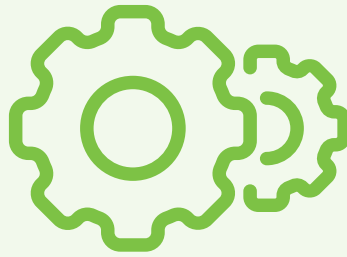




Ministry of Education  
and Higher Education



# National Policy Framework on Student Assessment

October 2024



## Table of Contents

<b>Introduction &amp; Objectives</b>	<b>03</b>
<b>Acknowledgments</b>	<b>04</b>
<b>Pillar 1. Assessment for system evaluation, policymaking, and accountability</b>	<b>05</b>
1.1 The Current Situation	05
1.2 The Way Forward	09
<b>Pillar 2. Assessment for school-level planning and improvement</b>	<b>11</b>
2.1 The Current Situation	11
2.2 The Way Forward	13
<b>Pillar 3. Assessment for formative and screening purposes</b>	<b>14</b>
3.1 The Current Situation	15
3.2 The Way Forward	16
<b>Pillar 4. Assessment for summative purposes</b>	<b>18</b>
4.1 The Current Situation	18
4.2 The Way Forward	20
<b>Glossary</b>	<b>22</b>

## Introduction & Objectives

Student assessment is a central element of every education system. It can shape how teachers teach and how children learn. Assessment can take many different forms and serve different purposes, from evaluating the performance of the overall education system to informing day-to-day teaching in the classroom. High-quality and timely information on student learning is a critical tool to strengthen teaching practices and accountability in the education sector. Hence, the purpose of gathering assessment data, whether at national, district, school or classroom levels, is to improve teaching and learning and move towards equitable learning outcomes for all students.

The Ministry of Education and Higher Education's National Policy Framework on Student Assessment ('the Framework') builds on existing strategies, manuals, and guidelines. The Framework lays out a pathway to strengthen student learning outcomes through better assessment at all levels. It outlines how student assessment data can be more effectively used at the central, district, school, and classroom level to inform strategic planning and policy decisions, teacher professional development programs, daily classroom practices, and other relevant aspects of the education system. Furthermore, this framework acknowledges the role of using computer-based assessment and its use in assessing the 21st century skills. To this end, the Framework consolidates the purpose, objectives, and Ministry's medium-term plans for different types of assessment in a single reference document.

The Framework is designed around four pillars:

- **Pillar (1): Assessment for system evaluation, policymaking & accountability**
- **Pillar (2): Assessment for school-level planning and improvement**
- **Pillar (3): Assessment for formative & screening purposes**
- **Pillar (4): Assessment for summative purposes**

Each pillar includes a brief description of the status quo and a set of key policy priorities on the way forward. While each pillar and its recommendations are presented separately, the pillars are interrelated and impact one another. For example, the outcomes of large-scale assessments (Pillar 1) can inform and affect school-level planning of principals, teachers, and district supervisors (Pillar 2). Similarly, the quality and depth of assessment for school-level planning and improvement (Pillar 2) is expected to affect teachers' practices as they relate to assessment for formative and screening purposes (Pillar 3), and assessment for summative purposes (Pillar 4).

The Framework does not intend to increase the assessment load on teachers and students. Rather, the focus is on making good use of existing assessment data in ways that benefit student learning and support equitable learning outcomes for all students.

Implementation of the policy priorities identified under the Framework will be severely affected by the Israeli aggression in Gaza that has caused unspeakable suffering. More than 625,000 children have been deprived of their right to education following October 2023, and schooling has also been severely disrupted in the West Bank. Learning losses are expected to be massive, and teachers will be faced with students with vastly differing needs in the classroom. In that context, equipping teachers with the necessary knowledge, skills, and tools for effective student assessment will be of particular importance in the upcoming academic years<sup>1</sup>.

The Framework builds on a series of focus group discussions and technical consultations within the Ministry of Education and Higher Education and with supervisors, and teachers. In addition to the situational analysis, the policy priorities have been informed by international best practice and latest research in the field of student assessment. The Framework is complemented by a one-page Executive Summary highlighting key elements and recommendations.

<sup>1</sup> The EiE and its associated assessments are discussed in another document.

## Acknowledgments

The development of the National Policy Framework on Student Assessment by the Ministry of Education and Higher Education was a collaborative effort of several departments, including the General Directorate for Examinations, and for School Education, the Center for Educational Research and Development (CERD), National Institute for Educational Training (NIET), and the Curriculum Development Center.

This activity was led by Eng. Jehad Draidi, Head of the National Center for Examination, Measurement and Educational Evaluation, Director General for Technical and Vocational Training, and focal point for assessment and under SERATAC. The following individuals have generously shared their time and expertise to provide invaluable contributions to this framework: Dr. Mohammed Matar, Tarek Alawneh, Dr. Eman Alnajjar, Dr. Omar Atwan, Dr. Khaled Bisharat, Dr. Samah Iriqat, Alaa Sajadeiah,

The Framework was developed under the umbrella of the Supporting an Education Reform Agenda to Improve Teaching, Assessment, and Career Pathways (SERATAC) project. Sincere gratitude is expressed to Nadine Njoun, Director of the SERATAC Project Coordination Unit (PCU), Dania Samer Hasarma, SERATAC Technical Expert for Literacy, STEM, and Assessment, and Carmen Kishek, Communications SERATAC Communications Specialist, for their continuous follow-up and support on the Assessment Framework.

## Pillar 1. Assessment for system evaluation, policymaking, and accountability

Under the first pillar, the Framework lays out the role of international and national large-scale student assessments in providing education stakeholders in Palestine with high-quality and timely data to strengthen teaching and learning. Large-scale assessments are typically sample-based and designed to monitor student performance in selected subjects and grades or age groups at the systems level.

### 1.1 The Current Situation

#### A. International Large-Scale Assessments

The Ministry of Education and Higher Education (MoEHE) is strongly committed to improving the availability of high-quality, internationally comparable data on learning outcomes. After more than a decade-long absence, the MoEHE has re-engaged in international large-scale assessments in 2022. In line with the Ministry's strategic vision and commitment to evidence-based policymaking, Palestine has partnered with the International Association for the Evaluation of Educational Achievement (IEA) and the Organisation for Economic Co-operation and Development (OECD) to administer the well-established Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS).

Palestine's first-ever participation in PISA 2022 and its participation in TIMSS 2023 for the first time since 2011 represent major milestones. Both assessments were delivered to students by computer. PISA results have been published by the OECD in December 2023 and TIMSS results are expected to be available in late 2024.

