



STUDENT ASSESSMENT

SABER Country Report
2017

Key Policy Areas for Student Assessment

1. Classroom Assessment

In Cabo Verde, classroom assessment is supported by formal, system-level documents that provide information on how classroom assessment should be carried out by teachers and how it should be used to evaluate student performance. The use of classroom assessment is well established in the country, and teachers are engaged in positive assessment practices in the classroom. At the same time, there are no mechanisms in place to ensure alignment between classroom assessment activities and the national curriculum, and there is a need to improve the quality of and access to resources and training opportunities to support teachers' classroom assessment practices.

Status

Established

2. Examinations

The Provas Concelhias (Regional Tests) and the Provas Gerais Nacionais (PGN, General National Tests) are administered to students in grades 2, 4, 6, and 8 (in the case of Provas Concelhias), and in grade 12 (in the case of the PGN). In combination with other classroom assessment activities, Provas Concelhias determines if a student can move to the next grade, while the PGN also certifies completion of secondary education. Although there is no official unit within the Ministry of Education in charge of examinations, an individual from the Ministry leads a team to organize and coordinate each of the examination programs every year. Although both examination programs are intended to be based on the official curriculum, in practice these examinations are more aligned with what teachers taught during the school year. No mechanism is in place to ensure the quality of the examinations or their impact on the education system and students.

Emerging

3. National Large-Scale Assessment (NLSA)

Cabo Verde's National Large-Scale Assessment exercise, the Prova de Aferida (Aferida), was administered in 2010 (to grade 6 students) and in 2014 (to students in grades 2, 4, and 6). The Aferida assessed students in Portuguese and math. For these administrations of the Aferida, individuals from the Ministry of Education and the teaching staff were brought together to carry out assessment activities as there is no permanent unit responsible for running the national assessment program. Although the individuals responsible for the Aferida had most of the appropriate resources to carry out the assessment, issues arose with how key assessment activities were implemented, and limited assurance procedures ensured the quality of the Aferida.

Emerging

4. International Large-Scale Assessment (ILSA)

Cabo Verde has not yet participated in an ILSA, and no concrete plans are in place to participate in a future ILSA.

Latent



Table of Contents

Introduction	3
What Is SABER-Student Assessment?	3
Education in Cabo Verde.....	5
Classroom Assessment in Cabo Verde	7
Examinations in Cabo Verde	10
National Large-Scale Assessment in Cabo Verde	14
International Large-Scale Assessment in Cabo Verde	16
Appendix 1: Assessment Types and Their Key Differences	17
Appendix 2: Summary of the Development Levels for Each Assessment Type	18
Appendix 3: Methodology for Assigning Development Levels.....	19
Appendix 4: SABER-Student Assessment Rubrics for Cabo Verde	21
Acknowledgments.....	71
Reference	71

Introduction

Cabo Verde 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. To gain a better understanding of the strengths and weaknesses of its existing assessment system, Cabo Verde decided to benchmark this system using standardized tools developed under the World Bank's Systems Approach for Better Education Results (SABER) program. SABER is an evidence-based program to help countries systematically examine and strengthen the performance of different aspects of their education systems.

What Is SABER-Student Assessment?

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

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

- (i) Providing information on levels of student learning and achievement in the system
- (ii) Monitoring trends in education quality over time
- (iii) Supporting educators and students with real-time information to improve teaching and learning and
- (iv) Holding stakeholders accountable for results.

SABER-Student Assessment Methodology

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

Assessment Types and Purposes

Assessment systems tend to comprise 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, or using results from the assessment; the availability of sufficient and stable sources of funding; and the presence of trained assessment staff.

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

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

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

Table 1: Framework for Building an Effective Assessment System, with Indicator Areas

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

Source: World Bank.

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

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

Levels of Development

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

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

Rubrics are used to evaluate data collected using the standardized questionnaires. 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. 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 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 2.

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. Although 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 3.

Education in Cabo Verde

Cabo Verde is a lower-middle-income country consisting of 10 islands located off the coast of West Africa. Approximately 88 percent of its population (0.5 million) lives on four of the 10 islands.

Cabo Verde has made substantial development progress and is currently the richest country in West Africa and the ninth richest in Sub-Saharan Africa (SSA). In 2016, Cabo Verde's gross national income per capita was US\$ 2,970, almost six times what it was in 1982. The country has also witnessed a sustained decline in

the number of poor, with the incidence of poverty falling from 58 percent in 2001 to 35 percent in 2015, and extreme poverty falling from 30 percent to 10 percent during this same period. Anchored in stable political institutions, the country's economic performance is attributable to significant investment in infrastructure linked to the promotion of the country as a tourist destination.

Cabo Verde has made significant progress in expanding access to education and has achieved nearly universal access to basic education over the past decade. Although preschool education is not compulsory, 85 percent of children in Cabo Verde are enrolled in preschool programs. Basic education is mandatory and free. The Net Enrollment Rate (NER) in primary and secondary levels has improved from 91.7 percent and 59.2 percent in 2006 to 97 percent and 70.3 percent, respectively, in 2015. Cabo Verde is in the process of expanding its basic education system beyond grade 6, to grades 7 and 8, with the goal of achieving 100 percent net enrolment rate (NER) from preschool to grade 8 by 2021.

After completing basic education, students can continue their secondary education in institutions of general secondary education, institutions of technical education, and institutions of professional-vocational education. Students who continue through general secondary education and complete grade 12 must pass the *Prova Geral Nacional* (National General Test) to receive a certificate of general secondary school completion.

Despite having made major progress, the education system in Cabo Verde faces several challenges. Results from the 2010 *Prova de Aferida* (*Aferida*) show that only 27 percent of 6th graders achieved a satisfactory score on the assessment, and that half performed at a level that is considered concerning. The island of residency and the household living conditions are also significant variables affecting student learning outcomes at the primary level. At the secondary level, low internal efficiency is a major concern and is considered the weak link of the education system: it is estimated that 87.5 percent of young people access secondary education, but only 45 percent complete it. Although student retention is high at the primary level (93 percent), it declines in secondary education (85 percent). At the same time, repetition rates are moderate at the primary level (9 percent) and are considerably higher at the secondary level (22

percent). Despite many young people leaving the secondary level without qualification and facing difficulties in the job market, Cabo Verde's Technical and Vocational Education and Training (TVET) system accounts for only 5 percent of enrollment at the secondary level. The current TVET system is characterized by a disperse offering of professional and technical trainings, but it does not constitute a true system that is well articulated with student flows from the secondary level and the job market demands.

In response to these and other challenges, Cabo Verde is in the process of developing its 2017–2021 Education Strategic Plan, which is built around the following three main priorities: (1) gradually increasing universal access to preschool, basic, and secondary school; (2) improving quality and relevance of education services; and (3) improving the efficiency and management of the education sector. Within basic education, the focus is on curriculum reform and the revised education structure (to include grades 1 through 8). At the secondary level, the main objective is to increase access to relevant secondary education aligned with the economic development of the country. As part of the process to improve the quality of education, Cabo Verde is interested in improving its assessment system to ensure that it has better data on student performance that can inform further education reforms and policies.

Detailed information was collected on Cabo Verde's student assessment system using the SABER-Student Assessment questionnaires and rubrics to benchmark it against best practices. Specifically, a local consultant with in-depth knowledge of, and experience with, the education system in Cabo Verde oversaw the completion of the four SABER-Student Assessment

questionnaires: one questionnaire each for Classroom Assessment, Examinations, National Large-Scale Assessment, and International Large-Scale Assessment. The data to complete these questionnaires were obtained through interviews and focus groups with key stakeholders and a review of existing official and technical documents. The information in the completed questionnaires was then applied to the SABER-Student Assessment rubrics (one rubric for each assessment type), and the conclusions of this report were determined on the basis of this analysis. 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 Cabo Verde, particularly by teachers and students in schools. The following sections discuss the findings by each assessment type, accompanied by suggested policy options. The suggested policy options were determined in collaboration with key local stakeholders based on Cabo Verde's immediate interests and needs. Detailed, completed rubrics for each assessment type are provided in Appendix 4.

Classroom Assessment in Cabo Verde

Level of Development: ESTABLISHED

Several formal, official system-level documents provide guidelines for classroom assessment in Cabo Verde, including the “*Programas*,” which outline what students are expected to learn at each subject and each grade level. However, these *Programas* have not been updated since 1999, and it is not clear how they are aligned with the curriculum (which, at the time of data collection, was undergoing significant reform). The *Programas* also do not specify the desired performance level for the subjects and grade levels, and it is not clear if they are used by teachers to help guide the development of classroom assessment activities.

Other official and publicly available documents provide more specific guidance for classroom assessment activities. For example, the *Sistema de Avaliação-Ensino Básico* (2003) and the *Sistema de Avaliação-Ensino Secundário* (2003) provide overall general information on the use of classroom assessment for primary education and secondary education, respectively. In addition, each year, teachers receive a document, *Orientações para a Organização das Provas de Avaliação Final*, with information on what should be assessed, how the information should be used, guidelines for the format of assessment questions, how the assessments should be administered, the subjects to be tested, how assessments should be corrected, and other logistics on carrying out assessment activities. At the time of data collection, the assessment system in Cabo Verde was undergoing a reconfiguration, and a new assessment policy document, *Decreto-lei 71 O Novo Sistema Nacional da Avaliação das Aprendizagens dos Ensinos Básico e Secundário, 2015*, had been authorized introducing significant changes to the assessment system at all levels of education. However, at the time of data collection, this new system had not yet been implemented, and it was unclear how or when the changes reflected in the new policy document would take place. Therefore, the data collected for the SABER-Student Assessment questionnaire reflect the current status of the student assessment system that was authorized by the *Sistema de Avaliação-Ensino Básico/Ensino Secundário* in 2003.

The school year in Cabo Verde is divided into three trimesters. At the end of each trimester, students receive a trimester score that is intended to include information from summative tests (*testes sumativas*) that are developed by a group of teachers from the same grade and subject level in a *conselho* (regional level) and administered two times per a trimester (and collectively make up approximately 80 percent of the trimester grade) and a mix of “other elements of assessment” (OEAs), which consist of attendance, group work, individual work, and class participation. Based on trimester scores, students receive an annual score, the *Classificação Anual*; the first and second trimester scores contribute approximately 30 percent each to the annual score, and the third trimester score contributes approximately 40 percent to the annual score. The third trimester score is weighted more heavily because it includes the marks from the final summative assessment (*Provas Concelhias, Prova Geral Interna [PGI]/Prova Geral Nacional [PGN]*) that takes place during this trimester (for more information, see the examination section below.) The final annual score is calculated out of 20 points.

In practice, however, very few teachers use or apply these OEAs, and there are no specific standards or guidelines concerning how to incorporate the data from these assessments into trimester scores. Consequently, trimester and annual scores are largely determined by the summative classroom assessment information.

Officially, teachers are required to use classroom assessment information to diagnose student learning issues, provide continuous feedback to students as part of instruction, evaluate student performance, and plan next steps in instruction. However, in practice, this does not seem to always occur, and teachers report insufficient training on how to use available classroom assessment information.

Also under this scoring system, classroom assessment activities help to determine whether students have successfully completed certain grade levels. Specifically, at the end of the second, fourth, sixth, and eighth grades, before transitioning to next grade level, students must have an overall annual score of at least 10 out of 20 points. At the end of secondary school (the 12th grade), students must have an overall annual score of at least 10 out of 20 to graduate. Classroom assessments are not, however, used as selection mechanisms for entering higher levels of education.

All schools are required to report on individual student performance in a manner that is accessible to the student and the parents. Grades from the classroom assessments (*testes sumativas*) are posted on the walls in the classrooms, where students and parents can view the results. In addition, after each *teste sumativa*, teachers return the tests to the students and are required to go through the test, review the questions, and correct answers with the students. Although the students' classroom assessment results are not sent to the Ministry of Education, their annual grades (which are based largely on classroom assessment) are sent to the Ministry's National Directorate for Education.

Two formal mechanisms at the system level support developing teachers' competencies in classroom assessment: preservice teacher training and in-service teacher training. Preservice is mandatory and is available annually, albeit perceived to be of low quality. In-service training is ad hoc and also perceived to be of low quality. Some resources are available to teachers for use in their classroom assessment activities, which include the guidelines (*Orientações*), mentioned above, and the *Programas*, specifying what students are expected to learn. Teachers also receive examples of previous classroom assessments (*testes sumativas*) and general scoring criteria. However, teachers report insufficient training and resources on how to use and follow up on classroom assessment information, particularly with regard to OEAs.

Summative assessment is developed at the regional level (the *conselho*) in coordination with a group of teachers from the same subject/grade level. As such, the same summative assessments are given to all students within each grade/subject area in the same *conselho*. The teachers meet regularly to determine their planning schedule (including the teaching schedule and the assessment schedule, because teachers in the same *conselho* follow the same schedules), develop the summative assessments, and establish the grading and scoring criteria used within the *conselho*. However, it was noted that teachers do not receive sufficient guidelines to support the development of these summative assessments to ensure they are aligned with the curriculum. Although teachers are regularly supervised in their use and application of classroom assessment (*testes sumativas*), no mechanism is in place to ensure that the *testes sumativas* being used to evaluate students learning are actually aligned with the curriculum.

