Papua New Guinea

STUDENT ASSESSMENT

SABER Country Report
2014

Key Policy Areas for Student Assessment	Status
1. Classroom Assessment The National Assessment and Reporting Policy (2003) and the National Curriculum Statement (2003) provide high-level guidance on assessment. In addition, teachers receive in-service training at the primary level; they are also provided with syllabi and corresponding teaching guides at the lower-primary, upper-primary, and secondary levels. There are some formal mechanisms in place to monitor the quality of classroom assessment practices, including reviews of how teachers design and use classroom assessment. Nonetheless, classroom assessment practices are known to be weak and vary significantly between schools and classrooms.	Emerging
2. Examinations Papua New Guinea's Higher School Certificate (HSC) examination has been operating on an annual basis for over 20 years. It assesses students in Year 12 in various subjects, including language and literature, mathematics, science, history, information and communication technology, and computer studies. The HSC provides certification for secondary school and is used for student selection into tertiary education. It is also used to inform teachers and pedagogical practices. The unit overseeing the HSC is insufficiently staffed to fulfill its mandate. There are also no mechanisms in place to monitor the consequences of the examination in terms of its positive and negative impacts on students, teachers, and schools.	Emerging
3. National Large-Scale Assessment (NLSA) The Curriculum Standards Monitoring Test (CSMT), which assesses numeracy and literacy, was piloted in 2003–2004 and fully administered to Year 5 and 7 students in 2008 and 2010. Its purpose is to monitor education quality and student progress towards goals defined in the national curriculum. Several official policy documents outline the objectives of the CSMT as well as its implementation, monitoring, and reporting standards. However, there are no formal mechanisms in place to ensure its quality. In addition, the unit in charge of the CSMT had only a few of the required facilities to carry out the assessment.	Emerging
4. International Large-Scale Assessment (ILSA) In 2012, Papua New Guinea took part for the first time in the Pacific Islands Literacy and Numeracy Assessment (PILNA). A total of 14 countries in the Pacific region participated. There were minimal quality problems identified with its implementation in Papua New Guinea. The country complied with all required technical standards for PILNA. Therefore, it is it is expected that the country's results will be presented in the main section of the international report when it is released.	Emerging

Introduction

Papua New Guinea 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, Papua New Guinea decided to benchmark this system using standardized tools developed under The World Bank's Systems Approach for Better Education Results (SABER) program. SABER is an evidence-based program to help countries systematically examine and strengthen the performance of different aspects of their education systems.

What is SABER-Student Assessment?

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

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

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

SABER-Student Assessment methodology

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

Assessment types and purposes

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

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

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

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

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

Quality drivers of an assessment system

The key considerations when evaluating a student assessment system are the individual and combined quality of assessment activities in terms of the

adequacy of the information generated to support decision making. There are three main drivers of information quality in an assessment system: enabling context, system alignment, and assessment quality.

Enabling context refers to the broader context in which the assessment activity takes place and the extent to which that context is conducive to, or supportive of, the assessment. It covers such issues as the legislative or policy framework for assessment activities; institutional and organizational structures for designing, carrying out, and/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, the 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. 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 aspire beyond that only 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 Papua New Guinea

Papua New Guinea is a lower-middle-income Pacific Island nation that occupies the eastern half of the island of New Guinea. Its GDP per capita is US$ 2,184 and its annual economic growth in 2012 was 8 percent. Although Papua New Guinea is a rich and diverse country with substantial cultural and natural resources, it faces relatively low levels of human development due to challenges associated with high population growth, the spread of the rural population across a difficult terrain, language barriers, and high prevalence of HIV/AIDS.

As a result of high population growth, the number of school-aged children in Papua New Guinea has grown dramatically over the last decade. Between 2006 and 2009, the number of students in basic education expanded by 31 percent. Such a surge in school enrollment has created challenges for the education sector in terms of absorption, as well as improving the quality of service delivery. However, the primary net enrollment rate of 53 percent reflects that a large portion of children are either still out of school or not enrolled at all. Of those children who do enroll, almost 60 percent do not complete the basic education cycle.

Given these challenges, Papua New Guinea's Universal Basic Education Plan (UBE) for 2010–2019 identifies several long-term strategic policy directions for improving access and retention. The UBE lays out future goals, such as increasing the primary net enrollment rate to 90 percent by 2019 and having 90 percent of Year 8 graduates meet the minimum standards of the UBE learning objectives by the same year. In order to achieve such goals, the plan identifies five key education results: (i) improved access in 2019; (ii) enhanced retention; (iii) improved quality of education; (iv) enhanced basic education management; and (v) enhanced equity.

Detailed information was collected on Papua New Guinea's student assessment system using the SABER-Student Assessment questionnaires and rubrics in 2013. It is important to remember that these tools primarily focus on benchmarking a country's policies and arrangements for assessment activities at the system, or macro, level. Additional data would need to be collected to determine actual, on-the-ground practices in the schools of Papua New Guinea, particularly by teachers and students. The following sections discuss the findings for each assessment type, accompanied by suggested policy options. The suggested policy options were determined in collaboration with key local stakeholders based on the country's immediate interests and needs. Detailed, completed rubrics for each assessment type in Papua New Guinea are provided in appendix 5.

Classroom Assessment

Level of development

The National Assessment and Reporting Policy (2003) and the National Curriculum Statement (2003) provide high-level guidance on assessment (both classroom assessment and assessment in general), including objectives, guiding principles, reporting and evaluation approaches, and roles and responsibilities associated with assessment activities.

There are syllabi for each subject area at the lower-primary, upper-primary, and secondary levels, which are complemented by Teaching Guides that indicate expected learning outcomes. The Guides include sample learning and assessment plans.

Pre- and in-service training that addresses competencies in classroom assessment is available to teachers. In-service teacher training is managed at the provincial level through the Standards and Guidance Division within the Provincial Division of Education. Standards officers are responsible, together with schools, for identifying training needs to be addressed by a given provincial division. This decentralization results in training disparities across provinces and from school to school, together with inconsistencies in the specific criteria and standards used by teachers in conducting assessments. At the national level, in-service training is provided by the Department of Education (DoE) on an as-needed basis determined by the Provincial Division of Education. Additionally, the Certificate of Basic Education Assessment and Examinations Handbook provides guiding principles for markers, who are always classroom teachers, for the Year 8 examination. Finally, the Measurement Service Branch relies on teachers to develop questions for the assessment.

In Papua New Guinea, national policy indicates that assessment results should be used to inform and enhance teaching and learning practices, as well as to provide feedback to students and parents. However, classroom assessment practices are generally considered weak. It is common for classroom assessment activities to rely mainly on multiple-choice, selection-type questions and to focus mainly on the recall of information. Grade inflation and uneven application of standards when grading student work are both of great concern. It is also common to observe errors in the grading of student work. At the same time, parents tend to be well informed about their children's grades and classroom assessment activities are rarely used as administrative tools. Senior teachers at the primary level and department heads at the secondary level are responsible for reviewing the design and use of classroom assessment by other teachers on their respective staffs. Standards officers also evaluate teachers based on the DoE Teacher's Inspection/Personal Report, which includes a section on assessment, recording, and reporting.

National policy states that schools must formally report individual student performance. Classroom assessment results for individual students are recorded at the primary level in graded student workbooks; for Year 7 and below, in teachers' record books; and for Year 8 and above, in aggregated subject assessment period and summary sheets.

Suggested policy options:

1. Make resources for classroom assessment widely available to teachers. For instance, provide sample activities and scoring guides aligned with the national curriculum. Scoring guides could be accompanied by examples of scored and unscored student responses. In addition, scoring guides could be accompanied by an analysis of common mistakes, as well as examples of the types of feedback that a teacher may provide to a student. Videos could be used to demonstrate best practices in using classroom assessment to improve learning.

2. Improve the quality of classroom assessment practices by, for example, requiring schools to use the national assessment resources made available by the Department of Education. In addition, set up a moderation panel where teachers can discuss and agree on student scores.

3. Ensure that teachers develop competencies in classroom assessment. For example, require pre-service teacher programs to include a core classroom assessment course in their curriculum. In addition, set up a forum where teachers can develop assessments and share them with teachers in different schools. At the same time, provide supervisors with clear guidelines on how to monitor and support teachers in classroom assessment practices.

Examinations

Level of development

Papua New Guinea's Higher School Certificate (HSC) has operated annually for over 20 years to certify secondary school completion and select students into tertiary education. Examination results are also used to inform teachers and pedagogical practices. Since 2003, oversight and administration of the HSC has been managed by the DoE. The examination is administered in Year 12 and evaluates students' skills in English, mathematics, and over 15 other subjects.

The HSC is officially authorized by the National Assessment and Reporting Policy, which is available online and distributed to schools. Some procedures are in place to ensure standardization of the examination. For example, the HSC Examinations Handbook is distributed to schools and provides guidance on assessment design, administration, scoring, and reporting. The handbook defines marking procedures, including specific responsibilities for members of a marking panel. In addition, each year the DoE trains provincial examination coordinators, who are required to train the school examination supervisor and subject-specific invigilators. However, there is no national training for teachers on the HSC, although ad-hoc workshops are occasionally offered and schools can request the Measurement Service Branch (MSB) of the DoE to conduct training on the examination.