As TIMSS tests students in Grade 8 and PISA tests 15-year-old students the Ministry has complemented those studies with assessments in lower grades. The IEA's Literacy and Numeracy Assessment (LaNA) study was administered to students at the beginning of Grade 5 in November 2023. While LaNA is still in its piloting phase, it is expected that the results can be used to express students' performance in reading and math on the same scale as PIRLS and TIMSS. This will facilitate cross-country comparisons and the analysis of different proficiency levels. The LaNA results can be used to generate Palestine's first estimate of Learning Poverty<sup>2</sup>, that is the share of 10-year-old children unable to read and understand a simple text.

**Table 1. Overview of International Assessments by Year**

Test	Subjects	Grade	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	
PISA	Reading	10 (15-year-olds)																			
	Math														✓			✓			✓
	Science																				
TIMSS	Math	8																			
	Science		✓													✓				✓	
LaNA	Literacy Numeracy	5													✓						

Indicates assessment has taken place

Indicates assessment is planned to take place in the future

<sup>2</sup> Full study available: <https://www.worldbank.org/en/topic/education/brief/what-is-learning-poverty>



## B. National Large-Scale Assessments

Palestine has a long history of implementing national large-scale assessments (NLSAs). With the exception of 2020 due to the COVID-19 pandemic, Palestine has implemented NLSAs at semi-regular intervals since 1998. Originally administered in Grades 4 and 10, the assessment was shifted to Grades 5 and 9 in 2016. The NLSAs in Grades 5 and 9 cover the three core subjects of Arabic, math, and science. They are implemented with a nationally representative sample including public and private schools as well as schools operated by the United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA). Test administration time per subject is 90 minutes, and the test is composed of both multiple-choice questions and short-answer questions. In 2018, the NLSAs were expanded to cover Information and Communications Technology (ICT) skills in Grade 10, and English in Grade 6. Results can be disaggregated by the national variables, including the school's supervision authority (government, private, or agency), geographic location (northern and southern governorates), and school type (all boys, all girls, or co-ed).

Classical Test Theory is used to evaluate the quality of test items and to scale achievement data. One test form is used for each subject and grade level. Results are presented as absolute scores and not mapped to pre-defined competency levels. The test is accompanied by a questionnaire on relevant contextual information such students' socioeconomic background. As the lack of a comprehensive technical report on the NLSA methodology poses challenges for quality assurance, the Ministry is planning to strengthen the design and documentation of upcoming NLSA cycles and will explore the feasibility of computer-based implementation. In addition, the Early Grade Reading Assessment (EGRA) is planned to be implemented in Grade 2 in 2025 for the second time since 2014.

**Table 2. Overview of National Assessments by Year**

Subject	Grades	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029
Arabic	4+10		✓		✓		✓		✓				✓			✓		✓		✓
Math	until 2016 5+9		✓		✓		✓		✓				✓			✓		✓		✓
Science	since 2017		✓		✓		✓		✓				✓			✓		✓		✓
English	6								✓				✓			✓				
ICT Skills	10								✓				✓			✓				
ICT Skills	5								✓							✓				
EGRA	2				✓												✓			

 Indicates assessment has taken place  
 Indicates assessment is planned to take place in the future

## C. Interpretation and Use of Results

The underlying design of each assessment needs to be taken into account when interpreting the results. The National Assessments and TIMSS assess curriculum-based skills and processes, while PISA assesses knowledge and skills deemed to be important for students' future lives rather than curricular knowledge.

LaNA and EGRA are designed as shorter, less demanding<sup>3</sup> assessments that are expected to be better aligned with students' current knowledge and skill levels than PISA and TIMSS.

Weak performance of students on external assessments in Palestine might be attributed, among other factors, to students' lack of motivation, as test results do not contribute to their school grades. However, International experiences and recent PISA results suggest it is more likely that students appear unmotivated because there is a strong mismatch between students' abilities and the tasks they are expected to complete. According to PISA 2022, 80 percent of students in Palestine do not meet minimum proficiency levels in mathematics. Similarly, 77 percent do not meet minimum proficiency in reading, and 72 percent in science.

These figures suggest that most external assessment items are too difficult for students to engage and highlight the need for a strong focus on equipping students in Palestine with essential knowledge and skills in core subjects. These results may also point to challenges in teaching, learning, and assessment methods in Palestinian schools, as well as reliance on international standards and requirements that do not fully align with the curriculum content and educational and assessment methods used in Palestine, in addition to the impact of surrounding circumstances on student motivation.

Strong capacity for data analysis and a clear communications strategy are important to ensure the effective use of the results from large-scale assessments. Building on its rich experience with NLSAs, the MoEHE has in the past formed data utilization teams at the district level to support analysis and help translate findings into action. Through its participation in PISA 2022, the MoEHE has further strengthened its capacity to lead large data collection efforts, conduct complex statistical analyses, and draft a comprehensive national report. In addition, the MoEHE has established partnerships with local universities to further explore the collected data and encourage local research projects.

## D. Availability of Data from Large-Scale Assessments

A review of latest available assessments by grade level and subject shows the great progress achieved by the MoEHE in recent years to collect high-quality, comparable data on student learning outcomes (see Table 3). With LaNA covering the end of primary education and students' knowledge and skills at the secondary level assessed through TIMSS and PISA, the MoEHE is collecting rich and comprehensive information. Nevertheless, a scarcity of learning data for the early primary grades persists, a critical period to monitor if students acquire foundational skills that they can build on as they progress through the school system.

<sup>3</sup> The International Association for the Evaluation of Educational Achievement (IEA), through readings in its research and publications, defines less demanding assessments as those that do not require intensive cognitive engagement or complex problem-solving skills from students. These assessments often focus on basic knowledge and skills rather than higher-order thinking skills.

**Table 3. Overview of National and International Assessments by Grade**

Subject	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12
Arabic		EGRA (2014)			LaNA (2023) NLSA (2022)				NLSA (2022)	PISA <sup>4</sup> (2022)		
Math					LaNA (2023) NLSA (2022)			TIMSS (2023)	NLSA (2022)	PISA (2022)		
Science					NLSA (2022)			TIMSS (2023)	NLSA (2022)	PISA (2022)		
English						NLSA (2022)						
ICT Skills					NLSA (2018)					NLSA (2022)		

Indicates subject is covered by national assessment

Indicates subject is covered by international assessment

Indicates subject is covered by both national and international assessment

( ) Year in brackets indicates latest year when respective assessment was administered

## 1.2 The Way Forward

**Table 4. Key Policy Priorities under Pillar 1**

Policy Priority	Brief Description
A. Ensure continuity and leverage synergies across assessments	Palestine's long-term commitment to participate in international assessments is a critical achievement that will allow the MoEHE to identify benchmarked trends in learning outcomes. By leveraging data from these international assessments, which focus on both curriculum-based knowledge and practical application, alongside existing national assessments, the MoEHE can develop a holistic understanding of students' performance in core subjects and the long-term impact of its reform efforts.
B. Strengthen national assessments based on international experience	To ensure that national assessments provide reliable and replicable data from cycle to cycle, a strong technical report with detailed information on the applied methodology for test design, sampling, and data analysis needs to be developed. In addition, innovations such as new scaling approaches, new types of test items, and computer-based administration should be introduced on an incremental basis building on international experience.
C. Invest in effective use of results and capacity of assessment staff	The implementation of large-scale assessments requires significant resources, including researchers' access to appropriate software and tools, and ongoing training of assessment staff to acquire and maintain highly specialized knowledge and skills. It is crucial that the analysis and reporting of national and international large-scale assessment data are guided by policy needs and accompanying research questions. Strategic partnerships with local universities and other relevant stakeholders will help support data analysis and communications.