Suggested Policy Options

1. As part of the ongoing curriculum revision/update, introduce an official, system-level document that:
 - a. Provides information on what students are expected to learn (particularly in language and math) at all age and grade levels and
 - b. Specifies the desired performance levels (that is, how well students should learn the specified content or skills)

Make this document available to all teachers through their schools as well as via preservice and in-service teacher training and professional development programs.

2. Ensure alignment between the curriculum, the document outlining desired student learning outcomes, and the development of classroom assessments. Ensure that there is alignment between how classroom assessment is intended to be used by teachers and how it is actually used through improved teacher monitoring and teacher performance evaluation.
 3. Provide high-quality training and resources to teachers to help them follow appropriate procedures in developing and implementing classroom assessment activities by:
 - a. Reviewing current pre- and in-service teacher training options (via live, distance learning, or prerecorded courses delivered via computer) focused on building teacher competencies in classroom assessment to ensure they are of high quality and available on an annual basis (especially with the release of the new curriculum)
 - b. Providing clear, official guidance and guidelines to teachers on using assessment results and information (both summative and formative) at the classroom level to improve teaching and student learning outcomes and

- c. Improving teacher monitoring and performance evaluation with a specific focus on how classroom assessment is intended to be used by teachers and how it is actually used at the classroom level.

Examinations in Cabo Verde: *Provas Concelhias*

Level of Development: EMERGING

The *Provas Concelhias* is an examination offered at the end of each two-year education cycle in the second, fourth, sixth, and eighth grades. In each cycle, students take the *Provas Concelhias* in language, math, and science. At the end of the sixth grade, students also take the examination in social sciences. At the end of the eighth grade, French, English, physics, chemistry, and the history and geography of Cabo Verde are part of the testing requirements.

The *Sistema de Avaliação-Ensino Básico* (2003) is the formal, publicly available policy document that authorizes the *Provas Concelhias*. Guidance on the *Provas Concelhias*, including who should be assessed, what should be assessed, the frequency, date, and length of time to administer the examination, number of test questions, structure of the examination, breakdown of multiple-choice versus open-ended questions, and how examination results should be communicated are provided through *Orientações para a Organização das Provas de Avaliação Final*.

No official unit within the Ministry of Education is in charge of examinations or student assessments in general. Within the Núcleo de Ensino Básico (the Basic Education Unit) one person is designated to organize and coordinate the *Provas Concelhias*. This person works with members of other departments within the Ministry of Education, particularly the Núcleo de Gestão e Orientações Escolares (the Unit for School Management and Guidance). This interunit team has primary responsibility for running the *Provas Concelhias*, but it is not exclusively dedicated to the examination activities.

The *Provas Concelhias* are developed, organized, administered, and corrected and the results published at the regional level by Delegações Concelhias. Delegações Concelhias are regional-level delegations consisting of teachers and school directors. *Provas Concelhias* are implemented on a regional level, and thus the examinations differ from each other depending on which *concelho* or regional grouping the school belongs to. Although the *Provas Concelhias* are intended to be based on the official curriculum, in

practice they are more aligned with what teachers actually taught during the school year. Thus, the *Provas Concelhias* differs among schools and *conselhos* depending on whether what was taught is aligned with the official curriculum. There are no official reviews in place to ensure alignment with the curriculum and while the Directorate of National Education is supposed to validate the *Provas Concelhias*, it is unclear what exactly this validation entails.

The *Provas Concelhias* are well aligned with classroom assessment activities in terms of the content and skills being measured. The *Provas Concelhias* follow the format and structure of other trimester and end-of-year summative assessments given throughout the entire education system. Students' *Provas Concelhias* results are incorporated into their third trimester score, which is combined with their first two trimester scores to calculate their overall annual score. In grades 2, 4, 6, and 8, students who receive an annual score of at least 10 out of 20 may advance to the next grade level. Students who achieve an annual score below 7 must retake the subjects in which they scored below a 7. Students who score above 7, but below 10 out of 20, may take a "second chance" test, called the *Provas de Recurso*. If these students do not pass the *Provas de Recurso*, they must repeat the subject or grade. It is important to note that the *Provas Concelhias* are not considered high-stakes. Rather the results of the *Provas Concelhias* are used in part to make a decision regarding the future selection and promotion of students through the education system.

Suggested Policy Options

1. Document the methods and procedures used during the examination. Specifically,
 - a. Document the processes used to construct, analyze, and score questions, items, or tasks, and to set cutoff scores.
 - b. Clarify the validation process used by the Ministry of Education. Outline the criteria used by the Ministry to validate the examination. Document the objectives of the validation process to promote transparency and enhance stakeholder confidence in the credibility of the exam results.
2. Expand upon quality assurance procedures.

- a. Implement an official review of the examination, perhaps by ensuring that the Ministry of Education's validation entails orienting the exam to address the official curriculum.
 - b. Introduce additional quality assurance procedures, such as ensuring that proctors/administrators are trained according to a standard protocol; questions, items, or tasks are piloted before the official examination is administered; and scorers are trained to ensure high interrater reliability. Such procedures will help better ensure the validity, reliability, and comparability of the examination results.
3. Institute mechanisms to monitor the impact of the examination on education quality, equity, and student outcomes, such as introducing oversight committees, expert review groups, funding for independent research and studies (for example, predictive validity studies), and focus groups or surveys of key stakeholders. Findings from these activities could be used to enhance the design, analysis, or use of the examination to maximize its positive impact on students and the education system.
 4. Create a semiautonomous unit/department within the Ministry of Education dedicated to student assessment:
 - a. Ensure that this unit is accountable to an external body.
 - b. Clearly outline the roles and responsibilities of this unit/group with regard to the key types of student assessment activities (classroom, examinations, large-scale assessment).
 - c. Provide training and other learning opportunities to ensure that this unit can adequately lead work in further developing the student assessment system.

Examinations in Cabo Verde: *Prova Geral Nacional*

Level of Development: ESTABLISHED

The *Prova Geral Nacional*—the General National Test—is an examination that students take at the end of the 12th grade. It was first administered in 2003 and has been administered annually thereafter. The purpose of the PGN is to certify students' completion of secondary education. Students in public and private schools take the examination.

The *Sistema de Avaliação-Ensino Secundário* (2003) is the formal, publicly available policy document that authorizes the PGN. Guidance on the PGN, including who should be assessed, what should be assessed, frequency of the examination, date of the examination administration, length of time to administer the PGN, and how examination results should be communicated, are provided in the *Orientações para a Organização das Provas de Avaliação Final*.

No official unit within the Ministry of Education is in charge of examinations or of student assessment in general. Within the Núcleo de Ensino Secundário Geral e Técnico (Technical and General Secondary Education Unit) one person is designated to organize and coordinate the PGN. This person works with members of other departments within the Ministry of Education, particularly the Núcleo de Gestão e Orientações Escolares (Unit for School Management and Guidance). This interunit team has primary responsibility for running the PGN, although it is not exclusively dedicated to the PGN.

Although the PGN is intended to be based on the official curriculum, in practice it is more aligned with what teachers actually taught during the school year. Secondary school teachers propose inputs, such as examination content, to the Ministry of Education's Technical and General Secondary Education Unit, which then uses these inputs to develop a common exam to be administered nationally by all secondary schools. Once the examination has been developed and shared with the secondary schools, teachers are responsible for administration, scoring, and reporting the results to the National Directorate of Education within the Ministry. It has been noted that sometimes secondary schools adapt the PGN to more closely reflect what was taught during that year. These

adaptations are often not shared with the Ministry, and as such, it is not clear in practice how well the PGN is aligned with the curriculum.

For both staff within the Ministry of Education and teachers who administer the PGN, opportunities to learn about the PGN or to build capacity in student assessment in general are very limited. Some workshops are offered to Ministry staff on an ad-hoc basis, and peer-to-peer learning among teachers exists because the PGN is a well-established part of the education system (although this peer-to-peer learning is not systematic.) However, no university/nonuniversity courses, programs, or workshops are available regarding student assessment or topics relevant to the PGN.

The PGN is well aligned with classroom assessment activities in terms of the content and skills being measured. The PGN follows the format and structure of other trimester and end-of-year summative assessments given throughout the education system. Along with the scores from other trimester tests, the PGN results factor into students' "cycle scores" (11th grade final score + 12th grade final score) to determine if they may graduate. The PGN accounts for 22 percent of the cycle score. Students who score above 14 out of 20 on the other components of their cycle score may graduate without taking the PGN. All other students must take the PGN. After taking the PGN, students achieving a cycle score of at least 10 out of 20 may graduate. Students achieving a cycle score of between 7 and 10 are eligible to take a "second chance" test called the *Provas de Recurso*. Students achieving a cycle score below 7 must retake the subject(s) in which they scored below a 7.

Suggested Policy Options

1. When designing the PGN, the team might place more emphasis on ensuring that the examination's content is closely aligned with, and adequately reflects, the official curriculum. (If the official curriculum does not reflect what teachers actually teach, consider revisiting the curriculum and/or what is being taught in the classroom.)
2. Document existing examination methods and procedures, including the construction of test specifications, questions, items, or tasks; test

- assembly; scoring of open-ended questions, items, tasks, or essays; and the setting of cutoff scores.
3. Introduce additional quality assurance procedures, such as training examination administrators according to a standard protocol; piloting questions, items, or tasks before the official examination is administered; and training scorers to ensure high interrater reliability. Such procedures will help better ensure the validity and reliability of the examination results.
 4. Institute mechanisms to monitor the impact of the examination on education quality, equity, and student outcomes, such as introducing oversight committees, expert review groups, funding for independent research and studies (for example, predictive validity studies), and focus groups or surveys of key stakeholders. Findings from these activities could be used to enhance the design, analysis, or use of the examination to maximize its positive impact on students and the overall system.
 5. Create a semiautonomous unit/department within the Ministry of Education dedicated to student assessment:
 - a. This unit needs to be accountable to an external body.
 - b. Clearly outline the roles and responsibilities of this unit/group in regard to all types of student assessment activities (classroom, NLSA, etc.)
 - c. Provide training and learning opportunities to ensure that this unit can adequately lead work in further developing the student assessment system
 6. In line with the updated/revised curriculum, develop an official document that outlines desired student learning outcomes and performance in all subject areas at all grade levels, and ensure alignment between these documents and all forms of student assessments being used in Cabo Verde (*testes sumativas, Aferida, PGN, etc.*)
 7. Ensure that high-quality training opportunities on use of student assessment (particularly classroom assessment) is available on a regular basis (both via pre- and in-service teacher training), with particular emphasis on the upcoming changes related to curriculum and new assessment system. For example:
 - a. Review preservice teacher training options for building competencies in classroom assessment to ensure they are of high quality and available on an annual basis.
 - b. Institute annual in-service teacher training opportunities on classroom assessment (including via live and prerecorded courses delivered via computer) to ensure that teachers have the ability to hone their classroom assessment knowledge and skills every year as needed.
 - c. Ensure that distance learning options are available to all teachers.

National Large-Scale Assessment (NLSA) in Cabo Verde

Level of Development: EMERGING

Cabo Verde's National Large-Scale Assessment exercise, known as *Prova de Aferida*, was administered in 2010 and 2014. The 2010 round was administered to a representative sample of public school students in the sixth grade, and the 2014 round was administered to a sample of public school students in the second, fourth, and sixth grades. Both rounds assessed Portuguese and math. Assessment results were used to monitor education quality, inform policy, and inform pedagogy, although no system is in place to ensure or monitor this impact.

The *Sistema de Avaliação do Ensino Básico Decreto Lei N. 43* (2003) provides an overall definition of a national large-scale assessment and a brief description of how it can be used to inform policy. However, the document does not authorize the *Aferida*, and it does not provide any guidelines such as on the frequency of the assessment, grade levels to be assessed, or the use of results.

The *Aferida* aims to assess overall levels of student achievement in relation to the national curriculum; however, no regular review system is in place to ensure that the *Aferida* is aligned with the curriculum. The format and structure of the *Aferida* are considered similar to the *Provas Concelhias* (the end-of-year summative assessments); however, because the *Provas Concelhias* are developed at the local/regional level, they are more aligned with what teachers actually taught in class, whereas the *Aferida* is developed at the national level and more closely aligned with the curriculum. In theory, if teachers are following the curriculum, all students in Cabo Verde should be exposed to the content and skills measured by the *Aferida* via regular course instruction.

In both 2010 and 2014, schools in the sample were provided with a framework document that provided general information about the *Aferida*, such as when it would take place, materials to be used, duration of the assessment, and basic instructions for teachers on how to administer it. However, preparatory information on the *Aferida* (e.g., sample questions, examples of scoring criteria, information on how to teach the

content to be assessed by the *Aferida*) was not provided to schools ahead of the test administration.