The MSB, which is part of the Curriculum Development and Assessment Division of the DoE, oversees the development of the HSC examination (it is responsible for the agenda, management, and budgeting for the Years 8, 10, and 12 examinations). General oversight of examinations is the responsibility of the Board of Studies (BoS), a semi-autonomous group composed of representatives from a broad range of stakeholders, including five ex-officio DoE members, two heads of secondary schools, one superintendent of secondary school inspections, one PNG Teachers Association representative, one church representative, four university representatives, two community representatives (appointed by the Minister of Education), and two business representatives (also appointed by the minister). Regular funding for the HSC covers design, administration, data analysis, and data reporting and planning. Funding is provided through the DoE budget at the discretion of the government and partners.

All students may take the HSC. However, policy does not mandate that HSC preparatory materials be made available to students. Instead, teachers are expected to review previous exam papers with students to prepare them for the types of questions and formats that will be on the examination. Students who do not perform well on the examination can attend remedial education offered through universities. They cannot retake the examination or repeat a year. There are no mechanisms in place to monitor the consequences of the examination in terms of its positive and negative effects on students, teachers, and schools.

There are some formal mechanisms in place to ensure the quality of the examination. During its administration, exam booklets are numbered and color coded, and all subjective questions are double marked. Examination results are regarded as credible by most stakeholders. National and international universities use HSC results to determine admissions, and employers often request the results of job applicants.

Suggested policy options:

1. Consider reducing the size of the examination program in Papua New Guinea. For instance, the examinations at Years 8 or 10 could be dropped, or fewer subject areas could be included for the examinations at Year 12. This would free up resources (human and financial) for conducting the national large-scale assessment. The freed-up resources could also be used to improve other aspects of the examinations.

2. Monitor the consequences of the examinations. For example, focus groups could be conducted with teachers and students in order to gather information on the examination, addressing questions such as: How does the examination influence pedagogy? How effective is the examination in selecting students into tertiary education?

3. Make preparation materials for the examination widely available at the national level. For example, the MSB could publish sample test questions, which could also be distributed to schools. This material could then be used by both teachers and students.

National Large-Scale Assessment (ILSA)

Level of development

The Curriculum Standards Monitoring Test (CSMT) was first introduced in 2008 after being piloted in 2003–2004; it was administered again in 2010. Although policy states that it should be administered every two years, it was not implemented in 2012. The CSMT uses multiple-choice and open-ended questions to monitor education quality and student progress towards the learning goals outlined in the national curriculum.
Several DoE policy documents pertain to the CSMT. The National Education Policy 2005–2014 outlines what the CSMT should evaluate and when it should be administered. The National Assessment and Reporting Policy (2003) indicates that the MSB is responsible for monitoring and reporting standards. Universal Basic Education 2010–2019: Achieving Universal Education for a Better Future recognizes the purpose of the assessment as a measure and mechanism to ensure quality of learning. Additionally, Education Sector Strategic Plan 2011–2030: A Roadmap to the Future refers to the CSMT as one of the primary strategies for monitoring the quality of curriculum implementation on a regular basis.

The development and implementation of the CSMT is led by DoE management (specifically, the CSMT section within the MSB and the BoS) and external donors. The Australian Council for Educational Research (ACER) provided support during the pilot and the first cycle of the CSMT. Funding was sufficient for the initial piloting process, as well as for the assessment's implementation in 2008 and 2010. In addition, once-off funding and independent research was provided by ACER to monitor the CSMT. Funding to carry out CSMT activities would have been available in 2012; however, the MSB did not have sufficient human capacity to implement the assessment. No annual opportunities to learn about the CSMT are offered in the country.

There are few formal mechanisms in place to ensure the quality of the CSMT. A "train-the-trainer" model is used by the provincial officer in each province to train teachers how to administer the CSMT in participating schools. The Handbook for Head Teachers and Teachers of Year 5 and Year 7 of the CMST provides further guidance. In addition, test booklets are numbered, double-data processing is instituted, and open-ended questions are double scored.

Results for the CSMT have been disseminated to participating schools only on a limited basis, such as through an official technical report reviewing the 2008 CSMT, Curriculum Standards Monitoring Test Analysis of Data (December 2009), which was prepared by ACER. However, educational practitioners found the report too technical and abstract for use in the classroom.

Results from the 2010 CSMT have not been analyzed and the 2012 CSMT was not conducted. Despite the nonavailability of the 2010 CSMT report, DoE used the 2008 CSMT report and preliminary 2010 data to develop the new National Education Plan (NEP) and the Universal Basic Education Plan (UBE) at the national level.

Suggested policy options:

1. Secure stable funding and human resources for the NLSA. This could be done by diverting resources from the examinations program to the CSMT.

2. Ensure effective communication of findings. Secure the participation of educators in the design of school reports; design the reports to answer simple, yet relevant, policy questions, such as how student performance varies across regions.

3. Align the assessment with the national curriculum. This could be done, for example, by involving the individuals who developed the curriculum in the development of the assessment framework and questions.

International Large-Scale Assessment (ILSA)

Level of development

In 2012, Papua New Guinea participated in the Pacific Islands Literacy and Numeracy Assessment (PILNA). A total of 14 countries in the Pacific region participated. PILNA is a regional tool developed by the South Pacific Board for Educational Assessment (SPBEA, now known as the Education Quality and Assessment Programme, or EQAP) with the participation of UNESCO. PILNA is designed to measure benchmarking indicators that were agreed upon by participating countries and endorsed at the Forum Education Ministers Meeting (FEdMM). The international large-scale assessment will be implemented in Papua New Guinea every three years, with the next assessment scheduled for 2015.

There is currently no country-level policy document that addresses participation in PILNA. However, in 2006, SPBEA received a mandate from FEdMM to develop PILNA. As a result, two documents have been created to guide implementation: Pacific Islands Forum Secretariat Education Initiatives (2006) and PILNA Implementation Manual (2012). The latter outlines the purpose of the assessment, its framework and approach, as well as the responsibilities associated with its implementation.

Funding for the administration of PILNA is supported by Papua New Guinea through its examinations program budget. Scoring activities are funded by the Australian Agency for International Development (AusAID) through the Pacific Benchmarking for Education Results (PaBER) project. Funding covered processing, analysis of data, and reporting and dissemination of assessment results.

Implementation of PILNA at the national level is overseen by a national coordinator. A school coordinator manages the administration of the text at the school level for selected schools, and test supervisors (teachers) administer the assessment at the classroom level in the selected schools. For each assessment area—literacy and numeracy—a panel is responsible for marking student responses. There were significant issues during the initial implementation of PILNA, largely due to transport, weather, and geographical issues. These issues resulted in a low participation rate: approximately 50 percent of selected schools implemented PILNA in full.

There were several opportunities to learn about PILNA in Papua New Guinea. MSB staff gave presentations about the assessment to senior standard officers from all provinces of the country. In addition, key personnel participated in workshops organized by SPBEA for all PILNA-participating countries. SPBEA also provided in-country training on scoring to panel leaders and members.

PILNA was sufficiently aligned with Papua New Guinea's learning goals. Classroom lessons, textbooks, and learning resources cover content similar to the content covered by PILNA. As a result, students had extensive previous exposure to the type of content and skills measured by PILNA.

Papua New Guinea complied with all technical standards for PILNA and it is expected that its results will be presented in the main section of the international report.

Suggested policy options:

1. Develop and authorize a policy document that formalizes the participation of Papua New Guinea in international assessments.

2. Secure sufficient resources for a thorough implementation of PILNA. For example, divert resources (human and financial) from the examinations program to PILNA.

3. Ensure high participation rates in PILNA so that results may be used to make inferences at the national level. Develop an implementation plan that considers the weather and geographical constraints that may impact assessment administration.

4. Ensure effective dissemination and use of PILNA 2012 results. Produce reports and brochures for different audiences, including teachers and policy makers. Respond to key policy questions, such as how the students of Papua New Guinea perform compared to other countries.

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 performance	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 no history of participation in an ILSA nor plans to participate in one.	Participation in an ILSA has been initiated, but there is still a 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 three dimensions on which a hypothetical country receives the following scores:

- Dimension A = 2 points
- Dimension B = 2 points
- Dimension C = 3 points.

