some mechanisms are in place to ensure that teachers develop skills and expertise in classroom assessment. For example, school inspection and teacher supervision include a component focused on classroom assessment. While there are no pre-service teacher training opportunities, teachers and supervisors are trained each year on implementing classroom assessment within planned training programs as part of professional development programs, and on-going follow-up is conducted by regional supervisors from each regional educational directorate.

Classroom assessment activities take a variety of forms. For grades 1 to 4, activities include oral tests, quizzes, and observation. For grades 5 to 12, activities include oral tests, quizzes, observation, periodic reports, and quarterly tests. Although it is uncommon for classroom assessment activities to rely on multiple-choice or selection-type questions, they do tend to be about recalling information. In addition, it is commonly believed that classroom assessment activities provide little useful feedback to students and are mainly used as administrative tools rather than as pedagogical inputs.

Systematic mechanisms are in place to monitor the quality of classroom assessment practices. For example, classroom assessment is a required component of a teacher's performance evaluation, and system-wide reviews of the quality of education include a focus on classroom assessment. Testing and auditing teams, referred to as moderation teams, from regional educational directorates visit schools to verify the accuracy of classroom assessments conducted in grade 12 classrooms. In addition, there is an external moderation system that reviews the difficulty of classroom assessment activities and the appropriateness of scoring criteria.

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

### Suggested policy options

1. Continue efforts to improve the *Scope and Sequence Matrix*, which specifies the objectives that students must achieve at each grade level in order to pass to the next grade, to clarify the level of performance that students must demonstrate for various learning objectives.

2. In order to improve the quality of classroom assessment practices, review existing in-service teacher training opportunities and adjust them so that they better support teachers in building knowledge and capacity on the use of classroom assessment. Introduce policies for instituting yearly in-service teacher training opportunities on classroom assessment and make them available to all teachers on a regular basis.

3. Improve the availability of systematic mechanisms to monitor the quality of classroom assessment practices. Make funding available for research on the quality of classroom assessment and how to improve it.

# Examinations

## Level of development

**Established**

The General Education Diploma Examination (previously referred to as the General Secondary Certification Examination) was first administered in 1980 with the main purposes of student certification for school cycle completion and student selection to higher-education institutions. Students in grade 12 are assessed in core and elective subjects from among the following: Islamic Education, Arabic Language, English Language, Physics, Chemistry, Biology, Pure Mathematics, Applied Mathematics, Social Studies (History), Geography, Science and the Environment, Musical Skills, Physical Education, and Fine Arts. A formal, system-level document, the *Regulations of the Committees for General Diplomas Examinations,* authorized the examination in 2010.

Regular funding for the examination is allocated by the government, and covers all core examination activities as well as research and development.

The Department of Tests and Examinations Management, which is part of the Directorate General for Educational Assessment within the Ministry of Education, is responsible for running the General Education Diploma Examination. Although the Department of Tests and Examinations Management has all of the required facilities to carry out the examination, it does not have enough full-time staff to meet the needs of the examination. While specialized senior supervisors and teachers from directorates, such as the General Directorate for Curriculum Development and the Directorate General for Human Resources Development, are utilized for exam preparation, the examination is managed and administered by specialized committees.

Oman offers a wide range of opportunities to prepare individuals for work on the examination. Options in place include university graduate programs at Sultan Qaboos University and non graduate training courses on

educational measurement and evaluation. In addition, there is funding for attending international programs and workshops on educational measurement and evaluation, as well as internships in the examination office.

Compulsory courses and workshops on the examination are available to teachers, but are not regularly updated. Teachers are involved in tasks related to the General Education Diploma Examination, including administering and scoring the examination, acting as judges, and supervising examination procedures.

Internal and external reviews are in place to systematically ensure the quality of the examination. For example, a review of the examination by a committee external to the Ministry of Education is currently underway. The examination questions are also reviewed by specialized technical committees and then presented to the assessment committee (composed of experienced specialists) prior to the administration of the examination.