No permanent unit or group is responsible for the *Aferida*. In 2010 and 2014 a temporary group was formed within the Ministry of Education, consisting of full-time staff from various departments within the Ministry who helped to design, lead, and implement the *Aferida*. In general, staff working on the *Aferida* have few relevant qualifications in NLSA, and no formal training associated with the *Aferida* is available. The 2010 round benefited from significant financial and technical support from UNESCO-Dakar, the Brazilian Cooperation, and the University of Beira Interior from Portugal. For the 2010 *Aferida*, UNESCO-Dakar organized a series of capacity-building exercises for Ministry of Education staff; however, there are no ongoing opportunities within the system to learn about NLSA/*Aferida* (such as university courses, workshops, or other programs.)

Although some issues were related to errors in item development, administration (such as fewer students attending the second day of the assessment administration than the first day), and delays in data processing, the 2010 round was considered successfully implemented, and stakeholders perceived the results to be credible. The 2010 results were published in the 2011 RESEN (Report of the State of the National Education System in Cabo Verde) by performance levels at the national level (i.e., percentage of students who performed high, medium, or low), which was made available to the general public. The report did not include scores or performance levels at the individual, school, or regional levels. A report documenting the methods and procedures used for the 2010 round was drafted with support from UNESCO-Dakar; however, it was never finalized or published. The 2014 round was administered fully in-house. This round faced many issues, and its results have not been released.

Suggested Policy Options

1. Prepare an official, system-level policy document authorizing the *Prova de Aferida* and establishing its objectives. This could entail updating the *Sistema de Avaliação do Ensino Básico Decreto Lei N. 43* to include a formal authorization of the NLSA, including the grade levels and topics to be

- assessed. In addition, ensure that this document is made available to the general public (for example, by disseminating hard copy and publishing online).
2. Ensure that there is a permanent unit with primary responsibility for the NLSA, and that it is accountable to an external body. Clearly outline the roles and responsibilities of this unit/group as well as of all individuals responsible for key NLSA activities.
 3. Provide key stakeholders, including staff and teachers, with training and learning opportunities to ensure that they can adequately implement NLSA-related tasks, including administration of the instruments.
 4. Establish regular reviews to ensure that the instruments used for the NLSA are adequately aligned with the school curriculum.
 5. Introduce procedures to monitor the quality, credibility, and impact of the NLSA program in general and the NLSA administration in particular. This could involve bringing in external observers, soliciting feedback from stakeholders and experts, and introducing quality assurance mechanisms such as numbering booklets and introducing double data scoring.

International Large-Scale Assessment (ILSA) in Cabo Verde

Level of Development: LATENT

Cabo Verde has not yet participated in an ILSA. The Ministry of Education has expressed interest in participating in one in the future. The 2017–2021 Education Strategic Plan broadly mentions “developing [an] external assessment model”; however, it does not explicitly mention any plans for the country to participate in an ILSA program. There is no unit within the Ministry directly responsible for student assessment.

Suggested Policy Options

1. Create an enabling context for Cabo Verde to successfully carry out ILSA activities in the mid- to long term:
 - a. Prepare a formal policy document that authorizes ILSA activity and explains its role in supporting improved education quality and learning in Cabo Verde, or update the 2017–2021 Education Strategic Plan to specifically authorize and justify participation in an ILSA.
 - b. Develop a medium- to long-term funding plan for ILSA activities and ensure that sufficient funding is available to carry out all ILSA activities, particularly those essential to the technical integrity of the assessment and the utility of the results.
 - c. Conduct a needs assessment of the organizational (e.g., computers, software, storage facilities, building security, communication tools) and human (e.g., specialists, translators) resources that will be needed to carry out ILSA activities.

Appendix 1: Assessment Types and Their Key Differences

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

Appendix 2: Summary of the Development Levels for Each Assessment Type

Assessment Type	LATENT	EMERGING	ESTABLISHED	ADVANCED
	<i>Absence of, or deviation from, the attribute</i>	<i>On way to meeting minimum standard</i>	<i>Acceptable minimum standard</i>	<i>Best practice</i>
Classroom Assessment	There is no system-wide institutional capacity to support and ensure the quality of classroom assessment practices.	There is weak system-wide institutional capacity to support and ensure the quality of classroom assessment practices.	There is sufficient system-wide institutional capacity to support and ensure the quality of classroom assessment practices.	There is strong system-wide institutional capacity to support and ensure the quality of classroom assessment practices.
Examinations	There is no standardized examination in place for key decisions.	There is a partially stable standardized examination in place and a need to develop institutional capacity to run the examination. The examination typically is of poor quality and is perceived as unfair or corrupt.	There is a stable standardized examination in place. There is institutional capacity and some limited mechanisms to monitor it. The examination is of acceptable quality and is perceived as fair for most students and free from corruption.	There is a stable standardized examination in place and institutional capacity and strong mechanisms to monitor it. The examination is of high quality and is perceived as fair and free from corruption.
National (or System-Level) Large-Scale Assessment	There is no NLSA in place.	There is an unstable NLSA in place and a need to develop institutional capacity to run the NLSA. Assessment quality and impact are weak.	There is a stable NLSA in place. There is institutional capacity and some limited mechanisms to monitor it. The NLSA is of moderate quality and its information is disseminated, but not always used in effective ways.	There is a stable NLSA in place and institutional capacity and strong mechanisms to monitor it. The NLSA is of high quality and its information is effectively used to improve education.
International Large-Scale Assessment	There is no history of participation in an ILSA nor plans to participate in one.	Participation in an ILSA has been initiated, but there still is 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 3: 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 has to be made based on expert judgment. For example, the aforementioned hypothetical country has an “Enabling Context” score of 2.33, corresponding to a preliminary level of development of “Emerging or Established.” Based on qualitative information not captured in the rubric, along with expert judgment, the country team chooses “Emerging” as the most appropriate level.

6. Scores for certain key dimensions under “Enabling Context” (in the case of EXAM, NLSA, and ILSA) and under “System Alignment” (in the case of CLASS) were set as ceiling scores, i.e., the overall mean score for the

particular assessment type cannot be greater than the score for these key dimensions. These key variables include formal policy, regular funding, having a permanent assessment unit, and the quality of assessment practices.

Appendix 4: SABER-Student Assessment Rubrics for Cabo Verde

This appendix provides the completed SABER-Student Assessment rubrics for each type of assessment activity in Cabo Verde. In each row of the rubric, the relevant selection is indicated by a shaded cell. The selection may include a superscript number that refers to the justification or explanation for the selection. The explanation or justification text is located in the “Development-Level Rating Justifications” section at the end of each rubric. If a row includes a superscript, but no shading, this means that insufficient information was available to determine the relevant selection in the row.

Cabo Verde
Classroom Assessment

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Curriculum/Standards	There was no official document at the system level that outlined what students were expected to learn.	There was an official document at the system level, but it provided limited and insufficient information on what students were expected to learn. ¹	There was an official document at the system level that provided sufficient, but not extensive information on what students were expected to learn.	There was an official document at the system level that provided extensive and comprehensive information on what students were expected to learn.
Policy Document	There was no document at the system level that provided guidelines for classroom assessment.	There was a document at the system level that provided guidelines for classroom assessment, but it was either unofficial, not publicly available, or provided limited guidance.	There was an official and publicly available document at the system level that provided sufficient, but not extensive, guidelines for classroom assessment.	There was an official and publicly available document at the system level that provided extensive and comprehensive guidelines for classroom assessment. ²
Resources	There were no resources available to teachers in the system for their use in classroom assessment activities.	There were resources available to teachers in the system for their use in classroom assessment activities, but these resources were not of high quality and were limited in number or availability. ³	There were sufficient high-quality resources available to all or almost all teachers in the system for their use in classroom assessment activities.	There was an extensive number of high-quality resources available to all or almost all teachers in the system for their use in classroom assessment activities.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Teacher Development	There were no formal mechanisms at the system level that supported the development of teachers' competencies in classroom assessment.	There was a minimum number of formal mechanisms at the system level that supported the development of teachers' competencies in classroom assessment, or else the available formal mechanisms were not of high quality or were limited in their availability. ⁴	There were sufficient formal mechanisms of high quality at the system level that supported the development of teachers' competencies in classroom assessment.	There was an extensive number of high-quality, formal mechanisms at the system level that supported the development of teachers' competencies in classroom assessment.
Quality Monitoring	There were no formal mechanisms at the system level to monitor the quality of classroom assessment practices.	There was a minimum number of formal mechanisms at the system level to monitor the quality of classroom assessment practices.	There were sufficient formal mechanisms at the system level to monitor the quality of classroom assessment practices, including inspection/supervision. ⁵	There were extensive formal mechanisms at the system level to monitor the quality of classroom assessment practices, including inspection/supervision.
Report to Stakeholders	Schools were not required to report individual student performance on classroom assessments to any stakeholders.	Schools had minimum requirements to report individual student performance on classroom assessments.	Schools were required to report individual student performance on classroom assessments to the student and their parents. ⁶	Schools were required to report individual student performance on classroom assessments to a variety of relevant stakeholders.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Report Content	Schools were not required to report individual student performance on classroom assessments in particular subject areas.	Schools were required to report individual student performance on classroom assessments in one or two subject areas.	Schools were required to report individual student performance on classroom assessments in more than two subject areas. ⁷	This option does not apply to this indicator.
Report Format	There were no requirements for schools to use specific formats for reporting individual student performance on classroom assessments.	There were requirements for schools to use specific formats for reporting individual student performance on classroom assessments to the students and their parents, but the formats specified did not include written reports and teacher/school meetings.	There were requirements for schools to use specific formats for reporting individual student performance on classroom assessments to the students and their parents, including written reports and teacher/school meetings. There also were requirements for schools to report this information to the school district, Ministry of Education, or equivalent although the format for reporting to these entities was not specified. ⁸	There were requirements for schools to use specific formats for reporting individual student performance on classroom assessments to students, parents, and the school district, Ministry of Education, or equivalent. These included written reports and (in the case of students and parents) teacher/school meetings.
Required Uses	There were no system-level requirements for teachers to use classroom assessment information.	Teachers were required to use classroom assessment information, albeit in a minimal number of ways.	Teachers were required to use classroom assessment information in a sufficient number of ways. ⁹	Teachers were required to use classroom assessment information in extensive ways.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Selection and Certification	At the secondary level, classroom assessment information was not required as an input for certification decisions or for selection to the next level of the education system.	This option does not apply to this indicator.	At the secondary level, classroom assessment information was required to be used as an input for certification decisions or for selection to the next level of the education system. ¹⁰	This option does not apply to this indicator.
Positive Uses	Classroom assessment information was used in positive ways by a marginal number of teachers.	Classroom assessment information was used in positive ways by some teachers.	Classroom assessment information was used in positive ways by most teachers. ¹¹	Classroom assessment information was used in positive ways by all or almost all teachers.
Poor Practices	All or almost all teachers engaged in poor classroom assessment practices.	Many teachers engaged in poor classroom assessment practices.	Only some teachers engaged in poor classroom assessment practices. ¹²	A marginal number of teachers or no teachers engaged in poor classroom assessment practices.

Classroom Assessment: Development-Level Rating Justifications

1. Official documents at the system level outline what students are expected to learn at each subject and each grade level: *Programa de Disciplina de Química* (1999/2000, Ministry of Education), *Programa de Disciplina de Biologia* (1999/2000, Ministry of Education), and *Programa de Disciplina de Historia* 1999/2000, Ministry of Education). These documents do not, however, specify the desired performance level for these subjects and grade levels.
2. Formal, publicly available documents provide guidelines for classroom assessment, including what should be assessed, how the information should be used, guidelines for the format of assessment questions, how the assessments should be administered, the subjects to be tested, how assessments should be corrected, and other logistics on test administration. These formal, publicly available documents are the following: *Orientações para a Organização das Provas de Avaliação Final* (Ministry of Education, 2016); *Sistema de Avaliação-Ensino Secundário* (Ministry of Education, 2003); and *Decreto-lei 71-O Novo Sistema Nacional da Avaliação das Aprendizagens* (Republic of Cabo Verde, 2015).
3. There are high- and medium-quality resources available to teachers for use in their classroom assessment activities. These resources are available to all or almost all teachers. They include a document outlining what students are expected to learn in different subject or topic areas at different grade and/or age levels (of high quality, part of the national curriculum); teacher guides (of medium quality, provided each year to teachers); and scoring criteria for grading student work (of high quality, which the Ministry of Education's National Directorate of Education department sends annually to teachers for the grading of final tests). Most teachers also have access to student textbooks that provide support for classroom assessment (of high quality, most student are provided with textbooks and workbooks that follow the curriculum).

For more information, see *PGN Economia 1a Chamada* and *Grelha PGN Economia 1a Chamada* (for secondary schools) and *Exemplo Matriz Prova 2º ano* and *Exemplo Critérios Classificação Mat. 2º ano* (for primary schools).

