جانب المدير العام للتربية

جانب رئيس المركز التربوي للبحوث والإنماء

الموضوع: الموافقة على اعتماد "الإطار العام للاستراتيجية الوطنية لتقييم التعلم لدى التلامذة"
National Student Learning Assessment Framework (NSLAF)

المرجـــع: - كتاب المركز التربوي للبحوث والإنماء رقم ٢٢٠/م تاريخ ٢٠٢١/٤/١٤.
- كتاب المديرية العامة للتربية رقم ٢٠٢١/٣/٤٨١٦ تاريخ ٢٠٢١/٦/٢٨.

بالإشارة إلى الموضوع والمرجع المبيّنين أعلاه،

استنادا إلى رأي كل من المركز التربوي للبحوث والإنماء والمديرية العامة للتربية بشأن "الإطار العام
للاستراتيجية الوطنية لتقييم التعلّم لدى التلامذة" National Student Learning Assessment Framework
(NSLAF) (المرفقة ربطاً) التي جرى إعدادها بالتنسيق والتعاون بين كل من المركز التربوي للبحوث
والإنماء والمديرية العامة للتربية ومشروع QITABI المموّل من الوكالة الأميركية للتنمية الدولية
(USAID)، وبدعم تقني من مؤسسة Cambridge Assessment International Education،

لذلك،

فإننا نبلغكم موافقتنا على اعتماد "الإطار العام للاستراتيجية الوطنية لتقييم التعلّم لدى التلامذة" لدى وزارة
التربية والتعليم العالي، على أن يصار إلى عقد ورش عمل تضم المعنيين في كل من المركز التربوي للبحوث
والإنماء والمديرية العامة للتربية ترمي إلى النظر بالتوصيات المدرجة ووضع الخطط الإجرائية اللازمة
لتفعيل العمل، وتيويم هذه الخطة دورياً ٠/٠

وزير التربية والتعليم العالي

طارق المجذوب

مرفق ربطاً:

National Student Learning Assessment Framework (NSLAF)
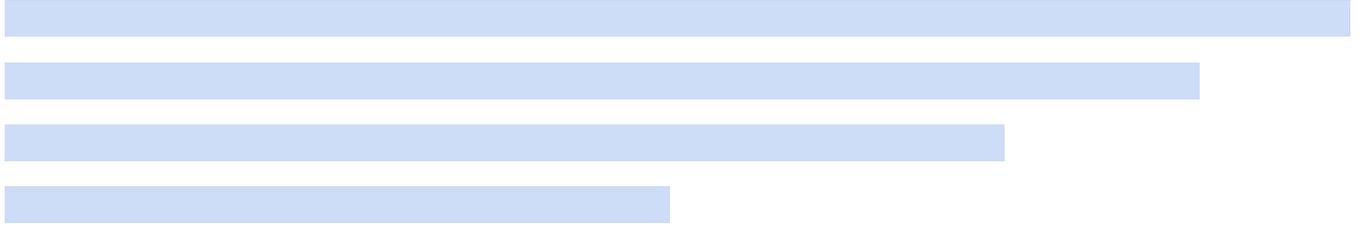
National Student Learning Assessment Framework (NSLAF)

Lebanon
Date: November 2020

Table of contents

Executive summary

The Lebanese government is committed to reforms that will develop a modern, effective and coherent school system in Lebanon. A key part of these reforms is introducing a National Student Learning Assessment Framework (NSLAF). The NSLAF sets out a plan to provide a coherent framework of assessment of students in Lebanon.

The Framework that is presented here aims to establish a coherent system of assessment that provides stakeholders with reliable assessment data and other information. This assessment data can be used to help individual students progress with the support that they need, make fairer judgments about student performance and allow policy makers to identify areas of the education system that could be improved.

The Framework is based on basic sound principles that form the basis, in education systems all over the world, for the fair and evaluation of student achievement. These principles are that the assessment processes must be curriculum aligned, equitable, interpretable and dependable, and manageable. In addition to these principles the NSLAF has been designed to indicate how the assessment framework can be sufficiently flexible to take account of planned curricula reforms.

The review of the current assessment landscape produced these general findings in relation to the current assessment landscape:

- *Classroom assessment:* High-stakes decisions about whether students should progress to the next grade are made based on a combination of ongoing assessment and teacher-set end of semester summative written tests. Procedures are currently in place to support effective teacher assessment, but these are applied unevenly and with some inconsistency. This can result in variations in the level of difficulty between different teachers and different schools.
- *National examinations:* The Brevet and the Baccalaureate require a high level of reliability and security to ensure fairness for students and confidence in the outcomes. The experts from Cambridge International, whilst acknowledging the rigour of the present arrangement, have proposed some improvements to the current system related to question paper authoring, marking and maintenance of standards.
- *Coherence and continuity:* There are occasions and opportunities for students' performance to be assessed at all stages of their school life. At present there is more national testing for older students, and currently no routine screening of children when they initially enrol into the school system. Equally, Lebanon does not take part in international comparative studies and surveys for students in the elementary school grades. A clear picture of how standards in the Lebanese education system compare to those in other countries only emerges when the children are around 13-14 years old.

To address these features of the current assessment landscape the following recommendations are made, based on the findings of the review by the Cambridge experts:

Initial screening: A national screening programme should be introduced for children in the first year of full-time schooling. The screening should assess each child's readiness to participate in primary education against minimum criteria levels. The results of the screening will allow teachers to identify children with developmental needs in particular areas and put in place the interventions necessary for them to reach the required level.

Classroom assessment: A national professional development programme beginning with a series of national and regional workshops should be initiated by CRDP with the aim of improving teachers' assessment literacy. These will develop teachers' skills and understanding of the use of formative assessment and improve teachers' capacity to and consistency in setting school-based summative assessments. Assessment literacy professional development programmes will also strengthen teachers' skills in using other forms of alternative and authentic assessment techniques and strategies for evaluating student skills, competencies and attitudes as well as their knowledge and understanding.

National examinations: The examination specification, formulated by CRDP, and the operation of the examination system by MEHE-GD should be improved to ensure consistency and comparability between subjects, the reliability of the markers and the maintenance of the standard over time. The report contains details of the recommended professional development programmes to achieve this.

Coherence and continuity: a programme of initial screening should be introduced in grade 1 to assist in the early identification of children with developmental needs for intervention purposes; consideration should be given to participation in the TIMSS study for grade 4 students in order to provide a better understanding of national achievement standards in the primary school grades; the NSLAF should inform the development of a data strategy to ensure that the data generated is used more effectively for student progression and for channelling.

During these times of acute economic distress in the public sector in Lebanon, the cost implications of addressing these proposals in full may be unaffordable. Difficult decisions and compromises must, therefore, be taken, about, for example, the cost-effectiveness of participation in international assessment studies, and the introduction of the TIMSS tests at grade 4; yet the data generated from this study could be strongly justified as, among other benefits, the TIMSS tests, which are currently only administered in grade 8, if also taken in grade 4 would provide vital cohort progression information in those critical school years between grades 4 and 8. This information would create a firm foundation on which to build a more sophisticated interpretation and analysis of the findings from the PISA study, relating to students at the age of 15 years.

1. Introduction

As part of reforms to create a modern, effective and coherent school system in Lebanon, the education institutions of Lebanon, guided by the Centre de Recherche et de Développement Pédagogiques (CRDP) and the General Directorate of the Ministry of Education and Higher Education (MEHE-GDE). have identified the need for a long-term learning assessment strategy to support Lebanon's educational goals for pre-university education.

This Framework is the product of a collaboration between colleagues associated with the USAID-funded QITABI programme, CRDP, MEHE-GDE, DOPS with technical guidance provided by Cambridge Assessment International, a department of Cambridge University, UK.

The research and consultations that underpin the findings, which are aligned with the World Bank's S2R2 initiative, ran between December 2018 and December 2019. Consultants from Cambridge International provided their technical expertise to the QITABI project and there was close collaboration with CRDP and MHE-GDE. In addition, the project team undertook a range of other stakeholder engagement activities including school visits; meetings with teachers, principals and initial teacher training providers; and observation of national examinations processes. The NSLAF is based on and builds upon the existing assessment practices, systems and infrastructure that have already been developed over many years. The Framework is designed to:
- be broad in scope
- address student assessment, system evaluation, comparability and equivalence
- have the potential to support school improvement activities as well as data-based educational research.

When fully implemented, it will enable:
- accurate indications of early levels of basic literacy and numeracy as well as
- psychosocial development
- dependable evidence-based systems for monitoring student attainment and progress in all school grades
- informed choices for student destinations, particularly at system transfer points
- reliable data from the range of assessment sources to guide policy formulation and evaluation
- comparability indicators of student achievement to higher education providers in Lebanon and abroad
- system strengthening and enhanced accountability.

The NSLAF will serve a range of different assessment need needs including initial screening; ongoing classroom assessment; certification of learning for national qualifications; and providing data for system monitoring and evaluation with particular reference to the indicators for the United Nations' Sustainable Development Goal for Quality Education [1].

---

[1] https://sustainabledevelopment.un.org/sdg4

2. Vision for the National Student Learning Assessment Framework

A well-designed assessment system can provide valuable information to help improve educational outcomes for students, inform education policy decisions and monitor their impact. The National Student Learning Assessment Framework (NSLAF) encompasses a series of high-quality assessments for schools in Lebanon. It is designed to provide reliable data that can be used to support better learning outcomes for individual students, make fairer judgments about student performance. It will provide the basis for establishing an equivalency across different examinations and systems, for example national, international and vocational. It will allow policy makers to make better informed decisions and improved future interventions into the education system.

To achieve this vision, the NSLAF sets out how a coherent system of assessments that align with the wider education system in Lebanon can be achieved. It has been designed to reflect the aims, knowledge, competencies and goals of the curriculum and is compatible with proposals for future curriculum updates.

Benefits to students, parents and teachers

The NSLAF will support students' learning throughout their school education, maximizing their chances of success and recognizing their achievements. It sets out a system of high-quality assessments that will be timely, appropriate in form, and fair and equitable. It will:

- Establish a strong and reliable structure within which the value and credibility of the certification of students' learning achievements is maintained; this will support public confidence in the awarding of national diplomas.

- Give parents clear, reliable and concise information about their children's achievements, their strengths, their weaknesses and their progress in comparison with expected norms – depending on local circumstances the information could be provided through termly or annual written reports or through electronic media

- Enable students and families to be make informed choices for their future learning at key points in the education system

- Identify students' learning needs early, so that these can be addressed effectively

- Unify the language of assessment for educators eg. diagnostic, formative and summative assessment, classroom assessment; survey and screening tools; international assessment surveys

- Inform the priorities, the focus and the content of professional development programmes for educators.

Benefits to higher education institutions and businesses

The NSLAF indicates progression routes for students and will help ensure that national examinations are well constructed, effectively administered and reliably marked. It will also align with curriculum reform by improving curriculum coverage, developing the competencies of future students and complementing their deep subject knowledge. It will enable:

- higher education institutions and employers to have an improved understanding of the school education system and its diplomas
- university and college admissions staff, both within Lebanon and abroad, to have greater confidence in the national examinations when using the results to select students
- employers to be more confident that students with national qualifications will have the skills needed to meet the demands of the 21st century workplace
- national examinations to better reflect the range of practical skills and competencies needed to succeed in different subjects.

Benefits to policy makers

The NSLAF will enrich and extend the information provided by the assessment system. Policy makers will benefit from the availability of a broader range of higher-quality, policy-relevant data, gathered as cost-effectively as possible and with minimum disruption in schools. It will enable:

- better system monitoring and evaluation by providing evidence of impact
- improved accountability throughout the system by producing more reliable assessment outcomes
- the information and data generated to inform related educational research.

3. Context

3.1 Curriculum considerations

3.1.1 The elementary phase: Cycle 1, grades 1–3, and Cycle 2, grades 4–6

A 'basic curriculum' in Cycles 1 and 2 includes Arabic, a foreign language (English or French), mathematics, science, civics, physical education, technology and arts and crafts, among other subjects. Mathematics and science are both taught in Arabic, English or in French from grade 6.

Learning development in each area is assessed informally by class teachers on an ongoing basis throughout the school year.

Learning achievement is summatively assessed at the end of the school year, as it is in every grade throughout schooling. This is based on a combination of teacher assessment and performances on teacher-produced monthly tests and end of semester tests.

3.1.2 The intermediate phase: Cycle 3, grades 7–9

In Cycle 3, students following an academic route continue to study the same subjects as in Cycles 1 and 2, but expanded with the introduction of separate sciences and the addition of some new subjects including a second foreign language (French or English, as appropriate) informatics and history. Learning development is partially supported through ongoing teacher assessment, with more formal summative assessment at the end of the school year.
Students can also enrol on a technical or vocational track, leading to the Certificat d'Aptitude Professionel (CAP) for vocational students and the Brevet Professionel (BP) for technical students.

For most students in Cycle 3, this curriculum provides preparation for the Brevet examinations which are taken at the end of grade 9. However, some students depart from school for work before taking the Brevet, leaving them without the formal recognition of knowledge, understanding, skills and aptitudes that may be required by future employers. Of the 80 per cent or so of students who remain to take the Brevet examinations, just over 80 per cent currently achieve the qualification, i.e. around two-thirds of the original cohort.

Compulsory education finishes at the end of Cycle 2. Some students progress into secondary education to study for either the General or Technical Baccalaureate. Others leave school to enter the workforce.

3.1.3 The secondary phase: secondary cycle – First till Third Secondary

In the first year of their three-year secondary education, students intending to study for an academic or 'general' Baccalaureate follow a common curriculum, extending the education followed in Cycle 3. For the second year they proceed into one of two study streams or 'tracks': humanities or sciences. In the third year, they divide further into four streams: humanities and literature or sociology and economics for humanities students and general sciences and life sciences or sociology and economics for science students.

Students in all four tracks also study a common core of subjects in greater or lesser depth: Arabic, French or English, philosophy, geography, civics, history, mathematics, physics, chemistry and life science. Note that the subjects of sociology and economics are only taught in that specifically named stream.

At the end of Third Secondary, studies culminate in examinations leading to the General Baccalaureate: the school-leaving diploma that provides access to university study. Curriculum subjects studied are differently weighted when calculating average subject marks, with greater weight being given to those subjects most clearly related to the focus of the relevant academic track.

Students in the secondary phase can also choose to study for a Technical Baccalaureate.

3.2 Assessment principles

Assessment systems must satisfy a set of general principles in order to be fit for purpose. The system must be curriculum-aligned; equitable; interpretable and dependable; manageable and sustainable; and, in the context of anticipated curriculum reform, 'future-proofed' as far as possible. Assessments should follow international good practice, while also reflecting the history, traditions and current constraints of the Lebanese education system.

3.2.1 Curriculum-aligned

Each assessment system must be aligned with the relevant curriculum, for example the grade 2 numeracy curriculum, the science curriculum at grade 7, the Baccalaureate humanities curriculum and so on. Alignment should reflect the expected progression of students within a subject over time and also between subjects. To enable this alignment, every curriculum will need to have appropriate taxonomies, vocabulary and progression models.

Some practical and performance skills cannot always be assessed effectively through written examinations and such skills and competencies are often best addressed through other forms of assessment such as teacher assessments and project-based assessments. Evidence from these other types of assessment can be considered alongside that gathered from written assessments. It is important to consider how evidence from these different types of assessment is combined to produce a grade or assessment outcome that balances reliability.

It is possible to achieve high levels of reliability through externally set and administered written examinations, though these are not always able to assess practical skills or competencies such as creativity or collaboration in a valid way. Conversely, teacher assessments and project-based assessments can provide more valid assessments of practical skills and competences but may not be as reliable as without robust harmonization and moderation processes there is likely to be significant variation in task demand and marking between individual teachers and across schools.

3.2.3 Equitable

The curriculum and its assessment should be accessible to every eligible young person. Furthermore, every assessment system must, as far as possible, be equitable – that is inclusive and fair to all students. Assessments must be designed and written to minimize bias and provide opportunities for all students to achieve, irrespective of gender; disability or special educational need; or social, linguistic and cultural background.

3.2.4 Interpretable and dependable

Assessment system outcomes must be clearly interpretable for all stakeholders, including students, parents and employers. Assessments must also be accepted as valid by the stakeholders that use the data generated by the assessment. For example, universities need to have confidence that the Baccalaureate provides a dependable measurement of students' ability and readiness to undertake higher education and therefore acts as a suitable entry requirement. This means that the assessments must provide a dependable measure of the content and competences in the curriculum. Data outcomes (test scores and teacher ratings) must be valid, but they also need to be technically reliable.

3.2.5 Manageable and sustainable

Assessment must be manageable in schools and classrooms, creating minimal disruption to normal teaching and learning and being cost efficient and practical in terms of time and resources. It should require as little additional workload as possible for teachers as assessment managers, invigilators and markers and should be supported by an effective data management system.

Assessment systems must also be sustainable in the longer term in order to maximize the return on investment. This implies adequate support in terms of finances and other resources for institutions producing and administering the assessments and an ongoing capacity building programme for teachers and other education and assessment professionals.

3.2.6 Future-proofed

During the design of any assessment system it is essential to anticipate possible changes in the learning and assessment landscape and ensure that, as far as possible, these will not adversely affect its processes or render it out of date. This is often referred to as 'future proofing'. While it is not possible to plan for every eventuality, it is, for example, highly likely that advances in technology will continue to have a significant impact on classroom practices and on how assessment data are processed and reported. These advances may also change the nature of the interaction of students with assessment instruments – for example on-screen delivery of summative assessments.

The NSLAF has been designed to describe principles that remain relevant whichever changes to curriculum, education policy and assessment technology are implemented and, as such, should be used as a reference for future assessment policy and design decisions.

3.3 Private Schools

More than two thirds of all Lebanese students attend private schools. Some of these schools are commercial operations, others are voluntary not-for-profit establishments. The relationship between the public and private schools is a complex one, and there is not a simple binary distinction. Some private schools are completely reliant on parents paying the full student fees, a few receive direct subsidies from the government, so called "free-private" schools; for others, the government provides subsidies for civil servants and security personnel to enable them to pay reduced school fees. There is not a wide public perception that private schools are always superior to public schools. Indeed, in the senior years of schooling there is some traffic from private schools into public schools with parents opting for the knowledge and expertise of teachers in the state sector in the critical preparation years for national examinations.

Most private schools follow the Lebanese curriculum as outlined in the previous section and students take the same examinations for the Brevet and Baccalaureate. In addition, private schools often offer students the opportunity to study another, international, curriculum in parallel. This allows them to take examinations for an international school leaving diploma, most typically the International Baccalaureate or the French Baccalaureate, but other options also feature in the system. The language of instruction is either English or French, depending on the school.

The combination of future Lebanese curriculum reform and the NSLAF will help ensure better alignment of the assessment of the content and skills of the scientific subjects with those in the international curricula. A closer alignment between the curriculum and assessment systems in the public and private sectors would ease the transition of those students who need to move between the two sectors. Additionally, improvements in the reliability of the assessment of the Lebanese qualifications will help ensure that students' achievements are better recognized as being dependable indicators of their ability by higher education providers in Lebanon and abroad. This may reduce the perceived need for private school students to sit for international examinations.