There are some options for students who do not perform well on the examination, including retaking the examination (if they fail in three or fewer subjects), repeating the grade, or attending remedial or preparatory courses (offered in the private sector) before they retake the examination.

A permanent oversight committee is the only mechanism in place to monitor the consequences of the examination.

### Suggested policy options

1. Introduce more mechanisms to ensure the positive impact of the examination on student learning, particularly in terms of the systematic use of examination results by teachers, school level leaders, and regional subject supervisors. For example, conduct regular focus groups or surveys of key stakeholders to discuss exam results and how they can be used to further student learning.

2. Provide teachers with regular opportunities, such as up-to-date courses or workshops, to learn about the examination.

3. Increase the number of qualified staff dedicated to carrying out the General Education Diploma Examination in the Department of Tests and Examinations Management.

## National Large-Scale Assessment (NLSA)

### Level of development

**Established**

The Cognitive Development Tests have been administered on an annual basis since the 2007-2008 school year as a way to monitor education quality at the system level, support schools and teachers, and inform policy design, evaluation, and decision making. The assessment is administered to all students in grade 7 and covers Mathematics, Sciences, and Environmental Geographic concepts. A formal, system-level document, the *Cognitive Development Manual,* authorized the Cognitive Development Tests in 2009. In addition, there is a written plan for the NLSA for the coming years, specifying who will be tested and in which subject areas.

Regular funding for the Cognitive Development Tests is allocated by the government. Funding covers all core NLSA activities, including assessment design and administration, data analysis and reporting, and staff training. Funding also covers annual prizes for students who exhibit outstanding performance on the assessment. However, funding does not cover long- or medium-term planning of program milestones or research and development activities.

The Directorate General of Educational Evaluation is a permanent unit responsible for running the assessment and is adequately staffed to carry out the NLSA effectively. At the same time, there have been some issues related to poor training of test administrators, and unclear instructions and guidelines for administering the assessment.

Oman offers some opportunities to prepare individuals for work on the NLSA. These include university graduate programs on educational measurement and evaluation at the Sultan Qaboos University, internships or short-term employment in the large-scale assessment office, and funding for attending international programs or workshops on educational measurement and evaluation.

In addition, regular internal and external reviews by qualified experts are in place to ensure the alignment between the assessment instrument and what it is supposed to measure. Other mechanisms to ensure the quality of the NLSA include the training of all test administrators according to a protocol, and the preparation of a standardized manual for administrators.

NLSA results are disseminated within 12 months of the assessment being administered. Reports are made available for all stakeholder groups and the results are published on the Ministry's Education Portal. Workshops and presentations are also used to share results with key stakeholders. However, the results are not featured in the media—newspapers, magazines, radio, or television—and there is no media briefing.

There are some mechanisms in place to help monitor the impact of the NLSA in terms of its effects on policy dialog, design, and decision making. These include a permanent oversight committee, themed conferences that provide a forum to discuss research and other data on the assessment, and expert review groups. However, funding for independent research on the impact of the NLSA, or regular focus groups and surveys of key stakeholders are not in place.

### Suggested policy options

1. Review the intended purposes and uses of the NLSA and evaluate options to ensure that those purposes and uses are achieved, including the optimal frequency of administering the NLSA, the subject matter to be assessed, and the desired technical characteristics and quality of the NLSA.

2. Utilize the funding that is made available by the Research Council (a public body that provides grants and other support for research and innovation activities across sectors) for research and development on the NLSA.

3. Introduce a variety of mechanisms to ensure the technical quality of the NLSA. For example, introduce double scoring of data or external certifications or audits.

# International Large-Scale Assessment (ILSA)

## Level of development

**Established**
●●●○

Oman participated in Trends in International Mathematics and Science Studies (TIMSS) in 2007 and 2011, and Progress in International Reading Literacy Study (PIRLS) in 2011. Oman has taken concrete steps to participate in PIRLS in 2016 and TIMSS in 2015. A formal policy document, *Ministerial Decision Number 140/2009*, authorized by the Ministry of Education in 2009, addresses participation in ILSAs and the preparation and publication of relevant national reports.