Teachers do not have access to the following resources: document(s) outlining the performance level(s) that students are expected to reach in different subject or topic areas at different grade and/or age levels; item banks with examples of questions and tasks to be used for classroom assessment activities; and computer-based testing with instant reports on student performance.

4. Two formal mechanisms at the system level support developing teachers' competencies: preservice teacher training and in-service teacher training that address competencies in classroom assessment. Preservice training is mandatory and available annually, albeit perceived to be of low-quality. In-service training is ad hoc and perceived to be of low quality.

The following are not available as formal system-level mechanisms to support developing teachers' competencies: online resources on classroom assessment (for example, modules, rubrics, criteria for scoring); opportunities to participate in conferences and workshops on classroom assessment; and opportunities to participate in the development or scoring of test questions for large-scale assessments or examinations.

5. Classroom assessment is a required component of school inspections, teacher supervision, and teachers' performance evaluations. An external moderation system is also in place.

At least twice per trimester, summative assessments are administered to all students. Final summative assessments are also administered to all students at the end of the school year. These summative assessments are developed at the regional level (the *conselho*) in coordination with a group of teachers from that *conselho* of the same subject/grade level. Teachers meet regularly to determine their planning schedule, develop the summative and final assessments, and to establish the grading and scoring criteria used within the *conselho*.

Teachers are regularly supervised and evaluated as a requirement of their day-to-day work, and, as such, they also undergo supervision while administering and grading/scoring summative assessments. Specifically, the school director administers classroom observations under the direction of the General Inspectorate of Education.

6. All or almost all schools are required to report on individual student performance in a manner that is accessible to the student and the parent. Cabo Verde accomplishes this by posting grades from classroom assessments on the walls in classrooms, where students and parents can view the results. In addition, after each of these assessments, the teachers return the tests to the students and are required to go through the test, review all questions, and correct answers with the students.

Schools are not required to report results to the school district or Ministry of Education.

7. Schools are required to report individual student performance on classroom assessment in all subjects, which include language and math. Schools are required to report to the student and the student's parent or guardian.
8. Schools are required to provide written reports on individual student performance to both students and their parents/guardians. The students' classroom assessment results are not sent to the Ministry of Education; however, students' final annual grades are sent to the Ministry's National Directorate of Education.
9. Teachers are required to use classroom assessment information to diagnose student learning issues, provide continuous feedback to students as a part of instruction, and to evaluate students' performance. Teachers are not required to use classroom assessment information to plan the next steps in instruction.
10. Classroom assessments are key inputs for student grades. Each year, students earn an annual score, based upon their scores each trimester and the students' final exams. The third trimester score is weighted a bit higher (around 40 percent) to account for the final summative assessments, which take place during the third trimester.

The trimester scores are based largely on classroom assessments. Officially, trimester scores are supposed to be based upon classroom assessments (about 80 percent of the trimester grade) and other elements of assessment (about 20 percent of the trimester grade), which include attendance, participation, group work, and individual work. In practice, however, very few teachers use or apply these other elements of assessments, and no specific standards or guidelines are in place concerning how to incorporate these assessments in trimester scores.

Consequently, trimester scores (and thus annual scores) are largely determined by classroom assessments. Under this scoring system, classroom assessments help measure whether students have successfully completed primary and secondary schools. At the end of the eighth grade, before transitioning to secondary school, students must have an annual score of at least 10 out of 20, which—as described above—is largely based upon classroom assessment results. At the end of secondary school (the 12th grade), students must have an annual score of at least 10 out of 20 to graduate. Classroom assessments are not, however, used as selection mechanisms for entering higher levels of education.

11. All or almost all teachers use classroom assessment to evaluate student performance, most teachers use classroom assessment to diagnose student learning issues, and some teachers use classroom assessment to continuously provide feedback to students as part of instruction and to plan the next steps in instruction.
12. Most teachers use and apply the multiple trimester summative assessments correctly. Teachers receive sufficient guidelines to support the development of assessments to ensure the assessments are aligned with the curriculum. (Summative assessments are developed by teachers at the regional level; teachers from the same region convene to develop, implement, and grade the assessments.) There are no problems with teachers administering the summative tests throughout the year as required. Teachers also receive sufficient guidelines to support the development of other assessments throughout the school year to ensure the assessments that the teacher develops for his or her class are aligned with the curriculum.

Two areas merit attention. First, classroom assessments are intended to include “other elements of assessment,” including individual work, group work, attendance, and participation. In practice, however, few teachers apply or use these other elements of assessment, in part because no standards are in place about how to measure or include them in the overall calculation of classroom assessments. Second, major discrepancies in 2010 sixth-grade *Aferida* test results and 2010 sixth grade classroom assessments have called into question the accuracy of both exams, but especially the 2010 sixth-grade classroom assessment results. Students passed the classroom assessments at greater rates than the *Aferida*, but there is little evidence as to why this happened. It is important to understand why the two tests produced different results, because they are directly linked to sixth-grade graduation rates.

CABO VERDE
Examinations: Provas Concelhias

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Program Stability	No examination program existed at the system level.	An examination program existed at the system level, but it was not sufficiently stable. ¹	A stable examination program had been in place for several years.	A stable examination program had been in place for 10 years or more.
Clarity of Purpose	There were no policy-mandated purposes of the examination.	The examination had clear policy-mandated purposes, but these did not include student certification or selection. ²	The examination had clear policy-mandated purposes that included student certification, selection, or both.	This option does not apply to this indicator.
Policy Document	No policy document authorized the examination program.	An informal/draft policy document authorized the examination program.	A formal/official policy document authorized the examination program, but the document was not available to the general public.	A formal/official policy document authorized the examination program and was available to the general public. ³
Program Guidelines	No official document provided guidelines for the examination program.	An official document provided guidelines for the examination program, but it was missing some key guidelines. ⁴	An official document provided key guidelines for the examination program.	This option does not apply to this indicator.
Stability of Organization	There was no unit with primary responsibility for running the examination program.	There was a unit(s) with primary responsibility for running the examination program, but the unit(s) was temporary or had been in place for less than 5 years. ⁵	There was a permanent unit(s) with primary responsibility for running the examination program that had been in place for 5 or more years.	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Accountability of Organization	There was no unit with primary responsibility for running the examination program, or else the unit responsible was not accountable to a clearly recognized body.	The unit(s) with primary responsibility for running the examination program was accountable to a clearly recognized body within the examination unit. ⁶	The unit(s) with primary responsibility for running the examination program was accountable to a clearly recognized body within the same institution as the examination unit.	The unit(s) with primary responsibility for running the examination program was accountable to a clearly recognized external body.
Organization Resources	The examination unit did not have the appropriate resources.	The examination unit had some of the appropriate resources.	The examination unit had most of the appropriate resources.	The examination unit had all of the appropriate resources. ⁷
Qualifications of Staff	There were no individuals responsible for completing key examination activities.	Some of the individuals responsible for completing key examination activities had the relevant qualifications.	Most of the individuals responsible for completing key examination activities had the relevant qualifications.	All or almost all of the individuals responsible for completing key examination activities had the relevant qualifications. ⁸
Effectiveness of Staff	There were no individuals responsible for completing key examination activities.	The responsible individuals completed key examination activities, but there were significant issues in how these activities were completed.	The responsible individuals completed key examination activities, with only some issues in how these activities were completed.	The responsible individuals completed key examination activities, and there were no issues in how these activities were completed. ⁹
Source of Funding	There was no funding available for examination activities.	The source of funding for the majority of examination activities was loans, credits, grants, or equivalent. ¹⁰	The source of funding for the majority of examination activities was the government's internal funding sources or student fees.	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Activities Funded	There was no funding available for examination activities.	Funding was not sufficient to cover all core examination activities. ¹¹	Funding was sufficient to cover all core examination activities.	This option does not apply to this indicator.
Staff/Teacher Opportunity to Learn	There were no opportunities to learn about the examination.	Opportunities to learn about the examination were minimal, or not of high quality, or did not benefit all key stakeholder groups. ¹²	There were sufficient high-quality opportunities to learn about the examination that were available to key stakeholder groups.	Opportunities to learn about the examination were extensive, of high quality, and benefited key stakeholder groups.
Teacher Participation	Teachers did not perform examination-related tasks.	Teachers performed a minimal number of examination-related tasks.	Teachers performed a sufficient number of examination-related tasks. ¹³	Teachers performed an extensive number of examination-related tasks.
Measuring What Intended	It was not clear what the examination was intended to measure.	There was weak alignment between the examination and what it was meant to measure, or there was no regular review process in place to verify that alignment existed. ¹⁴	The examination measured official learning standards or curriculum, and officially mandated reviews to verify this alignment took place during most examination rounds.	The examination measured official learning standards or curriculum, and officially mandated reviews to verify this alignment took place during all or almost all examination rounds.
Alignment with Other Assessments	The examination was poorly aligned with other types of assessment activities in the system.	The examination was somewhat aligned with other types of assessment activities in the system.	The examination was very aligned with other types of assessment activities in the system. ¹⁵	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Availability of Preparation Materials	There were no materials available to students to prepare for the examination. ¹⁶	Materials to prepare for the examination were available to some or a marginal number of students.	Materials to prepare for the examination were available to most students.	Materials to prepare for the examination were available to all or almost all students.
Quality of Preparation Materials	There were no materials available to students to prepare for the examination. ¹⁷	Minimal material was available to students to prepare for the examination, or the material available was not of high quality.	Sufficient and high-quality material was available to students to prepare for the examination.	Extensive and high-quality material was available to students to prepare for the examination.
Reasons for Not Taking the Examination	All or almost all individuals could not take the examination due to one or more non-examination-relevant reason(s).	Most or some individuals could not take the examination due to one or more non-examination-relevant reason(s).	There were no non-examination-relevant reasons that prevented individuals from taking the examination. ¹⁸	This option does not apply to this indicator.
Quality Assurance	No formal procedures were in place to ensure the quality of the examination.	Formal procedures to ensure the quality of the examination were minimal in nature or not required. ¹⁹	Formal procedures to ensure the quality of the examination were sufficient in nature and required.	Formal procedures to ensure the quality of the examination were extensive in nature and required.
Standardization	The examination was not standardized at the system level. ²⁰	The examination was partially standardized at the system level, or minimal or no procedures were in place to ensure standardization.	The examination was fully standardized at the system level, and sufficient procedures were in place to ensure standardization.	The examination was fully standardized at the system level, and extensive procedures were in place to ensure standardization.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Quality Processes	Many errors or delays in activities took place that affected the examination to a great extent. ²¹	Errors or delays in activities affected the examination to a significant level.	Any errors or delays in activities had only a minimal effect on the examination.	Errors or delays in activities did not affect the examination.
Inappropriate Behavior	Inappropriate behavior compromised the credibility of the examination to a great extent.	Inappropriate behavior took place and compromised the credibility of the examination somewhat.	Inappropriate behavior was low and did not compromise the credibility of the examination.	Inappropriate behavior, if any, was marginal and did not compromise the credibility of the examination. ²²
Credibility of Results	The results of the examination were perceived as credible by very few stakeholder groups.	The results of the examination were perceived as credible by some stakeholder groups.	The results of the examination were perceived as credible by most stakeholder groups.	The results of the examination were perceived as credible by all or almost all stakeholder groups. ²³
Confidentiality of Results	There was no official policy to keep student results confidential, and student results were not kept confidential in practice. ²⁴	Confidentiality of student results was partially accomplished.	There was an official policy to keep student results confidential, and student results were kept confidential in practice.	This option does not apply to this indicator.
Official Recognition of Results	Examination results were not officially recognized by educational institutions or employers in other countries. ²⁵	This option does not apply to this indicator.	Examination results were officially recognized by educational institutions or employers in other countries.	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Post-Examination Options for Students	No options were available to students after they had taken the examination.	Minimal options were available to students after they had taken the examination.	Sufficient options were available to students after they had taken the examination. ²⁶	Extensive options were available to students after they had taken the examination.
Methods and Procedures Documentation	There was no documentation on the methods and procedures used during the examination. ²⁷	There was minimal documentation on the methods and procedures used during the examination, or the documentation that existed was not public.	There was sufficient and public documentation on the methods and procedures used during the examination.	There was extensive and public documentation on the methods and procedures used during the examination.
Impact Monitoring	No mechanisms were in place to monitor the impact of the examination. ²⁸	Minimal mechanisms were in place to monitor the consequences of the examination, or the mechanisms took place in only some or a few examination rounds.	Sufficient mechanisms were in place to monitor the impact of the examination, and the mechanisms took place in all or almost all examination rounds.	Extensive mechanisms were in place to monitor the impact of the examination, and the mechanisms took place in all or almost all examination rounds.
Readiness to Start an Examination Program	The system was weakly prepared to start an examination program in the future.	The system was somewhat prepared to start an examination program in the future.	The system was well prepared to start an examination program in the future.	This option does not apply to this indicator. ²⁹

Examinations: Development-Level Rating Justifications

1. The *Provas Concelhias* is an examination offered once per year at the end of each education cycle in the second, fourth, sixth, and eighth grades. (In higher grades, there is an examination called the *Prova Geral Interna*, which is administered at the end of each year of secondary education.) Students take the *Provas Concelhias* in each key subject area, including language, math, and science. Additional areas tested vary across education cycles: in the sixth grade, the *Provas Concelhias* also tests social sciences, and in the eighth grade, the *Provas Concelhias* also tests French, English, history and geography of Cabo Verde, physics, and chemistry. Students' *Provas Concelhias* results are incorporated into their third trimester score, which is combined with the first two trimester scores to calculate students' final scores for the school year, their "annual scores."