The hypothetical country's overall score for this quality driver would be: (2+2+3)/3 = 2.33

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

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

For scores that allow a margin of discretion (i.e., to choose between two levels of development), a final decision must 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 examinations, NLSA, and ILSA) and under "System Alignment" (in the case of classroom assessment) were set as ceiling scores, that is, 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 Papua New Guinea

PAPUA NEW GUINEA
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***			
(Q1) There is no country-level document that provides guidelines for classroom assessment.	(Q1-3) There is an informal or draft country-level document that provides guidelines for classroom assessment.	(Q1-3) There is a formal country-level document that provides guidelines for classroom assessment, but the document is not available online to anybody interested.	(Q1-3) There is a formal country-level document that provides guidelines for classroom assessment, publicly available online to anybody interested.[1]
ENABLING CONTEXT AND SYSTEM ALIGNMENT 2: ***Aligning classroom assessment with country learning goals***			
(Q4) There are no country-wide resources for teachers for classroom assessment.	(Q4) There are very few country-wide resources for teachers for classroom assessment.[2]	(Q4) There are some country-wide resources for teachers for classroom assessment.	(Q4) There are a variety of country-wide resources for teachers for classroom assessment.
(Q5) There is no official curriculum or standards document.	(Q5) There is an official curriculum or standards document, but it is not clear what students are expected to learn.	(Q5) There is an official curriculum or standards document that specifies what students are expected to learn, but the desired level of performance is not clear.	(Q5) There is an official curriculum or standards document that specifies what students are expected to learn and the desired level of performance.[3]
ENABLING CONTEXT AND SYSTEM ALIGNMENT 3: ***Having effective human resources to carry out classroom assessment activities***			
(Q6) There are no formal country-level mechanisms to ensure that teachers develop competencies in classroom assessment.	(Q6) There are very minimal formal country-level mechanisms to ensure that teachers develop competencies in classroom assessment.[4]	(Q6) There are some formal country-level mechanisms to ensure that teachers develop competencies in classroom assessment.	(Q6) There are a variety of formal country-level mechanisms to ensure that teachers develop competencies in classroom assessment.

ASSESSMENT QUALITY

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ASSESSMENT QUALITY 1: ***Ensuring the quality of classroom assessment***			
(Q7) Classroom assessment practices are very weak, or there is no information available on classroom assessment practices.	(Q7) Classroom assessment practices are known to be weak. [5]	(Q7) Classroom assessment practices are known to be of moderate quality.	(Q7) Classroom assessment practices are known to be of high quality.
(Q8) There are no formal country-level mechanisms to monitor the quality of classroom assessment practices.	(Q8) There are minimal formal country-level mechanisms to monitor the quality of classroom assessment practices.	(Q8) There are some formal country-level mechanisms to monitor the quality of classroom assessment practices. [6]	(Q8) There are varied formal country-level mechanisms to monitor the quality of classroom assessment practices.
ASSESSMENT QUALITY 2: *Ensuring effective uses of classroom assessment*			
(Q10) There are no required uses of classroom assessment.	(Q10) There are minimal required uses of classroom assessment.	(Q10) There are varied required uses of classroom assessment.	(Q10) There are varied required uses of classroom assessment, including its use as an input for selection or certification. [7]
(Q11) Schools are not required to report information on individual student performance.	(Q11-12) At least some schools are required to report information on individual student performance.	(Q11-12) All schools are required to report information on individual student performance to parents. [8]	(Q11-12) All schools are required to report information on individual student performance to parents and other key stakeholders.
(Q11) Information on student performance is not required to be reported.	(Q13-14) Minimal information on student performance is required to be reported.	(Q13-14) Some information on student performance is required to be reported in school report cards. [9]	(Q13-14) A variety of information about student performance is required to be reported in school report cards.

Classroom Assessment: Development-Level Rating Justifications

1. Several Department of Education (DoE) documents are used to inform and guide classroom assessment and are publicly available. The National Assessment and Reporting Policy (2003) provides high-level perspective on the purposes, guiding principles, assessment, reporting and evaluation approaches, and the roles and responsibilities associated with assessments (not only in the classroom). The National Curriculum Statement (2003) complements the National Policy and provides an overarching rationale, goals, specific aims, a curriculum overview, and assessment and reporting requirements for the education system as a whole. Assessment and Reporting: Lower and Upper Primary (2006) is specific to primary education and provides teachers with guidance, resources, and examples with respect to assessing, recording, and making judgments about student achievement, as well as reporting to parents and guardians. Finally, the Syllabus and its corresponding Teaching Guide detail learning and teaching approaches for each subject and year, including subject requirements and learning objectives, performance standards, sample assessment tasks, learning activities and assessment tasks, recording and reporting methods and requirements, and resources.

2. A compilation of documents is used to outline what students are expected to learn in each subject area. In addition to the National Curriculum Statement—which provides an overarching rationale, goals, specific aims, a curriculum overview, and assessment and reporting requirements for the education system as a whole—there is a syllabus for each subject area at the lower-primary, upper-primary, and secondary level. Associated with each syllabus is a Teacher Guide, which details learning outcomes and expectations. The subject syllabus and teaching guideline outline the level of performance that students are expected to reach at each grade level. The DoE also provides textbooks and materials to schools that provide support for classroom assessment.

3. There is a National Curriculum, associated syllabus, and Teacher Guide for each subject. The Teacher Guide highlights the expected learning outcomes and presents performance standards. The Teacher Guide also provides examples of learning plans and examples of assessment plans.

4. There is no standardized course structure specific to a Bachelor's of Education degree, which is the requirement to become a teacher. Each university designs its own program, resulting in some variance across universities, although all university programs in Papua New Guinea are aligned with similar foundational learning concepts, norms, and skills related to the teaching profession. Specific courses or components of courses pertain to assessment, and in particular, to the use of classroom assessment techniques. In-service training is primarily managed at the provincial level through the Standards and Guidance Division of the Provincial Division of Education. Housed in each Provincial Division of Education are standards officers, who are responsible, in collaboration with schools, for identifying in-service training needs. Because training is decentralized to the provincial level, there are variances among in-service training programs, both within a province and nationally. In addition, training at the national level is provided by the DoE on an as-needed basis (determined by each respective Provincial Division of Education). This training generally accompanies the introduction of a new curriculum or syllabus. In-service trainings are often poorly funded, leaving teachers responsible for funding many of their own costs, including transportation and per diem. In-service training opportunities provided by Provincial Divisions of Education are mandatory for teachers. The Certificate of Basic Education Assessment and Examinations Handbook provides guiding principles for marking the Year 8 exam.

5. The approach to classroom assessment is decentralized to the school level (and to teachers in particular); therefore, different approaches are used in different schools. It is common for classroom assessment activities to rely on multiple-choice and selection-type questions and for teachers to submit assessments to the assessment supervisor for review. It is difficult to measure how rigorously specific criteria or standards are applied when assessing student work. Classroom

assessment activities are not always designed to be administered continuously, nor are they designed to be used by teachers to provide feedback on a regular basis. They are rarely used as administrative or control tools.

6. At the primary level, senior teachers, and at the secondary level, heads of departments for each subject, are responsible for reviewing a teacher's design and use of classroom assessment. In addition, standards officers (who are housed at the provincial level and report to the national DoE) carry out a DoE Teacher's Inspection/Personal Report for each teacher. This report encompasses a performance evaluation and constitutes a required component of teacher supervision. The "report" covers multiple sections, one of which is entitled "Assessment, Recording, and Reporting."

National reviews of the quality of education include a focus on classroom assessment. When the DoE reviews the National Curriculum, strong emphasis is placed on the review of syllabi and subject matter, and inter alia, learning outcomes and expectations. These reviews include a strong review of performance standards and the purpose of classroom assessment. In 2013 the DoE tested and introduced an approach to school assessment entitled the "Whole School Quality Assessment Form (WSQAF)." The WSQAF is a holistic approach to school assessment and includes five sections: (i) interview with the head teacher or teacher in charge and the Board of Management chairman; (ii) observing a taught lesson in the classroom, which includes a component on classroom assessment; (iii) interviews with primary school students and teachers, (iv) an elementary reading test; (v) interviews with parent and community members; and (v) Awarding the school a total score out of 220 marks and giving the school a Whole School Quality Assessment Recommendations Report.

7. The National Assessment and Reporting Policy is not specific to classroom assessment, but rather, is a higher-level policy document that does not distinguish between types of assessment. National policy does not clearly state that classroom information should be used to diagnose student learning issues; however, it does indicate that teachers should use assessment information to inform and enhance their teaching and learning practices, which should respond to the learning needs of students. According to national policy, classroom assessment information should also be used to provide feedback to students and inform their parents on their learning. Aside from Year 8, 10, and 12 examinations, classroom assessment is the only type of assessment used in PNG. As such, it is the primary means for determining grade advancement. For Years 8, 10, and 12, classroom assessment results are used in conjunction with examination results to select students into higher education institutions.

8. The National Assessment and Reporting Policy clearly outlines that schools must formally report on individual student performance (although the policy is not specific to classroom assessment). Apart from the nationwide examinations at Years 8, 10, and 12, classroom assessment is the only formal assessment type.

9. Information collected through classroom assessment is required as part of student performance reporting in all subject areas. Classroom assessment results for individual students are recorded at the primary level in each individual student's workbook. For Year 7 and below, results are recorded in a teacher's record book, and for Year 8 and above, in respective subject assessment period sheets for each term and in subject assessment summary sheets for each academic year.