Regular funding for international assessment participation is allocated at the discretion of the government. Although funding covers all core activities, it does not cover research and development.

A national coordinator and team (called the International Studies Program) carry out the ILSA activities. The ILSA office is adequately staffed to carry out the work. At the same time, although the team has previous experience working on international assessments, and has attended some international workshops and meetings, it does not have all of the necessary training and experience to carry out required assessment activities effectively.

Oman offers a wide range of opportunities to learn about ILSAs. For example, workshops and meetings are held on using international assessment databases, funding is made available for attending international workshops or training on international assessments, and online courses are available. However, there are no university courses on the topic of international assessments.

Oman has met all technical standards required to have its data presented in the main displays of the international report. In addition, country-specific results are widely disseminated in the country. For example, after previous ILSA exercises, a national report as well as brochures and PowerPoint presentations with Oman's results were made available online, copies of the

national and international reports were distributed to key stakeholders, and the country's results were communicated through a press release. Products providing information to schools and educators were also made available. In addition, after the release of the country's results for TIMSS and PIRLS 2011, a special seminar was held to discuss the results, which included participation by stakeholders from the Sultanate.

ILSA results have been used to inform decision making in Oman. For example, results have been used to track the impact of reforms on student achievement levels and to inform curriculum improvement, teacher training programs, and other assessment activities in the system. However, it is not clear that decisions based on ILSA results have had a positive impact on student achievement levels in the system, as there was no improvement in Oman's scores between TIMSS 2007 and 2011.

### Suggested policy options

1. Ensure that Oman's ILSA results are used to inform decision making around school- and classroom-level instructional support, assessment practices, and curricular review.

2. Strengthen technical expertise on ILSAs by collaborating with universities to introduce university courses on educational measurement and evaluation.

## 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 Oman

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

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

SABER COUNTRY REPORT |2013

## 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. [6] * | Classroom assessment practices are known to be of moderate quality. | Classroom assessment practices are known to be generally of high quality. |
| There are no mechanisms to monitor the quality of classroom assessment practices. | There are ad hoc mechanisms to monitor the quality of classroom assessment practices. | There are limited systematic mechanisms to monitor the quality of classroom assessment practices. | There are varied and systematic mechanisms in place to monitor the quality of classroom assessment practices. [7] * |
| **ASSESSMENT QUALITY 2:** *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. | 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. |

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

*Classroom Assessment*: Development-level rating justifications

1. In Oman, formal system-level documents such as *The General Document for Learning Assessment of Students in Grades 1 to 12*, authorized by the General Directorate of Educational Assessment in the Ministry of Education in 2012, provide guidelines for classroom assessment.

2. *The General Document for Learning Assessment of Students in Grades 1 to 12* is widely available, including at the Sultanate of Oman Educational Portal, in schools, teacher training colleges, and in the educational assessment departments in the educational governorates, Directorate General of Curricula Development, Directorate General of Human Resources Development, and Directorate General of Private Schools.

3. There are a variety of system-wide resources available for teachers to engage in classroom assessment activities. Specifically, there are documents that outline what students are expected to learn in different subject areas at different grade/age levels and the level(s) of performance that students are expected to reach in different subject areas at different grade/age levels. In addition, the *Scope and Sequence Matrix* document specifies the objectives for each subject that the students must achieve in each grade in order to pass to the next. There are also textbooks or workbooks that provide support for classroom assessment, scoring criteria or rubrics for students' work, item banks or pools with examples of selection/multiple-choice or supply/open-ended questions, and online assessment resources available. However, there is no computer-based testing with instant reports on students' performance.

4. An official curriculum document, *Scope and Sequence Matrix* document for each education subject, specifies what students are expected to learn, but the level of performance required is not clear.