The National Student Learning Assessment Framework
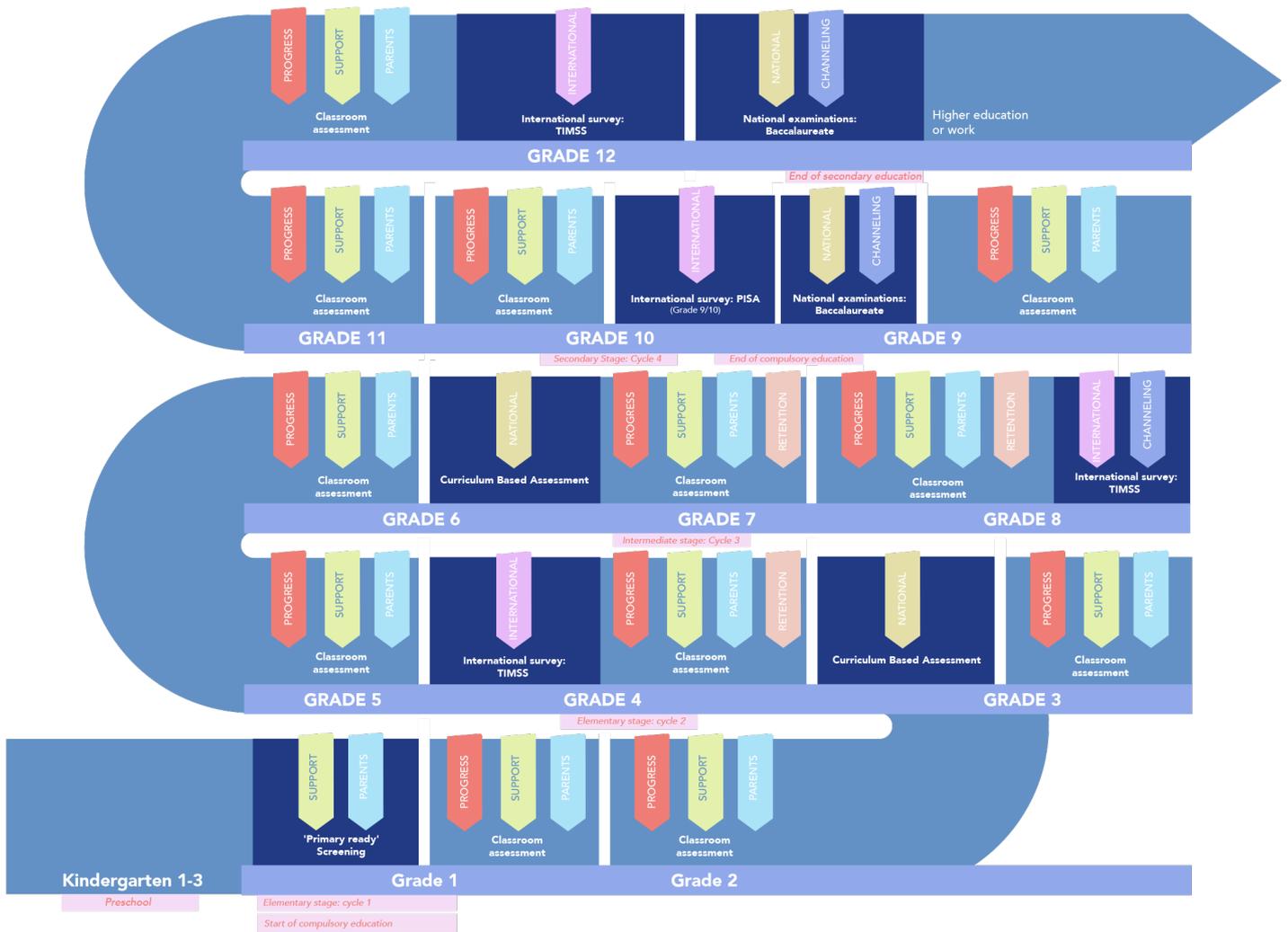
PURPOSES OF ASSESSMENT

PROGRESS Progress check	**PARENTS** Parent/guardian feedback	**NATIONAL** National system evaluation
SUPPORT Identify support needed	**INTERNATIONAL** International surveys	**CHANNELING** Channeling into next stage of education of work
RETENTION Grade retention decision		

KEY
Classroom assessment:
ongoing formative use of assessment
and end-of-grade summative assessment

Functional Model of the Assessment Systems and their Purposes in Lebanese schools

The functional model is designed, in a single graphical representation, to show the purposes of each assessment tool at the various stages of a student's journey through the pre-university education system:

- *Progress* ➜ students' progress is measured at each grade and phase

- *Support* ➜ the progress check will indicate whether or not there is need for learning assistance and support

- *Retention* ➜ the formal assessment tools will indicate whether a student has made sufficient progress during a school year system to be promoted to the next grade or be retained in the current grade

- *Parents* ➜ at every grade and phase parents are informed of the achievement of their child

- *National* ➜ the official examinations of Brevet and Baccalaureate are administered nationally to enable the formal recognition of student achievement

- *Channeling* ➜ at critical points in the school system the assessments are used to channel students into particular specialisms in general education or into vocational education and training

- *International* ➜ at recognized grades and stages international assessment surveys are administered to check on the age-related achievement of students in Lebanon compared with their peers internationally

Conceptual Model of the Assessment Systems and their Purposes in Lebanese schools

At a conceptual level, the Framework can be seen as a series interlocking circles at the centre of which is the learning achievement of the student. The capacity of children, their early skills and their learning potential, at the beginning of the school journey is best measured by screening tests which cover linguistic and numerical competencies as well as cognitive and behavioral ones. Once a student is within the school system their progress and achievement are underpinned by ongoing classroom assessment, some of which will be developmental and formative whereas at set points in the school year the classroom assessment will be summative so as to establish an individual's performance relative to national expectations. At two important points, in the primary education cycles, national surveys are conducted by CRDP in the form of Curriculum Based Assessments to gain a performance measure of a sample students against national benchmarks.

Crucially, the individual performance and achievement of each student is verified and validated by the official examinations at the critical key education transition points at the end of compulsory schooling and prior to university entrance. The national standards are evaluated against international benchmarks through participation in the international surveys of TIMSS and PISA.

Comparing the achievement of students in Lebanon against international standards.

Formally validating and certifiying student achievement.

Assessing cohort peformance data and age-related progression by sampling.

Assessing learning I checking on progress I measuring achievement.

Assessing readiness for primary education.

Student Achievement

Screening

Classroom Assessment

Curriculum Based Assessment

National examinations

International Surveys

A Conceptual Model of the Assessment Systems and their Purposes in Lebanese schools

Summary of the Assessment Systems in NSLAF

The assessment systems are listed sequentially in the tables below.

Initial screening (grade 1)

Purpose and application

Conducted each school year near the beginning of grade 1,the purpose is to assess each child for readiness to engage immediately with primary curriculum, as indicated by a minimum criterion-referenced level of development in literacy and numeracy skills. Students showing inadequate literacy and numeracy development are identified for additional individualized teaching support until their developmental levels reach those required.

Implementation infrastructure, systems and organisations

Class teachers or itinerant trained assessors should carry out the assessments for all students on a one-to-one basis, using appropriate assessment instruments (EGRA and EGMA, for example, or equivalent alternatives).

Assessment results are recorded electronically by the assessors and incorporated into the data management system for later access by authorized personnel within and outside the school.

Technical quality of the information produced

Provided the assessors conduct the assessments competently and diligently, the assessment instruments produce dependable outcomes, the technical quality of the attainment data is good – but should also be reconfirmed periodically.

Impact and interaction with the wider education system

Identifying students who have developmental needs affecting their ability to effectively access the curriculum will enable more effective interventions to be put in place.

Classroom assessment (grades 1 to 8)

Purpose and application

- Diagnostic assessment: Teachers, as needed, will administer tests to evaluate specific learning needs of individuals.
- Formative assessment: Teachers support learning throughout the school year through constructive feedback on student performance
- Summative assessment: Teachers assess subject achievement at the end of semester and school year, for the benefit of students, and for reporting to parents, other teachers and school principals.
- The results from summative assessment identify students with weak achievement who are not ready to be promoted to the next school grade.

Implementation infrastructure, systems and organisations

Class teachers currently devise their own tests for the purpose of summative assessment.

Technical quality of the information produced

It is difficult to generalize about the technical quality of the summative assessment information produced by class teachers. Moderation practices are not widely in operation, so there is likely to some variance which detracts from the validity or reliability achieved.

Impact and interaction with the wider education system

Where the grade retention rate is high it has a negative impact on the cost-efficiency of pre-university education and is likely also to have a negative impact on the learning motivation of those students who are held back.

Curriculum Based Assessment - (CBA) at grades 3 and 6

Purpose and application

- This sample-based assessment provides data on student attainment in literacy, numeracy and other subjects.
- It identifies gaps in the attainment of selected student sub-groups in each subject at each grade, and change over time. Groups to be defined by gender, type of school, region, SEND, etc.
- It records age-related attainment progression in each subject studied, i.e. change in subject attainment levels from grade 3 to grade 6 and change in attainment gaps by subgroup.
- It addresses the data needs of SDG 4.1.1.
- It identifies policy intervention needs, and to evaluate intervention effects over time.
- It indicates priorities for research on subject attainment.

Implementation infrastructure, systems and organisations

CRDP is experienced in in implementing and administering Curriculum Based Assessment; an appropriate infrastructure for survey implementation is in place. Technical/statistical support is available to assure the quality and programme design, data validation, analysis and summarization of these surveys.

Technical quality of the information produced

The technical quality of the information produced is supported by the quality of the underlying programme design and the degree of professionalism in programme implementation (assessment instrument developers, teachers, administrators, markers, data analysts, statisticians, and others).

The assessment is conducted on a defined stratified sample of the student population. It is not a high-stakes assessment for the individual student or school. The assessment is precisely calibrated psychometrically using pre-testing or the reuse of 'anchor' items.

Impact and interaction with the wider education system

CBA allows the monitoring of age-related progression over time, this is a powerful tool that measures the impact and effectiveness of the curriculum and its delivery. This enables policymakers to evaluate the effectiveness of education spending and make decisions about the most cost-effective use of financial and other resources to support learning and system outcomes

Brevet examinations (grade 9)

Purpose and application

- The examination certifies students' academic achievement at the end of compulsory education.
- Provides access to academic secondary education leading towards the Baccalaureate.

Implementation infrastructure, systems and organisations

MEHE-GDE is responsible for the construction and administration of the Brevet examination system. CRDP has responsibility for the specifications, the test blueprints and for ensuring curriculum alignment. Examinations are held annually and are currently in the form of external written tests; important subject-specific competence elements of the curriculum are not assessed, e.g. oral language skills and laboratory skills in science.

Question papers are constructed directly before administration, with assessment materials for some subjects produced in up to three languages. Student responses are marked by teachers arranged in regional teams. Scores in electronic form are recorded centrally and analysed within MEHE-GDE to produce students' final average marks and grades (mentions). Student results are issued some weeks after administration. A second session is available for students who fail.

Technical quality of the information produced

The technical quality of the information produced is dependent on the quality of the assessment materials and processes, and on the effectiveness of marker familiarization and harmonization. Marking quality is checked through double marking, when markers' independent mark allocations are electronically compared with reference to a pre-identified 'tolerance margin'; mark differences beyond tolerance are addressed by averaging.

There is no post-assessment manipulation of mark distributions. This practice raises issues to do with the comparability of achievement standards across component subjects in any year, and over time.

Impact and interaction with the wider education system

Pass rates for the Brevet are relatively high at over 80% of students taking the examinations; these are around 60% of a year group (around 20% of students leave school before this point and do not, therefore, take Brevet examinations).
The fact that subject-specific performance skills are not assessed in the Brevet has consequences for curriculum coverage in secondary education, and Baccalaureate assessment.

Baccalaureate examinations (Third Secondary)

Purpose and application

- To certify students' academic achievement at the end of secondary education (the school leaving certificate enabling entry into higher education).

Implementation infrastructure, systems and organisations

MEHE-GDE is responsible for the construction and administration of the Baccalaureate examination system. CRDP has responsibility for the specification the test blueprints and for ensuring curriculum alignment.

Examinations are held annually and are currently in the form of external written tests; important subject-specific competence elements of the curriculum are not assessed, e.g. oral language skills, laboratory skills in science, research skills in history, field work skills in geography, and so on.

Question papers are constructed directly before administration, with assessment materials for some subjects produced in up to three languages. Student responses are marked by teachers arranged in regional teams. Scores in electronic form are recorded centrally and analysed within MEHE-GDE to produce students' final average marks and grades (mentions). Student results are issued some weeks after administration. A second session is available for students who fail.

Technical quality of the information produced

The technical quality of the information produced is dependent on the quality of the assessment materials and processes, and on the effectiveness of marker familiarisation and harmonisation. Marking quality is checked through double marking, when markers' independent mark allocations are electronically compared with reference to a pre- identified 'tolerance margin'; mark differences beyond tolerance are addressed by averaging, with different weighted average calculations for the four academic streams.

Impact and interaction with the wider education system

The Baccalaureate is the school-leaving diploma that potentially opens access to higher education in Lebanon and internationally
Curriculum coverage in the Baccalaureate must be sufficiently broad for entry into universities and include opportunities for students' success in higher education through facilitating their' higher order thinking skills.

International surveys: TIMSS (grades (4)*, 8 and 12)

Purpose and application

- To monitor population levels of attainment in mathematics and science in each grade, at four-year intervals, to confirm stability or detect change.
- To compare Lebanese national student performance with that of other countries worldwide at the same grades.
- To identify gaps in the attainment of selected student subgroups in each subject at each grade, and gap change over time. Groups to be defined by gender, type of school, region, SEND, etc.
- To record age-related attainment progression in each subject studied, i.e. change in subject attainment levels from grade 4 to grade 8, and potentially to Third Secondary for students in scientific Baccalaureate streams.
- To gain questionnaire-based information about schools, teachers and students on a range of issues, including resources, and teacher and student attitudes.
- To identify policy intervention needs, and to evaluate intervention effects over time.
- To identify research potential in the field of comparative subject attainment.

Implementation infrastructure, systems and organisations

TIMSS surveys at grade 8 and TIMSS Advanced surveys at Third Secondary have regularly been carried out in Lebanon in recent years.

CRDP has acted as Lebanon's national centre, reviewing test items for curriculum appropriateness within the country, organising test distribution to the sample schools, along with test and questionnaire administration and test session management within the schools, arranging test marking and questionnaire response recording, forwarding electronic response data to TIMSS for analysis, and, most recently, producing national survey reports (for TIMSS 2015 grade 8 and TIMSS Advanced 2015). As a result, Lebanon can clearly claim to have an appropriate implementation structure in place, with systems and organisations ready and identified.

Technical quality of the information produced

The technical quality of the information produced can be assumed to be high, given the scale, long-established status and professionalism of this international enterprise.

Impact and interaction with the wider education system

Politicians and policy makers will already have benefited from this powerful assessment and evaluation tool, which provides a wealth of information about curriculum, student attainment, and learning backgrounds in Lebanon and in every other participating country.

---

*The TIMSS at G4 are not currently taking place, but the NSLAF recommends that it should be introduced.

International surveys: PISA (15-year-olds: grades 9 and 10)

Purpose and application

- To monitor population levels of achievement in reading, mathematical and scientific literacy at age 15 (i.e. at the end of compulsory schooling in many countries)
- To compare Lebanese national student performance with that of other countries worldwide at the same age.
- To identify differences in the attainment of selected student subgroups in each literacy domain at age 15, and change over time. Groups to be defined by gender, type of school, region, SEND, etc.
- To gain questionnaire-based information about schools, teachers and students in Lebanon and other countries worldwide, on a range of issues, including resources, and teacher and student attitudes.

Implementation infrastructure, systems and organisations

Lebanon participated in PISA in 2015 and 2018 with CRDP acting as Lebanon's national centre. This has involved, among other tasks, reviewing test items for curriculum appropriateness within the country, organising test distribution to the sample schools, along with test and questionnaire administration and test session management within the schools, arranging test marking and questionnaire response recording, forwarding electronic response data to the OECD for analysis, and producing a PISA national report (2015 survey). As a result, Lebanon can clearly claim to have an appropriate implementation structure in place, with systems and organisations identified and ready.

Technical quality of the information produced

The technical quality of the information produced for PISA is high. However, the validity of population performance estimates for Lebanon's 15-year-olds could be in doubt, given the fact that they are constrained to attempt all test questions, including reading literacy tasks, in French or English (as chosen by their schools). This is because reading literacy tasks are mixed in test booklets with mathematics and/or science items. Lebanon's PISA performance has been recognized as being particularly low in 2018, when reading literacy was the major domain.

Impact and interaction with the wider education system

Politicians and policy makers will already have benefited from this powerful assessment and evaluation tool, which provides a wealth of information about curriculum, student attainment, and learning backgrounds in Lebanon and in every other participating country at the end of compulsory education. However, the language issue noted above should be considered when evaluating Lebanon's education system on the basis of PISA survey findings.

4.1 Initial Screening

4.1.1 Purpose

In many countries young children are frequently and unobtrusively assessed by their teachers and others during their early years of education. Internationally, there is a growing consensus that a key point at which to carry out checks, such as on the development of basic skills in literacy and numeracy, is at beginning of compulsory schooling. These checks can be formal or informal and are used, additionally, to monitor physical, behavioral and cognitive development to confirm that the early years student is meeting normal expectations, and, if not, to identify intervention needs. England[2] and France[3] are two examples of countries that have primary-ready survey programmes in place. The purpose of this screening process is to assess each child's readiness to participate in primary education against minimum criteria levels. This allows teachers to identify children with developmental needs in particular areas and put in place the interventions needed for them to reach the required level.

Currently in Lebanon not all children of a young age attend preschool. Those who do are monitored by preschool staff for learning development in a range of cognitive, behavioral and attitudinal domains. Intervention in the form of appropriate support is then given where necessary. However, this does not necessarily provide information for primary schools as to whether the children are ready to progress to formal schooling in grade 1.

The developmental assessments should be carried out by the students' class teacher early in grade 1, though not immediately on entry. This will give the students a chance to settle into the school routine and the teachers an opportunity to form first impressions. It is important to make sure that teachers have a clear understanding of the purposes of the assessment results and understand that they will not be used for accountability processes. If this were the case it could lead to 'teaching to the test' which would inflate and therefore invalidate the results, potentially denying some students the additional support they need.

While it would be possible to commission a bespoke assessment tool, it would be much more cost effective to select from existing tools based on relevance for purpose, manageability and technical quality.

Early Grade Reading Assessment (EGRA)[4], is a well-developed early literacy assessment tool, which has been used in sample-based pre-test/post-test research to evaluate progression in early years students' reading skills after teaching input.

---

[2] STA (2018). 2019 national curriculum assessments. Key Stage 1. London: Standards and Testing Agency. https://assets.publishing. service.gov.uk/government/uploads/system/uploads/attachment_data/file/754076/Key_st   age_1_assessment_and_reporting_ arrangements.pdf

[3] DEPP (2019). Evaluations repères 2018 de début de CP: premiers résultats. Note d'E d@information 19.13. https://www.education. gouv.fr/cid139130/evaluations-reperes-2018-de-debut-de-cp-premiers-resultats.html

[4] RTI (2016). Early Grade Reading Assessment (EGRA). TOOLKIT, Second Edition. https://www.globalreadingnetwork.net/sites/default/files/resource_files/EGRA%20Toolkit%20Second%20Edition.pdf See also ACER (2014). The Early Grade Reading Assessment: Assessing children's acquisition of basic literacy skills in

In preparation for the trialling and intervention research, teachers in the participating schools were trained to use this tool. EGRA therefore seems a logical choice for the grade 1 literacy screening, though further exploration may be needed to confirm that it is also suitable for use in grade 1 in Lebanon. Lebanon's EGRA research teacher participants would be a good source of opinion, so should be consulted.

Early Grade Mathematics Assessment (EGMA) [5] EGRA's partner tool, could also meet the requirements for early numeracy assessment in grade 1. For this early stage, EGMA covers simple addition and subtraction, number comparison, number patterns (missing numbers), and simple word problems.

The assessment tools for both reading and numeracy should be administered to early years students individually. To ensure manageability, the time required per student should be relatively short (10– 15 minutes). The assessor would offer a series of very short test items to the student, orally, on paper or on-screen. These different options would allow schools flexibility depending on their ICT capacity. The student's responses would then be recorded by the assessor either on a specially designed paper-based recording sheet, or directly onto an electronic form. The advantage of the latter is that there could be inbuilt instant validation checks that would serve to avoid any clearly detectable errors in recording (in particular, illegitimate response codes).

4.1.3 Role and responsibilities

The professional development for teachers of grade 1 students will be provided by The Professional and Inservice Training Bureau (PITB) at CRDP and will build upon the expertise developed in recent years by several teachers as part of the EGRA research. Ongoing support and mentoring of the teachers will be provided by DOPS to ensure continuity, in the event of staff changes, and consistency between different teachers in different schools.

The data from the screening will be gathered by the school and maintained at a local level; it will, additionally, be collated by CRDP in order to assess the needs, at a national level, for additional learning support interventions for those children who do not reach the expected levels of achievement.

---

developing countries. Assessment Gems Series no.2. https://www.acer.org/files/AssessGEMs_EGRA.pdf

[5] RTI (2014). Early Grade Mathematics Assessment (EGMA) Toolkit.
https://iercpublicfiles.s3.amazonaws.com/public/resources/EGMA%20Toolkit_March2014.pdf

4.2 Classroom assessment

4.2.1 Purpose

Formative Assessment:
Is part of the instructional process.

. Quizzes
. Observations
. Creating T-Charts, Venn diagrams & other
  student learning evidence
. Classwork/Homework
. Writings & exit tickets
. Helps teacher modify future lesson
   planning based on learner needs

. Both are ways to assess
. Questioning strategies need
   to be addressed
. Both need to be used to
  evaluate a student effectively
. Both can be used for student
  feedback
. Assist in future lesson
  planning

Summative Assessment:
Used to determine at a particular point
in time what students know and do not know

. State assessments
. District benchmark or interim assessments
. End-of-unit or chapter tests
. End-of-term or semester exams
. Scores that are used for accountability
. SAT or ACT-type tests

Formative and Summative Assessment [6]

Throughout the school year class teachers in elementary schools and subject teachers in both intermediate and secondary schools are continuously engaged in student assessment. This includes both ongoing assessment for learning purposes and more formal assessment reviewing learning at the end of the semester. In each year summative assessment consists of a combination of ongoing classroom assessment and teacher-written tests taken at the end of the first semester and the end of the school year. These assessments are combined at the end of the year to produce an average achievement mark for each student.