PAPUA NEW GUINEA
Examinations

ENABLING CONTEXT

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 1: *Setting clear policies for the examination*			
(Q1_III-IV) There is no examination.	(Q1_III-IV) The examination has been operating on an irregular basis.	(Q1_III-IV) The examination has been operating regularly. [1]	This option does not apply to this dimension
(Q3) There is no examination, or there is no policy pertaining to the examination.	(Q3-5) There is an informal or draft policy that authorizes the examination; or there is a formal policy that is not available.	(Q3-5) There is a formal policy that authorizes the examination, available upon request or with restricted access.	(Q3-5) There is a formal policy that authorizes the examination, publicly available online to anyone interested. [2]
(Q6) There is no examination, or the examination is not standardized.	(Q6-7) The examination is partially standardized.	(Q6-7) The examination is fully or partially standardized, with at least some procedures in place to ensure standardization. [3]	(Q6-7) The examination is fully standardized, and a variety of procedures are in place to ensure standardization.
ENABLING CONTEXT 2: *Having leadership for the examination*			
(Q8-9) There is no examination, or the country does not have leadership for the examination.	(Q8-9) The country has weak leadership for the examination.	(Q8-9) The country has strong leadership for the examination, from an individual person or from a stakeholder body.	(Q8-9) The country has strong leadership for the examination from both an individual person and a permanent stakeholder body. [4]

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 3: *Having regular funding for the examination*			
(Q10) There is no examination, or there is no funding allocated for the examination.	(Q10-11) There is irregular funding for the examination, or the funding is not allocated by law or regulation. [5]	(Q10-11) There is a regular funding for the examination that is allocated by law or regulation.	This option does not apply to this dimension.
(Q12) There is no examination, or there is no funding coming from the government, student fees, or donors.	(Q12) The funding for the examination comes primarily from donors or loans.	(Q12) The funding for the examination comes primarily from the government or student fees. [6]	This option does not apply to this dimension.
(Q13) There is no examination, or there is no funding to cover activities.	(Q13) There is funding to cover at least some of the core activities.	(Q13) There is funding to cover all or most core activities. [7]	(Q13) There is funding to cover all core activities, plus research and development.
ENABLING CONTEXT 4: *Having institutional capacity for the examination*			
There is no examination, or there is no examination unit.	(Q14-15) There is a temporary unit, or a unit with minimum experience, in charge of the examination.	(Q14-15) There is a permanent unit with some experience in charge of the examination.	(Q14-15) There is a permanent unit with vast experience in charge of the examination. [8]
(Q16) There is no examination, or it is not clear to which body the examination unit is accountable.	This option does not apply to this dimension.	(Q16) The examination unit is accountable to a clearly recognized body.[9]	This option does not apply to this dimension.
(Q18) There is no examination, or the examination unit does not have facilities to carry out the examination.	(Q18) The examination unit has only a few of the required facilities to carry out the examination. [10]	(Q18) The examination unit has all of the required facilities to carry out the examination.	(Q18) The examination unit has up-to-date versions of all required facilities to carry out the examination.

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 5: ***Having human resources for the examination***			
(Q19-20) There is no examination, or there is no staff allocated to the examination unit.	(Q19-20) The examination unit has an inadequate number of staff to carry out the examination. [11]	(Q19-20) The examination unit has an adequate number of staff to carry out the examination, with some quality problems.	(Q19-20) The examination unit has an adequate number of staff to carry out the examination, with no quality problems.
(Q21-22) There is no examination, or the country offers no annual opportunities to learn about the examinations.	(Q21-22) The country offers very few annual opportunities to learn about the examinations. [12]	(Q21-22) The country offers some annual opportunities to learn about examinations to the examination staff.	(Q21-22) The country offers a wide range of annual opportunities to learn about examinations. These opportunities are available to a broad audience, including the examination staff.
(Q23-24) There is no examination, or teachers have no opportunities to learn about the examination, and are not involved in examination-related tasks.	(Q23-24) Teachers have at least some opportunities to learn about the examination, or are involved in at least some examination-related tasks. [13]	(Q23-24) Teachers have at least some opportunities to learn about the examination, and are involved in at least some examination-related tasks.	(Q23-24) Teachers have opportunities to learn about different aspects of the examination, and are involved in most examination-related tasks.

SYSTEM ALIGNMENT

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

LATENT	EMERGING	ESTABLISHED	ADVANCED
●○○○	●●○○	●●●○	●●●●
SYSTEM ALIGNMENT 1: *Aligning the examination with learning goals*			
(Q26-27) There is no examination, or the examination is not aligned with official learning goals or curriculum.	(Q26-27) The examination is weakly aligned with official learning goals or curriculum, or there are no regular reviews to ensure alignment.	(Q26-27) The examination is at least sufficiently aligned with official learning goals or curriculum, and there are regular reviews of the examination take place to ensure alignment. [14]	(Q26-27) The examination is fully aligned with official learning goals or curriculum, and regular external reviews take place to ensure alignment.
(Q28-29) There is no examination, or there the material to prepare for the examinations is available to a small number of students at most.	(Q28-29) The material to prepare for the examinations is accessible to at least some students. [15]	(Q28-29) There is comprehensive material to prepare for the examination that is accessible to most students.	(Q28-29) There is comprehensive material to prepare for the examination that is widely accessible to all or almost all students.
(Q30) There is no examination, or the examination is not consistent with other assessment activities. [16]	(Q30) The examination is minimally consistent with other assessment activities.	(Q30) The examination is sufficiently consistent with other assessment activities.	(Q30) The examination is fully consistent with other assessment activities.

ASSESSMENT QUALITY

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

LATENT	EMERGING	ESTABLISHED	ADVANCED
●○○○	●●○○	●●●○	●●●●
ASSESSMENT QUALITY 1: *Ensuring the quality of the examination*			
(Q31) There is no examination, or there are no formal mechanisms in place to ensure the quality of the examination.	(Q31) There are minimal formal mechanisms in place to ensure the quality of the examination.	(Q31) There are some formal mechanisms in place to ensure the quality of the examination. [17]	(Q31) There are a variety of formal mechanisms in place to ensure the quality of the examination.
(Q32) There is no examination, or there is no documentation about the technical aspects of the examination. [18]	(Q32) There is some documentation about the technical aspects of the examination.	(Q32) There is a comprehensive technical report about the examination available upon request or with restricted access.	(Q32) There is a comprehensive technical report about the examination publicly available online.
ASSESSMENT QUALITY 2: *Ensuring fairness*			
(Q33) There is no examination, or the majority of the students may not take the examination because of language, gender, or other equivalent barriers.	(Q33) A significant proportion of students may not take the examination because of language, gender, or other equivalent barriers.	(Q33) A small proportion of students may not take the examination because of language, gender, or other equivalent barriers.	(Q33) All students can take the examination; there are no language, gender or other equivalent barriers. [19]
(Q37; Q33) There is no examination, or student results are not confidential, or inappropriate behavior surrounding the examination is high.	(Q37; Q33) Student results are confidential. [20]	(Q37; Q33) Student results are confidential, and inappropriate behavior surrounding the examination is low.	(Q37; Q33) Student results are confidential, and there is no inappropriate behavior surrounding the examination.

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ASSESSMENT QUALITY 3: ***Ensuring appropriate uses of the examination***			
(Q34; Q37) There is no examination, or student results are not perceived as credible, or are not recognized by any broader certification or selection system.	(Q34; Q37) Student results are perceived as credible by at least some stakeholders.	(Q34; Q37) Student results are perceived as credible by most stakeholders, and are nationally recognized.	(Q34; Q37) Students' results are perceived as credible by most stakeholders, and are internationally recognized. [21]
(Q39) There is no examination, or there are no options in the education system for students who do not perform well on the examination.	(Q39) There are very few options in the education system for students who do not perform well on the examination. [22]	(Q39) There are some options in the education system for students who do not perform well on the examination.	(Q39) There are a variety of options in the education system for students who do not perform well on the examination.
(Q40) There is no examination, or there are no mechanisms in place to monitor the examination. [23]	This option does not apply to this dimension.	(Q40) There are some mechanisms in place to monitor the examination.	(Q40) There are a variety of mechanisms in place to monitor the examination.

Examinations: Development-Level Rating Justifications

1. The Higher School Certificate (HSC) examination has been administered on an annual basis for over 20 years. Its purpose is to certify student completion of the school cycle, to inform teachers and pedagogy, and to select students into tertiary education. The examination is administered in Year 12 and can include assessments of accounting, applied English, advanced mathematics, applied natural resource management, applied science, biology, business studies, chemistry, design and technology, economics, general mathematics, geography, geology, history, information and communication technology, computer studies, language and literature, legal studies, music, personal development, physical education, physics, theatre, tourism studies, and visual arts. The only compulsory examination is that for language and literature.

2. The National Assessment and Reporting Policy officially authorizes the examination at the country level. It has been managed by the DoE since 2003. The policy document is available online on the DoE website and distributed to schools for use by teachers and parents. The HSC Examinations Handbook is also distributed to schools.

3. The examination is fully standardized (i.e., assessment design, administration, scoring, and reporting are the same for all students taking the examination). According to policy, all HSC examinations are fully standardized. All examination papers are standard in each subject area. Training for examination administrators is not specifically captured in the National Policy. Provincial examination coordinators in each province are trained on an annual basis by the DoE. Within each province, these coordinators train the school examination supervisors and subject-specific invigilators. Quality-control mechanisms are captured under the Administration and Conduct sections of the HSC Examinations Handbook and include the physical environment (i.e., adequate desks and space), procedures (i.e., candidate numbers on desks, timing and process for exams), materials (i.e., specified calculators, rulers, etc.), as well as the objectivity and accountability of invigilators. The HSC Handbook clearly defines procedures for marking examinations. There are three primary roles in this process: marking coordinator, chief marker, and markers. Collectively, these members form a Marking Panel, which must convene on the Monday of the week after the National Examinations are administered. The HSC Handbook states that the matriculation form (which shows all grades) must be given to students.