5. There are some system-level mechanisms in place to ensure that teachers develop skills and expertise in classroom assessment. For example, school inspection or teacher supervision includes a component focused on classroom assessment. Although there are no pre-service teacher training opportunities, as part of professional development programs, teachers and supervisors are trained each year on implementing classroom assessment within planned training programs, and ongoing follow-up is conducted by regional supervisors from each regional educational directorate. In addition, online resources on classroom assessment are made available, as well as opportunities to participate in conferences and workshops. Opportunities to participate in item development for, or scoring of, large-scale assessments or exams are also available.

6. While classroom assessment practices are generally considered to be weak, some classroom assessment activities are carried out in several forms. For grades 1 to 4, activities include oral tests, quizzes, and observation. For grades 5 to 12, activities include oral tests, quizzes, observation, and periodic reports as well as quarterly tests. Although it is uncommon for classroom assessment activities to rely mainly on multiple-choice, selection-type questions, it is very common that they are mainly about recalling information. It is common for teachers to not use explicit or a priori criteria for scoring or grading students' work, and the uneven application of standards for grading students' work and errors in the scoring or grading of students' work are serious problems. It is rare for parents to be poorly informed about students' grades or for classroom assessment activities to be misaligned with pedagogical or curricular work; yet classroom assessment activities commonly provide little useful feedback to students and are mainly used as an administrative or control tool rather than as a pedagogical resource.

7. There are varied and systematic mechanisms in place to monitor the quality of classroom assessment practices. Classroom assessment is a required component of a teacher's performance evaluation and school inspection or teacher supervision. In addition, national or other system-wide reviews of the quality of education include a focus on classroom assessment. There are also testing and auditing teams which visit schools to verify the accuracy of classroom assessments. However, there is no external moderation system that reviews the difficulty of classroom assessment activities or appropriateness of scoring criteria. There is also no government funding available for research on the quality of assessment activities or how to improve classroom assessment.

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

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

# OMAN
## *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

| LATENT | EMERGING | ESTABLISHED | ADVANCED |
|---|---|---|---|
| No standardized examination has taken place. | The standardized examination has been operating on an irregular basis. | The examination is a stable program that has been operating regularly.[1] * | This option does not apply to this dimension |
| There is no policy document that authorizes the examination. | There is an informal or draft policy document that authorizes the examination. | There is a formal policy document that authorizes the examination.[2] * | This option does not apply to this dimension. |
| This option does not apply to this dimension. | The policy document is not available to the public. | The policy document is available to the public.[3] * | This option does not apply to this dimension. |
| This option does not apply to this dimension. | This option does not apply to this dimension. | The policy document addresses some key aspects of the examination.[4] * | The policy document addresses all key aspects of the examination. |

### ENABLING CONTEXT 2: Having strong leadership

| LATENT | EMERGING | ESTABLISHED | ADVANCED |
|---|---|---|---|
| All stakeholder groups strongly oppose the examination or are indifferent to it. | Most stakeholder groups oppose the examination. | Most stakeholders groups support the examination.[5] * | All stakeholder groups support the examination. |
| There are no attempts to improve the examination by stakeholder groups. | This option does not apply to this dimension. | There are independent attempts to improve the examination by stakeholder groups.[6] * | There are coordinated attempts to improve the examination by stakeholder groups. |
| 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. |

OMAN | SABER-STUDENT ASSESSMENT

(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. | This option does not apply to this dimension.[10] * | 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)