Importantly, from grade 4 onwards students achieving an average mark from the combined teacher assessments that is lower than 10 out of 20 are generally held back to repeat the grade. Levels of grade retention are relatively high – in 2018–19, for example, the retention rates for grades 4 and 7 were just under 20%.[7] This has a negative impact on the cost efficiency of pre-university education and also adversely affects the motivation and confidence of those students who are held back.

---

[6] © Dr Justin Tarte http://www.justintarte.com/2014/09/have-summative-assessments-become.html

[7] https://www.crdp.org/pdf/uploads/230082019_Passing_rates.pdf

As these combined assessments have high stakes, it is essential that teachers are able to produce good quality ongoing classroom assessments and end of semester tests and mark them accurately. However, there are significant concerns as to how valid and reliable the judgments from the combined assessments are. This is a result of the variable assessment literacy of teachers and a lack of guidance on how to construct the end of semester tests.

Much of the variability can be seen to be due to inconsistencies in the role description and activities of the assessment coordinator in the school. The internal system of public schools requires the presence of a coordinator for each subject with the number of hours required for each subject in the school, but at least two issues compound the problem of consistency, namely – there is a lack of specific criteria to choose a coordinator; the job role description and required duties are not clearly delineated.

Even with the variable quality of assessment coordinators, there are, nonetheless, ,some systems in place to support effective teacher assessment. Teachers receive guidance on assessment as part of their initial teacher training and through continuous professional development by CRDP. This training though, is not mandatory and quite a number of teachers choose not to benefit from these professional development opportunities. Furthermore, while the end of the first semester tests for grade 9 and Third Secondary, are submitted by teachers to coaches at the Department of Guidance and Counselling within the General Directorate of Education (MEHE-GDE / DOPS) for feedback, the consistency of this process varies significantly and there is a lack of coordination between CRDP and DOPS. The absence of a common set of documents setting out clear assessment requirements together with the roles and responsibilities of all the key players raises more challenges for consistency.

Taken together, these factors represent a particular large risk that the level of demand of these assessments could be somewhat different between schools. This will in turn have an effect on grade retention and repetition decisions. For example, a student in one school may take challenging tests set by their teacher, not meet the required average mark and be held back a grade; a student of similar ability in another school where the teacher's tests are less demanding would pass and progress to the next grade.

4.2.2 Tools and procedures

As indicated, the tools to ensure consistency, in principle, exist but they need to be applied in a more regular and systematic way in order to deliver dependable assessment outcomes. Not only should all teachers receive extended and improved training based on consistent guidance, but at a system level, there should be more opportunities for better liaison, sharing and moderation of teacher assessment practice. This will help ensure fairness for all students and develop teachers' understanding of what makes a valid and reliable test and how the outcomes from assessment can be used to help students progress.

Using assessment outcomes to inform learning is in its infancy and is not always well understood or applied in practice. Consequently, developing teachers' understanding of the uses of assessment would have a significant impact on improving students' progress, as well as the wider assessment system's ability to function effectively. The effect of improving the assessment literacy of teachers should be carefully monitored through targeted impact assessment studies.

In addition, to increase the level of fairness in the end of semester summative assessments, a greater level of control is necessary to ensure the comparability of teachers' marking standards within and across schools. Currently the lack of moderation and harmonisation practices means there is little data to establish the level of assessment validity or reliability achieved both within and across schools. Consequently, a system of moderation and harmonisation between teachers both within their own school and across similar schools is greatly needed.

Finally, rather than being aggregated into a single score, the Cambridge International experts recommended that the results of the end of first and final semester summative tests and teachers' ongoing classroom assessments should be reported separately. In the longer term, following planned curricula reform, this would allow the ongoing teacher assessment to be used to assess practical skills through methods such as projects and portfolios and more traditional knowledge-based learning to be assessed by summative tests. Separating the reporting in this way would produce a more detailed and accurate picture of students' strengths and areas for improvement and avoids combining two very different types of assessment into a single score.

The focus on improving assessment literacy recommended so far will support the development of dependable and consistent summative assessment practices in Lebanon. However, it also provides an opportunity to extend teachers' understanding of assessment for learning and the ways they could make use of it in their teaching practices.

As teachers' assessment literacy grows, assessment practice will improve and confidence and expertise will develop. This ongoing process for developing assessment literacy creates the opportunity for teachers to move beyond using assessment mainly to rank and measure student performance and start to make greater use of day-to-day, often informal, assessments to decide how best to help students develop their understanding and improve.

4.2.3 Roles and responsibilities

To be fully effective, there must open and transparent cooperation between the different government entities with their roles and responsibilities clearly delineated. For example, the establishment of what age-related achievement should be in each subject is a task that MEHE, CRDP and DOPS should agree about collectively. The responsibility for the maintenance of the standard should, rightfully, rest with MEHE ; for the professional development of teachers needed to apply the measures, liaise and moderate their assessment practices rests with CRDP; the day to day mentoring of teachers, the support and the local / district / regional organisation should be operated by DOPS.

4.3 The National Examinations – setting, marking and maintaining the standard

4.3.1 Purpose

Students take externally [8] set and marked national examinations in the intermediate and secondary phases (Cycles 3 and Secondary Cycle) at the end of grades 9 (Brevet) and Third Secondary (Baccalaureate). These qualifications certify student achievement at the end of basic education and secondary education respectively. The purpose of these examinations is to certify students' academic achievement at significant end points in their education.

Examinations are held annually in the form of formally administered written assessments; there is currently no assessment of practical or performance skills. Question papers, each in up to three languages depending on the subject, are constructed by subject committees which draw on questions from an item bank of moderated and standardised questions that has been drawn up over time. In practice, at the time of examination paper construction, when the representatives from particular subjects draw the questions from the bank in the presence of Director General Education at MEHE, they, frequently, are not satisfied with the quality of the question and many of the questions, up to 80 per cent, are modified by the assessment committee.

Marking is completed regionally in a series of residential marking centers. This system should be effective in identifying outliers in the marking process for those subjects where there are consistent or widespread discrepancies beyond tolerances. However, it could be made more effective by improving the consistency of the original markers' judgments.

Currently, responsibility for the specification, creation and administration of the Brevet and Baccalaureate is dispersed between a number of different departments in CRDP and MEHE-GDE. The quality of national examinations and the capacity to deliver NSLAF fairly, effectively and consistently would benefit from greater clarity, consistency and coherence in the system of administration, reporting and monitoring performance.

There is a potential tension arising between the current dominance of formally administered written assessments, which support public confidence in the security of national examinations, and more authentic performance and skill assessments. Future curriculum reforms which will, most likely, require the evaluation of skills and competences that are best assessed by other potentially less secure forms of performance evaluation such as practical activity assessment and portfolio development will require a form of moderation to ensure equity and fairness. Moderation of student performance using this wider and more authentic, range of performance evidence, is possible, but it does require a great amount of open and frank liaison between the key players coupled with a public information campaign to build public confidence in a wider more current and relevant range of assessment measures.

---

8 Formally set and marked, at a national level, by subject experts, independent of particular schools

4.3.2 i) Tools and procedures - setting the examinations

National qualifications, such as the Brevet and Baccalaureate, require a high level of reliability and security to ensure fairness for students and to provide confidence in the outcomes. During their consultations the experts from Cambridge International identified several issues in the current system related to question paper authoring, marking and maintenance of standards. These issues are likely to lower the confidence of parents, and stakeholders more generally, about the value, the credibility and the comparability of these qualifications, particularly in the international context.

To improve the security of the assessment and reduce the possibility of leaks, question paper construction and printing of papers take place the night before the examination. However, the complex logistics and extremely tight schedule make it challenging to conduct sufficient quality assurance checks on the question papers within this highly-compressed time frame. This leads to increased potential for errors or inconsistencies to occur, particularly when the same papers may need to be translated into several different languages.

The advisory team from Cambridge International gathered a panel of international assessment experts to scrutinize the quality of the Brevet and Baccalaureate assessment materials including question papers, mark schemes (barème) and the question paper specifications. The panel reviewed a sample of 10 Brevet and 17 Baccalaureate 2019 question papers and mark schemes, as well as the question item specifications on which they are based (the 'tawseef' documents). The panel used the review to identify issues that could have a negative impact on the validity of individual questions and whole question papers and undermine the reliability of the marking process.

i) The National Examination specification

The Brevet and Baccalaureate examinations are produced according to specifications provided by CRDP. The main purpose of a question paper specification is to communicate the scheme of assessment clearly and coherently to the subject committee and other relevant stakeholders. CRDP provides subject specifications across both the Brevet and Baccalaureate on key aspects such as paper structure, duration, the number of questions required, question types, and marks per question, as well as the expected balance of marks across assessment objectives or skill domains.

Owing to the difference between the specific characteristics of each subject, there is some inconsistency in the format and level of detail contained in the specifications across subjects.

A subject examination, such as the mathematics Baccalaureate, for example, lists the general principles related to content and skills : "the three main levels of knowledge should be well-balanced: acquisition, application and analysis". Whilst some other subjects contain full points-based mark schemes for essay questions, such as, the philosophy and civilizations Baccalaureate; or detailed information on layout, numbering, fonts (science Baccalaureate).

Furthermore, the specifications for some subjects are supplemented by 'explanatory guides', and these vary in the type and amount of information they include. For example, the English language Baccalaureate guide contains relatively detailed guidelines for question writers, such as on how to select a reading passage; the guide for biology includes lists of the action verbs to be used in order to check and assess the students' higher order thinking skills.

The specifications are design so that the correct skill domains will be addressed with the right balance across all subjects, potentially making the difficulty of the papers vary from year to year.

ii) Validity at question level construct-irrelevant variance

Valid questions provide evidence of students' relevant knowledge, understanding and skills. One of the main threats to validity is 'construct-irrelevant variance' or the introduction of an element that is unrelated to the construct being introduced. It is, therefore, important that the questions in the 'assessment bank' are of high quality, and they are valid and varied. This will reduce the need for the examination and the committee to change, at the last minute, questions selected from the bank.

Command words, in particular are designed to be clear, precise and consistent in order to reduce construct irrelevant variance. The aim is to have a clear and shared understanding of what is expected in terms of the scope, nature or depth of a student's answer in response to a command word (action verb); and this should be reflected mark scheme. To produce a valid question, command words must relate to the skill assessed and be used consistently. Command words are used to give students an indication of the depth or extent of response expected.

iii) Allocation of marks

Marks allocated to questions are designed to be commensurate with the cognitive demands of the task and the skill being assessed; the number of marks allocated to a question reflecting the complexity. Many inconsistencies in the allocation and distribution of the marks arise due to rushed practice of final exam setting, where the papers are hurriedly compiled in the 12 hours before implementation – this does results in questions being selected and then edited from the question bank that have low and variable quality and validity.

iv) Construct under-representation in question papers

Having designed valid questions, it is equally important to ensure validity at question paper level, and so an additional level of scrutiny is required during the construction of a paper from individual questions. Inadequate paper construction systems can lead to construct under-representation – in which areas of the curriculum being assessed are not sufficiently covered or are even absent in a question paper and so examination results fail to reveal a student's abilities in that respect. If questions in the item bank are linked to the specific skills, knowledge and understanding they address, they can be selected for inclusion in the question paper during construction according to a balance or 'weighting' stated in the question paper specifications.

The most significant example of construct under-representation is that of non-written skills, such as speaking and listening skills in languages, which are not assessed despite being included in the curricula objectives. While there are legitimate reasons why these skills are excluded from national examinations (such as cost, logistics, manageability and reliability) their absence points to a lack of coherence between curriculum and assessments and may lead to some curricula objectives becoming marginalised or even not being covered at all.

The experts from Cambridge International recommended that improvements in the construction of papers could be made by following these standardised steps:

- review the quality assurance process before questions are approved for selection in question paper construction
- allocate more time for the question paper construction process to allow for more rigorous checking of papers before going to print
- make improvements to templates for the layout of questions and question papers to increase the accessibility of question papers
- provide additional training for question writers and professionals responsible for constructing question papers and creating question paper specifications. This should include helping writers to judge the difficulty of questions and address higher order skills such as reasoning and evaluation where appropriate.

The Cambridge International experts' review of question papers, mark schemes, answer keys and official examination specifications underlined the need to improve the quality assurance process in both the production of questions and the construction of question papers, as well as to reconsider the current timeframe in place for this process. Furthermore, the review indicated a need for the assessment literacy of question writers and the professionals involved in question official exam paper construction should be developed further. This will assist in the development of the robust process that is necessary for producing and quality assuring the questions to be included in the item bank.

Quality improvements in the setting of the examinations should start by revising the methods for formulating test specification and follow the processes and procedures taking account of all the steps including test production and feedback. Checks to produce valid questions for the item bank are necessary to ensure that the process of constructing the question papers is as smooth as possible. It is of critical importance to ensure the validity examinations that enough time is allocated for the question paper construction process to allow for more rigorous checking of the papers before they are printed.

4.3.2ii) Tools and procedures - marking of the examinations

Student answer booklets are marked by teams of teachers and there is a well-delineated system of double marking to check the comparability of marker judgments. Scores are recorded centrally, and results are issued several weeks after administration. A second session is available for students who fail. An historic mark-to-grade (mention) mapping is applied for each subject, to maintain subject standards year on year, but the yearly mark distributions are not modified, as they are in some countries. The practice of adjusting mark distributions each year is used, in these countries, to avoid unusual and extraneous year on year variance and to ensure that the proportion of marks in each achievement category (eg. High, median, low) remains broadly constant.

The current double marking algorithm

The resource-heavy system of double marking is designed to improve the reliability of markers' judgments in both the Brevet and Baccalaureate. This is illustrated in the figure below:
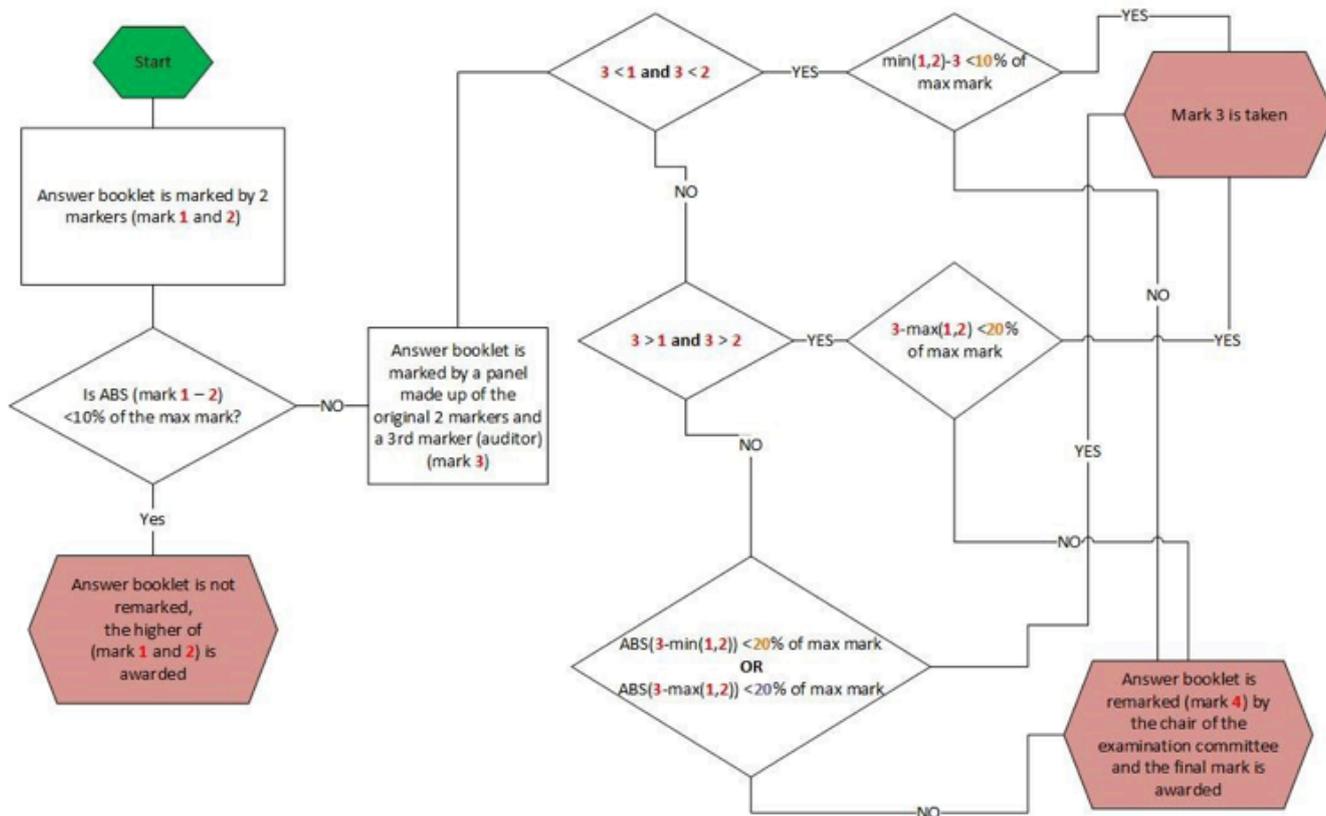
All answer booklets are marked by two markers. If there is a difference in the marks awarded, a system of tolerances is used to determine whether an auditor needs to carry out a remark. A different tolerance is then used to decide whether the Chair of the Examinations Committee needs to carry out a further remark if agreement has still not been reached.

This process is designed to identify 'outliers' – i.e. marks that either contain significant errors or where the marker's judgment is substantially different from the student's true mark.

While the double marking system represents best practice in terms of identification of marking outliers, it can be made less effective by poor inter-marker reliability:

It is necessary to use very high tolerances between markers to avoid overloading the system. For example, the answer paper is only passed to the auditor if the two initial marks are not within 10per cent of the maximum mark of each other.

Making this tolerance lower would lead to improved reliability but it would only be practical if marking quality was also improved to reduce the number of answer papers that were outside tolerance. Reducing the tolerance without improving marking quality would lead to more answer papers being remarked than the system could cope with.

4.3.2 iii) Tools and procedures - maintaining the national standard in examinations

The systems currently in place for maintaining standards rely heavily on the question paper construction process and subjective judgments of difficulty. There is no system for pre-testing questions to provide more objective data on how they will perform nor an effective mechanism of adjusting examination outcomes from year to year based on differences in difficulty. This will inevitably lead to variations over time in the standard of performance required to achieve a particular mark. This is a significant threat to the fairness and reliability of the examinations.

In line with international best practice in relation to maintaining the national standard the following procedures should be considered by MEHE-GDE.

i) Establish a specification for each subject

A specification for each subject to be monitored in the programme should be developed by CRDP. Each specification will identify the elements of the particular subject curriculum at the relevant grade that should be included in each survey test. Ideally, these 'elements' will be defined in the school curriculum in the form of learning objectives or learning outcomes. In any curriculum there will be certain competences and skills that it will not be manageable to assess on a large scale. Reasons will vary depending on the competence or skill concerned, but could include:

- unacceptably high financial costs
- challenging logistics
- ethical considerations
- an absence of essential physical resources in schools
- a very limited testing time per student
- general issues concerned with assessment quality generally.

Some examples might be the assessment of group collaboration skills (hard to assess reliably), tasks involving the monitoring of plant growth (given the time involved), computer-based tasks (where computer use is unfamiliar or there are insufficient computers in classrooms), and so on.

For these and other reasons, the curriculum assessed in any large-scale survey test will not fully represent the curriculum as intended to be delivered in schools.

ii) Build up a bank of assessment materials using question descriptors and taxonomies

Once the specifications have been agreed by all key stakeholders, a bank of assessment items and tasks addressing the assessable domain in each subject and target grade should be built up; at present the test bank from which the examination questions are drawn is too small and the quality is too variable. Improvements could be achieved by identifying relevant existing questions and tasks that have been used within the country and then supplementing these to fill any gaps in coverage by developing new ones.

iii) Identify a set of question descriptors and taxonomies to guide item build-up

To ensure the questions and tasks, whether existing or newly written, are appropriate, a set of question descriptors should be developed to define question style such as providing standard formats for questions and common wordings for question stems. Appropriate taxonomies, identifying features such as level of cognitive demand and what the question or task is assessing, should also be adopted for use in tagging items with essential metadata. These will support question development, by providing clear guidelines to ensure consistency and paper construction by making it easier to identify appropriate questions for selection. The question descriptors and taxonomies will also provide a structure for the valid interpretation of test survey results. In addition, these materials could form the basis for materials to train teachers as question writers.

a) Quality check items and associated resources

All potentially suitable questions, whether pre-existing or new, should be reviewed for quality against the question descriptors and tagged with relevant metadata based on the taxonomies. It is, however, important to recognize that this process should not be expected to result in only a set of 20–30 items for use in the first test survey.

b) Establish an appropriate item banking system to hold questions and their characteristics, including histories of survey use

All assessment materials should be entered into electronic item banks, along with resource descriptions, marking schemes, and usage statistics. Item records should be created within the item bank before any empirical trialling. A tag can be added after trialling to Indicate whether it is appropriate to be used in a test survey.

iv) Identify a measurement model to guide item trialling and survey analysis

Cambridge International recommends that multiple versions are created for each administration of the survey test to avoid students copying each other's answers.