4. The HSC has existed since the late 1970s. As such, no one individual or group of individuals in PNG is/ are recognized as the leader of the examination. Rather, the Measurement Service Branch (MSB), part of the DoE's Curriculum Development and Assessment Division, is the champion of examinations in the country. The MSB pushes for the development of the HSC examination, is recognized as the assessment authority in education, and is tasked with determining the examination agenda. It thus manages Year 8, 10 and 12 examinations. Further, the MSB manages the budget for all examination activities. In 2011, the Assessment Examinations Certification Advisory Committee was developed to act as an expert body on examination issues; however, it has only had one meeting.

The Board of Studies (BoS) is composed of representatives of a broad range of stakeholders, including five ex officio DoE members, two heads of secondary schools, one superintendent of secondary school inspections, one PNG Teachers Association representative, one church representative, four university representatives, and two community and two business representatives appointed by the minister of education. Among other responsibilities, BoS provides oversight over and input into curriculum content, standards, and examinations.

5. Regular funding for the examination is provided at the discretion of the government and its development partners. It is generally allocated based on the previous year's expenditure.

6. There are no fees for students to sit for the exams. All funding is provided by the DoE budget. Donors contribute significant sums to education in PNG, making it difficult to distinguish the origin of funding (i.e., donors or government revenues) for examinations.

7. Funding covers examination design, administration, data analysis, data reporting, and planning.

8. The MSB has been is in charge of examinations since the inception of the HSC and has implemented the exam over five times.

9. The MSB is directly accountable to the Curriculum, Development, and Assessment Division and to the secretary of the DoE. The BoS is tasked with general oversight of examinations on an as-needed basis. The BoS is a semi-autonomous group.

10. Although computers are up to date and servers are sufficient, there are not enough computers for all staff. Furthermore, staff needs to travel and there are an insufficient number of laptops to support their work program. In addition, security for the building has been an issue in past; additional storage space and a more organized approach is required.

11. As part of the PaBER project, the Australian Council for Educational Research recently conducted an institutional capacity analysis of the MSB. Findings indicate that the body is insufficiently staffed to fulfill its mandate. Time-consuming procurement procedures have, moreover, several times adversely impacted its performance in several areas. In addition, the following problems have affected a subset of schools or all schools in recent years: (i) exams are designed by local experts (from local universities) under the guidance of the MSB, which has created process management challenges for the MSB and delays when exam writers have missed deadlines; (ii) issues with cheating and delays with the government printing office led the DoE to print all examinations offshore, which added a layer of logistics and created procurement challenges; (iii) the MSB had insufficient funding to implement the HSC in 2012 , a period that coincided with elections, which compounded the problem; and (iv) delays in data submission and errors in marking have been observed, which impacted the timeline for reporting results.

12. There is no standardized course structure in PNG specific to a Bachelor's of Education (the requirement to become a teacher in the country). Rather, each university designs its own degree requirements. As a result, there is some variance across universities. Nevertheless, they all align with the same foundational learning concepts, norms, and skills related to the teaching profession. Course components and specific courses pertain to assessment, particularly, the use of examinations. Historically, there has been some funding for workshops; however, due to inadequate staff numbers and high workloads, this funding has not been used. Similarly, courses and workshops on assessment are offered from time to time, but these are both infrequent and not necessarily available to all educators. University students are the sole beneficiaries of opportunities to learn about examinations on an annual basis.

13. There is no national uniform training that offers teachers opportunities to learn about either the content and skills measured by the HSC examination or other aspects of the examination. Rather, PNG has a highly decentralized approach to teacher training. The school level is tasked with identifying the main areas of need and designing a strategy to build capacity in these areas. Schools could request that MSB staff conduct training on the HSC examination, but this would represent

one-off training at the school level. According to the HSC Handbook, the only task that is mainly performed by teachers is the supervision of examination procedures. Markers tend to be teachers, but this is not a requirement.

14. The National Assessment and Reporting Policy states that the HSC should reflect students' achievement of the learning outcomes described in the syllabus. In practice, the HSC is aligned with learning goals and outcomes. The examination for each subject is prepared by a panel of experts from local universities, based on the subject syllabus. As documented in the HSC Handbook, the MSB verifies as much as possible that the examination follows syllabus specifications.

15. There is no national policy or set of requirements about making materials on examinations available to students. Rather, this task is the responsibility of teachers. At the school level, teachers are expected to review previous exams with students to prepare them for question types and formats, as well as to familiarize them with the overall nature of exam writing. In addition, the school examination supervisor is required to ensure that all candidates are aware of conduct rules for the examination. Granting access to materials is largely the responsibility of local schools. As such, it is difficult to accurately quantify the level of access to the materials needed to prepare for the examination, although most students appear to have access to them.

16. Classroom assessment is designed, managed, and implemented at the school level. Results are not aggregated at the central level. For this reason it is impossible to tell whether the HSC examinations (or PILNA, the international large-scale examination) are fully consistent with classroom assessment practices.

17. Provincial examination coordinators are responsible for training the school examination supervisor, external invigilators, and in-school invigilators. All booklets are numbered using candidate ID numbers and color coded according to subject. According to the HSC Handbook, all subjective questions must be reviewed by an independent marker to confirm the accuracy and consistency of the first marker(s). There is an established process to ensure high inter-rater reliability**.** At the outset of the exam, markers who share a question are required to confer over the marking of the first 30 questions (or so) to establish common ground. In addition, the chief marker spot checks all marking. There is also double processing of data; however, this is not documented in policy.

18. Examination reports are prepared for each subject, but are not technical in nature. They tend to focus on logistics and other factors that influence implementation of the HSC. No examination report was produced for the 2011, 2012 or 2013 HSC exams, and no copy of the 2010 version of the report could be obtained during data collection.

19. All students are eligible to take the HSC. There are no specific costs associated with examinations; however, there are general school fees.

20. Individual matriculation forms (i.e., final results) are given to all students and/or guardians, as well as to their respective schools. The HSC examinations are recognized in PNG for the purposes of admitting students to tertiary education and certifying completion of secondary school. Results of the exam are recognized in certain Asia Pacific countries (e.g., Australia) for the purpose of admission to tertiary school.

Leakage (i.e., circulation of all or part the exam prior to its administration) has been an issue in recent years. The precise cause or origin of leakage is not clear; it is likely a combination of leakage at the school, provincial, and DoE levels. Investigations into the issue have generated recommendations, but no concrete action has been taken to date. Cheating has also been an issue for various reasons. The HSC Examinations Handbook outlines various measures to curb the prevalence

of this practice. In addition, a few instances of intimidation of examination officials have been observed. Examples of forged HSC certificates have also been observed.

21. Results are regarded as credible by most stakeholders. For example, national and international universities use results to determine admissions, and employers often refer to results when considering candidates. Examinations are recognized in PNG for the purposes of admitting students to tertiary education and certifying completion of secondary school. Results are also recognized in certain Asia Pacific countries (e.g., Australia) for admission to tertiary institutions.

22. The only option for students who do not perform well on examinations is to attend remedial education, which is held at nearby tertiary and vocational schools.

23. No formal mechanisms are in place to monitor the examination in terms of its impact, acceptance, and credibility.

PAPUA NEW GUINEA
National (or System-Level) Large-Scale Assessment (NLSA)

ENABLING CONTEXT

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 1: ***Setting clear policies for the NLSA***			
(Q3_III) No NLSA has taken place in the country.	(Q3_III, IV) The NLSA has been operating on an irregular basis. [1]	(Q3_III, IV) The NLSA has been operating regularly.	This option does not apply to this dimension.
(Q5) There was no NLSA, or there was no policy document pertaining to NLSA. [2]	(Q5-7) There was an informal or draft policy document that authorized the NLSA. [2]	(Q5-7) There was a formal policy document that authorized the NLSA, available upon request or with restricted access.	(Q5-7) There was a formal policy document that authorized the NLSA that is publicly available online to anyone interested.
(Q8) There was no NLSA, or there was no assessment schedule for future NLSAs.	(Q8-9) There was a common, informal understanding that there would be an NLSA in the future.	(Q8-9) There was an official assessment schedule for future NLSAs, albeit lacking in details. [3]	(Q8-9) There was a publicly available official assessment schedule for future NLSAs, specifying when (year), who (grade level) and what (subject areas) would be assessed.
ENABLING CONTEXT 2: ***Having leadership for the NLSA***			
(Q10-11) There was no NLSA, or the country did not have leadership for the NLSA.	(Q10-11) The country had weak leadership for the NLSA.	(Q10-11) The country had leadership for the NLSA from an individual person or from a stakeholder body.	(Q10-11) The country had leadership for the NLSA from both an individual person and a permanent stakeholder body. [4]