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

SABER COUNTRY REPORT |2013

| 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. | 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* | | | |
| It is not clear what the examination measures. | This option does not apply to this dimension. | There is a clear understanding of what the examination measures.[17] * | This option does not apply to this dimension. |
| What the examination measures is questioned by some stakeholder groups. | This option does not apply to this dimension. | What is measured by the examination is largely accepted by stakeholder groups.[18] * | This option does not apply to this dimension. |
| Material to prepare for the examination is minimal and it is only accessible to very few students. | There is some material to prepare for the examination that is accessible to some students. | There is comprehensive material to prepare for the examination that is accessible to most students. | There is comprehensive material to prepare for the examination that is accessible to all students.[19] * |
| **SYSTEM ALIGNMENT 2:** *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. | There is a comprehensive technical report but with restricted circulation. | There is a comprehensive, high quality technical report available to the general public.[22] * |
| 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.[24] * | Inappropriate behavior surrounding the examination process is marginal. |
| The examination results lack credibility for all stakeholder groups. | The examination results are credible for some stakeholder groups. | The examination results are credible for all stakeholder groups.[25] * | This option does not apply to this dimension. |
| The majority of the students (over 50 percent) may not take the examination because of language, gender, or other equivalent barriers. | A significant proportion of students (10 percent-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. | 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 Education Diploma Examination was first administered in 1980 with the main purposes of student certification for grade or school cycle completion and student selection to university or other higher education institutions. Students in grade 12 are assessed in Islamic Education, Arabic Language, English Language, Physics, Chemistry, Biology, Pure Mathematics, Applied Mathematics, Social Studies (History), Geography, Science and the Environment, Musical Skills, Physical Education, and Fine Arts.

2. The Ministry of Education, through the *Regulations of the Committees for General Diplomas Examinations* document, authorized the examination in 2010.

3. *The Regulations of the Committees for General Diplomas Examinations* document is available to the public.

4. *The Regulations of the Committees for General Diplomas Examinations* document addresses many of the key aspects of the examination. For example, it outlines governance, distribution of power, and responsibilities among key entities, describes the purpose of the examination, states funding sources, outlines procedures to investigate and address security breaches, cheating, or other forms of inappropriate behavior, specifies who can sit for the examination, identifies rules about preparation, and explains the format of the examination questions. However, it does not describe the authorized uses of results, outline procedures for special/disadvantaged students, or explain alignment with curricula and standards.

5. Most stakeholder groups support the examination. Policy makers, universities, and employers all strongly support the examination, while educators, parents, and media support it. Although it is unclear whether teacher unions support or oppose the examination, students, think tanks, and NGOs are neutral to it.

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

7. The leadership in charge of the examination generally welcomes efforts to improve it.

8. Regular funding is allocated for the examination by the government.

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

10. Funding for the General Education Diploma Examination covers research and development.

11. The Department of Tests and Examinations Management is responsible for running the General Education Diploma Examination. The Department of Tests and Examinations Management is part of the Directorate General for Educational Assessment, a division of the Ministry of Education.

12. The Department of Tests and Examinations Management is not accountable to an external board or agency.

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

13. Examination results are officially recognized by certification or selection systems in Oman, as well as in countries worldwide.

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

15. The Department of Tests and Examinations Management has permanent and full-time staff; however, the number of staff available is insufficient to meet the needs of the examination. While specialized senior supervisors and teachers from directorates such as the General Directorate for Curriculum Development and the Directorate General for Human Resources Development are utilized for exam preparation, execution and management of the examination is carried out by specialized committees. Although poor training of test administrators and unclear guidelines for administering the examination have been raised as issues with the performance of the human resources responsible for the examination, issues such as frequent errors in the examination questions and data processing have not been identified.

16. Oman offers a wide range of opportunities that prepare for work on the examination. University graduate programs at Sultan Qaboos University and non graduate training courses on educational measurement and evaluation are in place. In addition, funding for attending international programs, courses, or workshops on educational measurement and evaluation, as well as internships in the examination office, are available.

17. There is a clear understanding that the General Education Diploma Examination measures the national school curriculum guidelines or standards.

18. Stakeholder groups largely accept what is measured by the examination.

19. Comprehensive material to prepare for the examination is available and accessible to all students. Examples of the types of questions that are on the examination are accessible online at the Ministry of Education website and distributed to students during the school year. Additionally, 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 all publicly available.

20. Although there are compulsory courses or workshops on the examination available to teachers, these courses are not regularly updated.