The *Provas Concelhias* are offered two times per year, in two "waves." The first wave is mandatory for all students, but if students miss the first wave, they have the opportunity to take the test during the second wave. If students do not take the *Provas Concelhias* (if they miss both waves), they will receive a zero on the exam, and this will be incorporated into their final grade for the school year.

The *Provas Concelhias* are developed, organized, administered, and corrected and results are published at the regional level by the Delegações Concelhias, which are groups of schools arranged by region and proximity. (Each school in Cabo Verde belongs to a Delegação Concelhia.) To accomplish these tasks, each Delegação Concelhia leader convenes teams made up of teachers and school directors within the Delegação Concelhia. Together, they develop, organize, administer, correct, and publish the examination and its results.

The dates for the exam are set at the national level by the Ministry of Education's National Directorate of Education. The Directorate is also responsible for developing the general orientation of the examination (such as establishing the examination date, duration, and materials permitted to be used on the examination), general information on the structure of the examination structure (including subjects to be tested in each grade level, the number of test questions/items, and the breakdown of multiple choice vs. open-ended questions), and validation of the examination. (It is not clear, however, what this validation entails.)

Information on when this examination was administered for the first time is not available.

2. The policy-mandated purpose of the *Provas Concelhias* is to be an input for third trimester grades, which factor into students' annual grades. Students' annual grades, awarded at the end of the school year, determine whether students may pass to the next grade level.

Policy-mandated purposes of the exam do not include the following: certifying student completion of primary or secondary education, selecting students into tracks as they pursue further education (for example, academic, vocational); selecting students into tertiary education; monitoring education quality; holding the government, schools, or teachers accountable; evaluating interventions aimed at improving student learning; and informing policy and pedagogy.

3. The *Sistema de Avaliação-Ensino Básico* is a system-level policy document that authorized the *Provas Concelhias*. This document was authorized by the Ministry of Education in 2003. It is a formal/official document and is available to the general public. This document is currently undergoing changes according to the 2015 Regulation of the New Assessment System.
4. An official document, *Orientações para a Organização das Provas de Avaliação Final*, provided guidelines for the examination program. These guidelines addressed the governance of the examination, what should be assessed, who should be assessed, the frequency of the examination administration, and how results should be communicated to stakeholders. The document did not offer guidelines on how assessment results should be used.
5. No official unit holds primary responsibility for running the examination program; however, people from different departments work together to run the examination. In the Núcleo de Ensino Básico (Primary School Department within the Ministry of Education), one person serves as a focal point who leads the work and organization of the *Provas Concelhias* within the Ministry. This person works with other departments within the Ministry, particularly the Núcleo de Gestão e Orientações Escolares (School Management Department) to provide general administrative guidelines, overarching guidelines on test structure, and validation of the tests. (It is not clear, however, what this validation entails.)

At the local level, in Delegações Concelhias, teachers and school directors develop the *Provas Concelhias* examinations to align with what has been taught during the year. The Delegações Concelhias teams are in charge of the elaboration of the exam, sending in the exams for validation by the Ministry of Education, organizing the exam's administration, correcting the exam, publishing the exam's results, and treating/cleaning the exam's final results.

6. There is no official examination unit; however, individuals within the Ministry of Education and at the local level work together on the examination.
7. Individuals responsible for examination activities had all the appropriate resources, including computers for all technical staff, software, building security, storage facilities, computer servers, and communication tools.
8. Individuals responsible for the examination consisted of permanent staff from different units, all or almost all of whom had relevant qualifications, and teachers completing examination activities as part of their job responsibilities, most of whom had relevant qualifications.
9. Individuals were responsible for completing key examination activities, specifically permanent staff within the Ministry of Education (which does not have an official examination unit) and teachers completing examination activities as part of their job responsibilities. No significant issues were found with their effectiveness, including in the organization, design, administration, or grading of the examination or in disseminating the results.
10. Funding was available for examination activities; however, no information is available about the source of funding, the amount of funding, what the funding covers, or the regularity of funding.

11. Funding was available for examination activities; however, no information was available about the source of funding, the amount of funding, what the funding covers, or the regularity of funding.
12. Opportunities were offered to learn about the examination in preservice training. Teachers receive a general training in assessment, which addresses the *Provas Concelhias*. In addition, internal workshops on assessment within the Ministry of Education are occasionally organized.
13. Teachers perform sufficient examination-related tasks, including selecting and/or creating the examination questions, items, or tasks; selecting and/or creating examination scoring guides; administering the examination; scoring the examination; and supervising the examination process. Teachers do not, however, act as judges (for example, in orals) or resolve inconsistencies between examination scores and school grades (moderation).
14. The examination is intended to measure the official curriculum. No official or officially mandated reviews are in place.

In practice, whether the examination is aligned with the official curriculum can vary given that it is aligned with what teachers taught during the school year, and whether or not what they taught aligned with the official curriculum. Although the Ministry of Education is responsible for validating the examination, it does not seem to entail ensuring that the examination is aligned with the official curriculum. (In fact, there is little clarity about what the validation process entails exactly.)

15. The *Provas Concelhias* are very aligned with classroom assessments—specifically the trimester summative assessments given throughout the education system.
16. No materials are available to students to prepare for the examination; Cabo Verde considers the *Provas Concelhias* to be a regular part of its classroom assessment system and does not provide special materials so that students can prepare for this exam.
17. No materials are available to students to prepare for the examination; Cabo Verde considers the *Provas Concelhias* to be a regular part of its classroom assessment system and does not provide special materials so that students can prepare for this exam.
18. No non-examination-relevant reasons prevent individuals from taking the examination; all students can participate in the *Provas Concelhias*.
19. A standardized manual for examination administrators, called the *Orientations for the End of the Year Assessment*, was the only formal quality assurance procedure in place to ensure the quality of the examination.

Other forms of quality assurance procedures that were not present include proctors/administrators trained according to protocol; questions, items, or tasks piloted before the official examination was administered; all booklets numbered; double data scoring; scorers trained to ensure high interrater reliability; double processing of data; and external or internal observers.

20. The examination was not standardized at the system level. Assessment design, administration, and scoring varied across students in the same examination round. This was because the examination was standardized at the regional levels, rather than at the national level. Thus, students within the same region experienced similar conditions, but students across regions did not necessarily experience similar conditions.

The following procedures were not in place to ensure standardization of the examination:

- Examination papers and questions, items, or tasks were not the same or equivalent for all students
- Examination administrators were not trained to ensure that all students took the examination under the same conditions
- Quality control monitors and/or observers were not used to ensure the same administration conditions in all locations where the examination was administered
- The same scoring criteria were not used to correct the examination questions, items, or tasks and
- Examination results were not computed using the same procedures for all students.

21. There is not enough information to answer this because all *Provas Concelhias* are implemented on a regional level and differ.

As mentioned above, the Ministry of Education (and, specifically, the National Directorate of Education) is responsible for developing administrative guidelines for the examination, general guidelines for the examination structure, and validation of the examinations. It is unclear, however, what exactly this validation entails. At the regional level, Delegações Concelhias are in charge of exam elaboration, validation, organization, results publication, and the treating/cleaning of final results.

22. Inappropriate behavior did not take place during the examination round.

23. The results of the examination were perceived to be credible by all or almost all stakeholder groups.

24. Student names and results were supposed to be (and were, in practice) publicly accessible. Students' grades were posted in the school lobby. Keeping students results confidential is neither an official policy nor implemented in practice.

25. Examination results were not officially recognized by educational institutions or employers in other countries.

26. If students did not pass the examination, they were not required to leave the education system. Instead, students could attend remedial education and then take a "second chance" test called the *Provas de Recurso*, which is available to students whose overall, final score for the school year is lower than 10 but greater than 7. (A score of 10 is needed to progress to the next grade, and a score below 7 means that the student must retake the subject in which they scored less than 10). If a student is eligible for the *Provas de Recurso*, the second chance test, they also have the right to take two weeks of remedial classes, organized by the school, to prepare for the *Provas de Recurso*. After the student takes the *Prova de Recurso*, their end-of-year score is recalculated using the score from the *Prova de Recurso*. Students could also repeat the grade.

The following options were not available to students after they took the examination and received their results:

- Students could not apply to tertiary education institutions
- Students could not apply to secondary education institutions
- Students could not retake the exact same examination and
- Students could not take courses in preparation for the *Provas Concelhias*.

27. There was no documentation that specified the methods and procedures used during the examination round.

Specifically, the following information on the methods and procedures were not documented:

- Test specifications
- Construction of questions, items, or tasks
- Pilot testing of questions, items, or tasks
- Analysis of piloted questions, items, or tasks
- Test assembly
- Marking or scoring of open-ended questions, items, tasks, or essays
- Scoring of examination questions, items, or tasks
- Reliability
- Scaling and
- Setting cutoff scores.

28. No mechanisms are in place to monitor the impact of the examination.

Specifically, the following mechanisms to monitor the impact of the examination were not in place:

- Oversight committee(s)
- Expert review groups
- Funding for independent research on the examination
- Studies (for example, predictive validity) on the examination and
- Focus groups or surveys of key stakeholders.