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 3: *Having regular funding for the NLSA*			
(Q12) There was no NLSA, or there was no funding for the NLSA.	(Q12-13) There was funding for the NLSA. [5]	(Q12-13) There was regular funding for the NLSA that was allocated by law or regulation.	This option does not apply to this dimension.
(Q12) There was no NLSA, or there was no funding for the NLSA from internal or external sources.	(Q14) The funding for the NLSA came only or primarily from loans or external donors. [6]	(Q14) The funding for the NLSA came primarily or only from the country's internal funding sources.	This option does not apply to this dimension.
(Q12) There was no NLSA, or there was no funding.	(Q15) There was funding to cover at least some minimum core activities of the NLSA.	(Q15) There was funding to cover all or most core activities of the NLSA. [7]	(Q15) There was funding to cover all core activities, plus research and development.
ENABLING CONTEXT 4: *Having institutional capacity for the NLSA*			
(Q16) There was no NLSA, or there was no NLSA team.	(Q16-18) There was a unit or team with at least one person in charge of the NLSA.	(Q16-18) There was a permanent team, at least nationally recognized, with at least some experience in NLSA. [8]	(Q16-18) There was a permanent team, internationally recognized, with vast experience in NLSA.
(Q19-20) There was no NLSA, or it is unclear to which body the NLSA unit was accountable.	This option does not apply to this dimension.	(Q19-20) The NLSA unit was accountable to a clearly recognized body. [9]	This option does not apply to this dimension.
(Q21) There was no NLSA, or the NLSA unit did not have facilities to carry out the assessment.	(Q21) The NLSA unit had only a few of the required facilities to carry out the assessment. [10]	(Q21) The NLSA unit had all of the required facilities to carry out the assessment.	(Q21) The NLSA unit had up-to-date versions of all required facilities to carry out the assessment.

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 5: ***Having human resources for the NLSA***			
(Q16) There was no NLSA, or there was no NLSA staff.	(Q22-23) The NLSA team had an inadequate number of staff to carry out the NLSA. [11]	(Q22-23) The NLSA team had an adequate number of staff to carry out the NLSA, with some quality problems.	(Q22-23) The NLSA team had an adequate number of staff to carry out the NLSA, without quality problems.
(Q24-25) There was no NLSA, or the country did not offer annual opportunities to learn about NLSA. [12]	(Q24-25) The country offered very few annual opportunities to learn about NLSA.	(Q24-25) The country offered some annual opportunities to learn about NLSA, albeit only to the NLSA team members.	(Q24-25) The country offered a wide range of annual opportunities to learn about NLSA. These opportunities were available to a broad audience, including the NLSA team members.
(Q26) There was no NLSA, or teachers did not have annual opportunities to learn about the NLSA. [13]	This option does not apply to this dimension.	(Q26) Teachers had annual opportunities to learn about the content and skills measured by the NLSA.	(Q26) Teachers had annual opportunities to learn about different aspects of the NLSA.

SYSTEM ALIGNMENT

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

LATENT	EMERGING	ESTABLISHED	ADVANCED
●○○○	●●○○	●●●○	●●●●
SYSTEM ALIGNMENT 1: *Aligning the NLSA with learning goals*			
(Q27-28) There was no NLSA, or the country did not have official learning goals or curriculum, or the NLSA was not aligned with the official learning goals or curriculum.	(Q27-28) The NLSA was minimally aligned with official learning goals or curriculum. [14]	(Q27-28) The NLSA was sufficiently aligned with official learning goals or curriculum, and a regular internal review took place to ensure alignment.	(Q27-28) The NLSA was fully aligned with official learning goals or curriculum, and a regular external review took place to ensure alignment.
(Q29) There was no NLSA, or students had no previous exposure to the type of content and skills measured by the NLSA.	(Q29) Students had limited previous exposure to the type of content and skills measured by the NLSA.	(Q29) Students had sufficient previous exposure to the type of content and skills measured by the NLSA. [15]	(Q29) Students had extensive previous exposure to the type of content and skills measured by the NLSA.
(Q30) There was no NLSA, or the NLSA was not consistent with other assessment activities.	(Q30) The NLSA was minimally consistent with other assessment activities.	(Q30) The NLSA was sufficiently consistent with other assessment activities. [16]	(Q30) The NLSA was fully consistent with other assessment activities.

ASSESSMENT QUALITY

Degree to which the assessment 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*			
(Q31-32) There was no NLSA, or there were no mechanisms to include all student groups in the NLSA.	(Q31-32) There were informal or ad hoc mechanisms to include all student groups in the NLSA. [17]	(Q31-32) There were some formal mechanisms to include all student groups in the NLSA.	(Q31-32) There were a variety of formal mechanisms to include all student groups in the NLSA.
(Q31-32) There was no NLSA, or there were no formal mechanisms in place to ensure the quality of the NLSA.	(Q33) There were very few formal mechanisms in place to ensure the quality of the NLSA. [18]	(Q33) There were some formal mechanisms in place to ensure the quality of the NLSA.	(Q33) There were a variety of formal mechanisms in place to ensure the quality of the NLSA.
(Q34) There was no NLSA, or there was no technical documentation about the NLSA.	(Q34) There was some documentation about the technical aspects of the NLSA.	(Q34) There was a comprehensive technical report for the NLSA, available upon request or with restricted access. [19]	(Q34) There was a comprehensive technical report for the NLSA, publicly available online.

(CONTINUED)

ASSESSMENT QUALITY 2: *Ensuring effective uses of the NLSA*			
(Q35-37) There was no NLSA, or country results and information were not disseminated.	(Q35-37) Country results and information were disseminated using some communication strategy. [20]	(Q35-37) Country results and information were disseminated using a variety of communication strategies, including dissemination to some schools.	(Q35-37) Country results and information were disseminated using a variety of communication strategies, including dissemination to most schools.
(Q38-39) There was no NLSA, or NLSA results and information were not covered by the media. [21]	(Q38-39) NLSA results and information had hardly any coverage in the media.	(Q38-39) NLSA results and information were covered by some media outlets.	(Q38-39) NLSA results and information were covered by a wide variety of media.
(Q40-41) There was no NLSA, or results from the NLSA were not used to inform decision making in the country.	(Q40-41) Results from the NLSA were minimally used to inform decision making in the country. [22]	(Q40-41) Results from the NLSA were used in some ways to inform decision making in the country.	(Q40-41) Results from the NLSA were used in a variety of ways to inform decision making in the country.
(Q42) There was no NLSA, or there were no mechanisms in place to monitor the NLSA.	This option does not apply to this dimension.	(Q42) There were some formal mechanisms in place to monitor the NLSA. [23]	(Q42) There were a variety of formal mechanisms in place to monitor the NLSA.
(Q43) There was no NLSA, or there is no clear evidence or consensus about the positive impact of the NLSA on education quality. [24]	This option does not apply to this dimension.	(Q43) There is a general consensus about the positive impact of the NLSA on education quality.	(Q43) There is evidence of the positive impact of the NLSA on education quality.

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

1. The Curriculum Standards Monitoring Test (CSMT) was first introduced six years ago and is intended to be administered every two years. It uses multiple-choice and open-ended questions to monitor educational quality at the country level. CSMT results are also used to inform education policy design, evaluation, and decision making. The assessment was piloted in 2003–2004 in Years 3, 5, and 7, which were tested for both numeracy and literacy (i.e., reading and writing). The first full implementation of the assessment took place in 2008 for Years 5 and 7 (also for numeracy and literacy). The CSMT was implemented in 2010 (following the same format used in 2008), but not in 2012.

2. The CSMT is captured and articulated through various policy documents authorized by the DoE, which are available online. The National Assessment and Reporting Policy (2003) states that the Measurement Services Branch within DoE is responsible for ensuring that "standards are monitored and reported at appropriate levels." Building on this, the National Education Policy (NEP) 2005–2014 is the first policy document that notes the CSMT. Specifically, the NEP notes that, "A CSMT, which will sample performance for literacy and numeracy, is being piloted in Years 3, 5 and 8. Results will be used to monitor national levels of literacy and numeracy every two years." Universal Basic Education 2010–2019: Achieving Universal Education for a Better Future recognizes the existence and purpose of the CSMT, in particular, its importance as a measure of and mechanism for ensuring quality learning. The Education Sector Strategic Plan 2011–2030: A Roadmap into the Future refers to the CSMT as one of the primary strategies for regularly monitoring the quality of curriculum implementation.

3. As noted above, although the CSMT is scheduled to be implemented every two years, it was not administered in 2012 largely due to insufficient staff. In addition, analysis and report writing were not completed for the 2010 implementation. The DoE intends to continue administering the CSMT. During the 2010 round, the implementation schedule was not made publicly available. The public does, however, have access to the DoE policy documents that state that the CSMT should be implemented every two years.