21. Teachers are involved in some tasks related to the General Education Diploma Examination. Specifically, teachers are involved in administering and scoring the examination, acting as a judge, and supervising examination procedures. However, they are not involved in selecting or creating examination questions or scoring guides, or resolving inconsistencies between examination scores and school grades.

22. A comprehensive, high quality technical report is available to the general public.

23. Internal and external review or observers are in place to ensure the quality of the examination. Prior to administration, the General Education Diploma Examination questions are reviewed by specialized technical committees and presented to the assessment committee, which is composed of experienced

specialists. In addition, an assessment of the examination by committees external to the Ministry of Education is underway. However, there are no external certifications or audits, pilots or field testing, or translation verification.

24. Inappropriate behavior surrounding the examination process is low. While there have been occurrences of impersonation (when an individual other than the registered candidate takes the examination), copying from other candidates, using unauthorized materials such as prepared answers and notes, and collusion among candidates via mobile phones or passing of paper, certain mechanisms have been put in place to address such behaviors. For example, disciplinary measures are taken against impersonators, and their exam results are cancelled. Candidates are forbidden from bringing cell phones into examination halls, and if found in violation of this rule, their exam results are also subject to cancellation. Other behaviors, such as the leakage of the content of an examination paper prior to the examination, intimidation of examination supervisors, markers, or officials, the issuing of forged certificates or altering of results information, and provision of external assistance via the supervisor or mobile phone, have not been reported as typically occurring during the examination process.

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

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

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

28. Student results are confidential. 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, including retaking the examination (if they fail in three or fewer subjects), repeating the grade, or attending remedial or preparatory courses (offered in the private sector) prior to retaking the examination.

30. There is only one mechanism in place, a permanent oversight committee, to monitor the consequences of the examination. There are no studies that are updated regularly or funding for independent research on the impact of the examination. Additionally, regular focus groups or surveys of key stakeholders, or expert review groups are not in place.

# OMAN

## *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. | 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.[8] * | 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.[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.[11] * | The NLSA office is accountable to a clearly recognized body. | This option does not apply to this dimension. |

(CONTINUED)

SABER COUNTRY REPORT |2013

| 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. | The NLSA office is adequately staffed to carry out the NLSA effectively, with minimal issues. [12] * | The NLSA office is adequately staffed to carry out the NLSA effectively, with no issues. |
| 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. |

## 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. [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* | | | |
| There are no courses or workshops on the NLSA. | There are occasional courses or workshops on the NLSA. [17] ✱ | There are some courses or workshops on the NLSA offered on a regular basis. | There are widely available high quality courses or workshops on the NLSA offered on a regular basis. |

## ASSESSMENT QUALITY

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

| LATENT ○○○○ | EMERGING ●●○○ | ESTABLISHED ●●●○ | ADVANCED ●●●● |
|---|---|---|---|
| **ASSESSMENT QUALITY 1:** *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. [19] * | There are a variety of mechanisms in place to ensure the quality of the NLSA. |
| There is no technical report or other documentation about the NLSA. | There is some documentation about the technical aspects of the NLSA, but it is not in a formal report format. | There is a comprehensive technical report but with restricted circulation. | There is a comprehensive, high quality technical report available to the general public. [20] * |
| **ASSESSMENT QUALITY 2:** *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. | NLSA information is used by all stakeholder groups in a way that is consistent with the purposes and the technical characteristics of the assessment. [22] * |
| There are no mechanisms in place to monitor the consequences of the NLSA. | This option does not apply to this dimension. | There are some mechanisms in place to monitor the consequences of the NLSA. [23] * | There are a variety of mechanisms in place to monitor the consequences of the NLSA. |

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

1. The Cognitive Development Tests have been administered to monitor education quality at the system level, support schools and teachers, and inform policy design, evaluation, and decision making. The assessment program was first carried out in 2007-2008 and is administered to all students in grade 7 to assess Mathematics, Sciences, and Environmental Geographic concepts.