29. This indicator does not apply to this rubric.

CABO VERDE
Examinations: Prova Geral Nacional

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Program Stability	No examination program existed at the system level.	An examination program existed at the system level, but it was not sufficiently stable.	A stable examination program had been in place for several years.	A stable examination program had been in place for 10 years or more. ¹
Clarity of Purpose	There were no policy-mandated purposes of the examination.	The examination had clear policy-mandated purposes, but these did not include student certification or selection.	The examination had clear policy-mandated purposes that included student certification, selection, or both. ²	This option does not apply to this indicator.
Policy Document	No policy document authorized the examination program.	An informal/draft policy document authorized the examination program.	A formal/official policy document authorized the examination program, but the document was not available to the general public.	A formal/official policy document authorized the examination program and was available to the general public. ³
Program Guidelines	No official document provided guidelines for the examination program.	An official document provided guidelines for the examination program, but it was missing some key guidelines. ⁴	An official document provided key guidelines for the examination program.	This option does not apply to this indicator.
Stability of Organization	There was no unit with primary responsibility for running the examination program.	There was a unit(s) with primary responsibility for running the examination program, but the unit(s) was temporary or had been in place for less than 5 years.	There was a permanent unit(s) with primary responsibility for running the examination program that had been in place for 5 or more years. ⁵	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Accountability of Organization	There was no unit with primary responsibility for running the examination program, or else the unit responsible was not accountable to a clearly recognized body.	The unit(s) with primary responsibility for running the examination program was accountable to a clearly recognized body within the examination unit.	The unit(s) with primary responsibility for running the examination program was accountable to a clearly recognized body within the same institution as the examination unit. ⁶	The unit(s) with primary responsibility for running the examination program was accountable to a clearly recognized external body.
Organization Resources	The examination unit did not have the appropriate resources.	The examination unit had some of the appropriate resources.	The examination unit had most of the appropriate resources.	The examination unit had all of the appropriate resources. ⁷
Qualifications of Staff	There were no individuals responsible for completing key examination activities.	Some of the individuals responsible for completing key examination activities had the relevant qualifications.	Most of the individuals responsible for completing key examination activities had the relevant qualifications.	All or almost all of the individuals responsible for completing key examination activities had the relevant qualifications. ⁸
Effectiveness of Staff	There were no individuals responsible for completing key examination activities.	The responsible individuals completed key examination activities, but there were significant issues in how these activities were completed.	The responsible individuals completed key examination activities, with only some issues in how these activities were completed.	The responsible individuals completed key examination activities, and there were no issues in how these activities were completed. ⁹
Source of Funding	There was no funding available for examination activities.	The source of funding for the majority of examination activities was loans, credits, grants, or equivalent. ¹⁰	The source of funding for the majority of examination activities was the government's internal funding sources or student fees.	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Activities Funded	There was no funding available for examination activities.	Funding was not sufficient to cover all core examination activities. ¹¹	Funding was sufficient to cover all core examination activities.	This option does not apply to this indicator.
Staff/Teacher Opportunity to Learn	There were no opportunities to learn about the examination.	Opportunities to learn about the examination were minimal, or not of high quality, or did not benefit all key stakeholder groups. ¹²	There were sufficient high-quality opportunities to learn about the examination that were available to key stakeholder groups.	Opportunities to learn about the examination were extensive, of high quality, and benefited key stakeholder groups.
Teacher Participation	Teachers did not perform examination-related tasks.	Teachers performed a minimal number of examination-related tasks.	Teachers performed a sufficient number of examination-related tasks. ¹³	Teachers performed an extensive number of examination-related tasks.
Measuring What Intended	It was not clear what the examination was intended to measure.	There was weak alignment between the examination and what it was meant to measure, or there was no regular review process in place to verify that alignment existed. ¹⁴	The examination measured official learning standards or curriculum, and officially mandated reviews to verify this alignment took place during most examination rounds.	The examination measured official learning standards or curriculum, and officially mandated reviews to verify this alignment took place during all or almost all examination rounds.
Alignment with Other Assessments	The examination was poorly aligned with other types of assessment activities in the system.	The examination was somewhat aligned with other types of assessment activities in the system.	The examination was very aligned with other types of assessment activities in the system. ¹⁵	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Availability of Preparation Materials	There were no materials available to students to prepare for the examination.	Materials to prepare for the examination were available to some or a marginal number of students.	Materials to prepare for the examination were available to most students.	Materials to prepare for the examination were available to all or almost all students. ¹⁶
Quality of Preparation Materials	There were no materials available to students to prepare for the examination.	Minimal material was available to students to prepare for the examination, or the material available was not of high quality. ¹⁷	Sufficient and high-quality material was available to students to prepare for the examination.	Extensive and high-quality material was available to students to prepare for the examination.
Reasons for Not Taking the Examination	All or almost all individuals could not take the examination due to one or more non-examination-relevant reason(s).	Most or some individuals could not take the examination due to one or more non-examination-relevant reason(s).	There were no non-examination-relevant reasons that prevented individuals from taking the examination. ¹⁸	This option does not apply to this indicator.
Quality Assurance	No formal procedures were in place to ensure the quality of the examination.	Formal procedures to ensure the quality of the examination were minimal in nature or not required. ¹⁹	Formal procedures to ensure the quality of the examination were sufficient in nature and required.	Formal procedures to ensure the quality of the examination were extensive in nature and required.
Standardization	The examination was not standardized at the system level.	The examination was partially standardized at the system level, or minimal or no procedures were in place to ensure standardization.	The examination was fully standardized at the system level, and sufficient procedures were in place to ensure standardization. ²⁰	The examination was fully standardized at the system level, and extensive procedures were in place to ensure standardization.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Quality Processes	Many errors or delays in activities took place that affected the examination to a great extent.	Errors or delays in activities affected the examination to a significant level.	Any errors or delays in activities had only a minimal effect on the examination. ²¹	Errors or delays in activities did not affect the examination.
Inappropriate Behavior	Inappropriate behavior compromised the credibility of the examination to a great extent.	Inappropriate behavior took place and compromised the credibility of the examination somewhat.	Inappropriate behavior was low and did not compromise the credibility of the examination.	Inappropriate behavior, if any, was marginal and did not compromise the credibility of the examination. ²²
Credibility of Results	The results of the examination were perceived as credible by very few stakeholder groups.	The results of the examination were perceived as credible by some stakeholder groups.	The results of the examination were perceived as credible by most stakeholder groups.	The results of the examination were perceived as credible by all or almost all stakeholder groups. ²³
Confidentiality of Results	There was no official policy to keep student results confidential, and student results were not kept confidential in practice. ²⁴	Confidentiality of student results was partially accomplished.	There was an official policy to keep student results confidential, and student results were kept confidential in practice.	This option does not apply to this indicator.
Official Recognition of Results	Examination results were not officially recognized by educational institutions or employers in other countries. ²⁵	This option does not apply to this indicator.	Examination results were officially recognized by educational institutions or employers in other countries.	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Post-Examination Options for Students	No options were available to students after they had taken the examination.	Minimal options were available to students after they had taken the examination.	Sufficient options were available to students after they had taken the examination. ²⁶	Extensive options were available to students after they had taken the examination.
Methods and Procedures Documentation	There was no documentation on the methods and procedures used during the examination. ²⁷	There was minimal documentation on the methods and procedures used during the examination, or the documentation that existed was not public.	There was sufficient and public documentation on the methods and procedures used during the examination.	There was extensive and public documentation on the methods and procedures used during the examination.
Impact Monitoring	No mechanisms were in place to monitor the impact of the examination. ²⁸	Minimal mechanisms were in place to monitor the consequences of the examination, or the mechanisms took place in only some or a few examination rounds.	Sufficient mechanisms were in place to monitor the impact of the examination, and the mechanisms took place in all or almost all examination rounds.	Extensive mechanisms were in place to monitor the impact of the examination, and the mechanisms took place in all or almost all examination rounds.
Readiness to Start an Examination Program	The system was weakly prepared to start an examination program in the future.	The system was somewhat prepared to start an examination program in the future.	The system was well prepared to start an examination program in the future.	This option does not apply to this indicator. ²⁹

Examinations: Development-Level Rating Justifications

1. The *Prova Geral Nacional* examination (PGN, the General National Test) was first administered in 2003 and has been administered once a year since then.
2. The policy-mandated purposes of the examination are to (1) certify student completion of secondary education and (2) make pass/fail decisions in transitioning students from one grade to another.
3. *Sistema de Avaliação-Ensino Secundário* is a system-level policy document authorized by the Ministry of Education in 2003. This document is formal/official and is available to the general public.
4. An official document, *Orientações para a Organização das Provas de Avaliação Final*, provides guidelines for the examination, including guidelines on the governance of the examination, what should be assessed, who should be assessed, the frequency of the examination administration, and how results should be communicated to stakeholders. It does not provide guidelines on how results should be used, on funding for the examination, on confidentiality of results, or on how results can be accessed by stakeholders.
5. Within the Ministry of Education's Núcleo de Ensino Secundário Geral e Técnico (Technical and General Secondary Education Unit, part of the Ministry), one person leads a team to organize and coordinate the PGN. This person works with members of other departments within the Ministry, particularly the Núcleo de Gestão e Orientações Escolares (the School Management and Guidance Unit). This interunit team has primary responsibility for running the PGN examination program. Although this team is not exclusively dedicated to the PGN, it is permanent. The team has overseen all examination rounds since 2003.
6. There is an interunit team led by an individual in the Núcleo de Ensino Secundário Geral e Técnico, which holds primary responsibility for running the PGN. This team is accountable to the Ministry of Education. It is not accountable to an external board or committee that is institutionally separate from the team in charge of the examination.
7. The examination unit had all the appropriate resources: computers for all technical staff, software (such as statistical packages), building security, storage facilities, computer servers, and communications tools (phone, internet, email).
8. Two types of individuals were responsible for completing key examination activities: (1) permanent staff from different units within the Ministry of Education (including the Núcleo de Ensino Secundário Geral e Técnico and Núcleo de Gestão e Orientações Escolares) and (2) secondary school teachers completing examination activities as part of their job responsibilities. Almost all permanent staff of the different Ministry units were qualified, and most of the teachers had relevant qualifications as well.

Secondary school teachers are responsible for proposing exam content to the Ministry of Education, on an annual basis. The Ministry is responsible for developing the final PGN and setting the exam times and dates. Once the exam has been developed, the teachers are responsible for administering the exam, correcting the exam, and reporting exam results.

9. For both groups of individuals responsible for completing key examination activities (permanent staff in the Ministry of Education and teachers completing examination activities as part of their job responsibilities), no significant issues were found with their effectiveness or issues that affected the quality of specific examination activities or the overall quality of the examination.
10. Funding was available for examination activities; however, there is insufficient evidence to describe the funding's source or allocation. (It is likely that the sources of funding for the majority of examination activities are the government's internal funding sources and/ or student fees.)
11. Funding was available for examination activities; however, there is insufficient evidence to describe the funding's allocation or sufficiency.
12. Opportunities to learn about the PGN were available through organized internal workshops for Ministry of Education staff. These workshops, which were of medium quality, discussed the content and skills measured by the examination. In addition, there is peer-to-peer learning among teachers (addressing administering the PGN, grading, and interpretation of scores), because the PGN is a well-established part of the education system. (This peer-to-peer learning, however, is not systematic.)

The following opportunities were not available for learning about the PGN:

- University graduate programs (master's or doctoral level) on student assessment that include topics relevant to the examination (for example, test design, administration)
- University courses and/or workshops on the content and skills measured by the examination (for example, courses on curriculum)
- University courses and/or workshops on examination topics other than the content and skills measured by the examination (for example, test design, reporting)
- Non-university courses and/or workshops on examination topics other than the content and skills measured by the examination (for example, test design, reporting)
- Funding for attending international programs, courses, and workshops on student assessment that cover topics relevant to the examination
- Internships and/or short-term employment in the unit running the examination and
- Presentations about the examination (for example, presentations on test design, administration).

In addition to including more of the opportunities listed above, it is recommended that these opportunities extend to sufficient beneficiaries, including full-time staff on the Ministry of Education's examination team, primary school teachers, and secondary school teachers.

13. Teachers performed a sufficient number of examination-related tasks, including creating examination questions, items, or tasks; administering the examination; and scoring the examination. Specifically, teachers send proposals to the Ministry of Education’s Directorate of Education outlining items to include in each subject area. (The Directorate is in charge of primary and secondary education, including assessment.) The Directorate determines and creates the final PGN and its scoring guide. Teachers are responsible for administering the PGN and scoring the results under the supervision of the Ministry’s General Inspectorate department, which is in charge of evaluating, controlling, and supervising the overall functioning of the education system. After teachers score the PGN, students’ results are incorporated into their trimester and annual scores. The actual test is not returned to the students, but an answer guide is posted in the school lobby.

Examination tasks that are not performed by teachers include acting as judges (for example, in orals), supervising examination procedures, and resolving inconsistencies between examination scores and school grades (moderation).

14. The examination is intended to measure the official curriculum. In practice, however, the PGN is only somewhat aligned with the official curriculum, because the examination is based on the content that teachers propose and not necessarily on the national curriculum. Secondary schools submit proposals to the Directorate of Education, outlining items they would like to include on the PGN. Subsequently, the Directorate drafts the exam.

There are officially mandated, external, regular reviews of the PGN that have taken place during almost all examination rounds. The National Directorate of Education hires an external consultant to conduct a final review and verification of the PGN. The external consultant conducts this review after the Directorate receives proposals from the secondary schools and develops a common PGN. After the consultant reviews and verifies the draft PGN, the Directorate prepares the final PGN and sends it to the secondary schools to administer. This external review of the exam is systematic and takes place every year.

15. Classroom assessment activities are very aligned with the PGN. Specifically, the PGN is similar to the type of trimester and end-of-year summative assessments given throughout the entire education system.

16. Materials available to prepare for the examination were (1) the official framework document, which explained what was measured on the examination and (2) examples of the types of questions that were on the examination provided by the examination unit and/or authority overseeing the examination. These materials were available in school to all or almost all students. Specifically, each April, all schools publicly display the “reference matrix,” which outlines the content and objectives for each subject to be covered by the PGN, when the PGN will take place, and the materials required for the PGN. In addition, teachers use past PGNs to help students prepare for the examination.

17. Materials available to prepare for the examination were (1) the official framework document, of medium quality and (2) examples of the types of questions that were on the examination, which were of high quality.

18. All students can participate in the PGN. There are no non-examination-relevant reasons that prevent individuals from taking the examination. Students who scored above a grade of 14 out of 20 on their cycle score do not have to take the PGN, however. (The cycle score is the total 11th grade score plus the total 12th grade score.)

19. Minimal quality assurance procedures were in place to ensure the quality of the examination: a standardized manual was available for examination administrators, required for every examination round, and internal observers were present at some schools. The internal observation was conducted by the General Inspectorate of Education, which goes to some schools during the PGN to observe the administration of the assessment. Although observers are not at each classroom undertaking the PGN, PGN protocol ensures that each classroom undertaking the PGN is supervised by two supervising teachers.

Quality assurance procedures not in place include proctors/administrators trained according to protocol; questions, items, or tasks piloted before the official examination was administered; numbered booklets; double data scoring; training of scorers to ensure high interrater reliability; and double processing of data.

20. The examination was fully standardized at the system level. The PGN is designed by the Ministry of Education's National Directorate of Education, with input from secondary school teachers. The Directorate provides the PGN to all schools. All students take the PGN on the same day and at the same time. All schools use the same formula to calculate the students' score on the exam. Moreover, sufficient procedures were in place to ensure the standardization of the examination: (1) examination papers and questions, items, and tasks were the same for all students; (2) the same scoring criteria were used to correct the examination questions, items, and tasks; (3) examination results were computed using the same procedures for all students; and (4) examination results were reported to all students in the same way.