There are two possible measurement models: item response theory on the one hand, and domain sampling on the other. Item response theory is sophisticated mathematically and offers some powerful benefits if it meets certain criteria relating to the given attainment data. Domain sampling is simpler, has less stringent trialling requirements and produces more user-friendly attainment results. Further expert technical advice will be needed to decide which is the most appropriate strategy to use in Lebanon.

All test questions and assessment tasks should be empirically trialled with samples of students from the target grades before first survey use to confirm quality and to establish empirical

properties, such as facility and discrimination. Whichever measurement model is adopted for the survey programme, the student samples used for trialling should:

- be relatively large (200+ students)
- be representative of the target population
- be unlikely to have seen the same items beforehand
- attempt the items in similar circumstances to those they would meet in an actual survey.
-

Once a survey programme is underway, one way to meet these requirements is to include trial questions within the set of 'genuine' items during an actual test survey, while excluding their results from the final survey analysis.

Pre-survey preparation should include empirical studies of potential test reliability in the different subjects, assessment domains and target grades. It is vital that the reliability of test marking is investigated before the event and assured for surveys. This demands the organisation of multiple marker studies, as appropriate.

v) Use learning environment questionnaires in surveys to contextualize the results

Information from school principals and teachers about learning environments and from students about their attitudes to school, subjects and learning, is invaluable for contextualizing survey attainment results. This contextualization in turn enriches system performance interpretation and so supports effective policy development and evaluation.

To gain maximum benefit from the questionnaires, it is important that they are carefully constructed to yield the right information and that sufficient time and resources are given to develop them. Small scale piloting should follow with the questionnaire being administered to groups of target individuals to ensure that the questions are clear and interpreted as intended. Student questionnaires should use language that is at an appropriate level for the grade of student. The questionnaire formats should be available in adapted versions to allow students with disabilities to access them (for example large print versions for students with visual impairments). Where appropriate, common questions should be included in the questionnaires for different subjects across the target grades to enable research into subject-related differences and age-related evolution. Student attitudes to school subjects – their value to them and to society – and their assessment of their own subject abilities are but two examples of issues relevant to learning motivation.

vi) Validate data and automate analyses and reports

Student response sheets and mark sheets for all subjects across the target grades should be well designed to help ensure responses are recorded accurately. Where information is taken from paper-based response sheets and mark sheets, it should be double keyed (that is independently entered by two different individuals) to help ensure validity by avoiding human error in data transfer. Students' direct electronic input of responses, as in computer-delivered assessment, should be instantly validated.

The survey response data should be analysed in a number of standard ways, with data weighting where needed, to produce population and student subgroup outcomes (by gender, school, class, region, type of school or refugee status for example), with estimation error indications attached. Where possible reporting formats for disseminating the results should be predesigned and analyses and reports automated for regular application and fast delivery.

Reports for teachers for general use and professional development; short reports for parents; and press releases may use non-standard formats and cannot be automated.

vii) Prepare for programme launch and ongoing training needs

It will be essential to keep school principals, teachers and parents fully informed about the implementation of the national monitoring programme well before it is launched. Depending on their role, things they will need to know and understand include:

- why the programme is being introduced
- when student assessment will occur
- who will carry out the assessment
- what resources will be needed to deliver the assessment
- what their particular responsibilities within the programme will be
- what the potential benefits for the education system as a whole could be.

Clear guidance documents should be produced for schools, to ensure that all staff are fully aware of the upcoming survey each time and that they understand the importance of carrying out the exercise within their school. The guidance should be updated, as appropriate, before each survey.

Before the programme is launched, and on a continuous basis thereafter, a number of training requirements will need to be addressed and resourced. These include training for personnel involved in:

- assessment material development (all subjects)
- survey design, sampling and data analysis
- questionnaire development
- marking
- data interpretation and report writing
- research (using the cumulating data for policy-relevant research analyses)
- dissemination (describing and explaining survey findings for wide audiences of teachers, parents and others).

Specific recommendations to maintain the standard for Brevet and Baccalaureate

Both the Brevet and the Baccalaureate are high-stakes qualifications. The mark awarded to a student has far-reaching consequences for the future of their education and employment. As such the national examinations should provide a reliable measure of students' knowledge, understanding and skills and a dependable means of distinguishing between students' performance both within and across examination series.

The Brevet is taken by all students at the end of grade 9. Those students who progress in high school to Third Secondary will finish their schooling with the Baccalaureate examination. The Baccalaureate is a pre-requisite for university entry, but many universities also administer their own entrance examination additionally. A process to ensure that the Baccalaureate provides a

dependable a measure of student achievement and differentiates effectively between them would make it more useful to universities who may, then, have the confidence to abandon their own additional testing regimes.

Currently there is some inconsistency of the awarding standard of the Baccalaureate over examination series or between subjects, across language versions of the same question paper, or across option track eg. scientific vs. humanities. Here 'awarding standard' refers to the quality of work necessary to be awarded a pass mark or a mention, **taking into account the demand of the examination** (that is how hard the candidates found the assessments). Ensuring that qualifications provide a consistent measure of the awarding standard both within and across examination series is referred to as 'maintaining the standard'. Failure to maintain the standard means that the results achieved by students across examination series do not represent a consistent standard of performance and, therefore, their mark cannot be used as a dependable selection tool. Two students who both achieve a mark of 10 on the life science track in different years may not be equally knowledgeable and skillful, but their mark suggests that they are.

Methods of maintaining the standard
Ways to maintain the standard can broadly be divided into pre-examination methods, which focus on the production of consistent question papers across series, and post-examination methods that compensate for changes in the demand of the examination across series by adjusting the marks awarded to students.

Pre-examination methods
These aim to ensure that the demand of the examination is always the same over different examination series and therefore that the standard is maintained without any adjustments to marks being necessary. This can be achieved in the following ways:

a) Examination quality
The most straightforward method of maintaining the standard is to only use question papers and mark schemes of a consistent quality, including across question papers translated into multiple language versions, and to ensure that marking processes are robust.

b) Pre-testing questions
Pre-testing involves small groups of students answering the questions under examination conditions so that subsequent statistical and qualitative analysis of their responses can provide valuable information about how demanding the students found the questions. The data from pre-testing can be used while constructing the question paper to ensure that the level of demand is consistent with questions in the previous series. Where each question paper is in multiple language versions, pretesting must be undertaken in the language of assessment to ensure that differences in demand between the versions are highlighted. Using pre-testing data to construct question papers is effective in eliminating differences in the demand of the examination over different series but is costly and complex to manage and, unless a large number of alternative questions are trialled, it risks students' question papers containing questions they saw in pre-testing.

Post-examination methods
The term 'awarding' refers to methods used to maintain the standard after the examination has taken place and marking is complete. They focus on the comparative performance of the whole group of students taking the examinations in each series – awarding is not carried out at individual student level. These methods use data to decide whether changes in the performance of students are due to changes in the demand of the examination or variations in the ability of the cohorts of students in the different series (due to changes in the education system for example). If the data suggest the demand of the examination has changed then adjustments should be made to marks to compensate.
How the methods approach the decision about what is causing the change in students' performance differ:

a) Cohort referencing
Cohort referencing assumes that the students taking a subject within a track in a particular examination series have the same level of ability as the students taking the subject in the same track in the previous series. Therefore, any change in the performance of students between series is due to a change in the demand of the examination. Using this awarding process for the Baccalaureate would allow the same cumulative percentage of students in one year to achieve the pass mark and each of the three mentions in each track as achieved them in the year before. Adjustments to the marks for each subject would be made at awarding to regulate the cumulative percentage of students achieving the marks at track level.

Technically, this method of awarding is relatively simple to administer and is not time-consuming, but it does rely on the ability of the students being unchanged from year to year. In the short-term this assumption may be true. However, if efforts to improve the quality of education in Lebanon start to have an impact or the profile of students changes, for example in terms of their fluency in the language(s) of instruction, over time the assumption will become increasingly unsustainable.

As students' ability and marks change, the awarding process adjusts the marks at subject level to ensure that the whole group of students is not credited with better or worse performance at track level. Therefore, this method can be unfair to students and unhelpful to stakeholders who want to use the Baccalaureate results to differentiate between students over a number of years, taking into account changes in their ability.

b) Statistical modelling
Statistical modelling makes linear assumptions about both the demand of the paper and the ability of the cohort, assuming that they are broadly similar each year. Modelling using Rasch or comparison of mean marks across subjects in the track, for example, can be used to adjust the weighting of the raw marks at subject level to track level, operating within these two linear parameters. Technically this method of awarding is more complex to carry out and more time-consuming than cohort referencing, but not significantly so.

Significantly, the underlying assumption of this type of statistical modelling is that the subjects within each track together define a shared construct that is closely related to the constructs being measured by individual subjects. Therefore, any deviation in the performance of students in one subject from the other subjects on the track represents a difference in the standard of the examination. Given the diverse nature of the subjects in each track the assumption of a shared construct may not be true. Rasch does, however, provide a sound theoretical basis for the analysis and, as long as the interpretation of the results is justified, it can operate as a relatively straightforward method of maintaining the standard of the Baccalaureate over time.

c) Criterion referencing

Criterion referenced awarding measures the extent to which students have met set learning outcomes and assessment objectives. In its most pure form, the student must meet all the criteria to be awarded the grade. 'Weak' criterion referencing acknowledges that good performance in some areas might offset poorer performance in other areas.

Awarding through criterion referencing uses a mixture of quantitative and qualitative evidence to judge whether changes in the cumulative percentage of students at pass and each mention are due to changes in the demand of the examination, changes in demand across multiple language versions of the same question paper or changes in the ability of the students compared to last year. In contrast to cohort referencing and statistical modelling this method makes no assumptions about the ability of the students or the demand of the question papers over time and therefore is flexible enough to deal with changes to both. It does, however, rely on the assumption that the personnel carrying out awarding are impartial. If there are either positive or negative consequences for them raising or lowering student performance, their judgments may be compromised. Additionally, this method relies on the availability of suitably skilful personnel to carry out awarding as well as adequate IT systems and sufficient time for the process between the completion of marking and the release of results to students.

Equivalency of the Baccalaureate

The successful embedding of the processes to maintain the standard will improve the confidence of all stakeholders in the credibility and equivalence of the Baccalaureate. Additionally, equivalence procedures should be regularly conducted to benchmark the Lebanese terminal school qualification with international equivalents such as the French Baccalaureate or the International Baccalaureate (IB). Comparability studies can be carried out for vocational qualifications providing a suitable comparator is identified. Quantitative and qualitative evidence to support this comparison can be generated thorough a comparability study. A comparability study compares equivalency between qualifications in three interrelated standards:

a) The content standard: refers to the subject content that needs to be learnt in order to succeed in the qualification. The depth and breadth of the subject content and the assessment objectives are compared to establish the level of equivalency.

b) The demand standard: refers to the depth of knowledge, skills and competence required by the assessment instruments and their mark schemes. Question papers, including those in multiple language versions, should be compared in terms of their complexity and level of demand. Mark schemes are compared in terms of the degree of leniency they encourage when marking. Demand standard can also be influenced by the way markers apply the mark schemes to students' answers. Marked answer papers from both qualifications are compared in terms of how strictly markers have applied the mark schemes.

c) The awarding standard: refers to the underlying skills, knowledge and understanding that it can be inferred that students possess from their observed performance in assessments. Marked answer papers from both qualifications, including those in multiple language versions, are compared at pass level and at the minimum mark required to achieve each mention to establish whether students demonstrate similar performance.

The standard of the Lebanese Baccalaureate must be steady before a comparability study is conducted, otherwise the conclusions will become meaningless in the next examination series.

These three interrelated standards form a study of the assessment-related comparability of qualifications. Further studies outside the scope of this report and not related to assessment include a comparison of:

- teaching methods
- recommended study hours
- entrance requirements, particularly with reference to eligibility to apply to take qualifications which can have social equity repercussions.

4.3.3 Roles and responsibilities

CRDP has responsibility for the specification the test blueprints and for ensuring curriculum alignment. MEHE-GDE is responsible for the construction and administration of the Baccalaureate examination system.

Examinations are held annually and are currently in the form of external written tests; important subject-specific competence elements of the curriculum are not assessed, e.g. oral language skills, laboratory skills in science, research skills in history, field work skills in geography, and so on.

Question papers are constructed directly before administration, with assessment materials for some subjects produced in up to three languages. Student responses are marked by teachers arranged in regional teams. Scores in electronic form are recorded centrally and analysed within MEHE-GDE to produce students' final average marks and grades (mentions). Student results are issued some weeks after administration. A second session is available for students who fail.

The setting of the examinations

a. Definition of roles
The roles and responsibilities of key personnel involved in producing and banking questions and constructing materials should be clearly defined. These should include personnel with the following responsibilities:

- **Test specification team:** subject specialists who are knowledgeable about international trends and standards, so that the specification aligns with international expectations for age-related attainment
- **Result analysis team:** the team should be multi-disciplinary including statisticians, subject specialist and senior teachers as well as officials for MEHE-GDE and CRDP
- **Subject specialists:** Permanent members of staff with assessment expertise and subject knowledge, who are responsible for managing the overall production of question papers for their subjects. They brief question writers on their responsibilities and recommend specific areas where questions in the item bank need to be developed or replenished. They review questions and mark schemes and make decisions on whether these should be included in the bank. They also chair editing meetings attended by the question writers and revisers, where the questions are reviewed. Finally, they construct the question papers in accordance with the specifications and check the first and final proofs of the typeset question papers to ensure coherence and quality
- **Question writers:** subject experts who produce questions and their associated mark scheme, attend editing meetings and amend their questions as required
- **Revisers:** subject experts who review the questions, attend editing meetings and suggest changes
- **Vetters:** subject experts who have not been involved in the earlier stages of the question production. They check that the content of the constructed question paper meets the specification by answering the questions in the paper from the point of view of the student and then reviewing the mark scheme
- **Proofreaders:** language and quality assurance experts who check the questions and mark schemes for clarity, typographical errors and inconsistencies
- **Typesetters and artworkers:** members of staff who typeset the question papers and mark schemes, and who source or produce associated artwork where needed
- **Translators:** subject experts who translate the question papers and mark schemes and who edit and check translations done by others
- **Modifiers:** experts in SEND (Special Educational Needs and Disabilities) who modify question papers to accommodate students with disabilities, such as producing braille versions.

b. Quality assurance of the process
To improve the process, the Cambridge International experts recommended that the following steps should be considered to ensure that the different stages of the process are completed appropriately, and they improve the quality of the question papers:

- Well-documented processes for producing items, constructing question papers and translating should be designed. A simplified flowchart is shown in the figure below:

```
┌──────────────┐     ┌──────────────┐     ┌──────────────┐     ┌──────────────┐
│  Commission  │ ──> │Receive first │ ──> │Review        │ ──> │  Editing     │
│  questions   │     │draft         │     │questions     │     │  meetings    │
│              │     │of questions  │     │              │     │              │
└──────────────┘     └──────────────┘     └──────────────┘     └──────────────┘
       │
       v
┌──────────────┐     ┌──────────────┐     ┌──────────────┐     ┌──────────────┐
│Typesetting   │ ──> │  Questions   │ ──> │Select        │ ──> │ First proof  │
│and           │     │  banked      │     │questions     │     │              │
│artwork       │     │              │     │for paper     │     │              │
│              │     │              │     │construction  │     │              │
└──────────────┘     └──────────────┘     └──────────────┘     └──────────────┘
       │
       v
┌──────────────┐     ┌──────────────┐     ┌──────────────┐
│   Vetting    │ ──> │ Proofreading │ ──> │Approve for   │
│              │     │              │     │print         │
└──────────────┘     └──────────────┘     └──────────────┘
```
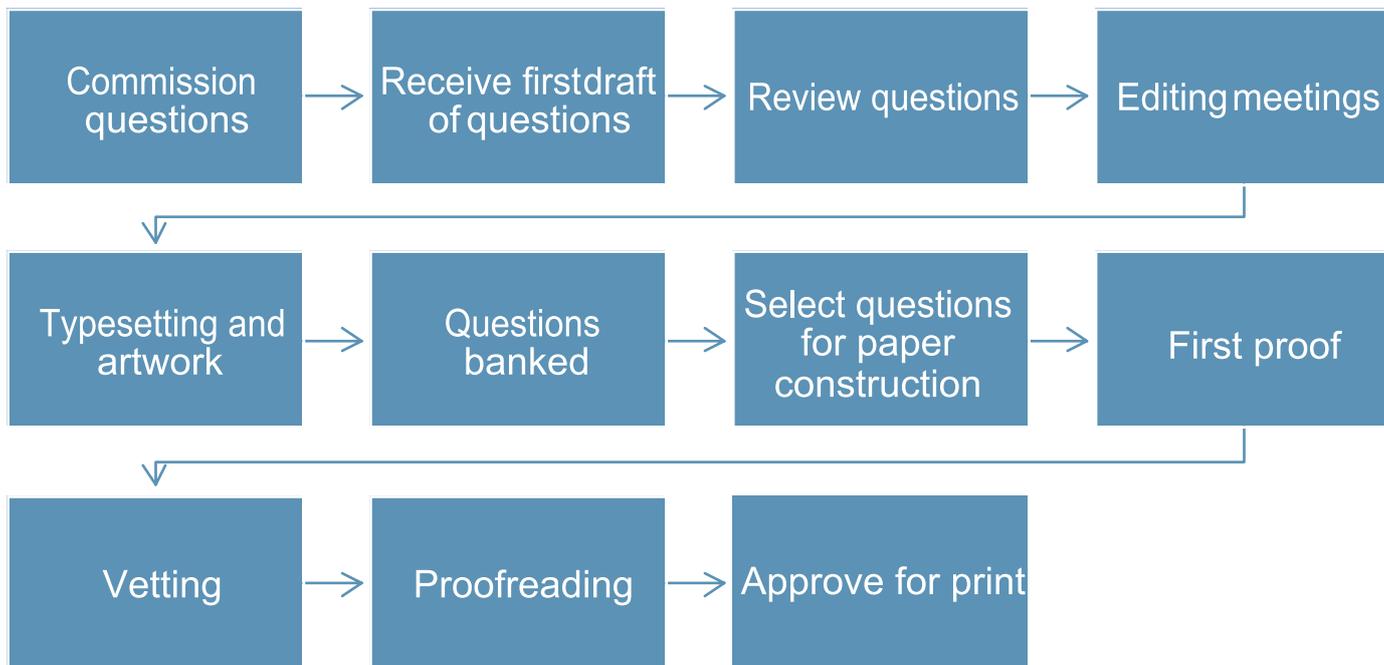
Question paper and mark scheme quality assurance process

The marking of the examinations

MEHE-GDE is responsible for marking of the scripts. Recent professional development initiatives have taken been undertaken by MEHE-GDE to provide effective training and harmonisation of markers both before and during the marking process. As this begins to have an impact and marking accuracy improves over time, MEHE-GDE would be able to implement smaller tolerances between markers and therefore improve the overall reliability of the marking[9].

The Cambridge consultants noted that marking could be made more efficient by using the data in the system to identify markers in real time whose judgments are inconsistent or to a different standard from other examiners. This would enable MEHE-GDE to retrain these markers or stop them from marking. The improvement of poor markers or their elimination from the system would also contribute to increased systematic accuracy that would allow MEHE-GDE to implement smaller tolerances.

Maintaining the standard of the examinations

MEHE-GDE has the responsibility of maintaining the standard of examinations over time. The consultants have recommended that national monitoring survey test development groups should be established for each subject included in the programme to develop the specifications.

---

[9] One such initiative was a Digital Bordereau process carried out in 20per cent of schools that increased marking accuracy while reducing the time and cost of the operations.

Their first task will be to identify the elements of each subject curriculum at each target grade that are suitable for large-scale assessment within the chosen test format, bearing in mind the issues outlined above. This is often referred to as the 'assessable domain'. The specification must define the assessable domain clearly, with curriculum exclusions noted, as this will be the basis for developing the monitoring assessment materials and for interpreting the results. In addition, the specification should include the relative percentage of marks for different elements of the assessable domain that the test should assess. This would allow content that is seen as more important to be awarded proportionally more marks. The percentage should be expressed as a range (e.g. 10–15per cent) to allow some flexibility when constructing papers.