2. The Cognitive Development Tests were authorized by the Ministry of Education's *Cognitive Development Manual* document in 2009.

3. The *Cognitive Development Manual* document is publicly available and accessible at the Ministry of Education's website.

4. There is a written NLSA plan for the coming years specifying who will be tested and in which subject areas. The plan is available to, and easily accessible by, the public.

5. Most stakeholder groups support the NLSA. While policy makers and universities strongly support it, educators, students, parents, media, and employers all support the NLSA. It is unclear whether teacher unions, think tanks, and NGOs support or oppose it.

6. There is regular funding for the Cognitive Development Tests allocated by the government.

7. Funding covers all core NLSA activities, including assessment design and administration, data analysis and reporting, and staff training. Funding also covers annual prizes for students exhibiting outstanding performance on the Cognitive Development Tests. However, funding does not cover long- or medium-term planning of program milestones.

8. Funding for the NLSA does not cover research and development activities.

9. The Directorate General of Educational Evaluation is a permanent unit responsible for running the assessment.

10. Political considerations never hamper technical considerations; large-scale assessment results have never been withheld from publication because of political reasons.

11. The Directorate General of Educational Evaluation is not accountable to a clearly recognized body.

12. The Directorate General of Educational Evaluation is adequately staffed with permanent and full-time staff to carry out the NLSA effectively, with minimal issues. However, there have been some issues related to the poor training of test administrators, and unclear instructions and guidelines for administering the assessment.

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

13. Some opportunities are offered in Oman to prepare individuals for work on the NLSA. University graduate programs at the Sultan Qaboos University on educational measurement and evaluation, internships or short-term employment in the large-scale assessment office, and funding for attending international programs, courses, or workshops on educational measurement and evaluation, are all available. However, there are no stand-alone university or non university courses or workshops on educational measurement and evaluation available in Oman.

14. The NLSA measures performance against national curriculum or learning standards.

15. Stakeholder groups largely accept what the NLSA measures.

16. Regular internal reviews and regular independent reviews by qualified experts are in place to ensure alignment between the assessment instrument and what it is supposed to measure.

17. Courses or workshops and presentations are occasionally offered on the NLSA. However, most teachers do not have access to live or online courses, and most courses do not provide teachers with resources relevant for use in their classrooms.

18. Different options are offered to include all groups of students in the NLSA. For example, special plans are made to ensure that the large-scale assessment is administered to students in hard-to-reach areas, and the large-scale assessment is offered in the language of instruction for almost all students. However, accommodations or alternative assessments are not provided for students with disabilities.

19. There are some mechanisms in place to ensure the quality of the NLSA. For example, all proctors or administrators are trained according to a protocol, there is a standardized manual for large-scale assessment administrators, all booklets are numbered, scorers are trained to ensure high inter-rater reliability, there is double processing of data, and there are internal reviewers or observers. However, there are no external reviewers or observers, external certification or audits, or double data scoring. In addition, there is no requirement for discrepancies to be recorded on a standard sheet and a pilot is not conducted before the main data collection takes place.

20. A comprehensive, high quality technical report on the NLSA is available to the general public.

21. NLSA results are disseminated within 12 months of the large-scale assessment being administered, and reports with results are made available for all stakeholder groups. The main reports on the results contain information on overall achievement levels and by sub-groups, and on trends over time overall and for sub-groups. However, the main reports on the results do not contain standard errors. There are workshops or presentations for key stakeholders on the results, and the results are published on the Ministry's Educational Portal. Results, however, are not featured in newspapers, magazines, radio, or television, and there is no media briefing organized to discuss results.