21. Errors in the printing of test booklets have somewhat affected the examination. Specifically, there are sometimes problems with the printing or formatting of the questions on the PGN such that material cannot be read easily. As a result, the PGN is resent to the national level for correction. This has sometimes caused administrative delays but never affected the exam timing; the corrections were all conducted in advance of the PGN administration and never caused schools to postpone the administration of PGN to a different day. All schools administered the tests on the same day and using the same examination booklets.

22. Inappropriate behavior did not take place during the examination round.

23. The results of the PGN examination were perceived to be credible by almost all stakeholder groups.

24. No official policy was in place to keep student results confidential, and, in practice, student results were not kept confidential. Student names and results were supposed to be and, in practice, were publicly accessible. Specifically, students' grades on the PGN are posted in the school lobby.

25. Examination results were not officially recognized by educational institutions or employers in other countries.

26. Sufficient options exist for students after they had taken the examination and received their results: students could ask for remedial education, or students could repeat the grade. Students who failed the PGN or did not reach a minimum level or score did not have to leave the education system. The specific guidelines are as follows:

- Students with an annual score of at least a 10 (out of 20) at the end of the year may graduate from whatever grade they are in. If students have an annual score of above a 10 at the end of 12th grade, they may graduate from secondary school (9th–12th grade) and apply to higher education institutions.
- Secondary school (9th–12th grade) students whose annual scores are between 7 and 10 are eligible to take a “second chance” test called the *Provas de Recurso*.
- Students whose annual scores are below 7 must retake the subject(s) in which they scored below 7.
- Students do not have the option to retake the PGN examination itself.
- The *Prova de Recurso* for 9th to 11th grade levels is developed at the school level. The *Provas de Recurso* for 12th grade are developed by the Directorate of Education, and students can take the *Prova de Recurso* in only three subjects at a time. All students have the right to take remedial classes for two weeks to prepare for the *Provas de Recurso*. Remedial classes are organized by the school. After students take the *Prova de Recurso*, the students’ annual grade in that subject is calculated as 40 percent of the student’s overall grade in that subject and 60 percent of the student’s grade in the *Prova de Recurso*.

27. No document specified the methods and procedures used during the examination round.

28. No mechanisms are in place to monitor the impact of the examination.

29. This indicator does not apply to this rubric.

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

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Program Stability	No NLSA program existed at the system level.	An NLSA program existed at the system level, but it was not sufficiently stable. ¹	A stable NLSA program had been in place for several years.	A stable NLSA program had been in place for 10 years or more.
Clarity of Purpose	There were no policy-mandated purposes of the NLSA.	The NLSA had clear policy-mandated purposes, but these did not include informing policy or pedagogy.	The NLSA had clear policy-mandated purposes that included informing policy or pedagogy. ²	This option does not apply to this indicator.
Policy Document	No policy document authorized the NLSA program. ³	An informal/draft policy document authorized the NLSA program.	A formal/official policy document authorized the NLSA program, but the document was not available to the general public.	A formal/official policy document authorized the NLSA program and was available to the general public.
Program Guidelines	No official document provided guidelines for the NLSA program.	An official document provided guidelines for the NLSA program, but it was missing some key guidelines. ⁴	An official document provided key guidelines for the NLSA program.	This option does not apply to this indicator.
Stability of Organization	There was no unit with primary responsibility for running the NLSA program.	There was a unit(s) with primary responsibility for running the NLSA program, but the unit(s) was temporary or had been in place for less than 5 years. ⁵	There was a permanent unit(s) with primary responsibility for running the NLSA program that had been in place for 5 or more years.	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Accountability of Organization	There was no unit with primary responsibility for running the NLSA program, or else the unit responsible was not accountable to a clearly recognized body. ⁶	The unit(s) with primary responsibility for running the NLSA program was accountable to a clearly recognized body within the NLSA unit.	The unit(s) with primary responsibility for running the NLSA program was accountable to a clearly recognized body within the same institution as the NLSA unit.	The unit(s) with primary responsibility for running the NLSA program was accountable to a clearly recognized external body.
Source of Funding	There was no funding available for NLSA activities.	The source of funding for the majority of NLSA activities was loans, credits, grants, or equivalent. ⁷	The source of funding for the majority of NLSA activities was the government's internal funding sources.	This option does not apply to this indicator.
Activities Funded	There was no funding available for NLSA activities.	Funding was not sufficient to cover all core NLSA activities. ⁸	Funding was sufficient to cover all core NLSA activities.	This option does not apply to this indicator.
Organization Resources	The NLSA unit did not have the appropriate resources.	The NLSA unit had some of the appropriate resources.	The NLSA unit had most of the appropriate resources. ⁹	The NLSA unit had all of the appropriate resources.
Qualifications of Staff	There were no individuals responsible for completing key NLSA activities.	Some of the individuals responsible for completing key NLSA activities had the relevant qualifications. ¹⁰	Most of the individuals responsible for completing key NLSA activities had the relevant qualifications.	All or almost all of the individuals responsible for completing key NLSA activities had the relevant qualifications.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Effectiveness of Staff	There were no individuals responsible for completing key NLSA activities.	The responsible individuals completed key NLSA activities, but there were significant issues in how these activities were completed. ¹¹	The responsible individuals completed key NLSA activities, with only some issues in how these activities were completed.	The responsible individuals completed key NLSA activities, and there were no issues in how these activities were completed.
Staff/Teacher Opportunity to Learn	There were no opportunities to learn about the NLSA. ¹²	Opportunities to learn about the NLSA were minimal, or not of high quality, or did not benefit all key stakeholder groups.	There were sufficient high-quality opportunities to learn about the NLSA that were available to key stakeholder groups.	Opportunities to learn about the NLSA were extensive, of high quality, and benefited key stakeholder groups.
Measuring What Is Intended	It was not clear what the NLSA was intended to measure.	There was weak alignment between the NLSA and what it was meant to measure, or there was no regular review process in place to verify that alignment existed. ¹³	The NLSA measured official learning standards or curriculum, and officially mandated reviews to verify this alignment took place during most NLSA rounds.	The NLSA measured official learning standards or curriculum, and officially mandated reviews to verify this alignment took place during all or almost all NLSA rounds.
Alignment with Other Assessments	The NLSA was poorly aligned with other types of assessment activities in the system.	The NLSA was somewhat aligned with other types of assessment activities in the system.	The NLSA was very aligned with other types of assessment activities in the system. ¹⁴	This option does not apply to this indicator.
Opportunities for Students to Be Exposed to Content and Skills	Students did not have opportunities to be exposed to the content and skills measured by the NLSA.	Students had limited opportunities to be exposed to the content and skills measured by the NLSA.	Students had sufficient opportunities to be exposed to the content and skills measured by the NLSA.	Students had many opportunities to be exposed to the content and skills measured by the NLSA. ¹⁵

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Preparatory Information for Schools	Official information on the NLSA was not made available to schools in the system.	A minimal amount of official information on the NLSA was made available to schools in the system, although not necessarily all schools.	A sufficient amount of official information on the NLSA was made available to most or almost all schools in the system. ¹⁶	An extensive amount of official information on the NLSA was made available to all or almost all schools in the system.
Quality Assurance	No formal procedures were in place to ensure the quality of the NLSA.	Formal procedures to ensure the quality of the NLSA were minimal in nature or not required. ¹⁷	Formal procedures to ensure the quality of the NLSA were sufficient in nature and required.	Formal procedures to ensure the quality of the NLSA were extensive in nature and required.
Standardization	The NLSA was not standardized at the system level.	The NLSA was partially standardized at the system level, or minimal or no procedures were in place to ensure standardization. ¹⁸	The NLSA was fully standardized at the system level, and sufficient procedures were in place to ensure standardization.	The NLSA was fully standardized at the system level, and extensive procedures were in place to ensure standardization.
Representativeness	A nonrandom sample or a convenience sample of students participated in the NLSA.	A random sample of students that was not representative at the country level participated in the NLSA.	All students in public schools, or a representative sample of students in public schools, participated in the NLSA. ¹⁹	All students in public and private schools, or a representative sample of students in public and private schools, participated in the NLSA.
Reasons for Not Taking the NLSA	All or almost all individuals could not take the NLSA due to one or more non-assessment-relevant reason(s).	Most or some individuals could not take the NLSA due to one or more non-assessment-relevant reason(s).	There were no non-assessment-relevant reasons that prevented individuals from taking the NLSA. ²⁰	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Quality Processes	Many errors or delays in activities took place that affected the NLSA to a great extent. ²¹	Errors or delays in activities affected the NLSA to a significant level.	Any errors or delays in activities had only a minimal effect on the NLSA.	Errors or delays in activities did not affect the NLSA.
Inappropriate Behavior	Inappropriate behavior compromised the credibility of the NLSA to a great extent.	Inappropriate behavior took place and compromised the credibility of the NLSA somewhat.	Inappropriate behavior was low and did not compromise the credibility of the NLSA.	Inappropriate behavior, if any, was marginal and did not compromise the credibility of the NLSA. ²²
Methods and Procedures Documentation	There was no documentation on the methods and procedures used during the NLSA. ²³	There was minimal documentation on the methods and procedures used during the NLSA, or the documentation that existed was not public.	There was sufficient and public documentation on the methods and procedures used during the NLSA.	There was extensive and public documentation on the methods and procedures used during the NLSA.
Publication of Results	NLSA results were not published. ²⁴	Limited information on the NLSA results was published, or the results were published using a minimum number of dissemination mechanisms.	Sufficient information on the NLSA results was published using an array of dissemination mechanisms.	Comprehensive information on the NLSA results was published using an array of dissemination mechanisms.
Credibility of Results	The results of the NLSA were perceived as credible by very few stakeholder groups. ²⁵	The results of the NLSA were perceived as credible by some stakeholder groups.	The results of the NLSA were perceived as credible by most stakeholder groups.	The results of the NLSA were perceived as credible by all or almost all stakeholder groups.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Impact Monitoring	No mechanisms were in place to monitor the impact of the NLSA. ²⁶	Minimal mechanisms were in place to monitor the consequences of the NLSA, or the mechanisms took place in only some or a few NLSA rounds.	Sufficient mechanisms were in place to monitor the impact of the NLSA, and the mechanisms took place in all or almost all NLSA rounds.	Extensive mechanisms were in place to monitor the impact of the NLSA, and the mechanisms took place in all or almost all NLSA rounds.
Readiness to Start an NLSA Program	The system was weakly prepared to start an NLSA program in the future.	The system was somewhat prepared to start an NLSA program in the future.	The system was well prepared to start an NLSA program in the future.	This option does not apply to this indicator. ²⁷

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

1. Cabo Verde has a system-level NLSA program, the *Prova de Aferida*, which was administered in 2010 and 2014. Both rounds assessed Portuguese and math. The 2010 NLSA was administered to sixth graders, and the 2014 NLSA was administered to second, fourth, and sixth graders.
2. The policy-mandated purposes of the NLSA are monitoring education quality, informing policy, and informing pedagogy. The following are not policy-mandated purposes: holding the government, schools, teachers, or students accountable; and evaluating interventions aimed at improving student learning.
3. There was no system-level policy document authorizing the NLSA program; however, there is a document that generally describes the purpose of an NLSA without authorizing any specific program. This document is the *Sistema de Avaliação do Ensino Básico Decreto Lei N. 43* and was authorized in 2003 by the Republic of Cabo Verde. This document is available to the general public.
4. “General Information Guides” and “Application Guidelines” were circulated to the schools participating in the NLSA sample. These documents provided guidelines on governance of the NLSA, what should be assessed, and who should be assessed. The guidelines did not address how assessment results should be used.
5. In 2014, a temporary *núcleo* was formed with technicians from various departments within the Ministry of Education to help lead and implement the 2014 *Aferida*. This *núcleo* had primary responsibility for running the NLSA program and completed one round. The unit was not held accountable to a clearly recognized body.

As a part of the *núcleo*, a team of teachers and Ministry of Education technical staff worked together on the elaboration of the assessment and planning for implementation. The Ministry staff involved included employees from the Directorate of National Education, the General Inspectorate of Education, and the Department of Statistics.

The Ministry of Education implemented the assessment and was responsible for scoring the assessment. The Department of Statistics within the Ministry was officially responsible for data processing. The *núcleo* no longer exists.

6. The Ministry of Education had primary responsibility for running the NLSA, but it was not held accountable to a clearly recognized body.
7. Internal funding was available; however, no information is available on the amount of funding or the source of funding.
8. Internal funding was available; however, it is unclear what NLSA activities were funded.
9. All Ministry of Education staff responsible for NLSA activities were properly equipped with needed computers and resources, including appropriate computers for all technical staff, appropriate software, appropriate building security, appropriate storage facilities, appropriate computer servers, and appropriate communication tools.

10. Individuals responsible for completing key NLSA activities were:

- a) Permanent staff from the Ministry of Education (including staff from the Directorate of National Education, the General Inspectorate of Education, and the Department of Statistics), 10–50 percent of whom were qualified, and
- b) Teachers completing NLSA activities as part of their job responsibilities, of whom less than 10 percent were qualified.