### A. Ensure continuity and leverage synergies across assessments

Palestine's long-term commitment to participate in international assessments is a critical achievement that will allow the MoEHE to identify benchmarked trends in learning outcomes. Through the implementation of PISA, TIMSS, and LaNA in 2022 and 2023, the Ministry has established a credible baseline against which the impact of its ambitious reform efforts on student learning can be assessed. The diverse focus of these assessments—curriculum-based knowledge and skills in TIMSS and LaNA, and the application of knowledge and skills in PISA—provides a valuable opportunity to gain a comprehensive understanding of students' strengths and challenges and allows educators to develop a more nuanced and complete picture of student performance. This integrated perspective enables a better identification of areas where students excel and where they may need additional support, ultimately informing targeted interventions and improvements in the educational system.

Since PISA and TIMSS were administered in both the West Bank and Gaza before October 7, they serve as a baseline of student learning and can help quantify the devastating impact of the war on learning and human capital across the Palestinian territories. The availability of international data as a source of transparency and accountability further sends an important signal to donors and development partners that may help mobilize additional financial resources to rebuild the education sector in Gaza and support the education emergency response in the West Bank.

The list of current and forthcoming assessments in Table 3 suggest a number of synergies. In particular, there is potential to develop a comprehensive report on Arabic reading literacy, drawing on the outcomes of EGRA (Grade 2), LaNA Reading Literacy (beginning of Grade 5), NLSA (Grade 5 and 9) and PISA Reading Literacy (15-year-olds, mostly Grade 10). Similarly, an overall report on mathematics could be based on LaNA Mathematics (beginning of Grade 5), NLSA Mathematics (Grade 5), TIMSS Mathematics (Grade 8), and PISA 2022 Mathematics (15-year-olds, mostly Grade 10). In the case of science, an overall report could be based

<sup>4</sup> 15-year-old students aged partake in this assessment, with the majority being in the tenth grade, with other 15-year-olds from grades 7 to 11.

on NLSA Science (Grades 5 and 9), TIMSS Science (Grade 8) and PISA Science (Grade 5 and 9, Age 15 – Modal Grade 10). Such overall reports could address the following topics by triangulating data across studies:

- **Performance and proficiency at different stages of schooling – overall and by subgroups (e.g., gender, socio-economic status, region),**
- **Changes in average performance and/or proficiency over time – overall and by subgroups**
- **Practices in teaching and learning (with comparative international data)**
- **Home practices and support related to Arabic literacy/numeracy/science**
- **Students’ attitudes, interests, and motivations, and how these relate to performance**
- **Access to resources for supporting literacy and numeracy, at home and at school**

## B. Strengthen national assessments based on international experience

National assessments need to be able to provide reliable data from cycle to cycle. One way to ensure this is to hold as many factors as possible constant between cycles. This includes a consistent approach to:

- **Sampling**, so that any observed changes in performance from cycle to cycle cannot be attributed to differences in sampling, Scaling, so that results are generated based on a consistent testing theory and comparable over time,
- **Standards or proficiency levels**, so that, for example, the proportions of students with reading or mathematical difficulties in the system or region can be identified and compared across cycles.

However, incremental changes to national assessments can and should be introduced over time to further strengthen its design and enhance the informational value for education stakeholders. These changes should be made gradually with Benefitting from its recent experience with international assessments in PISA, TIMSS, and LaNA, the MoEHE will explore the gradual introduction of the following approaches:

1. **Item Response Theory.** While both Classical Test Theory and Item Response Theory are used in psychometrics to design and evaluate the reliability and validity of tests, they differ in their underlying assumptions, focus, and methods of analysis. IRT is often preferred in situations where precise measurement of individuals’ abilities or traits is needed, while CTT is more straightforward and easier to apply in some contexts. In the next cycle of national assessments, the Ministry could consider using use both Classical Test Theory (CTT) and Item Response Theory (IRT). This will facilitate comparisons with previous cycles that used CTT while establishing a basis for future national assessments relying on IRT.
2. **Computer-based administration.** Many educational systems are transitioning to computer-based assessments, which can make the administration of large-scale assessments more efficient and reduce cost for printing, marking, etc. It further introduces options for adaptive testing, or efforts to match test content to students’ abilities, thereby making the test experience more relevant and less frustrating for students at all levels of ability.

## C. Invest in effective use of results and capacity of assessment staff

Many countries invest large sums in generating high-quality assessment data but fail at the ‘last mile’: effective use of the data to improve teaching and learning. Sufficient staff time and financial resources need to be dedicated to effective communication and use of assessment results. This includes the development of a strong communications strategy to convey the results and implications to Ministry officials at the central and district level, as well as principals, teachers, parents, students, and universities.

The design and implementation of large-scale assessments requires knowledge across a range of areas. Some of these are specific to subject-matter or content knowledge; others relate to knowledge about sampling and scaling, psychometrics, analysis of questionnaire and achievement data derived from an assessment, and the ability to write in a way that addresses the needs of different audiences. As such, assessment staff require continuous training and capacity building to implement the aspects of the national or international

assessment for which they are responsible. This may also include training for teams of teachers and other subject-matter experts to develop good items or examination questions for use in national assessments and examinations.

Building on the existing memoranda of understanding between the Ministry and local universities, university faculty and researchers have been trained between July and October 2023 through a set of technical support meetings on the varying levels of complexity in how data from international assessments are analyzed, ranging from reporting mean scores and simple percentages, to implementing multi-level modelling. Strengthening capacity on these aspects both within the Ministry and higher education sector is important to build a strong pool of qualified researchers who can support the Ministry in fully exploiting the rich data collected through large-scale assessments.

The Ministry intends to publish accessible and user-friendly summaries or dashboards of results from both national and international assessments on its website and other relevant channels. This will greatly enhance the transparency and accessibility of these findings for a broad audience. These summaries and dashboards will be designed to present complex information in a clear, concise, and visually engaging manner, allowing stakeholders to quickly grasp key trends and insights. By making these data readily available, the Ministry aims to foster greater understanding and engagement with the assessment results.