4.4 Assessing students with disabilities, special educational needs and those with exceptional talent

4.4.1 Purpose

The purpose of the arrangements to address the special requirements necessary for those students who fall outside the normal range of student learning and achievement seen in the general population is to meet the needs of exceptional students, fairly and equitably. Exceptional students occupy a broad spectrum of competency and prior attainment. At one end of the spectrum these may be students who have exceptional learning needs or who have some form of disability that requires special consideration; whilst at the other end they could be gifted and talented students who, necessarily, must be given appropriate opportunities to demonstrate their talents to their fullest extent. In order to ensure equity and equality for all, it is important to consider how the assessment process meets the needs of students with Special Educational Needs and Disabilities (SEND) as well as the gifted and talented. Improving the assessment literacy of teachers will support them in creating and delivering assessments that are inclusive and provide all students with the opportunity to achieve and demonstrate their learning. Adjustments to assessments to accommodate the requirements of students with SEND are referred to as 'access arrangements'. To improve access to assessments for SEND students, teachers need to understand that the arrangements will need to be flexible and should be varied to meet the specific needs of the student.

A purposeful and effective national assessment strategy must also embrace the needs of students who have the highest learning potential. High achievers typically, show similar characteristics:[10] they have a thirst for knowledge and a passion for learning; they show fortitude and authenticity in their approach to their studies; and they have a love of learning. In order to enable the expression of these special talents and abilities in formal assessments, the specifications, the mark schemes and the rubrics of the tests taken by all students must be carefully constructed. The assessment design must give opportunities for high achieving students to demonstrate their exceptional abilities through, for example, extension test papers, together with creative and open-ended questions and tasks. To give them the chance to further demonstrate their abilities, additional assessment tools can be carefully formulated in a way that stretches and further challenges them. These tools should be: purposeful in respect to the learning outcomes desired; effective in respect to technical adequacy; and advanced in respect to providing a learning challenge for the gifted learner.

Gifted students can be assessed in ways that suitably reflect their attainment levels and high learning potential through the use of performance-based assessment that is advanced, open-ended, focused on problem solving and thinking, and asks students to articulate thinking. Such performance assessments should be open-ended with multiple correct responses rather than one correct answer. Performance assessment addresses important concepts in a given discipline.

---

[10] Salmela, M., & Uusiautti, S. (2015). A positive psychological viewpoint for success at school–10 characteristic strengths of the Finnish high-achieving students. High Ability Studies, 26(1), 117-137.

4.4.2 Tools and procedures

The purpose of providing access arrangements is to remove unnecessary barriers to the assessment of students with SEND with the aim of achieving a valid assessment of the subject-related knowledge, skill and ability. However, it is important that the access arrangement does not change the demand of the assessment or compromise the assessment objectives of a particular assessment. An example of an access arrangement compromising the assessment objective would be the use of an assistant to read the text in an assessment designed to assess students' reading skills. The student should have the same chance of demonstrating their knowledge and skills as any other student, but must not gain an advantage over other students. Equally gifted and talented students should be given additional challenges or open-ended tasks to establish the extent of their aptitudes and skills over and above the norms.
Teachers should make reasonable adjustments to the end of semester summative assessments to accommodate students with SEND, but the type of adjustment will depend on the nature of the student's learning difficulty, disability or illness.
If a person is required to assist a student, e.g. a scribe, they should not normally be the student's teacher, or somebody close to the student.

When assessing an individual student's need for a particular access arrangement it must be recognized that students may have one of more types of need as follows:

- cognitive and learning
- communication and interaction
- medical
- physical and sensory
- social, mental or emotional.

Schools may become aware of needs by receiving information from people working in the school (teachers, SEND staff, pastoral staff etc.) or through the use of diagnostic test results or the diagnosis by a specialist in the type of need (doctor, psychologist, speech and language therapist etc.). There should be a person in the school who is responsible for coordinating the SEND provision. This coordinator should also be responsible for making teachers aware of a student's special educational need or disability and advising them on the access arrangements that should be made during the end of semester summative assessment. This coordinator will ideally have training in different types of need and how to assess students, and be responsible for maintaining records, but they may also make use of qualified specialist assessors.

---

[11] VanTassel-Baska, J. & Johnsen, S. (2015) Content acceleration: The critical pathway for adapting the common core state standards for gifted students. University of Iowa.

Types of access arrangement

Access arrangements can be divided into two groups. The first group consists of arrangements that can be prepared before the assessment takes place and the second group are interventions that take place during assessment.

Access arrangements carried out before assessment
The assessment materials may be modified to make them more accessible to students. These modifications should take into account individual student needs and should be part of the student's normal way of working, so that they have had opportunities to practise working with modified materials before assessment. Most types of modified materials are for students with visual impairments. The following are examples of how assessment materials might be modified:

- **Enlarged print assessments on the same size paper:** the print may be enlarged to 18 point bold, but the paper size remains the same as is used for the unmodified paper. Care has to be taken to ensure that questions are not split awkwardly across pages. Visual information is simplified if possible, providing it does not change the demand of the assessment. If there are diagrams that students have to refer to, it may be necessary to reproduce them on a separate insert. Scale diagrams will need to remain the same size
- **Enlarged print assessments on enlarged paper:** the font may be enlarged to 18 point bold and printed on larger paper than is used for the unmodified paper. Visual information is simplified without changing the demand of the assessment
- **Enlarged paper:** the standard assessment is enlarged to A3 paper, but there is no other modification
- **Coloured papers:** the assessment materials are unmodified but are printed on coloured paper
- **Electronic papers in non-interactive formats:** these can be read onscreen but cannot be edited. These may be used by students who need computer readers
- **Transcripts of listening tests or videos:** for students with hearing impairments, a transcript of a listening test or a video with subtitles can be provided. A transcript may also be necessary for students who are unable to follow speech at the speed of the listening test. Students with visual impairments may require an audio description of a video
- **Braille papers:** the layout and presentation of the materials need to be modified before being transcribed by an expert into Braille. This should also include tactile diagrams and graphs with Braille labels.

For certain subjects with scale diagrams, maps and other complex visual information, a human assistant may also be required in the examination room.

Access arrangements during assessment

The following are examples of types of access arrangements which occur during assessment. Some require a human assistant, and these assistants should be clear about their role and responsibilities and how to avoid giving students help. Assistants should not give advice on which questions to answer, tell the student when a question or task is finished, comment on the student's answers or give any other help that will give the student an advantage. If a student is using technology as part of their arrangements, the teacher should ensure that the technology does not aid the student, is in good working order, has sufficient power and that the student is familiar with it. With all of these, the arrangements should be part of the student's normal way of working.

- **Extra time to complete the assessment:** extra time of 25per cent should meet the needs of most students. This accommodation will be for students who have low reading and writing speeds, have difficulty with their processing or have other needs such as visual impairment. Students may also have modified assessment materials, as well as extra time. Some students may require extra time of more than 25per cent if, for example, their needs are extensive (multiple disabilities, severe illness etc.) or if they need to use Braille papers, or have to dictate letter by letter. Consideration should be given as to whether or not students need rest breaks instead of extra time and whether the extra time, if too long, could affect the student's condition.
- **Rest breaks:** rest breaks may be helpful to students even if they do not have extra time. These should be supervised and may be in or out of the examination room. Students should not have access to the assessment materials during their breaks, and the breaks are not part of the official examination time.
- **Use of a reader:** a reader will read the questions to the student. The reader should speak in a neutral tone and not give help to the student in the form of explaining or clarifying questions, identifying errors, adding in extra instructions, telling the student which questions to answer etc. The reader may read entire questions or only the words the student asks them to read. Readers may also read the student's work back to them or repeat a question, if asked. Readers should not normally be used for examinations where the skill of reading is being tested or for languages. They should not disturb other students and so the examination may have to take place in a different room.
- **Use of a practical assistant:** this is a person who helps a student perform practical tasks that the student is unable to do safely, unless the tasks are part of the assessment (for example laboratory practical skills or playing a musical instrument). The assistant will support the student, following the student's instructions, but not give extra help.
- **Use of a scribe:** a scribe will write down a student's answers exactly as the student dictates. This may be available to students whose writing speed is very slow, or their writing is illegible, or who are unable to write at all due to disability or injury. The scribe must record accurately what the student says or asks the scribe to do (such as annotate diagrams) and can only change answers at the request of the student.
- **Use of a prompter:** a prompter helps to keep a student focused on the task and the need to answer questions and move on. This is appropriate for students who have attention difficulties or little sense of time. They may sit next to the student but may not necessarily do so. A prompter may also function as a reader, scribe or practical assistant, but in all cases may not give the student additional help.

- **Use of a dictionary (with extra time):** this may be considered for students whose first language is not the language of the assessment. The student may be new to the country or to the language of instruction or assessment and may need this additional support, especially for subject specific vocabulary. The extra time may not be needed.
- **Transcript of student's work:** if a student works in Braille or records their spoken answers or their writing is illegible, a transcript of the work may be necessary. This should accurately reflect the student's answers with no changes. Words should not be added or removed or put in a different order. Spelling, punctuation, grammar and technical terms should not be corrected.
- **Student reads aloud:** this may be helpful to some students. They will need to take their examination in a separate room where they cannot be overheard by other students.
- **Use of a word processor:** a student who writes slowly or illegibly or who is unable to write using a pen may benefit from using a word processor. A word processor must not help the student in any way, such as predictive typing, checking grammar, defining words etc. It should not be connected to the internet or provide access to other software packages.
- **Use of a visual aid:** this may be a magnifier or other low vision aid.
- **Use of colour name assistant:** this is used for students who are colour blind – an assistant names colours if the student asks them to.

Special assessment arrangements for high attaining students

Classroom performance assessment is a particularly appropriate tool for guiding and assessing the learning of advanced learners. In the classroom, performance assessments are designed to promote multiple responses supported by evidence and to emphasize fluency and complexity over speed. Classroom performance assessment for advanced learners focuses on the exploration of advanced content, on higher-level thinking and problem solving, and on the use of reflective metacognition to help students internalize their learning.

As is true in all forms of alternative assessment, technical issues in terms of the validity of the task performed in relationship to the instructional purpose and the interrater reliability in evaluating student responses must be considered. The use of established research-based protocols and rubrics enhances effective assessment. Although the use of performance tasks and assessment requires additional classroom time, the open-ended format, the support for articulation of thinking, and the resulting development of more complex responses makes performance assessment particularly appropriate for advanced learners.

4.4.3 Roles and Responsibilities

The Professional and Inservice Training Bureau (PITB) at CRDP provides the training necessary to support the assessment of students with exceptional needs. CRDP also produces an adapted official examination specification and has successfully run pilot progammes in the use of screening.

4.5 National Monitoring – Curriculum Based Assessment

4.5.1 Purpose

National monitoring in in the form of Curriculum Based Assessment in currently being carried out by CRDP. Its purpose is to survey the levels of achievement of a representative cross-section of the students in grades 3 and 6. It uses a sample-based methodology and, as such, it clearly signals that the assessments are not high stakes for the teachers or students but are intended only to inform policy at a national level. This enables precise psychometric calibration of materials by pre-testing questions or by reusing some questions to provide an anchor between test versions. Developing the test versions using these techniques helps to provide the dependable information needed to inform policy making.

Running the assessment regularly has significant benefits in relation to development and allows the administration costs to be manageable. A three-year cycle provides the information needed to inform policy however, but it is not sufficiently frequent to generate the data required to monitor specific schools and teachers. This serves to improve the quality and scope of reliable assessment information to inform and evaluate policy and monitor levels of achievement in both student subgroups and the student population as a whole. Assessments should focus on literacy and numeracy in grade 3, and on literacy, numeracy and science in grade 6. A three-year cycle would support analysis of age-related achievement progression as built into the respective curricula, as well as measuring changes in educational achievement over time.

CBA provides the necessary data evidence to address the relevant indicator in Sustainable Development Goal 4, relating to education; it is compatible with plans laid out in S2R2. The programme would also complement the information provided by participation in international studies while offering greater flexibility to create assessments and data that are more directly relevant to the Lebanese context.

The primary purposes of CBA is to:

- monitor population levels of attainment in selected subjects over time in the selected grades, to confirm stability or detect change
- monitor gaps in the attainment of selected student subgroups over time defined by, for example, gender, school, region, SEN, or disability
- identify priorities for policy intervention
- evaluate the impact of policy interventions over time
- suggest research needs in relation to subject attainment.

The assessments are not high stakes for the individual student or school because they are only be conducted on a stratified sample of students and could therefore not be used for accountability purposes.

4.5.2 Primary Resources and outputs

CBA would occur at the end of each cycle; the imperative is for the assessment to take place in primary education schools by the end of grades 3 and 6.

The operation of CBA would include two phases: the field trial and the main study.

This assessment would be done every three years in grades 3 and 6.

As these types of tests would be happening for the first time, the assessments and administration should be underpinned by awareness workshops about the function, the goals, and importance of such assessment processes. The workshops should include principals, teachers, parents and students. The staff should be trained on administering CBA in schools and how all conditions must be well controlled to avoid any bias.

The processes for the administration of CBA must be open, clear and consistent. There should be, for example, the construction of a representative sample based on a suitable sample size, according to schools sector; according to different governorates, and according to first foreign language (French-English-trilingual). The scope should include a facility for the creating replacement samples in case the need arises.

Schools will require significant additional resources for the marking of scripts and for collecting the data for assessment and questionnaires responses. The funding necessary to support these technical and infrastructure needs is an important consideration during the formulation of long-term education budget cycles.

Establishing CBA will allow system monitoring over time in numeracy and literacy and other subjects. The presence of student, teacher, and principal questionnaires allows for different analyses of variables that are likely to be influential on achievement of students: such as, attitudes towards the different subjects as well as perceptions about its teaching. The school principal and teacher surveys will also enable their views about school climate and teaching resources to be tracked.

National reports: There should be analysis of the results and the questionnaire evidence.

Short reports: short reports are recommended for the teachers that are concise and focus on specific areas of knowledge and skills from the curriculum and provide examples of test questions with different question level data and commentaries. The short reports would help raise the profile of CBA tests in Lebanon and provide materials for classroom teaching.

Producing good quality national reports requires, the presence of good statisticians and quantitative methodologists in order to ensure the efficiency of sampling techniques and methods, and to ensure the validity and the reliability of the questionnaires/data analysis. These individuals should be hired and trained in order fully understand the needs at both system and school level. Appropriate measures are also needed to prepare the stakeholders so that they are aware of the limitations and meaning of the data and the content from the reports.

4.5.3 Roles and responsibilities

The establishment of CBA as a regular and routine procedure, and an integral part of the NSLAF will require the following actions:

- Constructing survey questions aligned with the curriculum of the cycle and subject concerned
- Constructing the school sampling for CBA
- Organizing survey distribution to the sample schools
- Constructing questionnaires for different stakeholders
- Constructing management tools within the schools
- Arranging marking and questionnaires responses
- Collecting response data for analysis and producing reports.

Given their great experience in conducting such complex research activities it is recommended that CRDP leads and coordinate this national testing regime.

4.6 International surveys

4.6.1 Purpose

For the purposes of comparability and calibration of student achievement, Lebanon is currently engaged with international studies, including TIMSS (Trends in International Mathematics and Science Study) and PISA (Programme for International Student Assessment). TIMSS is sample-based, with surveys on a four-year cycle with a focus on the mathematics and science achievements of students in grades 4 and 8. TIMSS Advanced is also sample-based and focuses on students following advanced mathematics or science courses in the upper secondary school. PISA takes place every three years and tests a sample of 15-year-old students in reading, mathematics and science. Underachievers in PISA are those pupils who fail to reach the minimum proficiency level necessary to participate successfully in society. PISA also makes it possible to analyse national performance by gender, socio-economic status and immigrant background, and contains other contextual information on the school environment and pupils' attitudes, such as their plans for further education, their assessment of their own exposure to bullying and sense of belonging at school. In recent years the achievement of students in Lebanon in the PISA tests has been disappointing. It has been suggested that one reason for this may be due to the three different languages of instruction making it difficult to find a statistical representative sample for the reading literacy tests, and that higher achieving students are under-represented.

Lebanon has regularly participated in TIMSS at grade 8 and TIMSS Advanced at Third Secondary. While this produces valuable information, it currently only related to students in the intermediate and secondary phases. Increasingly, there is recognition amongst education professionals that learning progress in the primary years is of critical importance in improving the life-chances of individuals; beyond the age of 10 years, it gets more difficult for students to catch up on earlier learning losses, so early identification of learning support needs is critically important. The proposal for the introduction of widespread Initial pre-Primary Screening tests in Lebanon is one aspect of the shifting the assessment focus to younger learners; equally, by participating in TIMSS at grade 4, the Government of Lebanon and MEHE would have a stronger evidence base upon which to build policymaking in respect of future education priorities.

Cambridge International recommends that Lebanon also participates in TIMSS at grade 4 in the elementary phase to provide better longitudinal evidence of trends and progress from Cycle 2, through Cycle 3 and beyond.

4.6.2 Tools and procedures

To take part in TIMSS at grade 4 the following tasks and procedures should be followed:

i) Carry out appropriate communications and training programmes to prepare primary schools, students and parents.

While TIMSS at grade 8 takes place in intermediate education schools and TIMSS Advanced in secondary education schools, TIMSS at grade 4 would take place in primary education schools in Cycle 2. This means that it is not possible to simply extend the current processes from TIMSS at grade 8 to grade 4. Primary schools recruited for the field trial and the main study assessments and administration of the tests, for example, will be taking part for the first time and are likely to be unfamiliar with the TIMSS survey. Therefore, good communication with principals, teachers, parents and students on the benefits and importance of TIMMS is key to ensuring the required participation rates. Training for staff administering TIMSS in schools will also be vital in ensuring that bias is not introduced into the study and that there is a smooth start to the new survey at grade 4.

ii) Ensure that adequate resources are available.

The processes needed to administer TIMSS at grade 4 in Lebanon, such as the recruitment of main sample and replacement sample schools and marking the assessment and questionnaire responses, will require significant additional resources. Additional funding will also be required by the International Association for the Evaluation of Educational Achievement (IEA) to cover the cost of the study at national level and to contribute to the costs of coordinating TIMSS internationally. Furthermore, in 2019 TIMSS was made available in digital format, following on from PISA making scientific literacy available digitally in 2015. The cost of implementing the digital version of TIMSS is less than of the traditional paper-based version, but brings with it technical and infrastructure considerations.

iii) Review and expand existing processes for administering TIMSS at grade 8.

Some existing processes will be transferable from TIMSS at grade 8 to TIMSS at grade 4, such as preparing the Lebanese versions of the test booklets. However, the overall resource requirements will increase. Other processes should be adapted for primary schools.

Reporting the outcomes of international surveys

Taking part in TIMSS at grade 4 will allow system monitoring over time in mathematics and science from Cycle 2 to Cycle 3 and in mathematics and Physics for the older students in the years in which TIMSS Advanced takes place. Monitoring over time will include both the Lebanese context independently, as well as comparison against international benchmarks and other participating countries. Depending on which surveys Lebanon takes part in, the inclusion of student questionnaires allows analysis of changes in factors that are likely to be influential on achievement in Lebanon, such as pupil attitudes towards mathematics and science and their perceptions of its teaching. The principal and teacher surveys will also enable their views about school discipline and teaching resources to be tracked. To maximize the usefulness and impact of this additional data a broader range of targeted reports is needed.

National reports: Currently the IEA analyses the international database of country results and the questionnaire evidence; in addition, some analysis of data from international surveys already takes place in Lebanon. However, to take full advantage of the additional information

provided by the introduction of TIMSS at grade 4 the Cambridge International experts recommended that a national report is produced and disseminated, along with separate reports for TIMSS at grade 8, TIMSS Advanced and PISA. Producing a national report at each stage would allow stakeholders to use the survey information when reviewing the curriculum and teacher training programmes etc.

Short reports: in addition to the national reports Cambridge International recommended producing short reports for teachers providing a digest of relevant information from the surveys. These should focus on specific areas of knowledge and skills from the curriculum and provide examples of test questions with question level data and a commentary. The short reports would help raise the profile of international tests in Lebanon and provide materials for classroom teaching.

Resources: producing good quality national reports for the TIMSS and PISA surveys requires input from experienced psychometricians who will understand the way the surveys are designed, including the use of complex sampling, data weighting and scaling. The report writers should also receive adequate training to ensure that they are also aware of the limitations and meaning of the data and the content that should be included in each report.

4.6.3 Roles and responsibilities

In previous rounds of international testing CRDP has acted as Lebanon's national centre for PISA and TIMSS. The tasks completed include:

- reviewing survey questions for curriculum appropriateness within the country
- organising survey distribution to the sample schools
- survey and questionnaire administration and survey session management within the schools
- arranging marking and questionnaire response recording
- forwarding electronic response data for analysis.

CRDP has also produced the national TIMSS and TIMSS Advanced reports (2015 surveys). Given their great experience in conducting such complex research activities it is recommended that CRDP continues to be the national coordinating centre for international testing as well as being the focal point for all research and comparative testing regimes in the future.