11. As mentioned above, individuals responsible for completing key NLSA activities were permanent staff from other Ministry of Education units, 10–50 percent of whom were qualified, and teachers completing NLSA activities as part of their job responsibilities, of whom less than 10 percent were qualified.

Significant issues were found with these individuals' effectiveness, which affected the quality of specific NLSA activities and compromised the overall quality of the NLSA. Specifically, delays occurred in reporting information and processing results, some remote schools that were intended to participate in the NLSA did not participate because of their remote locations, and some teachers did not administer the exam correctly because of confusion about the instructions they received.

12. No opportunities were available in the system to learn about the NLSA. Specifically, there were no university graduate programs (master's or doctoral level) on student assessment that include topics relevant to the NLSA (for example, test design, reporting); no university courses and/or workshops on the content and skills measured by the NLSA (for example, courses on curriculum); no non-university courses and/or workshops on the content and skills measured by the NLSA (for example, courses on curriculum); no university courses and/or workshops on NLSA topics other than the content and skills measured by the NLSA (for example, test design, reporting); no non-university courses and/or workshops on NLSA topics other than the content and skills measured by the NLSA (for example, test design, reporting); no funding for attending international programs, courses, and workshops on student assessment that cover topics relevant to the NLSA, no internships and/or short-term employment in the unit running the NLSA, and no presentations about the NLSA (for example, test design, administration).

13. The NLSA is intended to measure curriculum and is very aligned with what it intends to measure; the NLSA is based upon the official curriculum (*Programas*) for each grade. In 2015, an internal review took place. However, there are no officially mandated reviews—internal, external, or otherwise.

14. The NLSA is very aligned with classroom assessment.

15. School educators from both public and private schools provided students with the opportunity to be exposed to the content and skills measured by the NLSA. Students were exposed to this material as a part of regular course instruction at school. In both public and private schools, all or almost all students were exposed to the skills and content assessed by the NLSA.

16. Only the “*General Information Guide*” was available to almost all schools. This unofficial framework document provided information on the grade levels to be assessed, when the assessment would take place, the structure of the assessment and the types of questions that are part of the NLSA, and information on how

the NLSA will be administered at the school level. The *General Information Guide* did not include information on how to teach the content and skills assessed by the NLSA or examples of the criteria used for scoring open-ended and essay questions.

17. The two formal quality assurance procedures that were in place to ensure the quality of the NLSA were (1) a standardized manual for NLSA administrators and (2) internal observers. The standardized manual included information on the logistics of test administration, such as how to set up the classroom, the number of supervisors per classroom, how to read instructions to students, and how students should fill out the cover form (a sheet for recording student identification information such as class, school, and region). The internal observers were Ministry of Education staff, who observed the administration of the assessment to ensure that schools followed the protocols.

Quality assurance procedures did not include the following: ensuring that all proctors or administrators were trained according to a protocol; piloting questions, items, or tasks before the official NLSA was administered; numbering all booklets; double data scoring; training scorers to ensure high interrater reliability; double processing of data; and bringing in external observers (such as a representative of the community observing at administration sites).

18. The NLSA was partially standardized at the system level. Assessment design was consistent and overseen by the Ministry of Education. However, although all teachers received the same information and manuals, weak training led to differences in NLSA administration across the country. Additionally, NLSA papers and questions, items, and tasks were the same for all students, and quality control monitors were used to ensure consistent administration conditions. These quality control monitors, however, were not present in all locations.

The following measures were not used to ensure the standardization of the NLSA: NLSA administrators were not trained to ensure that all students took the NLSA under the same conditions; the same scoring criteria were not used to correct NLSA questions, items, or tasks; NLSA results were not computed using the same procedures for all students; and NLSA results were not reported to all students in the same way.

19. The 2014 NLSA was implemented in a random sample of public schools, representative of the overall target population at the system level for the grades. There was difficulty, however, concerning access to some of the schools in the sample, and some schools were excluded from the *Aferida* at the last minute. The *Aferida* was not administered in private schools.

20. No non-NLSA-relevant reasons prevented individuals from taking the NLSA, with the exception of the school cancellations mentioned above.

21. The following factors affected the NLSA round to a great extent: errors in administering the NLSA, poor training of NLSA administrators, delays in scoring student responses, delays in data processing, delays in reporting results, and failure to report results. Specific examples included problems with administration (some schools in the initial sample selection were excluded for participation for logistical reasons) and slow processing of results (as of this report, NLSA results have not been processed and there are no formal plans to publish them).

22. Inappropriate behavior did not take place during the 2014 NLSA.

23. No document specified the methods and procedures used during the 2014 NLSA.
24. NLSA results were not published.
25. The 2014 NLSA results have not been released.
26. No mechanisms are in place to monitor the impact of the NLSA.
27. This indicator does not apply to this rubric.

CABO VERDE
International Large-Scale Assessment (ILSA)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Stability of Participation	The system did not participate in an ILSA round in the last 10 years. ¹	The system participated in an ILSA round in the last 10 years but did not complete it.	The system completed one ILSA round in the last 10 years.	The system completed two or more ILSA rounds in the last 10 years.
Policy Document	No policy document authorized the ILSA program. ²	An informal/draft policy document authorized the ILSA program.	A formal/official policy document authorized the ILSA program, but the document was not available to the general public.	A formal/official policy document authorized the ILSA program and was available to the general public.
Stability of Organization	There was no unit with primary responsibility for running the ILSA program. ³	There was a unit(s) with primary responsibility for running the ILSA program, but the unit(s) was temporary or had been in place for less than 5 years.	There was a permanent unit(s) with primary responsibility for running the ILSA program that had been in place for 5 or more years.	This option does not apply to this indicator.
Accountability of Organization	There was no unit with primary responsibility for running the ILSA program, or else the unit responsible was not accountable to a clearly recognized body. ⁴	The unit(s) with primary responsibility for running the ILSA program was accountable to a clearly recognized body within the ILSA unit.	The unit(s) with primary responsibility for running the ILSA program was accountable to a clearly recognized body within the same institution as the ILSA unit.	The unit(s) with primary responsibility for running the ILSA program was accountable to a clearly recognized external body.
Source of Funding	There was no funding available for ILSA activities. ⁵	The source of funding for the majority of ILSA activities was loans, credits, grants, or equivalent.	The source of funding for the majority of ILSA activities was the government's internal funding sources.	This option does not apply to this indicator.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Activities Funded	There was no funding available for ILSA activities. ⁶	Funding was not sufficient to cover all core ILSA activities.	Funding was sufficient to cover all core ILSA activities.	This option does not apply to this indicator.
Organization Resources	The ILSA unit did not have the appropriate resources. ⁷	The ILSA unit had some of the appropriate resources.	The ILSA unit had most of the appropriate resources.	The ILSA unit had all of the appropriate resources.
Qualifications of Staff	There were no individuals responsible for completing key ILSA activities. ⁸	Some of the individuals responsible for completing key ILSA activities had the relevant qualifications.	Most of the individuals responsible for completing key ILSA activities had the relevant qualifications.	All or almost all of the individuals responsible for completing key ILSA activities had the relevant qualifications.
Effectiveness of Staff	There were no individuals responsible for completing key ILSA activities. ⁹	The responsible individuals completed key ILSA activities, but there were significant issues in how these activities were completed.	The responsible individuals completed key ILSA activities, with only some issues in how these activities were completed.	The responsible individuals completed key ILSA activities, and there were no issues in how these activities were completed.
Staff/Teacher Opportunity to Learn	There were no opportunities to learn about the ILSA. ¹⁰	Opportunities to learn about the ILSA were minimal, or not of high quality, or did not benefit all key stakeholder groups.	There were sufficient high-quality opportunities to learn about the ILSA that were available to key stakeholder groups.	Opportunities to learn about the ILSA were extensive, of high quality, and benefited key stakeholder groups.

(CONTINUED)

Indicator	LATENT Absence of, or deviation from, the attribute	EMERGING On way to meeting minimum standard	ESTABLISHED Acceptable minimum standard	ADVANCED Best practice
Alignment with Other Assessments	The ILSA was poorly aligned with other types of assessment activities in the system. ¹¹	The ILSA was somewhat aligned with other types of assessment activities in the system.	The ILSA was very aligned with other types of assessment activities in the system.	This option does not apply to this indicator.
Opportunities for Students to Be Exposed to Content and Skills	Students did not have opportunities to be exposed to the content and skills measured by the ILSA. ¹²	Students had limited opportunities to be exposed to the content and skills measured by the ILSA.	Students had sufficient opportunities to be exposed to the content and skills measured by the ILSA.	Students had many opportunities to be exposed to the content and skills measured by the ILSA.
Quality Processes	Many errors or delays in activities took place that affected the ILSA to a great extent. ¹³	Errors or delays in activities affected the ILSA to a significant level.	Any errors or delays in activities had only a minimal effect on the ILSA.	Errors or delays in activities did not affect the ILSA.
Inappropriate Behavior	Inappropriate behavior compromised the credibility of the ILSA to a great extent. ¹⁴	Inappropriate behavior took place and compromised the credibility of the ILSA somewhat.	Inappropriate behavior was low and did not compromise the credibility of the ILSA.	Inappropriate behavior, if any, was marginal and did not compromise the credibility of the ILSA.
Meeting Standards for Publication	ILSA results for the system did not meet the standards required for publication in the international report. ¹⁵	ILSA results for the system met sufficient standards to be presented beneath the main displays in the international report.	ILSA results for the system met all of the standards required to be presented in the main displays of the international report.	This option does not apply to this indicator.

(CONTINUED)

Publication of Results	ILSA results were not published in the system. ¹⁶	Limited information on the ILSA results was published in the system, or the results were published using a minimum number of mechanisms.	Sufficient information on the ILSA results was published in the system using an array of mechanisms.	Comprehensive information on the ILSA results was published in the system using an array of mechanisms.
Credibility of Results	The results of the ILSA were perceived as credible by very few stakeholder groups. ¹⁷	The results of the ILSA were perceived as credible by some stakeholder groups.	The results of the ILSA were perceived as credible by most stakeholder groups.	The results of the ILSA were perceived as credible by all or almost all stakeholder groups.
Use of Results	ILSA results were not used by stakeholders in the system. ¹⁸	ILSA results were used in minimal ways by stakeholders in the system.	ILSA results were used in sufficient ways by stakeholders in the system.	ILSA results were used in extensive ways by stakeholders in the system.
Readiness to Participate in an ILSA	The system was weakly prepared to participate in an ILSA program in the future. ¹⁹	The system was somewhat prepared to participate in an ILSA program in the future.	The system was well prepared to participate in an ILSA program in the future.	This option does not apply to this indicator.

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

1. The system has not participated in an ILSA in the last 10 years.
2. There is no system-level policy document that authorized participation in the ILSA program.
3. There is no unit responsible for running an ILSA program.
4. There is no unit responsible for running an ILSA program.
5. There is no funding allocated to ILSA activities.
6. There is no funding allocated to ILSA activities.
7. There is no ILSA unit.
8. N/A because there has not been an ILSA exam.
9. N/A because there has not been an ILSA exam.
10. N/A because there has not been an ILSA exam.
11. N/A because there has not been an ILSA exam.
12. N/A because there has not been an ILSA exam.
13. N/A because there has not been an ILSA exam.
14. N/A because there has not been an ILSA exam.
15. N/A because there has not been an ILSA exam.
16. N/A because there has not been an ILSA exam.
17. N/A because there has not been an ILSA exam.
18. N/A because there has not been an ILSA exam.
19. Cabo Verde has not participated in an ILSA in the past. However, the Ministry of Education has discussed the possibility of participating in an ILSA in the future. The current draft of the 2017–2021 Education Strategy “*Plano Estrategico 25 Abril*” broadly mentions “developing [an] external assessment model”; however,

there is no official mention of participating in any specific ILSA. Although no unit within the Ministry of Education is directly responsible for student assessment, temporary units and individuals from various departments within the Ministry have supported other assessment activities in the past. No specific opportunities are available for ILSA training; however, if Cabo Verde were to decide to move forward with an ILSA, funding could be made available for capacity building.

Acknowledgments

This report was prepared by the World Bank Cabo Verde Country Task Team (Kamel Braham, World Bank Lead Education Specialist, and Jem Heinzel Nelson, World Bank Education Consultant) and the SABER-Student Assessment Team (Marguerite Clarke, World Bank Senior Education Specialist, Julia Liberman, World Bank Operations Officer, and Tara Danica Siegel, World Bank Education Consultant). The Team is grateful for the feedback and support from the Cabo Verde Ministry of Education, in particular the Direção Nacional de Educação (National Directorate of Education), the Direção Geral de Planeamento, Orçamento e Gestão (Directorate-General for Planning, Budgeting and Management), and the Inspeção Geral da Educação (IGE) (General Inspection of Education Unit).

Reference

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

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 policy makers and business people—an accessible, objective snapshot showing how well the policies of their country's education system are oriented toward ensuring that all children and youth learn.

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

This work is a product of the staff of The World Bank with external contributions. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.