Furthermore, these resources will be leveraged as part of in-service teachers’ professional development programs (see Pillar 2). Integrating assessment data into training modules can provide teachers with practical examples and actionable insights, helping them to align their teaching strategies with identified trends and areas of need. This approach supports teachers in making data-informed decisions, enhancing their instructional practices, and ultimately improving student outcomes. Offering workshops or webinars on how to interpret and utilize these summaries and dashboards could further support educators in integrating assessment data into their daily practice. By promoting and facilitating the use of these materials, the Ministry aims to create a more data-driven education system that is responsive to the needs of both students and teachers.

## Pillar 2. Assessment for school-level planning and improvement

Under the second pillar, the Framework reviews the role of student assessment in guiding the work of central- and district-level supervisors. It further discusses existing efforts and opportunities for student assessment results to feed into school-level planning and improvement.

### 2.1 The Current Situation

#### A. Supervision of Schools and Teachers

Supervision of public and private schools and teachers in Palestine occurs under the auspices of the Directorate General of School Education. Through its 24 directorate-level offices and district-level supervisors, the Ministry implements supervision of primary and secondary schools, leading up to the Secondary School Diploma (Tawjihi).

Supervisors are typically responsible for a cluster of schools. They may also be involved in subject matter supervision as they visit schools and observe teachers of specific subjects. Within the supervision framework, there is a strong emphasis on supervising and supporting newly qualified teachers, a responsibility jointly shared by school leaders and supervisors. Newly appointed supervisors are typically accompanied by the head of supervision on their initial visits to schools to ensure that they can work effectively.

Supervision is currently provided to all teachers based on a plan prepared every semester at the district level to ensure at least one supervision visit for each teacher. There is no optimal number of visits that should be carried out each semester. Rather, supervision should be provided on a needs basis rather than a pre-defined frequency. Teachers may also request support from supervisors. Supervision visits do not follow a strict, systematic protocol. There is a focus on cooperation and a holistic approach, while trying to bring about specific improvements. Observations encompass time management, discipline, subject matter expertise,

and other relevant competencies. Supervision may include a review of tests or exams used by the teacher and students' performance on those, though it is not yet systematically integrated into the supervision routine. Other aspects of a teacher's work that may be evaluated by supervisors include general performance in the classroom in terms of classroom management and creating an appropriate learning environment. Following the observation, supervisors draft a report that is shared with the teacher and the school principal, documented in the school files, and delivered formally to teachers through the e-school portal.

Supervision may also involve peer supervision, with teachers observing each other, and documenting what they see. These peer supervision or cluster meetings may take place with or without the presence of a supervisor. Teachers also engage in self-evaluation; for example, they maintain a learning journal, where they record observations from the lessons they have taught (e.g., the beginning of the lesson, the ways students interacted with one another, what students have learned). Supervisors can ask to see teachers' learning journals.

The MoEHE applied school evaluation with comprehensive follow-up between 2011 and 2018. The school conducted self-evaluation involving a committee made up of the school administration and teachers; a report of the assessment was issued; the directorate completed the external evaluation report within three working days so that the internal and external evaluation decisions were discussed, and schools were categorized into four levels, each with its own follow-up and planning procedures.

Currently, there is a mismatch between the evaluations of teachers (with many teachers classified as 'excellent' and 'very good' by their supervisors) and the performance of students on assessments (learning outcomes), including national assessments. There is a need to revisit and recalibrate performance evaluation in schools, both in terms of the roles and responsibilities of supervisors and principals, as well as in terms of evaluation criteria or standards. The Education Sector Strategic Plan (ESSP) includes a dedicated goal on maintaining efforts to reform the supervision system and provide technical support to teachers within the framework of the school clusters and teachers' professional learning communities. The ESSP also notes a need to change the focus of the supervisor's role (i.e., to encompass mentoring and coaching as well as formal observations, and apply authentic assessment methods, as opposed to traditional ones as a key cornerstone of educational reform.

### B. Use of Assessment Data by Teachers and Schools for Planning

Teachers prepare plans for delivering their lessons based on students' performance, including their test scores. The level of the class determines the focus of the plan. Assessment guidelines are distributed to schools each year in all subjects. These cover mid-term exams, quizzes, final exams, and qualitative assessment. Districts also provide some support to teachers for developing exams, including tables of specifications and examples of good items.

Some teachers may benefit from a refresher course on assessment including testing with a focus on procedures for evaluating the process of learning, building a test based on a table of specifications, and distributing marks based on specified criteria. The National Institute for Educational Training (NIET) has developed training manuals on test preparation by teachers. NIET provides training programs for newly-appointed teachers, unqualified in-service teachers, school leaders, and supervisors. The modules offered to educators in these different groups all includes a focus on assessment, including the use of assessment data to inform planning for teaching and learning.

There is currently limited analysis of results on tests and examinations at the school level. Supervisors provide some support on aspects of assessment to schools and individual teachers or groups of teachers. However, the analysis of results from assessments at school level to inform planning and instructional decision making could be further strengthened. An earlier initiative (pre-2019, but now discontinued) involved 'data utilization teams' supporting schools and teachers to analyze assessment data to strengthen teaching and learning. It involved training personnel at district level to support schools in analyzing and interpreting their national assessment data.

## 2.2 The Way Forward

**Table 5: Key Policy Priorities under Pillar 2**

Policy Priority	Brief Description
A. Establish clear links between school evaluation, school planning and teaching, learning, and assessment	Supervision could be more streamlined, with a clear protocol laying out specific items that supervisors will observe, and teachers held accountable for. Supervision should transition from a semester plan built on pre-defined number of visits per teacher to a more needs-based approach, where struggling teachers receive more frequent visits and mentoring. Student assessment data could be better utilized to inform the various aspects of supervision, planning, and evaluation at the school level to form a coherent whole.
B. View school planning as an iterative process and update plans on a regular basis	School planning should occur on an ongoing basis, with plans updated at least annually, based on new information gathered at school, teacher and student levels. Revisions should be made to the plan in line with students' progress and learning needs. School plans should include central (ministerial) goals, indicators and targets related to assessment, with some indication of how the school can work towards achieving those goals.
C. Provide guidance and build capacity among school staff on how to leverage assessment data for planning purposes	Schools and teachers need more support by the district and central in analyzing the outcomes of both formative and summative assessments, so that they can develop evidence-based approaches to planning. This includes monitoring the impact of teacher professional development in assessment on teachers' assessment practices.

### A. Establish clear links between school evaluation, school planning, and teaching, learning, and assessment

There should be links between different components of evaluation and assessment. School evaluation (whether external or internal), teacher evaluation and assessment of students should all focus on this same goals. These might relate to implementing a new approach to assessment, or to using existing assessment data for new purposes. Either way, the assessment initiative should be included in school evaluation (external and internal), the school plan and teachers' plans for teaching and learning.

Some schools and teachers may struggle to bring about change in teaching, learning and assessment, based on clearly articulated school plans focusing on assessment. There should be clear consequences for such schools and teachers. These may include the provision of additional supports (by, for example, supervisors), additional external evaluation at school and class levels, more frequent observation of teachers by colleagues, and in-class demonstrations on improving teaching, learning and assessment.