4.7 A Focus on Data

The NSLAF will significantly extend the range of student attainment data available to the Lebanese government for education system management and improvement. To ensure data quality and policy value, it is important that the data are appropriately validated and stored. There is already a sophisticated data management infrastructure in Lebanon, however the current system can sometimes make it difficult to gather together the full range of information needed to make informed, coherent and timely decisions. It is therefore vital for Lebanon to have an integrated system for managing data. This will ensure that data from assessments are properly validated at the time of capture and appropriately stored, allowing authorized individuals and groups to have access to relevant subsets of data for legitimate purposes[12] such as school management and research. Output 3.1 of Pillar III of RACE II states that CRDP is capacitated to administer effective data management through a unified framework that allows understanding of the percentage of schools with functioning data management and their datasets for refugee enrolment per scholastic year.

When fully developed, the NSLAF will be dependent on the future data strategy as specified in the RACE II Strategy. Therefore, the NSLAF should inform the development of the data strategy to ensure the assessment data it provides are used effectively.

4.7.1 Data sources and analysis

When fully implemented the assessment landscape provided by NSLAF will include:

- assessments of students' literacy and numeracy development in the early primary grades
- test scores and teachers' marks from end of year summative classroom assessment
- results of assessment in national system monitoring surveys at the end of Cycles 1 and 2
- Brevet and Baccalaureate marks in grades 9 and 12 respectively, overall averages and subject marks
- students' scores achieved in TIMSS and PISA surveys
- student demographic data.

Each of these data sources is valuable individually, enabling routine analyses (summing, averaging, mapping etc.) leading to automated lists and reports for official distribution. For example:

- national examination marks and grades (e.g. student lists nationally and within schools)
- students' average marks in end of year summative classroom assessment, overall and by subject, with grade retention decisions if relevant
- student population performance estimates in national monitoring surveys with subgroup breakdowns (e.g. by gender, type of school or region of the country).

---

[12] Output 3.1 of Pillar III of the RACE II Strategy: MEHE (2016), Reaching All Children with Education: RACE II (20172021), p.16. https://planipolis.iiep.unesco.org/sites/planipolis/files/ressources/lebanon_race-ii_2017-2021.pdf

The data management system will monitor Lebanon's progress in meeting the SDG 4 indicators.[13]

The data management could provide a particularly powerful new service of data linking across assessment systems from grade 1 to Baccalaureate. This will increase the scope of the analyses supporting system improvement, for example through more effective resources allocation at school or regional level and through more effective educational research.

4.7.2 Data capture, validation, storage and retrieval

For the system to be efficient and manageable, it is essential that there is a straightforward and 'user-friendly' process to allow teachers, practical assessors, examination markers and others to transfer assessment data into the data management system as the data are gathered. Data validation is also crucial. Whether the information provided is national examination marks or teacher judgments, without a robust data validation process before and during data input, the value of that dataset and any analyses using it will be compromised.

In order to maximize the value of assessment outcomes, whenever possible data should be stored at the micro level, that is at the level of marks for individual questions. This is because data at this level can be summarised when necessary, whereas summarised data cannot be separated later into its component parts.

Once validated, data stored within the system should allow rapid and flexible retrieval by authorized users. These could include national examination analysts in MEHE-GDE, school support personnel in MEHE-GDE-DOPS.

4.7.3 Data management

A considered data strategy driven by efficient databases is essential for an enriched assessment landscape as indicated in this assessment framework to be implemented effectively. It is also critical in obtaining the maximum benefit from the assessment data it generates.

Demographic information is routinely held for students through their time in schooling. This is supplemented with contextual information about learning and the learning environment. Additional information is gathered through teacher and student questionnaires in the large- scale attainment surveys. This information will support the analysis and interpretation of student attainment data.

The experts from Cambridge International recognized that a sophisticated data management infrastructure already exists in Lebanon and that there are ongoing projects to improve these systems.

---

[13] Proportion of children and young people (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex UNESCO-UIS (2018). Quick Guide to Education Indicators for SDG 4. Table 2, p.19. http://uis.unesco.org/sites/default/files/documents/quick-guide-education-indicators-sdg4-2018-en.pdf

Enriching Lebanon's assessment landscape in the ways suggested in this report will have a significant impact on the range, depth and variety of student attainment data. This will add value in education system management in general as well as supporting policy guidance and system evaluation.

The enriched assessment landscape will include the continuation of:

- end-of-year summative classroom assessment
- national examinations for the Brevet in grade 9 and the Baccalaureate in Third Secondary
- international surveys (PISA for 15-year-olds, and TIMSS at grades 8 and 12).

In addition, Cambridge International proposes introducing:

- assessment of students' literacy and numeracy development at the start of Cycle 1
- additional TIMSS international surveys at grade 4.

It is important that attainment data are dependable, in other words both valid and reliable, as well as readily interpretable. In addition, all demographic and questionnaire-based information must be as free as possible from extraneous errors in data capture, data merger, and so on.

Optimize the quality of attainment data

The quality of assessment data can be optimized, if not always guaranteed through:

- assessment instrument design
- response recording procedures
- standardised marking and rating
- formal post-capture validation checks.

To ensure that the data are an accurate reflection of the assessment judgments made:

- Clerical errors can be minimized by ensuring that markers are not required to sum or average marks for multi-part questions or entire question papers. Data should be entered in a raw form and calculations completed automatically.
- Marks should be electronically stored in a form as close to raw data as possible, not as summary statistics. This will maximize the value of the data stored and facilitate sophisticated marker monitoring systems.

Assessment instruments need to be carefully designed to maximize validity in terms of knowledge and skills coverage, fairness in terms of candidate access, and ease of use for the assessor and student. Test and examination papers should reflect the learning focus in schools and classrooms and be presented so that questions are clearly worded.

i. Ensure the participant records are valid

Participant records relate to information on students, schools, school principals, teachers, parents/guardians, and others. The records will include names, addresses, age, grade, gender, etc.

It is vital that these records can be uniquely associated with individuals, so that information might be added to the correct record or records merged. This is critical to monitoring student progression over time.

The use of a unique identifier for students is currently being implemented in Lebanon and it is critical that this work is adopted as standard across all systems.

ii. Ensure a definitive version of data

It is important that there is a single definitive version of each type of data and information. If information is stored on multiple systems, there is a significant risk of the datasets on each system diverging. This would threaten the accuracy of analyses performed on these data. This would undermine the trust and perceived value of these analyses.

There must be clear validation and verification rules in each system where the definitive version of the data is owned. It is also important that there are clear protocols for automatically merging datasets, so avoiding error-prone manual processes, especially if the definitive versions of the datasets are stored on different systems.

4.7.4 Recommendations for facilitating data analyses

For most assessments in the NSLAF there is a set of required data analyses. These are needed to produce standard reports. Example reports might include:

- national examination results by region and school
- national examination results by gender, language of instruction, SEN, etc.
- national monitoring survey outcomes by year, region, gender, etc.
- summative classroom assessment marks by student, teacher, school etc.
- grade retention rates by region, school, teacher, etc.
- reading and mathematics achievement data relevant to SDG 4.1.1 – for grades 3, 6 and 9.

To ensure that the reports are as cost-efficient as possible, the Cambridge International experts recommended the following:

i) Identify the data analyses carried out as standard and plan for automated production
Most reports on assessment outcomes that are routinely produced can be automated. For reports where this is not already the case automation should be considered and budgeted for.

To maximize the flexibility of the reporting and to ensure that reports generated are as up to date as possible, systems should avoid storing pre-compiled reports or test forms. Wherever possible these should be generated on demand from the base data by authorized users. Automated reports that show the level of compliance with these key performance indicators should be designed and produced.

ii) Identify data processing requirements to support secondary analysis

It is important that there is a single definitive version of each type of data and information. If information is stored on multiple systems, there is a significant risk of the datasets on each system diverging. This would threaten the accuracy of analyses performed on these data. This would undermine the trust and perceived value of these analyses.

There must be clear validation and verification rules in each system where the definitive version of the data is owned. It is also important that there are clear protocols for automatically merging datasets, so avoiding error-prone manual processes, especially if the definitive versions of the datasets are stored on different systems.

* tracking Brevet to Baccalaureate results overall, and by subgroup
* correlating reading attainment in grade 3 with that in grade 6
* exploring variation in grade 4 to 5 retention rates across schools with reference to grade 1 'primary-ready' assessments
* comparing national examination outcomes overall with international survey results over time.

iii) Agree researcher responsibilities and data access rights

Policy-relevant secondary data analyses are carried out both within MEHE-GDE and CRDP. For clarity and consistency, it is important that access protocols for stakeholders within these organisations are agreed to facilitate these analyses.

There will also be many non-sensitive research analyses that could usefully be carried out by university researchers and others. These should be planned for and protocols governing data access rights and responsibilities established for such cases.

5. Continuous Professional Development Programmes (CPD)

5.1 CPD for Initial Screening

The initial screening of all children beginning their primary education would be a large and expensive national undertaking. Ideally the screening should involve not only assessing children's literacy and numeracy but also their cognitive and behavioral characteristics, to assess their all-round development and their readiness to commence formal classroom-based schooling. On practical and pragmatic grounds, therefore, it has been proposed by the experts from Cambridge International that, based on the considerable expertise that already exists in conducting Early Grade Reading Assessments (EGRA), the skills and experience of the existing trained assessors in Lebanon should provide the basis of a CPD programme national screening programme for literacy. EGRA's focus, to date has been upon grades 2 and 3, and it has been conducted as sample-based assessment, rather than as a tool for assessing all students; yet the basic principles are the same and building upon locally-based expertise is always the most successful way to implement system change amongst teachers.

Equally, the partner assessment tool for numeracy Early Grade Mathematics Assessment (EGMA), could meet the requirements for early numeracy assessment in grade 1. For this early stage, EGMA covers simple addition and subtraction, number comparison, number patterns (missing numbers), and simple word problems. After an initial piloting of the new EGRA screening assessment EGMA screening could be introduced in successive years. These basic skills checks for literacy and numeracy would then be supplemented by the wider cognitive and behavioral assessments to gain an all-round picture of each child's characteristics during their first year of school.

The training could be conducted on a cascade basis with a core of national 'tier one' trainers receiving their instructional foundations from The Professional and Inservice Training Bureau (PITB) at CRDP. These 'tier one' national trainers would, in turn, train local staff in the regional training centres to carry out the assessments. The first line assessors could be either class teachers or itinerant trained assessors who would travel from school to school to carry out the assessments for all students on a one-to-one basis, using the appropriate assessment instruments as indicated.

Assessment results would be recorded electronically by the assessors and incorporated into the data management system. The long-term aim would be to give a unique student number for each learner in the school system so that their attainment and progress could be tracked even as they moved from school to school across the country. The quality of data security and privacy are of the utmost importance here, as the pupil performance date is highly sensitive and access to it should be restricted only to individuals who have explicit authorization from MEHE / CRDP.

5.2 CPD for assessment literacy

To fully understand the power of assessment as an educational tool, teachers must always remember how it fits into the bigger educational structure. This can be achieved by making clear links between the curriculum, assessment and pedagogy. A strong connection between these three elements will maximize the positive impact in the classroom and also ensure that decisions made on the basis of the end of semester summative assessments are both reliable and fair to students.

Improving teachers' assessment literacy will help them to make sense of assessment in an organised and consistent way and ensure it has a positive impact on their students. Teachers at all career stages need to have a clear understanding of why we assess, what to assess, how to assess and how to use assessment information effectively.

Successful assessment of the current curriculum, and any newly emergent curriculum, is dependent on the assessment literacy of teachers and changes should be piloted and progressed in line with curriculum needs.

5.2.1 Formative assessment

Formative assessment, often seen as synonymous with 'Assessment for Learning', is part of good pedagogy. Throughout the world effective teachers use their regular dialogue with their students to check on understanding and progress and provide feedback in order to inform and improve future learning. This is natural part of the teaching and learning process. Traditional classroom assessment techniques – such as questioning, quizzes, homework, topic tests, projects and portfolios – are the most common type of assessment taking place in the Lebanese classroom and these tools help both teachers and students to understand, often in an informal way, how the learning is progressing.

The introduction of a more varied range of ongoing classroom-based assessment can involve a shift in the traditional classroom dynamic, as it often requires a significant level of agency on the part of the student to realize its full benefits. These benefits include improved student outcomes, a cooperative classroom culture and more independent and motivated students. In order for a teacher to make this shift, they will need to engage in the following:

- making learning intentions clear, but also sharing success criteria which show how students' success will be measured against the learning intentions
- replacing the idea that only the teacher can assess and suggest improvements with the expectation that students will also reflect on their work independently and take responsibility for their own learning
- encouraging students to take risks and attempt challenging tasks, safe in the knowledge that mistakes are valued and can be analysed as an opportunity for learning
- emphasizing personal improvement rather than the competitive nature of marks so that students gain more self-confidence and are more motivated to improve.

The following tables provide examples of strategies for classroom assessment that would help achieve the benefits outlined above. To support implementing these strategies, it is important that teachers participate in professional learning communities beyond the assessment literacy training described in this report. The professional development that takes place in learning communities does not have to be formal face-to-face training. It could include groups who meet regularly to share their knowledge, skills and experience for the improvement of lesson delivery, assessment and academic performance. These groups would provide a vehicle for training, allowing teachers to share strategies for effective classroom assessment and reflect on and improve their practice.

i) Effective questioning and feedback

Types of question

Strategy	Key benefit(s)	Notes and examples
Open questions	Requires students to consider their own ideas, attitudes and understanding and typically provide more than a one-word answer. Open-ended questions also encourage students to develop their language and vocabulary skills.	In a lesson, the teacher might ask 'Why do you think the author chose this particular word?' (open question). This encourages a more developed answer where the student needs to provide an opinion and justification. Follow-up questions such as, 'Can you tell me more about that?' or 'Why do you think that?' help to develop the discussion.
'Might' questions	Gives students greater opportunity to think and explore possible answers.	When questioning, insert the word 'might'. For example, instead of asking 'What is the meaning of democracy?', try asking 'What might the meaning of democracy be?' The first implies that there is a single, correct answer known by the teacher whereas the second encourages discussion and challenge.

High challenge	Encourages students to move beyond lower-order thinking skills (remembering, understanding, applying) to higher-order thinking skills (analysing, creating, evaluating).	Plan questions carefully using Bloom's Taxonomy. For example, the following question stem examples show how students can be moved towards higher order thinking skills: Remembering (How many …?) Understanding (How do you know …?) Applying (What would happen if …?) Analysing (What evidence proves …?) Evaluating (How effective is …?) Creating (If you had to find a new way to …?)
Hinge questions	Helps to check understanding of a particular concept so that the teacher can decide whether to move on or whether a concept needs revisiting.	An effective hinge question should be quick for students to answer and easy for teachers to assess all students' answers (i.e. multiple choice). It is more effective if the question is based on a common misconception and hard to answer correctly if the student hasn't understood the concept.
Bouncing	Encourages students to build on each other's ideas and engages more students in the discussion.	Bounce answers around the room. For example: 'Maya, what do you think of Charbel's answer?' 'Charbel, how could you develop Maria's answer to include more detail?' 'Maria, how might you combine everything we've heard into a single answer?'

Wait time	Gives students time to think before producing an answer. Not everyone in the class thinks at the same speed or in the same way – waiting allows students to build their thoughts and explore what has been asked.	After asking a question, wait before taking students' responses. Once the student has given an answer, wait before responding. This gives the student space to elaborate or continue – or for another student to add their ideas.
No hands up	Encourages all students to engage in the discussion and have an opportunity to participate and ensure no one dominates the discussion.	Write students' names on a piece of paper and put these in a box. Choose a name at random to answer a question. If a student doesn't know the answer, encourage them to give ideas about what they do know rather than skipping straight to the next student so that ideas, and not just answers, are valued.

Types of feedback

Strategy	Key benefit(s)	Notes and examples
Confidence level	Allows students to show their level of understanding and identify the areas where they need to focus their efforts.	Students respond by standing / sitting / hands up / fist of five / thumbs up depending on their level of confidence with a task. Students can also make their indications with their eyes closed, so it is only the teacher who sees the judgment. For example, at the end of a physical education lesson, the teacher conducts a plenary review of the main objectives of the lesson. Students indicate with thumbs up / thumbs down or hands wavering in the middle as to whether they believe they met the objectives. They then discuss one thing they could do next time to improve with a partner on the way back to the changing rooms. As they leave the changing rooms, they tell the teacher.

Traffic lights	Allows students to show their level of understanding and identify the areas where they need to focus their efforts.	At the start of a science lesson, students are asked a key question. Each student responds to the question by raising a red/amber/green card indicating whether they could answer the question with confidence. Throughout the lesson, students leave the card on their table, changing the card as the lesson progresses (turning the red card face up if they don't understand something at all or turning the green card over when they feel confident enough to answer the question). The task is repeated at the end of the lesson. If coloured card is unavailable, students can draw smiley faces on three pieces of card (11 for green / 11 for yellow / 11 for red)
Show me boards	Challenges common misconceptions.	Each student in the class has their own 'show me' board. These can be an A4 piece of paper folded to allow for answers to several questions. A teacher asks the whole class a question, gives students time to think and then indicates when to hold their answer/ideas up. Giving students questions that require careful consideration and challenge misconceptions is important and they need to be planned in advance.
ABCD cards	Challenges common misconceptions.	Give each student a set of cards with the letters A, B, C and D on them. A question is posed by the teacher who also gives four different responses. Students use their cards to identify what they believe is the correct answer. The teacher can quickly scan the room to see the spread of choices. Individuals can be asked to justify their choice. Following this, students may decide to change their letter.
Corners	Challenges common misconceptions.	Each corner of the room represents an answer to a problem posed. Each member of the class goes to the corner that they believe is correct. Once settled, the class is given a few minutes to discuss in their corner why they are correct and the others are incorrect. A representative from each corner is then allowed to put this to the class. The aim is that eventually all the students will have chosen the correct corner.

Exit passes	Provides a clear insight into what has been learnt from the lesson and this information can be used to guide the next lesson.	This activity can be presented in several ways. At the end of a lesson, students each have to volunteer something useful they have learnt from the lesson in order to leave the room. Repetition is not allowed and if a contribution is deemed to be too simple the teacher can ask for a second 'exit pass'. An alternative idea is that a challenging question is posed at the end of the lesson. Students write their solution down and show this to leave the room – whether correct or not it is collected.
Hand in, pass out	Encourages students to practice marking answers and gain further knowledge and understanding about a topic.	Ask students questions and, have them respond on notebook paper anonymously. Students then hand their papers in. Immediately, the teacher randomly gives them back to students for marking. Students should not know whose work they have. The teacher then takes an informal poll about how many questions students answered correctly.
Quizzes	Provides information to inform future learning activities, or to give feedback to the students.	Give students quizzes, which either the teacher or they mark. Different students can be given different types of quiz on the same topic (differentiation). One quiz might be multiple choice with four answers, and another might only have two choices.

Types of self-assessment and self-reflection

Strategy	Key benefit(s)	Notes and examples
Learning diary	Encourages students to reflect regularly how they are learning not just what they are learning.	At the end of each activity or lesson ask students to answer questions in their diaries to support them to reflect on how, rather than simply what, they've learnt. For example: What did I find easiest to learn his week? Why? What did I find most difficult this week? Why? What could I do differently next time? This needs careful scaffolding and modelling in the early stages as students tend to write descriptively rather than reflectively – modelling the language of reflection and building up this vocabulary is essential for success.
3–2–1	Encourages students to reflect on what they have learnt and provides information for the teacher to help plan the next lesson based on feedback from students.	<table><tr><td>3</td><td>Things I learned today</td></tr><tr><td>2</td><td>Things I found interesting</td></tr><tr><td>1</td><td>Question I still have</td></tr></table>

Personal portfolio	Supports students to document their progress and explain the differences between examples of their work at different points and stages of development. If portfolios are digital, they are easy to store and access	Students compile a series of files of their own pieces of work annotated in Arabic at different levels.
Pre- and post-task reflection	Allows students to demonstrate prior learning and enables them to create a baseline from which they can measure progress.	At the start of a unit of work on a science topic, students make a note of anything and everything they already know about it. They review their notes at the end of the unit of work and check to see how accurate they were and what else they have now learned. advance.
Self-assessment checklist	Encourages students to reflect on their work and whether it meets the success criteria.	Before handing in a piece of persuasive writing, students review their own work and suggest the mark they believe they should receive for the work. They use the success criteria for the task as a checklist and need to identify the evidence to support their judgment. It is most effective when there is a time-lag between completion of the work and the reflection point.
Highlighting success	Encourages students to reflect on their work and whether it meets the success criteria.	Ask students to use different colors to highlight on their work where they have shown evidence of different skills according to success criteria.
Exam wrappers	Supports students to review their performance in a test and think about how they can improve.	An 'exam wrapper' is a worksheet with a set of reflective questions. You could give the worksheet to students before they receive feedback with questions to prompt them to think about how they prepared for the exam and the types of study strategies they used. The worksheet can also be handed out after students have had their feedback. The questions could focus on categorising errors and reflecting on how they can prepare differently next time.