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

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

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

# OMAN

# *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. | The policy document is available to the public. [4] * | 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. [7] * | This option does not apply to this dimension. | This option does not apply to this dimension. | Funding covers research and development activities. |

| LATENT | EMERGING | ESTABLISHED | ADVANCED |
|---|---|---|---|
| ●○○○ | ●●○○ | ●●●○ | ●●●● |

**ENABLING CONTEXT 3:**
*Having effective human resources for ILSA*

| LATENT | EMERGING | ESTABLISHED | ADVANCED |
|---|---|---|---|
| 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. |

SABER COUNTRY REPORT |2013

## 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. | The country/system offers a wide range of opportunities to learn about ILSA. [12] * |
| 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. | There is some media coverage of the ILSA results. [18] * | 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. [19] * | 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. Oman participated in TIMSS in 2007 and 2011, and PIRLS in 2011.

2. Oman has taken concrete steps to participate in PIRLS in 2016 and TIMSS in 2015.

3. The formal policy document, *Ministerial Decision Number 140/2009*, authorized by the Ministry of Education in 2009, addresses participation in ILSAs and the preparation and publication of relevant national reports.

4. The document, *Ministerial Decision Number 140/2009*, is available to the public.

5. Regular funding for international assessment participation is allocated at the discretion of the government.

6. Funding covers all core activities of the ILSAs, including international participation fees, administration of the assessment exercise in Oman, processing and analyzing data collected from implementation of the assessment exercise, reporting and disseminating the assessment results in Oman, and attendance at international expert meetings for the assessment exercise.

7. Funding for the ILSAs does not cover research and development activities.

8. There is a team, the International Studies Program, and a national coordinator to carry out the ILSA activities.

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

10. The ILSA office is adequately staffed to carry out the ILSAs effectively, with minimal issues. Although the team is sufficiently staffed, has previous experience working on international assessments, and has attended some international workshops or meetings, it does not have the necessary training or experience to carry out the required assessment activities effectively.

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

12. Oman offers a wide range of opportunities to learn about ILSAs. There are workshops or meetings on using international assessment databases, funding for attending international workshops or training on international assessments, and online courses on international assessments. However, there are no university courses on the topic of international assessments.

SYSTEMS APPROACH FOR BETTER EDUCATION RESULTS

13. Opportunities to learn about ILSAs are available to a wide audience besides Oman's ILSA team members. Educational Supervisors in the Educational Governorates, teachers, and specialists from the Curricula and Educational Assessment Departments can benefit from these opportunities; however, they are not available to university students studying assessment, or to professionals and university staff interested in assessment.

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

15. Oman has contributed new knowledge on ILSAs.

16. Country specific results and information are regularly and widely disseminated in Oman. For example, a national report was made available online, copies of the national and international reports were distributed to key stakeholders, and Oman's results were communicated through a press release. Brochures and PowerPoint presentations with Oman's results were made available online, and products providing feedback to the schools or educators about the results were also made available.  In addition, a seminar was held to present and discuss the results of participating in TIMSS and PIRLS 2011 and included stakeholders from the Sultanate.

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

18. In addition to editorials or columns commenting on the international assessment results, the Ministry organized presentations on the radio and television, and interviews, to provide information on the results in TIMSS and PIRLS in 2011.

19. ILSA results are used in some ways to inform decision making in Oman. For example, results have been used in tracking the impact of reforms on student achievement levels, and in informing curriculum improvement, teacher training programs, and other assessment activities in the system. However, results have not been used to inform resource allocation.

20. It is not clear that decisions based on ILSA results have had a positive impact on students' achievement levels, as there was no improvement in student achievement between TIMSS 2007 and 2011.

## Acknowledgements

## References

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

Rassekh, Shapour. 2004. "Education as a Motor for Development: Recent Education Reforms in Oman with Particular Reference to the Status of Women and Girls." Geneva: UNESCO-IBE.

Sultanate of Oman Ministry of Education and the World Bank. *Education in Oman: The Drive for Quality.* 2012. Muscat: Sultanate of Oman.

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

World Bank. 2012. Oman Country Indicator Data. Washington, DC: World Bank. Data retrieved from http://data.worldbank.org/ on October 15, 2013.

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

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

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

**WORLD BANK GROUP**