### B. View school planning as an iterative process and update plans on a regular basis

School plans for assessment should be updated on a regular basis as new information about the impact of the current plan becomes available. Such information can come from school evaluation, teacher evaluation, and information that has been gathered on students' learning, including test data (e.g., the results of formal and informal tests administered by teachers).

Central goals, such as achievement targets for national and international assessments, should be reflected in school plans, and steps to achieve those goals at school and classroom levels should be identified in the school plan. For example, if the central goal is to increase the proportion of students reaching the Low Benchmark in TIMSS mathematics and science (and, by inference, reduce the proportion below the Low Benchmark), the school plan and teachers' own plans should indicate how this is to be achieved. Activities could include:

- **A stronger focus on problem solving in mathematics and on interpreting the outcomes of experiments in Science**
- **The inclusion of additional assessment activities to check on student learning during mathematics and science classes (these would be formative assessment activities, such as teacher questions, short tests, and activities in which students assess their own learning).**
- **Gathering of summative data on students' progress in mathematics and science on a regular (e.g., monthly) basis, though, for example, administering short tests. The results discussed among teachers in the context of determining if sufficient progress has been made.**

### C. Provide guidance and build capacity among school staff on how to leverage assessment data for planning purposes

If assessment is to become a central focus of school planning, school leaders, and teachers will need support in analysing and interpreting assessment data. Such data may arise from:

- **Students' responses to informal assessment activities, such as writing up a report on an experiment, or explaining how a particular mathematics problem was solved;**
- **Students' performance on teacher-made tests**
- **Students' performance on external tests, including unified tests**
- **Students' performance on national assessments.**

Data based on these assessments can then be used to update the school plan and teachers' own plans in ways that will further improve student performance. The process can be viewed as cyclical: teach, gather assessment data to see if teaching has been effective, analyze the assessment data, make adjustments to school and teacher plans, and start over again.

In parallel, the impact of preservice and inservice professional development in assessment needs to be monitored on a consistent basis, including:

- **Teachers' levels of satisfaction with the content and process of professional development on assessment, by administering surveys following professional development input**
- **Teachers' implementation of assessment strategies, by observing teachers in the context of teacher evaluation or school evaluation and discussing their assessment practices with them.**
- **The outcomes of these activities should feed into subsequent professional development, to ensure that it meets the needs of schools and teachers, and that it is applicable to their teaching contexts.**

## Pillar 3. Assessment for formative and screening purposes

Under the third pillar, the Framework explores the role of student assessment for formative and screening purposes. Formative assessments aim to provide actionable information to teachers and students, with results used to adjust learning activities as needed to teach students at the right level. Screening assessments, on the other hand, are typically brief assessments administered to all students at a given grade level. They are used to identify students who are at risk of or failing to achieve key learning outcomes in a specific skill or subject and provide them with tailored support.

## 3.1 The Current Situation

### A. Screening check for literacy development in the early grades

Building on successful experiences in other countries in the Middle East and beyond, the Ministry has developed new screening check for Arabic literacy. The test identifies children who are not on track to acquire early literacy skills such as letter knowledge, syllable reading, word reading, sound recognition and syllable deletion, by the end of grade 1. Building on this test, teachers can provide them with targeted support. The test was successfully piloted in December 2023 with Grade 2 students and in April 2024 in Grade 1 students in the West Bank. The Ministry intends to gradually roll out the screening check across all governorates.

### B. Types of formative assessment

Formative assessment is a crucial part of the learning process and provides ongoing feedback to both students and teachers to guide instruction and learning. It can also be described as assessment for learning. The role of these assessments in improving student learning is multifaceted. It provides immediate feedback and real-time information about student understanding, allowing for immediate adjustments in teaching to individual student needs and can guide teachers in planning future lessons based on student needs. Formative assessment also encourages reflection and prompts students to think about their own learning process and progress. This can help students learn how to monitor and direct their own learning. Formative assessments are low-stakes<sup>5</sup>, offering opportunities for students to practice and improve without the pressure of high-stakes testing. They also promote engagement and can make learning more interactive and engaging for students. Formative assessment can take many forms, including, for example:

- **Quizzes:** brief, low-stakes tests that can be given frequently to assess learning progress.
- **Exit tickets:** short, focused tasks given at the end of a lesson to check comprehension.
- **Observations:** teachers observing students as they work, taking notes on their
- **Self-assessment:** students reflecting on their own learning and identifying areas for improvement.
- **Peer assessment:** students providing feedback to each other on their work.
- **Think-pair-share:** students think about a question individually, discuss with a partner, then share with the class.
- **One-minute papers:** brief writing tasks where students summarize key points or express confusion.
- **Digital tools:** various apps and platforms that allow for quick, interactive assessments.

While data on how teachers in Palestine administer and use different types of formative assessment in their day-to-day classroom instruction is lacking, supervisors report that teachers deploy a wide variety of different strategies, including 'authentic' assessments and project-based work. However, both supervisors and NIET staff report that teachers' capacity in formative assessment strategies could be further strengthened. Areas for further improvement include the knowledge and application of different formative assessment strategies, as well as practical advice on how to act upon the results and provide struggling students with extra support and feedback designed to improve their performance.

Supervisors and NIET staff have highlighted the importance of collaborative learning, peer assessment, and self-assessment as particularly important because they require students to internalise what good work means and communicate the outcomes to themselves or other students. Research has also shown that these strategies can have a positive impact on learning. They involve students engaging in self-regulation, planning and monitoring their progress towards learning goals in a deliberate self-directed manner. Rubrics or checklists could be used more often to support the processes of peer assessment and self-assessment.

<sup>5</sup> Low-stakes assessments are those that do not have a high impact or significant consequences affecting the life course of the test-taker, such as short quizzes or educational tests aimed at enabling students to master a particular subject. On the other hand, high-stakes assessments have consequences on a person's life and their professional and academic career, with an example being the General Secondary Education Examination (Tawjihi)



Supervisors have observed some teachers effectively link collaborative learning to peer and self-assessment. This can entail group members working together to complete authentic assessment tasks, with each individual assigned a role, and assessed (by the teacher, by peers or by themselves) on how they perform that role.

NIET already provides training on aspects of formative assessment to teachers in Palestine. For example, techniques such as the use of exit cards at the end of a lesson have been introduced. However, training on assessment for learning should be further expanded, with particular emphasis on the range of formative assessment tools that can be deployed. NIET staff have also pointed to the need for more intensive training on how to administer and score authentic assessments. While teachers in Palestine have received several courses on authentic performance assessment, anecdotal evidence suggests that they need further support on implementing such assessments and using the results to design effective interventions.