Personal portfolio	Supports students to document their progress and explain the differences between examples of their work at different points and stages of development. If portfolios are digital, they are easy to store and access.	Students compile a series of files of their own pieces of work annotated in Arabic at different levels.

Types of peer assessment and peer feedback

Strategy	Key benefit(s)	Notes and examples
Two stars and a wish	Supports students to give effective feedback to their peers on the quality of their work and ideas for improvement.	Once students have completed a piece of work, ask them to swap with a partner. Students should make two positive comments (stars) on their partner's work and one specific area for improvement (a wish). Students are given the opportunity to respond to the feedback either by improving the work or by applying the wish next time.
Peer-assessment prompts	Supports students to give effective feedback to their peers on the quality of their work and ideas for improvement.	Put some helpful prompts on the board that students can use to give each other feedback on their work and model using them for your students. For example: - I like … because … - From your work, I have learned … - I think next time you should … because …

Types of self-assessment and self-reflection

Talk partners / response partners	Provides an overview of learning that has taken place.	Ask students to share with a partner: • three new things they have learnt • what they found easy • what they found difficult • something they would like to learn in the future.
WAGOLL **(What a good one looks like)**	Helps create a shared understanding of success criteria by highlighting the features that make a good piece of work.	Give students a piece of work that is not perfect but is of a standard that students can achieve. Students work in groups to assess the piece of work, using the success criteria. Alternatively, you could get students to use the example work to create their own success criteria by identifying what they think works well and why.
Observation checklists	Provides students with a focused task to do when evaluating the work For example, in a language lesson, ask students to create a checklist and then evaluate each other's oral presentations. of a peer. Helps students focus on specific aspects of performance. It focuses students on learning intentions and outcomes and ensures that their comments are objective and constructive.	*(see table below)*

Criteria	Student 1	Student 2
Uses present tense		
Uses past tense		
Uses future tense		

| Peer-feedback observation | Supports students to give effective feedback to their peers. | This is a useful follow-up activity to do after having established ground rules or key skills for peer assessment. Students work in threes – one person observing the peer assessment skills of the other two. After establishing that students need to actively listen, focus comments on the learning intentions and encourage their partners when giving feedback, an observer 'evaluates' the feedback. |

Criteria	Student 1	Student 2
Makes eye contact		
Picks up on comments and develops		
Keeps to the objectives

Backs up comments with evidence | | |
| Suggests improvements | | |

5.2.2 Alternative and authentic assessment

At present Lebanon's grade 1 to Third Secondary school curricula are subject-focused and rely strongly upon memorization with less emphasis upon cognition and analysis. The international experience of curriculum development, currently, shows a direction of travel which places an increasing emphasis on the development of 'competence' (skills and values), at both subject specific and transversal levels. These often might include practical skills in science and oral skills in language. The current core skills in the Lebanese curriculum of Creativity, Critical thinking and Problem Solving, Citizenship, Leadership, Digital Literacy, Communication and Collaboration, Cooperation, Empathy and Respect for Diversity. The status and value of these core skills will be raised when they become part of a formal assessment process.

High quality assessments reflect the intended curriculum. A stronger emphasis on the transversal competencies within any new curricula is reinforced by the implementing a wider range of assessment forms to supplement the currently used written tests. However, there some important considerations to be made before these new forms of assessment are introduced.

It is important that the higher order skills and competencies in a reformed curriculum are assessed in an appropriate way. If well implemented, these assessments can provide meaning and intrinsic motivation to students and help drive and structure the pedagogical developments needed to introduce higher order skills and competencies successfully into the classroom.[14] However, tensions related to validity, reliability and comparability must also be resolved.[15] There is significant potential for subjectivity in the assessment process of these competencies which can decrease confidence that external influences have been eliminated and an appropriate level of reliability has been achieved.

Competencies in any reformed curricula must be clearly conceptualized so that appropriate methods of assessment can be selected. However, a difficulty with assessment of these competencies is that some are too subjective and too ambiguous to be measured objectively. In these cases, it is sometimes possible to measure the product of the skill but not the skill itself. Suto and Eccles (2014) give the example of creativity: it 'involves every sense (sight, hearing, touch, smell and taste) and is almost infinite, it defies precise definition'. The point they make is that while a piece of art produced by a student may be readily assessable, the skills that form the creativity that helped produce it are not.

The following are methods of assessment are useful in assessing many of the higher order thinking skills and transversal competences that could be given more emphasis in reformed curricula. These would be used supplements to written tests:

i. Multiple choice tests

The assessment of higher order skills in written tests could be relatively simple. Multiple choice questions usually test recall and application of knowledge but can also be used to test higher order skills. Examples of multiple choice tests being used in this way include assessments for higher education entry such as the Bio-Medical Admissions Test and subject specific tests such as the Thinking Skills multiple choice tests used as part of the admissions process by the Universities of Cambridge and Oxford and University College London. However, although multiple choice questions have high levels of reliability and validity, writing them is a difficult task requiring significant training.

ii. Individual and group project

Project assessments take place over an extended period of time allowing meaningful problems and questions to be set that extend and challenge students' understanding. Problems can be open, complex and multifaceted. Examples of project work are a social science research investigation; developing an innovative and creative way of working; production of a piece of art etc. Project work allows for the possibility of students working outside the classroom,

---

[14] Suto, I. and Eccles, H. (2014). The Cambridge Approach to 21st Century skills: definitions, development and dilemmas for assessment. IAEA Conference, Singapore, 2014.
[15] Child, S. and Shaw, S. (2016). Collaboration in the 21st century: Implications for assessment. Research Matters: A Cambridge Assessment publication, 22, 17–22.

73

USAID funded program, Quality Instruction Towards Access and Basic education Improvement (QITABI), Tel/Fax +9611 983904/5/6
Downtown Beirut, Lebanon

extending the scope of what can be assessed, including ability to apply theoretical learning to real-world situations. Projects must be designed to measure the stated constructs, should be practical for the teacher to administer and should be sufficiently motivating for students.

Group projects offer the potential for assessment of skills such as collaboration and problem solving. The greatest validity for this type of assessment is achieved through teacher assessment, using rating scales and checklists, as it is the teacher who is best placed to observe the behaviours. Successful projects and group work that assess collaboration must contain the following features:

- The task must be sufficiently complex as overly simplistic or trivial tasks do not encourage group members to collaborate because there is little need to share.[16]
- The task should not have and obvious structure so that it cannot be solved by one capable group member. Tasks should be open with more than one plausible solution.[17]
- The placement of students in groups should encourage negotiation. It is important that students are placed in groups where there is likely to be a difference of opinion.
- The teacher should motivate students to work together. Meeting the criteria above should create an assessment that is motivating for students, but productivity can be improved when members are rewarded as a group, with the context of individual accountability.[18]

iii.   Authentic assessment

Authentic assessment is an assessment of students' skills, knowledge and understanding in a 'real world' context that can be used in addition to written assessments. It is often seen as a 'true test' of academic achievement or progress because it requires students to demonstrate deeper understanding and a range of higher order thinking skills. Authentic assessment requires learners to apply skills and competencies such as critical thinking, communication, collaboration and creativity in real-life global contexts.

Authentic assessment:

- focuses on deeper thinking and the application of skills using real or simulated situations, audiences and purposes
- values and assesses the process of learning and not just the outcome of learning
- uses flexible timescales dependent on the nature of the assessment and the students
- can involve collaboration with others and enables students to be resourceful before and during the assessment e.g. gathering information and/or resources.

---

[16] Brna, P. & Burton, M. (1997). The computer modelling of students collaborating in learning about energy. Journal of Computer Assisted Learning, 13, 193–204.

[17] Webb, N. M., Nemer, K. M., Chizhik, A.W., & Sugrue, B. (1998). Equity issues in collaborative group assessment: Group composition and performance. American Educational Research Journal, 35(4), 607–651.

[18] Bossert, S. T. (1988). Cooperative activities in the classroom. Review of Research in Education, 15, 225–250. Slavin, R.E. (1983). When does cooperative learning increase achievement? Psychological Bulletin, 94, 429–445

The GRASP model[19] is an approach that can be used to design authentic assessment tasks. The model uses the acronym 'GRASPS', which stands for:

- **G** – goal
- **R** – role or responsibility
- **A** – target audience
- **S** – situation or scenario (context)
- **P** – product or performance
- **S** – expected standards

These tasks should be planned and designed carefully to ensure that the depth and challenge are appropriate for all students within the group. Tasks are often multi-step and require collaboration and resourcefulness in gathering information that will support them.

This is an example of a primary GRASPS task:

You are a zookeeper at local zoo. You have been asked to write a caption that describes one of the animals. Your description will help visitors who come to the zoo to learn more about that animal.

This is how the teacher used GRASPS to design this task:

Goal: The students' goal is to write an informative and descriptive caption about an animal from the zoo.

Role: The student will be a zookeeper.

Audience: The audience of the writing will be people who visit the zoo.

Situation: The situation is simulated because the student needs to imagine that they are a zookeeper who knows a lot about different animals. The zookeeper has been asked to write an informative and descriptive caption about one of the animals in the zoo. They will need to find out important information about the animal and write it for a specific audience.

Product, Performance and Purpose: The student will write an informative and descriptive caption.

Standards and Criteria for Success: The writing will be assessed using a writing rubric devised by the teacher.

Authentic assessment can be used in a range of subjects. The following list provides examples of tasks:

Languages: you are a teacher. Your goal is to write a blog that will help and support teenagers across the globe with revising for their examinations and how to cope with examination stress.

Mathematics: you are the Accounts Manager for a large company. You need to reduce the amount the company is spending by 50per cent. Look at the budget provided by the CEO. Your challenge is to write a report to identify where savings can be made and the rationale for your decisions. Prepare to present your report to the company CEO.

---

19 Wiggins, G. & McTighe, J. (2005). Understanding by design. Expanded, 2nd ed. Alexandria, VA: Association for Supervision and Curriculum Development.

Science: you are an expert in nutrition and exercise. Your goal is to produce an information leaflet for the general public providing information about how they can eat a healthy and balanced diet and improve their fitness.

Geography: you work for an organisation that carries out recycling of materials. Your goal is to work as part of a team to create an eye-catching infographic about recycling in our country and how it compares with recycling in other countries. This will be shared with the public so it must appeal to both adults and children.

History: you work for a tourist information office. Your challenge is to work as a group to produce and record an interesting and exciting 10-minute podcast about the history of your local area in order to increase the number of foreign visitors and tourists.

iv. Practical science assessment

This type of assessment provides an opportunity to assess students' practical skills such as science laboratory experimental skills which are not easily assessed through written assessments. As well as providing an authentic method for assessing experimental skills, practical work can be motivating for students and assess their wider understanding of 'how science works'.[20] Science practical assessment can take the form of traditional laboratory work, investigations, projects and production of a portfolio of teacher-administered assessment.

As with all assessments it is important that teachers have a clear taxonomy setting out what will be assessed, such as capturing data, data analysis, production of diagrams, drawing (biology) experimental design, predicting outcomes and use of apparatus and techniques. Additionally, when designing science practical assessments it should be recognized that the more complex the assessment task the more complex they become administratively and teachers should consider this when designing assessments.

v. Oral assessment of languages

Assessment of oral performances typically include role plays, interviews, oral reports and summarising or paraphrasing a piece of text are assessed by teachers using rubrics and checklists. These alternative assessments can be used to supplement written assessment of reading and writing skills. Role plays can be used to assess students individually or as a group. Simulated role-play Interviews between teachers and students are a useful method of assessing early language learners as the assessor can provide a lot of visual clues such as asking the student questions about a picture.[21]

Student anxiety can have a significantly negative effect on student performance in oral assessments[22] and therefore it is important that teachers are aware of ways they can reduce anxiety as the assessor, thereby lessening the amount of construct-irrelevant variance

---

[20] Watts, A., (2013). The assessment of practical science: a literature review. Research Division of Cambridge Assessment.

[21] https://www.fluentu.com/blog/educator-english/assessing-english-language-learners/

[22] Young, D J., (1986). The relationship between anxiety and foreign language oral proficiency rating. Foreign Language Annals, Volume 19, Issue 5.

introduced to the assessment. Teachers can seek to reduce anxiety by, for example, having informal seating arrangements, positive and open body-language and maintaining eye contact as far as possible. The design of the task can also have a significant impact on oral performance with performance being more accurate and more fluent in structured tasks compared with less structured tasks.[23]

vi. Artistic performance assessment

The use of written assessment can be used to assess students' ability to carry out research but performance assessment should be used to assess artistic performance. Performance assessments include playing a piece of music, creating a piece of art and participating in a dramatic performance. While written assessment can be used to assess aspects of artistic performance, such as the ability to compose and notate a musical phrase, more practical assessments should be used to provide authentic assessment of students' mastery of discipline-based skills, [24] such as ability to perform a learned piece of music. In this case, authentic assessment refers to assessment that focuses on the observable evidence of what students know and can do. These assessments can be self-devised by students or text-based. Examples of discipline-based performance assessment for the visual arts could include drawing lines and textures to demonstrate understanding of the concepts or creating a work of art in a given medium, within set parameters and following clear instructions. Examples of discipline-based performance assessment for theatre could include group performance, improvisation and giving peer feedback. Portfolios can be used to bring together the written assessment tasks that are typically carried out over a period of time such as over the course of a semester, while discipline-based assessments can be carried out at set points over the same semester or can be assessed at the end of the semester. Evidence might include journals and logbooks, design portfolios for theatre, self-reflection, reviews of the performance of others, an exhibition of work etc.

Performance assessments are complex to administer consistently because of variation in performance sites, the requirement for group work, access to technical equipment and in the case of performing arts, the composition and reaction of any audience watching the performance. [25] Therefore, performance assessments should be feasible in terms of the available time and resources, must be amenable to reliable rating, assess both procedural and discipline-based knowledge and be of sufficient depth and breadth to allow valid generalisation about student performance.[26]

Assessment of performing arts requires the teacher or assessor to capture their thoughts about the quality of work as it occurs. During a performance the assessor is required to make judgments about the quality of work and to physically record their thoughts in relation to the assessment criteria balanced against their own criteria based on their personal experiences.[27] Training of the teacher to carry out this type of assessment is vital to ensure reliability.

[23] Tavakoli, P., (2009). Assessing L2 task performance: Understanding effects of task design. System, Volume 37, Issue 3.

[24] Armstrong, C. L. (1994). Designing assessment in art. Reston, VA: National Art Education Association.

[25] Oreck, B., Baum, S. (2004). Assessment of potential theatre arts talent in young people: The Development of a New Research-Based Assessment Process. Youth Theatre Journal. 18 (1).

[26] Don, C. M. (2003). Assessing art performance (MAAP): A K–12 project. Studies in Art Education, Volume 4, No. 4.

77

USAID funded program, Quality Instruction Towards Access and Basic education Improvement (QITABI), Tel/Fax +9611 983904/5/6
Downtown Beirut, Lebanon

vii. Fieldwork assessment

Fieldwork assessment lends itself to subjects such as geography, biology, economics, history etc. where assessment is more authentic if conducted in the natural environment rather than in semi-controlled environments such as a laboratory. Fieldwork provides a method of assessing students' understanding of the way scientific theories interact with real life. Common barriers to the use of fieldwork include issues of student security, difficulties with timetabling of other subjects and cost.[28] Because of these barriers fieldwork should be used to occasionally supplement written tests, not to replace them on a regular basis.

Fieldwork can be assessed through portfolios of work compiled over a period of time, possibly a semester or year. This extended time period means that portfolios are not strictly a summative assessment, although submission of evidence for assessment can take place at the end of the semester. The content of a portfolio of evidence can be varied, including reflective journals, written assignments, teacher observation records etc. Points for gathering evidence during fieldwork can be divided into research and planning for the fieldwork in the classroom, tasks to be carried out during fieldwork such as carrying out observations, recording data, and the summative assessment or final assignment. The final assignment could take the form of a reflective research report or an oral presentation.

viii. Investigative assessment in mathematics

Investigative and problem-solving assessments in mathematics aim to assess students' ability to work through mathematical processes rather than their acquisition of facts. Investigative assessments aim to go beyond students' knowledge of mathematical arithmetic skills to assess their ability to use mathematics in different social settings. Assessments are based around real- life or lifelike scenarios with open-ended problems. The assessments should provide opportunities for students to use multiple pathways to investigate the situation or problem. Mathematical investigative assessments may be posed as a problem to be solved, a question to be answered, a significant task to be completed or an issue to be explored[29] and substitute emphasis on routine technical assessment with more challenging and complex tasks.

Mathematical investigation can be built into existing written mathematics end of semester assessments where appropriate. A basic investigative assessment would be an open-ended question requiring thinking of multiple solutions or finding the same solution using different pathways. Increasing complexity increases the demand of the assessment as does increasing the range of skills, concepts and the connection of concepts. These more challenging

[27] Baptise, L. (2007). Managing subjectivity in arts assessments. In: Quamina-Aiyejina, L. (ed.) Reconceptualising the Agenda for Education in the Caribbean. 1st ed. St. Augustine: School of Education: University of West Indies. 28 Rogers, E., (2011). Fieldwork: Assessment in outside-class environment. 10.13140/RG.2.2.17366.86081. 29 Queensland Studies Authority (2005). Mathematics support materials: Thinking, reasoning and working mathematically in the classroom.

investigations should involve students in using higher order skills such as evaluation, creation and justification as set out in Krathwohl's revision of Bloom's Taxonomy.[28]

It is important that teachers receive training on how to develop students' writing skills in mathematics to ensure that students are able to communicate their mathematical ideas adequately. This teaching should take place before investigative assessments are introduced into the programme of end of semester summative assessment.

5.2.3 Arrangements for the training programmes

These training programmes should be based on a common understanding of the value and the status of the assessment that takes place in classrooms – both formative and summative. CRDP will take overall responsibility for the training, but there must be agreement by all the stakeholders on the key features as follows:

i) Schedule of training

In the long-term all teachers should receive assessment literacy training, but this would be a challenging task to perform in one phase. Priorities should be made for training so that training could be organised in phases. Groups could be based around different priorities, for example, primary level teachers; teachers of subjects where there is most need to supplement the end of semester assessments with alternative assessments; subjects where the need for reliable assessment data is priority such as mathematics.

ii) Fundamental concepts

CRDP, MEHE-GDE, Lebanese University and DOPS should agree on:

- a definition of assessment literacy
- a glossary of assessment terminology
- the methods of assessment that might be used to assess the reformed curricula access arrangements for students with SEND that should be available in schools

proposed definition of assessment literacy

Teacher assessment literacy, in its broadest sense, goes beyond the traditional focus on teachers' perceptions and their technical knowledge and skills in assessment with its emphasis on psychometric principles and test design. The Classroom Assessment Standards for PreK–12 Teachers, released in 2015 by the Joint Committee for Standards on Educational Evaluation (JCSEE) reflect a more modern view of assessment literacy where teachers exercise 'the professional judgment required for fair equitable classroom formative, benchmark and summative assessments for all students'.[29] The guidelines released by the JCSEE are divided into three key assessment processes:

- Foundations: guidelines related to the assessment purposes, design and preparation

---

[28] Krathwohl. D. (2002). A revision of Bloom's taxonomy: An overview. Theory into practice, 41 (4).

[29] Klinger, D. A., McDivitt, P. R., Howard, B. B., Munoz, M. A., Rogers, W. T., and Wylie, E. C. (2015). The Classroom Assessment Standards for PreK-12 Teachers. Kindle Direct Press.

- Use: guidelines related to examining students' work, providing feedback and reporting
- Quality: guidelines related to fairness, diversity, bias and reflection.

For the purposes of producing end of semester summative assessments to measure student achievement of the skills contained in a new or reformed curriculum Teachers should have a detailed and well-grounded knowledge of assessment instruments and how to construct them, including:

- A good understanding of the ways that assessment, pedagogy and learning are connected and link to the concept of validity
- The ability to design appropriate assessment instruments to effectively assess the skills and knowledge contained in the new curriculum reliably, validly and efficiently. Instruments should include alternatives to timed written tests, extending to instruments designed to assess students' practical skills and 21st Century skills as required, avoiding factors that are likely to have a negative effect on the validity of the assessment
- The ability to construct and adapt assessment instruments to ensure that students are treated fairly during the assessment process. This should include designing assessment instruments that are equally accessible to all students regardless of their national, cultural or social background or any special educational needs or disabilities (SEND), removing barriers to assessment without assisting students with the knowledge, understanding or skills being assessed.