4. DoE management and external donors led the development and implementation of the CSMT. In 2001, the same two groups utilized the existing Curriculum Reform Implementation Project (CRIP) to commission a study that investigated the feasibility of DoE monitoring curriculum standards in basic education over time. The recommendations of that study informed the pilot CSMT program. The DoE management team is recognized as the authority in education, has the power necessary to determine the CSMT agenda and the skill set and the experience to manage important assessment projects, and is responsible for obtaining funding. This permanent team of stakeholders led the development and implementation of the CSMT. Other stakeholders who were (and are) heavily involved in the exam's administration, and who provided a great deal of technical support, include the Australian Council for Educational Research and the National Research Institution based in Port Moresby.

5. Funding was sufficient for the CSMT pilot in 2008 and its full implementation in 2010. Funding would have been available in 2012; however, the MSB did not have sufficient human capacity to implement the CSMT.

6. It is difficult to distinguish the source of funding for the CSMT in both 2008 and 2010. The government of Australia (and of New Zealand to a lesser extent) provides substantial funding to DoE. Some funding is earmarked, other funding is for general policy and operation support.

7. No specific data breakdown was provided for this question during data collection.

8. The MSB within the DoE is in charge of the CSMT. This CSMT unit consists of three CSMT positions (senior officer, primary officer, and secondary officer). The MSB is recognized as the national authority in student assessment and has implemented the CSMT three times (including the pilot).

9. The team responsible for carrying out the CSMT was accountable to a clearly recognized, semi-autonomous body, the Board of Studies (BoS), which is composed of representatives from a broad range of stakeholders, including five ex officio DoE members, two heads of secondary schools, one superintendent of secondary school inspections, one PNG Teachers Association representative, one church representative, four university representatives, and two community and two business representatives appointed by the minister of education. The BoS is tasked with oversight of and input into curriculum content and educational standards and examinations, as well as general oversight of examinations on an as-needed basis.

10. Although computers are up-to-date and servers are sufficient for the nature of its work, there are not enough computers for all staff. Furthermore, staff need to travel and there are an insufficient number of laptops to support their work program. Security for the building has been an issue in the past. Additional storage space and a more organized approach are required. Notwithstanding the shortage of computers, the unit's communication tools are sufficient.

11. As part of the PaBER project, the Australian Council for Educational Research recently conducted an institutional capacity analysis of the MSB. Findings indicate that the body is insufficiently staffed to fulfill its mandate. The annual examinations place a great load on the MSB. The logistical and operational aspects of the examination cycle are demanding; the MSB is able to manage because all staff members step in to help, including those whose time should be spent on professional and technical activities. In addition, there were delays in data entry for some schools in both 2008 and 2010. To date, data analysis and report writing have not been completed for 2010.

12. There is no opportunity to learn about the CSMT on an annual basis.

13. There is no opportunity to learn about the CSMT on an annual basis.

14. One of the primary intents of the CSMT is to monitor the curriculum. The CSMT aligns with the content and skills areas of the curriculum, as well as with official learning goals and pedagogical approaches and activities. In 2008 and 2010, curriculum officers reviewed test items to ensure alignment of the assessment with the curriculum.

15. Schools use similar textbooks and learning resources, and teachers cover similar content and skills, although there is some variance across schools. Students were exposed to more advanced content and skills than those covered by the CSMT.

16. The classroom assessment section of this report explains that classroom assessment is designed, managed, and implemented at the school level. Results of such assessment are not aggregated back to the central level. For this reason it is not possible to tell whether the CSMT is fully consistent with classroom assessment

practices. The CSMT is somewhat consistent with PILNA (discussed in following ILSA section). Both are diagnostic tools; however, PILNA reflects general learning norms for the Pacific Region, whereas the CSMT is based on the curriculum specific to PNG.

17. Efforts were made to administer the CSMT in schools in hard-to-reach areas, but only sparse data was returned and it was of poor quality and thus not used. The CSMT was offered in English, which is the official language of instruction. Documentation is not available on the sampling method used. In total, approximately 2,000 students participated in both the 2008 and 2010 rounds.

18. A train-the-trainer model was used by the training officer for each province, who was generally selected from teachers or examination officers from the province. The training officer trained teachers in participating schools, using the CSMT Handbook for Head Teachers and Teachers of Year 5 and Year 7 for guidance.

The pilot was conducted in 2003. All student booklets were numbered to ensure accuracy and effective coordination, opinion and/or subjective questions were double scored by trained markers, and double processing of data was completed in 2008. The same protocol was supposed to be used for the 2010 data.

19. The Australian Council for Educational Research (ACER) conducted a review of the 2008 CSMT and prepared an official technical report entitled “Curriculum Standards Monitoring Test Analysis of Data December 2009.” No report was produced for 2010 implementation.

20. The ACER report mentioned in the previous note was disseminated to provinces and participating schools. General feedback from education practitioners was that the report was too technical and abstract for use in the classroom. Assessment results were sent to most of the participating schools.

21. CSMT results were not covered by the media.

22. Despite the non-availability of the 2010 CSMT report, DoE used the 2008 CSMT report and the preliminary 2010 results to develop the new National Education Plan and the Universal Basic Education Plan at the national level.

23. One-off funding and independent research was provided by the Australian Council for Educational Research in order to monitor the CSMT.

24. The 2008 pilot provided the baseline data for the CSMT and no analysis of results from either the 2010 and 2012 rounds was conducted. It is thus impossible to quantify the exam’s impact on national educational quality.

PAPUA NEW GUINEA
International Large-Scale Assessment (ILSA)

ENABLING CONTEXT

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 1: *Setting clear policies for the ILSA*			
(Q1, Q2) The country has not participated in an ILSA in the last 10 years.	(Q1, Q2) The country has participated in, but not completed, an ILSA in the last 10 years.	(Q1, Q2) The country has completed one ILSA in the last 10 years. [1]	(Q1, Q2) The country has completed two or more ILSAs in the last 10 years.
(Q3) The country is not currently participating in an ILSA or has not taken concrete steps to participate in an ILSA in the next 5 years.	This option does not apply to this dimension.	(Q3) The country is currently participating in an ILSA or 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.
(Q5) There was no country-level policy document that addressed participation in the ILSA. [3]	(Q5,Q6) There was an informal or draft country-level policy document that addressed participation in the ILSA.	(Q5,Q6) There was a formal country-level policy document that addressed participation in the ILSA that was available upon request or with restricted access.	(Q5,Q6,Q7) There was a formal country-level policy document that addressed participation in the ILSA that was publicly available online to anyone interested.
ENABLING CONTEXT 2: *Having sufficient funding for the ILSA*			
(Q8) There was no funding for participation in the ILSA, discretionary or otherwise.	(Q9) Funding for the ILSA activities was primarily allocated at the discretion of the country's government. [4]	(Q9) Funding for the ILSA activities was primarily allocated by law or regulation.	This option does not apply to this dimension.
(Q8) There was no funding from loans, external donors, or internal sources.	(Q10) There was funding only or primarily from loans or external donors.	(Q10) There was funding primarily from the country's internal funding sources. [5]	(Q10) There was funding only from the country's internal sources.
(Q8) There was no funding for core items or research and development.	(Q11) The ILSA funding covered at least minimum core items of the ILSA.	(Q11) The ILSA funding covered most core items. [6]	(Q11) The ILSA funding covered most core items, plus research and development.

(CONTINUED)

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
ENABLING CONTEXT 3: *Having effective institutional and human capacity for the ILSA*			
(Q12-16) There was no ILSA unit or team.	(Q12-16) There was at least one person in charge of the ILSA. [7]	(Q12-16) There was a recognized unit or team with at least some experience in international assessments that carried out the ILSA in an effective way.	(Q12-16) There was an internationally-recognized unit or team with vast experience in international assessments that carried out the ILSA in an effective way.
(Q17) There was no ILSA unit, or the unit did not have the required facilities to carry out the ILSA.	(Q17) The ILSA unit had only a few of the required facilities to carry out the ILSA. [8]	(Q17) The ILSA unit had all of the required facilities to carry out the ILSA.	(Q17) The ILSA unit had up-to-date versions of all of the required facilities to carry out the ILSA.
(Q18-20) The country offered no opportunities to learn about ILSAs.	(Q18-20) The country offered minimal opportunities to learn about ILSA. [9]	(Q18-20) The country offered adequate opportunities to learn about ILSA.	(Q18-20) The country offered adequate opportunities to learn about ILSA to a broad audience, including the ILSA team and educators.

SYSTEM ALIGNMENT

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

LATENT ●○○○	EMERGING ●●○○	ESTABLISHED ●●●○	ADVANCED ●●●●
SYSTEM ALIGNMENT 1: *Aligning the ILSA with learning goals for the country*			
(Q21) The ILSA was not aligned with the country's official learning goals, or the country did not have official learning goals.	(Q21) The ILSA was partially aligned with the country's official learning goals.	(Q21) The ILSA was sufficiently aligned with the country's official learning goals.[10]	(Q21) The ILSA was fully aligned with the country's official learning goals.
(Q22) Students were not previously exposed to the type of content and skills measured by the ILSA.	(Q22) Students had limited previous exposure to the type of content and skills measured by the ILSA.	(Q22) Students had sufficient previous exposure to the type of content and skills measured by the ILSA. [11]	(Q22) Students had extensive previous exposure to the type of content and skills measured by the ILSA.
(Q23) The ILSA was not consistent with the country's other assessment activities.	(Q23) The ILSA was minimally consistent with the country's other assessment activities.	(Q23) The ILSA was generally consistent with the country's other assessment activities. [12]	(Q23) The ILSA was fully consistent with the country's other assessment activities.