Curriculum plays a key role in formative (and summative) assessment, as the ultimate goal of teaching, learning and assessment is to ensure that students have achieved key curriculum outcomes or standards appropriate to their grade or developmental level. Efforts to link curriculum and assessment can be supported by ensuring that teachers and students are aware of key curriculum outcomes, and how achievement of those outcomes can be demonstrated. It will be importance to align the assessment in textbooks with both international and national assessments, as well as ensure that teacher preparation is consistent with current assessment methods. Moreover, standards and guidelines should be established for curriculum designers and developers regarding assessments that promote students' acquisition of knowledge and skills.

### 3.2 The Way Forward

**Table 6: Key Policy Priorities under Pillar 3**

Policy Priority	Brief Description
A. Make an Arabic reading screening check available to all schools	Building on the pilot experience, the Arabic reading screening check will be made available to all schools, accompanied by clear communications to principals, teachers, and parents around the objectives and expected use of the tool.
B. Strengthen teachers' mastery of formative assessment practices	Teachers' understanding of formative assessment, formative feedback, and use of formative assessment results could be further expanded and strengthened through targeted trainings. The MoEHE could further explore the establishment of a resource bank on formative assessment that provides teachers with best practice examples and fosters collaboration among teachers.
C. Empower teachers to adjust their instruction based on formative assessment results	Formative assessment's effectiveness hinges on teachers' ability to adapt their instruction based on assessment results. However, rigid curriculum expectations can hinder this flexibility, necessitating clear guidance from education authorities at all levels to empower teachers to focus on core knowledge and skills.
D. Explore the feasibility of remote formative assessment approaches	Continued disruptions to in-person schooling may require increased utilization of remote formative assessment approaches. Various digital tools for formative assessment exist but challenges related to availability of devices, connectivity, digital skills, and data security need to be considered. Alternative approaches such as phone-based assessments could be further explored. Resource banks can be provided as a source of formative assessment tools such as quizzes and tests for teachers and parents.

#### A. Make an Arabic reading screening check available to all schools

The primary purpose of the screening check is to help schools to identify children who may need additional support. A secondary benefit of the screening check is to provide data to monitor the effectiveness of the education system in providing on-time development of a small set of the most important early literacy skills. Given the primary purpose, the screening check needs to be short and quick to administer. Building on the pilot experience, the Arabic reading screening check will be made available to all schools through a gradual rollout. In parallel, targeted interventions to support students are being developed.

#### B. Strengthen teachers' mastery of formative assessment practices

Teachers in Palestine have already received some training in formative assessment that can be built upon to further expand and strengthen their understanding of formative assessment, formative feedback, and use of results. Given strong evidence on the benefits of student peer and self-assessment, teachers should be supported in implementing those to promote students' metacognitive skills and capacity for self-regulated learning. Trainings should also cover relevant pedagogical content knowledge as well as assessment literacy, which have been identified as necessary pre-requisites for effective formative assessment.

A key element of effective formative (and summative) assessment is the use of a range of question types in both oral and written contexts, including those that elicit higher-ordering thinking. The provision of opportunities to teachers to improve their skills in developing a range of question types (e.g., open and closed questions, inferential and evaluative questions) can lead to improved formative assessment, and might also be expected to transfer to summative assessment (e.g., end-of-term tests). Students can also be involved in the construction of questions, and in identifying the skills required to answer different types of questions. Questions drawn from international and national assessments can be used to demonstrate the range of question types that teachers might construct.

The MoEHE will continue exploring the gradual expansion of a resource bank on formative assessment that provides teachers with best practice examples and fosters collaboration among teachers. For example, the resource bank could include sets of authentic, performance-based assessments that could be administered to students, and then used as a basis for addressing students' learning needs. These assessments could be designed centrally, with district-level subject specialists/supervisors supporting teachers to implement them, interpret the outcomes, and change instruction based on those outcomes. The design and purpose of such a resource bank would have to be carefully considered and communicated to teachers to ensure it is appropriately integrated into instructional practices. Formative assessment is often gathered in a less-structured and more informal manner than other approaches to assessment and can derive from a range of different assessment sources, including observations, portfolios, practical demonstrations, paper-and-pencil tests, peer assessment, self-assessment, and dialogues. The resource bank should therefore not be perceived as a prescriptive and exhaustive source of formative assessment strategies and approaches.

#### C. Empower teachers to adjust their instruction based on formative assessment results

Formative assessment can only deliver if teachers feel fully empowered to adjust their instruction based on the results. In a system that follows relatively rigid expectations on covering certain materials and completing the textbook by the end of the academic year, teachers may not feel comfortable tailoring their instruction if students are falling behind. Clear communication and guidance from central MoEHE, district-level supervisors, and school principals are required for teachers to focus on core knowledge and skills in their respective subjects. This is particularly important in the current context in the West Bank, with students in public schools having received distance education for up to three days per week since October 2023. For Gaza, a comprehensive package of targeted education interventions outlined in the Ministry's Integrated Education Response Plan will be required.

## D. Explore the feasibility of implementing remote formative assessment approaches

Remote learning settings present unique challenges for formative assessment, but there are some tools that can be used. For example, this includes online quizzes and polls, digital exit tickets or discussion boards, screen recording tools through which students can record themselves explaining concepts or solving problems, collaborative documents, and other tools. When considering digital tools for formative assessment, it is critical to examine accessibility constraints that students and teachers may face, such as lack of digital devices, digital skills, and reliable connectivity, as well as data privacy and security concerns. Other options include phone-based approaches that leverage text messaging for formative assessments, and the use of take-home activities such as homework or open-book tests for formative assessment purposes.

## Pillar 4. Assessment for summative purposes

Under the fourth pillar, the Framework reviews the role of summative assessments at different grade levels, including the high-stakes secondary school leaving examination ‘Tawjihi’ at the end of Grade 12. The term ‘summative assessment’ is generally associated with an end-of-term or end-of-course mark or grade. The grade can be based on test performance, students’ performance on tasks set by the teacher, or a combination of both.

### 4.1 The Current Situation

#### A. Summative assessment in grades 1-4

In Palestine, formal tests are not administered to students in Grades 1-4. Instead, the assignment of grades by teachers is based on in-school tasks and coursework completed by students and on performance assessments. The outcomes of this work provide a basis for assigning grades ratings from Excellent to fair to students by their teachers. The grades represent students’ understanding of key content and processes. However, students who fail consecutive grades are still promoted. Teachers may administer some tests to their students in Grade 4. These tests do not inform the grades assigned to students. Rather, their purpose is to prepare students for subsequent grade levels. Although, in general, teachers working at the same Grade level or teaching the same subject do not meet to discuss the assignment of grades to students, teachers may meet to discuss students who are to be assigned a failing grade.