ASSESSMENT QUALITY

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

LATENT	EMERGING	ESTABLISHED	ADVANCED
			
ASSESSMENT QUALITY 1: *Ensuring the technical quality of the ILSA*			
(Q24) The country did not meet sufficient technical standards to have its data presented in the international report or an annex.	(Q24) The country met sufficient technical standards to have its data presented beneath the main display of the international report or in an annex.	(Q24) The country met all technical standards required to have its data presented in the main displays of the international report. [13]	This option does not apply to this dimension.
ASSESSMENT QUALITY 2: *Ensuring effective uses of ILSA*			
(Q25-27) Country results and information were not disseminated in the country.[14]	(Q25-27) Country results and information were disseminated using at least one communication strategy.	(Q25-27) Country results and information were disseminated using some communication strategies.	(Q25-27) Country results and information were disseminated using a variety of communication strategies, including dissemination to most schools.
(Q28-29) Country results and information were not covered by media in the country. [15]	(Q28-29) Country results and information were covered by one media outlet in the country.	(Q28-29) Country results and information were covered by some media outlets in the country.	(Q28-29) Country results and information were covered by a variety of media outlets in the country.
(Q30-31) Results from the ILSA have not been used to inform decision making. [16]	(Q30-31) Results from the ILSA have been used in a very limited way to inform decision making in the country.	(Q30-31) Results from the ILSA have been used in some ways to inform decision making in the country.	(Q30-31) Results from the ILSA have been used in a variety of ways to inform decision making in the country.
(Q32) There is no clear evidence or consensus about the positive impact of the ILSA on education quality. [17]	This option does not apply to this dimension.	(Q32) There is a general consensus about the positive impact of the ILSA on education quality.	(Q32) There is evidence of the positive impact of the ILSA on education quality.

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

1. PNG participates in the Pacific Island Literacy and Numeracy Assessment (PILNA), which has 14 participating countries in the Pacific region. The regional trial and analysis of PILNA took place on October 17–20, 2011. The first round of PILNA was then introduced in 2012. This round established a benchmark and PILNA is intended to be implemented every three years moving forward (the next round is scheduled for 2015). Since PILNA results for PNG have not been approved, they have not been disseminated throughout the country. However, these results have been discussed at the top management level and at certain meetings and conferences. The dissemination of results was planned to be carried out in March 2014.

Several quality problems impacted the introduction, administration, and utility of PILNA. Due to the geographic layout of the country, several logistical challenges were encountered. Sampling, for example, was designed to be nationally representative, but only about 50 percent of selected schools were able to implement PILNA in full. As such, the participation rate was quite low. Various other factors, such as weather, transportation, and a disconnect among standards officers at the provincial level also contributed to the poor participation rate. For some schools, materials were delayed due to funding and geographic location. Furthermore, PILNA was administered towards the end of November, when schools were giving priority to examinations. In addition, some schools were unable to return assessment materials to the MSB on time due to lack of transport, hence their participation was not included in the marking, scoring, and analysis process.

2. PNG agreed to participate in the Pacific Islands Literacy and Numeracy Assessment (PILNA) in 2012.

3. The Secretariat of the Pacific Board for Educational Assessment (SPBEA) is a regional body based in Suva, Fiji. It received a mandate from the Pacific Forum Education Ministers Meeting in 2006 to develop regional benchmarking indicators for literacy, numeracy, and life skills, known as PILNA. For this reason, Samoa and other individual countries do not have a specific policy for PILNA; rather, they have agreed to participate. SPBEA developed two central documents: the Pacific Islands Forum Secretariat Education Initiatives under the Pacific Plan (2006), which lays the foundation for the development of PILNA; and the PILNA Implementation Manual (2012), which outlines exam's purpose, framework, approach, and implementation responsibilities.

The official regional policy citation is the Pacific Islands Forum Secretariat Education Initiatives under the Pacific Plan, which is authorized by the Pacific Islands Forum Secretariat (2006). The official document is the PILNA Implementation Manual. The regional authorizing body for PILNA is the SPBEA, which was authorized in 2012.

4. Funding to administer PILNA was obtained from the budget examination allocation. Scoring was funded by the Austrian Aid Programme.

5. PILNA funding was provided by Austrian Aid Programme through its bilateral assistance for the Pacific Benchmarking Education Quality for Results (PaBER) Project. Administration of the assessment was funded by the individual countries who participated, using funds from their education budget allocation for examinations.

6. Funding covered implementation of the assessment exercise in the country, including printing and freighting of assessment materials to the countries. In-country transportation costs were covered by PNG.

Funding included the processing and analysis of data, as well as reporting and dissemination. Representatives from participating countries convened in Suva for pre-analysis of PILNA data from April 8–12, 2013. The pre-analysis involved the inspection of preliminary results based on the draft Data Analysis and Reporting Framework document. The main analysis was done by SPBEA—the specific cost of this exercise was not provided as it constituted part of SPBEA's larger work program. External technical assistance experts reviewed individual country reports and data analysis.

7. A national coordinator oversees PILNA at the national level in PNG, a school coordinator manages administration at the school level for selected schools, and a test supervisor (typically a teacher) administers the assessment at the classroom level in selected schools. The national coordinator resides within the Measurement Service Branch (MSB). In addition, there is a panel leader for each focus area (numeracy and literacy, respectively) at the MSB. Under the guidance of each respective leader, the panels are responsible for marking the exams. The MSB is recognized as the authority in student assessment in PNG.

8. Although computers are up to date and its servers are sufficient for the nature of its work, there are not enough computers for all staff. Furthermore, staff need to travel and there are an insufficient number of laptops to support their work program. In addition, security for the building has been an issue in past; additional storage space and a more organized approach are also required. Other than the lack of computers, the organization has a sufficient number of communication tools.

9. Several opportunities to learn about the introduction of PILNA in PNG were available. Specifically, presentations about the regional assessment were made to senior standards officers from all provinces by MSB staff involved in its administration. In addition, key MSB personnel participated in workshops organized by SPBEA for all countries participating in PILNA. Lastly, SPBEA provided in-country training on scoring to panel leaders and panel members. The primary beneficiaries of the aforementioned opportunities were MSB staff and teachers. A few university students also benefited because they were members of scoring panels.

10. PILNA is a regional tool developed by the SPBEA with the participation of UNESCO. It is designed based on benchmarking indicators agreed upon by each country and endorsed by the Forum Education Ministers Meeting (FEdMM). PILNA is a fairly accurate measure of the content areas and skills areas of the official curriculum and is aligned with official pedagogical approaches.

11. Textbooks and learning resources, together with teachers in school, cover content and skills similar to those covered by PILNA. This is supported by the fact that the content and skills measured by PILNA align closely with the content and skills measured by the CSMT.

12. Classroom assessment is designed, managed, and implemented at the school level. Results from these assessments are not aggregated back to the central level. For this reason, it is impossible to tell whether PILNA is fully consistent with classroom assessment practices. Both the CSMT and PILNA are diagnostic tools; however, the latter assessment reflects general learning norms for the Pacific Region, whereas the CSMT is based on the PNG curriculum.

13. The country met all technical standards required to have its data presented in the main displays of the international PILNA report. The final report is awaiting the approval of FEdMM.

14. Since PILNA results have not been approved, they have not yet been disseminated throughout the country. However, results have been discussed at the top management level and at certain meetings and conferences. Dissemination of results was expected to be carried out in March 2014.

15. The report outlining the 2011 baseline results of the PILNA pilot has also not been finalized or disseminated.

16. Because PILNA was recently implemented and its results have not been disseminated, it's too early to observe any impacts of its implementation.

17. No observed impact has been documented to date.

Acknowledgements

This report was prepared by The World Bank SABER-Student Assessment team in collaboration with Clark Matthews (World Bank Consultant). The report benefitted from the data collection efforts and insight of Adrian Alamu (PaBER Assessment Officer, South Pacific Board for Educational Assessment) and Seema Prasad (PaBER Assessment Officer, South Pacific Board for Educational Assessment), as well as input from the Department of Education of Papua New Guinea.

References

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

PNG (Papua New Guinea). Department of Education, National Executive Council. 2009. Universal Basic Education Plan 2010–2019. National Capital District, Papua New Guinea. http://www.education.gov.pg/QL_Plans/index.

UNESCO (United Nations Educational, Scientific, and Cultural Organization). Apia Office. 2009. UNESCO Country Programming Document: Papua New Guinea 2008–2013. Apia, Samoa. http://unesdoc.unesco.org/images/0018/001835/183587E.pdf.

UNICEF (United Nations Children's Fund). Papua New Guinea: Education. UNICEF, Port Moresby, Papua New Guinea. http://www.unicef.org/png/activities_4369.html.

World Bank. 2012. Papua New Guinea Country Indicator Data. World Bank, Washington, DC.

www.worldbank.org/education/saber

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

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

THE WORLD BANK