#### B. Summative assessment in grades 5-12

In Grades 5-12, teachers in Palestine administer a range of different summative assessments. These include quizzes, mid-term tests, authentic assessments and a final exam at the end of each semester. Tests or exams often take up a whole class period, while quizzes can be administered more quickly, leaving time for other teaching and learning activities. Specifications on the number and types of summative assessments to be completed are issued to schools by the district every year. Approximately 30% of the marks that count towards grades in each subject are based on in-school tasks and coursework completed by students. The remaining 70% is based on their performance on tests / exams.

Teacher manuals are available for the more qualitative assessments. They include tables of specifications regarding the assessment methods. Each unit includes activities that enrich students’ understanding of the unit. Each student is assigned some part of the activity or project. Some activities are implemented by groups of students working collaboratively. The manual includes assessment criteria for four levels of performance and points out links between the curriculum and the assessment activity. Practical activities (experiments) are implemented in STEM subjects such as Chemistry and Physics classes, and performance on these can also inform students’ grades, though the use of laboratories is not well systemized in Palestine yet and part of ongoing consultations around the national STEM education Framework.

Supervisors report that teachers allocate a considerable amount of time to summative assessment, and this is not limited to core subjects. For example, a Grade 8 student taking 8 subjects might take 4 exams or quizzes in each subject each semester, leading to 64 tests in total during an academic year. This could be at the expense of teaching and learning, especially when assessments are not of high quality.

A key role of supervisors is to ensure that teachers use assessment tools effectively – not only in the context of exams, but to monitor the progress of students on an ongoing basis. Supervisors occasionally review the tests given to students, including the content and distribution of marks.

Supervisors and NIET staff have expressed concerns about the quality of test items prepared by teachers. Anecdotal evidence suggests that exams prepared by teachers often include lower-order questions that may be linked to textbook content instead of high-order questions that measure such traits as creative thinking, evaluation, critical thinking etc. Some supervisors suggest that the low performance of Palestinian students may be partly attributable to the differences between the tests authored by teachers and administered in schools, and those used in international assessments, which were believed to include more items assessing problem solving and other aspects of higher-order thinking. On PISA 2022 mathematics, 80% of 15-year-olds in Palestine performed below Proficiency Level 2, meaning that they did not have sufficient skills to solve mathematical problems in real life, or to engage in further study of mathematics.

Teacher training on how to develop test items is provided by a wide range of groups, including universities, supervisors, non-governmental organizations, and NIET. There are also concerns about grade inflation, as students are awarded higher grades than their performance on external assessments (e.g., national assessments) might support. However, a data analysis of school grades in Grades 10-12 in the West Bank in 2021-2023 did not find strong evidence for that.

#### C. The Tawjihi exam

The Tawjihi plays a key role in Palestinian society and determines students’ career trajectory. Each year, around 80,000 12th grade students in Palestine sit for the secondary school leaving examination. On average, about 65-70 percent of students pass the exam during the first exam period in June/July, with students in the scientific stream typically having higher success rates than the humanities stream. Similar to other countries in the Middle East and around the world, the Tawjihi serves a dual purpose: (1) it certifies completion of secondary school, and (2) it determines admission into higher education. Students’ success is entirely determined by the Tawjihi, as grades from coursework completed during the upper secondary education cycle are not factored into students’ final grade.

The Tawjihi test items are designed by a central exam committee, which is formed of subject-matter experts from the Ministry’s district offices. The test design differs for each secondary school stream. In total, there are 9 Tawjihi streams: humanities, science, technology, entrepreneurship, industry, agriculture, hotel management, home economics, and religious studies. The majority of students are enrolled in the humanities stream and the scientific streams, accounting for around 62 and 27 percent of 12th graders, respectively. About 2-3 supervisors per subject are officially appointed by the Ministry of Education to develop test items for the Tawjihi. While supervisors are appointed as test item writers on an annual basis, the composition of the committees typically does not change over the years. Supervisors involved in the exam design are not allowed to offer any private tutoring or have any direct relative applying the test in that specific year. The test development is a highly confidential process, and all items are exclusively developed and reviewed by the supervisor committees.

The administration of the Tawjihi is a major logistical effort, costing around US\$7.6m annually. A strong set of procedures is in place to ensure the standardization of the examination. The MoEHE has issued a comprehensive procedural manual that provides a single reference document. Key elements of standardization include the exam questionnaires, scoring criteria, data processing, and reporting of results.

If students fail the Tawjihi exam for some subjects or want to improve their grade, they have the opportunity to take a make-up exam. Following the main exam in June/July, the make-up exam cycles are usually implemented in August and in December/January of the following academic year. Unless in special cases

such as sickness or death of a family member, students are allowed to sit for a make-up exam for a maximum of 4 subjects.

The quality of test items in the Palestinian Tawjihi has been the subject of a national debate for years. Like other high-stakes examinations in the region, the Palestinian Tawjihi is said to rely heavily on rote memorization. In addition, there have been challenges with the technical soundness of test items, including questions that were not clearly phrased or extremely difficult to solve, causing outrage among students and parents. The high-stakes nature of the Tawjihi is a key challenge as well. It fuels a large private tutoring industry that further exacerbates inequalities. Those who do not pass the Tawjihi lack any formal document certifying that they have completed 12 years of education and are left with bleak prospects for the job market.

## 4.2 The Way Forward

**Table 7: Key Policy Priorities under Pillar 4**

Policy Priority	Brief Description
A. Review the frequency of summative assessments	The frequency of summative tests and examinations should be reviewed to achieve a better balance between continuous assessment and summative assessment. The MoEHE could explore replacing some of them with project-based work and more formative assessments. The introduction of age-appropriate assessments in earlier grades will be considered.
B. Provide teachers with more support on how to design, grade, and utilize summative assessments	Support may include additional training to better balance testing of lower- and higher-order skills, the provision of annotated examples of students' work demonstrating the application of scoring rubrics, and support in moderating standards within schools. The MoEHE further intends to establish an item bank that enables teachers to generate tests with known difficulty levels including objective items and more open-ended performance tasks.
C. Continue consultations with all relevant stakeholders about the prospects of a gradual reform of the Tawjihi exam	The MoEHE has been considering a reform of the Tawjihi exam for several years to address challenges with its content and function. Consultations should be continued to move towards a gradual implementation of incremental changes that are well understood and supported by the Palestinian society.

### A. Review the frequency of summative assessments to achieve a better balance with formative assessments

Given the high assessment load placed on students in Grades 5-12, consideration should be given to reducing the numbers of quizzes, mid-term tests, and end-of-semester tests by replacing them some of them with performance-based assessments. This would also increase time available for quality feedback to students about their performance. While there is no hard and fast rule in terms of the desirable number of summative assessments that should be administered in a given school year, a re-balancing towards more and better continuous assessment across subjects should be considered. There may also be value in asking students to reflect in writing on their academic strengths and weaknesses, based on the assessments they have completed – whether formative or summative. Such an activity could help students to better regulate their own learning, by identifying aspects on which they need to improve. Introducing age-appropriate assessments that do not create undue pressure in earlier grades will be considered as they could help familiarize students with assessments and track student progress more effectively over time to support early intervention.

### B. Provide teachers with more support on how to design, grade, and utilize summative assessments

Teacher professional development is central to all efforts on improved student assessments. Item banks are often linked to great expectations in improving student assessment in the classroom. However, in practice it has often proven challenging in many countries to fully reap the benefits of item banks. As the Ministry is intending to establish an item bank, the OECD (2023) has identified a number of issues and prerequisites to consider, including:

- Defining the purpose, scope, and audience (e.g., grade levels and subject areas included; types of assessment targeted, such as formative and/or assessment; open vs. restricted access).
- Setting up the governance, including who has ongoing responsibility for maintaining and updating the bank, and allocating the needed financial resources.
- Developing the content and organisation (e.g., developing items or assessment tasks, gathering data on student performance (metadata), and providing additional resources such as scoring rubrics for open-ended questions, sample guidelines for providing feedback, and links to relevant instructional resources).
- Determining the workflow for item development and validation. This may involve peer review of items and tasks, piloting or small-scale trials, and expert review.
- Defining the functional requirements (e.g., the particular online platform selected, and the technical solutions underpinning it). This includes deciding which item types the platform will support. Ideally, the platform will coordinate item review and validation, test assembly, and test delivery, correction and reporting.
- Considering broader technical requirements and their implications for the underlying technology solution, including off-the-shelf software, hosting of the platform, and data security and privacy concerns.
- Supporting teachers in its use through targeted professional development.

### C. Continue consultations with all relevant stakeholders about the prospects of a gradual reform of the Tawjihi exam

Over the years, there have been various reform proposals to reduce the high-stakes nature of the Tawjihi, strengthen its content to assess skills that matter, and separate its dual function of secondary school certification and admission into tertiary education. The introduction of a Secondary School Certificate has been under consideration. The certificate would be conferred to every student who successfully passes Grade 12 (irrespective of their Tawjihi performance) to acknowledge their completion of 12 years of education. It could report on the following dimensions that would provide value to the Ministry and potential employers, including school attendance, extracurricular activities, and students' proficiency levels in selected competences, evaluated by teacher committees based on agreed-upon standards that can be informed by international reference frameworks. Attributing some weight to assessments and project-based work completed in Grades 10 to 12 towards the final Tawjihi grade has also been considered and could be further explored.

Effective stakeholder engagement on the Tawjihi reform is key to manage risks and shape the debate. Setting up participatory processes to engage the education community and the general public is critical. High-stakes examination reform is always contentious and emotional as it touches upon the interests of a diverse set of stakeholders and determines students' future prospects. Transparency and clarity about the goals of the envisaged reform, public consultations, and participatory decision-making are important to establish trust and confidence in the reform process.

Moving beyond the stakes and the function, enhancing the Tawjihi's effectiveness and relevance necessitates a focus on refining item writing and review processes to ensure that exam questions are clear, fair, and aligned with educational standards. This will include incorporating a diverse range of question types that evaluate both factual knowledge and higher-order skills. Additionally, piloting new question formats and items in a controlled environment before full-scale implementation can help identify and address potential issues. By adopting these measures, the Tawjihi exam will be better positioned and trusted to assess students' true capabilities, foster the development of higher-order skills, and provide a more accurate reflection of educational attainment.

## Glossary

**Early Grade Reading Assessment** is a standardized oral student assessment typically administered in the early grades of primary education. It evaluates early literacy skills such as letter recognition, reading comprehension, and listening comprehension.

**Formative assessment** comprises “all those activities undertaken by teachers and/or by their students [that] provide information to be used as feedback to modify the teaching and learning activities in which they are engaged” (Black & William, 1995, p. 7). To serve a formative purpose, assessment needs to provide actionable information for teachers and students (Shepard, 2005). Formative assessment is not necessarily based on a particular instrument or task, but is defined by its purpose – shaping or influencing student learning during the learning process. It can provide information about a student’s progress, thought processes and misconceptions.

**Literacy and Numeracy Assessment (LaNA)** is an international, large-scale assessment of literacy and numeracy skills by the International Association for the Evaluation of Educational Achievement (IEA) targeting students at the end of primary education in low- and middle-income countries administered on demand rather than following regular cycles. It is designed to be a shorter, less demanding assessment compared to TIMSS and PIRLS.

**Peer Assessment** is a student-centred assessment approach that allows students to develop a deeper insight into the quality of their own work through the assessment of peers’ work. Students take responsibility for assessing the work of their peers using the assessment criteria. Students’ understanding of the assessment criteria is a key component of successful peer assessment. (<https://teachinghub.bath.ac.uk/>, 2014).

**Performance assessment** is a task/event/performance designed to measure a student’s ability to directly demonstrate particular knowledge and skills. e.g., a student may be asked to demonstrate some physical or artistic achievement: play a musical instrument, create or critique a work of art, or improvise a dance or a scene. These kinds of assessments (e.g., tasks, projects, portfolios, etc.) are scored using rubrics: established criteria for acceptable performance. (State Collaborative on Assessment and Student Standards).

**Progress in International Reading Literacy Study (PIRLS)** is an international, large-scale assessment of reading targeting grade 4 by the International Association for the Evaluation of Educational Achievement (IEA) and administered every five years.

**Programme for International Student Assessment (PISA)** is an international, large-scale assessment by the Organisation for Economic Co-operation and Development (OECD) assessing 15-year-old students in reading, mathematics, and science every three years.

**School evaluation<sup>6</sup>** is a systematic, goal-oriented, and criteria-based process conducted by an external authority consisting of data collection (most often including site visits) and data feedback on school quality serving accountability/control purposes, enforcement of policy, and/or school improvement. (Hofer et al., 2020)

**Student Self-assessment** is the involvement of learners in making judgements about their achievements and the outcomes of their learning. It can support student learning, particularly when used formatively (Wride, 2017).

**Summative assessment** can be defined as ‘the assessment of students that occurs at the end of a period of instruction. [It] provides a holistic measurement of an individual’s knowledge, skills and dispositions. (Nicholas, 2018, p. 1634). Summative assessment generally compares student performance to some benchmark or standard. Examples of summative assessments include mid-term exams, final projects, papers and final recitals. Information from summative assessments can be used formally by teachers or students to guide future learning.

**Trends in International Mathematics and Science Study (TIMSS)** is an international, large-scale assessment in mathematics and science by the International Association for the Evaluation of Educational Achievement (IEA) targeting grades 4 and 8 and administered every four years.

---

<sup>6</sup> School evaluation differs from classroom evaluation in its scope and focus. While classroom evaluation emphasizes student performance and teaching and learning processes within classrooms, school evaluation addresses broader aspects such as leadership and management, infrastructure, overall school environment, effectiveness of school-wide programs, and the performance of the school as a whole in achieving its educational and pedagogical objectives.