Teachers will be able to interpret the results of the end of semester summative assessments effectively, including:

- Making informed inferences about the performance of students while understanding the limitations of the information contained in the result of an assessment. This includes limitations relating to measurement error or aspects of the curriculum the assessment does not or cannot assess
- Knowing the purpose(s) for which the outcomes of the assessment will be used and the possible effects of the results obtained. The ability to balance the information obtained from an assessment with other evidence about the performance of a student is also needed
- Being able to provide constructive feedback to students and to use that feedback to help students improve their knowledge and skills.

iii) Teacher guidance materials

CRDP should review and revise current teacher guidance materials as appropriate, wherever possible. The end result should be a pack of guidance materials which are suitable for application immediately but which indicate the broader approaches and tools that should be applied as the curriculum evolves to include a stronger emphasis upon transversal competencies including collaboration, project-based learning and problem solving. Examples of materials include a glossary of terminology, centralised protocols for the assessment process, and information about the full range of methods available for use in the end of semester classroom assessments. These materials should be developed by assessment professionals.

CRDP should write subject specific guidance materials as required to supplement existing materials. These should include assessment specifications, sample assessment materials and other subject specific materials to assist with developing assessments related to the new competencies in the curricula. These materials should be developed by subject expert assessment professionals.

Teachers should be provided with guidance on how to accommodate students with SEND in the end of semester summative assessments. This information should contain guidance on accommodations that could be made both before and during the examinations, how to assess students' needs for accommodations and examples of modified materials.

Centralised protocols and expectations for the assessment process to support teachers with assessment design, marking, moderation and harmonization.

It is important that teachers see assessment as an ongoing process rather than a one-off event. Improving assessment literacy will enable teachers to understand all parts of the assessment process and why they are important, both individually and when taken together. This will raise expectations and provide a clear framework for effective practice for all teachers.

Key parts of the assessment process that require protocols include assessment design, marking, moderation and harmonization. Teachers need to understand how to carry out each of these processes rigorously and consistently so that assessment is transparent, valid and reliable. Establishing these common ways of working will ensure that important aspects are not missed out.

Clear assessment specifications and sample assessment materials for end of semester tests across grades and subjects.

In order to produce high quality and consistent assessments for end of semester tests, it is important that teachers receive clear and unambiguous guidance about how to construct them. Clear written guidance on how to produce classroom assessments is currently only available for assessments in grades 9 and 12. This guidance should be extended to all teachers and should be in two parts:

1) A test specification for each subject, in each grade, including:

- the subject content to be sampled in an assessment
- the skills and competences being assessed and how many marks should be allocated to each one
- guidance on the different levels of performance that can be expected of students
- information about the different types and formats of questions that could be included.

2) Sample assessment materials

SAMS provide model question papers and mark schemes that exemplify best practice. Making these available to teachers will raise expectations and help them to understand what the process of effective assessment looks like in practice. In this way, creating a centralised bank of sample assessment materials across phases, subjects and sectors will support the development of assessment literacy in Lebanon.

The SAMS should include:

- assessment guidance for teachers linked to phase or subject
- sample assessments and questions
- sample assessment criteria and mark schemes
- assessed examples of student responses at different levels with a commentary
- a report from the assessor identifying strengths, common errors and key misconceptions from the student responses.

The use of SAMS would support key assessment processes such as designing tests and mark schemes. They would also enable teachers to moderate and harmonize their judgments and would demonstrate how teachers can begin to use summative assessments in a formative way.

Subject specific materials to assist with developing and implementing valid and reliable assessments in subjects that have a practical element such as science, art and languages.

Currently, assessment of the practical elements of some subjects are excluded from summative assessments in Lebanon. This means that important curriculum content such as oral skills in languages, practical skills in science and portfolio or project work in subjects like art are often not assessed. As a result, little attention is given to the development of the practical aspects of subjects at a classroom level. This in turn means there is a significant gap between the intended curriculum and the taught curriculum.

Improving assessment literacy in this area would build teachers' assessment experience and expertise in addressing the practical elements of subjects. Teachers would know what to assess and how to assess it. This would have a positive impact by allowing them to make a secure, holistic and balanced assessment of students' achievement across the whole curriculum thereby strengthening the links between the curriculum, assessment and learning.

Materials to assist with moving beyond knowledge and fact-based assessments towards assessments that incorporate higher order and 21st century skills.
Currently in Lebanon there is a strong curriculum emphasis on knowledge acquisition and memorization of facts in most subjects. As a result, teacher assessment and summative assessment often focus heavily on testing knowledge and recall. Assessment of skills, competences and higher order thinking is limited.

Improving assessment literacy in this area will develop teachers' understanding of how to assess skills, competences and higher order thinking in a transparent, valid and reliable way. Although

'higher order' command words such as 'analyse' and 'evaluate' are often in assessments, the nature and scope of assessing skills and competences needs further development. If the curriculum is to develop 21st century skills such as critical thinking, problem solving, creativity, innovation, collaboration and communication, then assessment needs to adapt to reflect this as indicated in the previous section on alternative and authentic assessment. An improved balance between assessing knowledge, skills and competences will result in deeper learning, the development of transferable skills and improved personal and academic outcomes for students.

iv) The model of training

The long-term aim is that the assessment literacy training programme should be available to all teachers at all grades. While different approaches can be taken to deliver this training, Cambridge International recommends that a cascade system is used.
Hayes (2000) [30] reflected on the challenges of effective cascade training for teachers and described the following five criteria for success:

- The method of conducting the training must be experiential and reflective
- The training must be open to reinterpretation; rigid adherence to prescribed ways of working should not be expected
- Expertise must be diffused through the system as widely as possible, not concentrated at the top
- A cross-section of stakeholders must be involved in the preparation of training materials
- Decentralisation of responsibilities within the cascade structure is desirable.

Cambridge International recommends that the cascade assessment literacy training for teachers should be divided into two phases as shown in the figure below

- Phase 1 equips assessment specialists with the required knowledge and understanding of the knowledge and skills they will cover when training the teachers in Phase 2.
- Phase 1 training should be carried out by master trainers who are a mixture of subject/ assessment experts and SEND/assessment experts
- Phase 2 equips teachers with the knowledge and skills they need to deliver reliable end of semester assessments taking account of the needs of students with SEND. Phase 2 training should be carried out by the assessment specialists trained in Phase 1.

---

[30] Hayes, D. (2000). Cascade training and teachers' professional development. ELT Journal Volume 54 (2), Oxford University Press, Oxford.

Master trainers (subject and assessment experts, SEND specialists)

Phase 1
Assessment specialist and SEND specialist trainers

Phase 2
Teachers
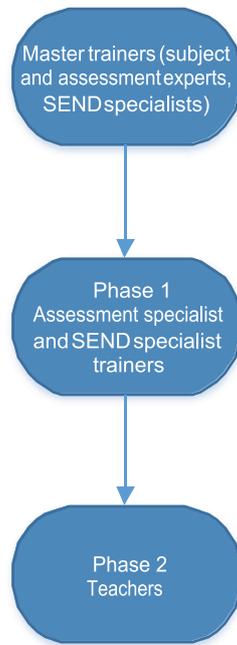
A Model for Cascade Training

Making training experiential and reflective

Wei et al. (2009) [31] conducted a study into teacher development and concluded that 'professional development is most effective when teachers engage actively in instructional inquiry in the context of collaborative professional communities, focused on instructional improvement and student achievement'. The training should be designed and the trainers briefed to ensure that the training is relevant to these communities and delivered in an active and engaging manner. The trainers or teachers will form a professional community for the duration of the training and will take part in many collaborative activities (including group and pair work, microteaching and role play).

All trainers and teachers should engage with the training and they should be encouraged to reflect on the concepts and approaches described and consider the implications for their assessment practice. Teachers should compare the approaches with their existing end of semester tests, discuss the effectiveness of the different approaches to assessment, and plan how to produce similar assessments in the future end of semester tests they produce.

Empowering the Phase 1 trainers

Phase 1 trainers will be provided with objectives and materials to train teachers in Phase 2. They will also be made responsible for delivering high quality training and making sure that all the objectives of the training are met.
The details of how the training is delivered should not be fully prescribed (Hayes, 2000). Phase 1 trainers will therefore need to adapt or reinterpret the details of the training delivery to meet

---

[31] Wei, R. C. et al. (2009). Professional Learning in the Learning Profession. A Status Report on Teacher Development in the U.S. and Abroad. Technical Report. National staff development council, Dallas.

the needs of the teachers within each course who will have different prior experiences of teaching and classroom assessment.

Identifying and training the Phase 1 trainers

An adequate number of Phase 1 trainers should be recruited to carry out the training for each subject and grade dependent on the number of teachers requiring training in Phase 2 and according to the agreed schedule. All Phase 1 trainers should be assessment and subject specialists and should have experience of teaching.

Training should use a mixture of written materials, face-to-face workshops and practical online activities (for example in a virtual learning environment). Online activities should be monitored to ensure participants are actively engaged with the course. To be effective, face-to-face training workshops should not usually have more than 25 trainees attending.

Trainers should experience activities they will deliver to teachers and should produce their own alternative assessment / adapted assessment materials as part of their training. They will reflect on the activities and materials they have produced and be given the opportunity to trial and refine their materials before they train the teachers.

Follow-up and support

As with Phase 1 training, delivery of training in Phase 2 could include written materials, face-to- face workshops and online training. Additionally, follow-up support and mentoring, by the trainers should be provided to teachers as they produce and use end of semester tests they have developed. This support could be provided externally or in schools. The maximum number of participants at face-to-face workshops should not usually exceed 25.

5.2.4 Framework for monitoring and evaluating the training programme

Cambridge International recommends that a monitoring and evaluation framework is developed to coordinate all the training activities; a monitoring and evaluation framework should include:

i. The purposes of monitoring

To assess whether or not the assessment literacy training programme is having the desired impact on the quality of the end of semester summative assessments.

ii. Key questions to be answered

Examples of key questions for the monitoring and evaluation programme are:

- Do teachers have a common understanding of the meaning of assessment literacy and its importance?
- Do the end of semester summative assessments make use of alternative and complementary methods of assessment to measure student achievement of 21st century and higher order skills?
- To what extent are assessments being adapted for students with SEND?

iii. Sources of evidence to be used

This will depend on the key questions but could include questionnaires for teachers, evaluation of the trainees by the trainers and sampling of end of semester summative assessments.

iv. Methods of collecting the evidence

These could include establishing a baseline before the training programme to measure performance against the key questions, and then using this baseline information to monitor the end of semester summative assessments produced after training is carried out. It is important that this section of the framework sets out who will carry out the monitoring activities and the percentage of teachers whose end of semester summative assessments will be sampled at the baseline and in each round of monitoring.

v. How the results will be reported

The data collected should be focused on establishing how far the training programme has met its stated aims. Analysis of this data should answer the key monitoring and evaluation questions and provide recommendations for improvement of the training programme in future rounds.

5.3 CPD for national examinations

5.3.1 CPD for examination setting

Experts at CRDP and MEHE have considerable experience in providing examination specifications to enable colleagues in MEHE to set examinations to meet the requirements of the specification. The experts from Cambridge International saw, at first hand, some of the processes taking place, additionally, they analyzed a great deal of documentation related to the national examinations. Accordingly, they have made the following recommendations which are aimed at improving the processes so as to maintain the essential underlying principles that the examinations should be fair, accurate, valid and accessible to all students.

- The timeframe for writing questions and mark schemes, constructing question papers and translating them should be carefully scheduled, allowing ample time for each stage, to support the reliability of the processes and avoid the introduction of errors. While ensuring the security of the assessments is paramount, the current practice of completing test construction, typesetting and proofreading on a single night does not leave sufficient time for all necessary quality checks. It is recommended that a minimum of two days is given to these processes
- Specific checklists should be produced for every stage and for every role (from question writing to final question paper approval). This will ensure the requirements of the tasks at each stage have been met. These should be kept for audit purposes
- Specification grids should be used at the question paper construction stage to record details of the questions chosen. This will help monitor the balance of marks across competencies to ensure they meet the question paper specifications
- Content coverage grids should be used to ensure adequate year-on-year coverage of the curriculum
- The accessibility of question papers would be improved by consistent formatting including the use of templates. A style guide or formatting guidelines document should be produced to ensure consistency in the layout and formatting of question papers and mark schemes across subjects. It would provide guidelines on what information to include on front and back covers, and specify standardised requirements for features such as fonts, font sizes, margins, use of bold, and how marks are displayed (for example down the right-hand margin).

i) Review question paper specifications and guidelines for question writers

There should be a consistent approach to developing question paper specifications across subjects. A precise test specification will improve the integrity and robustness of the assessment. It is best practice for these documents to contain key information such as a clear statement of purpose, information on the modes and structure of the assessment, and the balance of marks across assessment objectives or skill domains. The use of a standard template across subjects will ensure that the information included is consistent and comprehensive.

- Test specifications should be complemented by subject and grade-specific guidelines for question writers. These should include more detailed information on the most appropriate task types and question formats for testing the relevant skill domains, guidance on how to write them, examples of appropriate command words and instructions on writing reliable mark schemes.

ii) Train assessment professionals involved in the production of questions and the construction of question papers

The assessment professionals involved in different roles in the question production and question paper construction processes (described below) should receive further training to enable them to fulfil their roles more effectively. These are features of the training process:

a) Purpose of training
The purpose of the training would be to enable participants to consolidate and extend their knowledge and understanding of the skills and processes required for good assessment. In addition, the training will help identify participants who demonstrate the necessary skills to contribute, in different roles, to creating Brevet and Baccalaureate question papers.

b) Form of training
Training can be delivered in a combination of face-to-face instruction and practical online activities (for example in a virtual learning environment). Online activities would be monitored to ensure participants are actively engaged with the course.

c) Content of training
The content should cover both generic and subject-specific principles.
The **generic sessions** would develop participants' understanding of:
- fundamental assessment concepts such as validity and reliability
- the overall question writing and question paper construction process along with the importance of quality assurance checks
- the roles and responsibilities of different assessment professionals throughout the process
- the role of question paper specifications and specimen papers
- the relationship between the curriculum, assessment objectives, and task types/command words
- features of an effective mark scheme as a means to ensure accurate and reliable marking
- effective question paper construction
- other issues such as originality, copyright, security, typesetting and artwork.

d) **subject-specific sessions** applying the principles covered in the generic sessions to the individual subjects. In addition, they would provide participants with the opportunity to:

- reflect on the specific features and question types appropriate for their subjects
- write different question types and stimulus materials that are clear and effective
- produce mark schemes that promote accurate and reliable marking
- review and revise questions and mark schemes produced by other participants providing constructive comments
- revise their own questions and mark schemes in response to comments where appropriate
- develop their ability to judge and adjust the level of demand of questions and address higher order skills such as reasoning and evaluation where appropriate
- participate in discussions of the revised questions and mark schemes, including giving and receiving constructive feedback
- receive mentoring and feedback from the trainers on their work.

5.3.2 CPD for examination marking

The mark schemes (barème)

The main purpose of a mark scheme is to provide a framework for the allocation of marks to a student answer. Mark schemes should be easy for markers to use and enable marks to be consistently assigned to reflect student performance. A good mark scheme helps markers award marks accurately and improves inter-marker reliability.

The mark scheme should be written in sufficient detail to ensure that the marks allocated for a task are commensurate with its demand. This includes giving guidance to markers on how the different elements of an answer can be awarded partial credit (or part-marks). The Brevet and Baccalaureate mark schemes reviewed do not always provide sufficient guidance on how to award partial credit. For example, the mark scheme for the comprehension questions in the English language Baccalaureate gives model answers, some of them containing a number of ideas, but is not always consistent in indicating how students would be credited if they provided a partially correct answer.

The review also found that decisions on how to interpret and apply the mark scheme, including how partial credit should be awarded, are made immediately before the marking period starts (to produce the 'micro-barème'), rather than at the time of question writing. From an international perspective, it is more normal for mark schemes to be amended before marking begins. As has been noted earlier , as decisions related to the choice of questions for the examination papers are made so late and in such a highly-compressed timescale, the quality of the mark scheme and the validity of the assessment risks being compromised.

To ensure the reliability of marking, the characteristics of each question type should determine the best type of mark scheme to use for that question. Brevet and Baccalaureate mark schemes are primarily designed for objective questions (such as multiple-choice questions where the mark scheme gives a single correct answer) or points-based questions (such as short-answer questions where the mark scheme credits a specific number of relevant points).

Where extended-response or essay questions are included in question papers, there are two approaches to mark schemes:

- In the case of the writing tasks in language papers, the criteria are briefly given in the mark scheme. In the English language Baccalaureate, for example, these are content and organisation (3.5 marks), language and style (3.5 marks), tidiness and handwriting (1 mark). However, the mark scheme does not provide any guidance or further elaboration on how these criteria should be applied.
- For extended response questions in other papers, such as geography, economics and philosophy, the mark scheme gives a detailed outline of the expected answer and the marks to be awarded to each section or point given (including, for example, marks for writing a transition between the introduction and the main body of the essay).

Neither approach supports reliable marking: the first lacks guidance on how to award marks within each criterion; the second is too prescriptive and a student whose answer meets the demands of the question but does not follow the rigid structure given may not achieve full marks. Levels-based or 'banded' mark schemes are most suitable for questions that award multiple marks and where there may be a range of creditworthy responses, from those that show a partial or basic knowledge at the bottom to detailed and complete answers at the top. These mark schemes are divided into several bands, each describing a different level of performance and the corresponding mark (or mark range) it should be awarded.

Below is an example of a banded mark scheme for a history question worth 6 marks, where students have to provide a more developed response in order to reach the higher levels

Target: Demonstrating an understanding of historical explanation		Marks
Level 4	Explanation of at least two identified reasons.	6
Level 3	Explanation of one identified reason.	4–5
Level 2	Identifies and/or describes reasons. No valid explanations given for reasons.	2–3
Level 1	Valid general comment lacking specific subject knowledge.	1
Level 0	No evidence submitted or response does not answer the question.	0

Banded mark schemes will generally provide higher reliability than points-based mark schemes for questions with a high number of marks. They are most appropriate for questions that assess higher order skills (such as analysis and evaluation) or proficiency in a skill, (such as language writing). They are particularly appropriate where marking criteria are not specific to the content of the question but are generic and can therefore remain the same from year to year. This can help maintain comparable standards across examination series. These mark schemes are however normally supplemented by a list of 'indicative content' providing points that students are likely to include in their answers to that specific question.

Finally, Brevet and Baccalaureate mark schemes rarely specify whether, or which, alternative answers would also be acceptable, or which answers are likely to be given but are unacceptable.

Other issues in brief
Some further issues are summarised below:

- **Layout and formatting:** The review found some evidence of inconsistent formatting across question papers. Mark allocations are often displayed inconsistently: in some question papers they are shown in the margin, in others immediately after the question and often, they are not shown at all for sub-questions (notably in the mathematics and sciences Baccalaureate). This means that students sometimes are not provided with information that would help them prioritise questions and allocate their time accordingly. Furthermore, the cramped layout and minimal white space in many question papers may make it more difficult for students to follow complex sets of instructions or may cause them to miss questions. This is particularly the case in the mathematics and sciences question papers where, in addition, the sources (or 'documents') are wrapped by the text.
- **Ambiguous wording of texts, tasks or distractors:** The review found occasional instances where ambiguity of wording could cause confusion or elicit a legitimate answer that is not included in the mark scheme. In an English language Brevet question paper, for example, the ambiguity of a comprehension question is partly caused by conflicting information in the reading text. In the French language Brevet and Baccalaureate papers, some distractors in multiple-choice questions are arguably correct. One of the principles of multiple-choice questions is that they should have one unambiguously correct answer, while the distractors should be incorrect but plausible.
- **One written question paper per subject:** Brevet and Baccalaureate examinations comprise one written question paper per subject. The best international and more usual practice is to have more than one paper per subject. This mitigates the risk of the student having a 'bad day' and not performing to the best of their ability. The mark they receive would therefore not represent their true level of skill and ability. In addition, if different skills are to be assessed it may not be practical to use one assessment method for all, and a single question paper would not give a full measure of the student's level of skill in the subject.

The following recommendations are designed to improve the reliability and validity of the marking process:

i) **Reinforce the organisation of markers**

The current system of organising markers should continue to be used. However, to implement the recommendations that follow effectively, the lines of authority and responsibility between markers should be reinforced to strengthen the coordination and quality assurance of marking, both within each regional marking centre and across centers.

Ultimate responsibility for marking quality and ensuring marking is completed to schedule lies with the Chair of the Examinations Committee. In turn, each Supervisor is responsible for the work of markers in their team and they are accountable to the Chair of the Examinations Committee for this. The lines of communication run up and down the structure with little need for direct communication between the Chair of the Examinations Committee and markers.

ii) **Stability of marking teams**

The previous recommendations, as well as those that follow, will be more effective if marking teams are stable in terms of who is doing the marking and who is supervising them. Wherever possible markers should be placed with the same Supervisor in each examination series. Where there is a high turnover of markers, efforts should focus on retaining experienced markers.

iii) **Train all markers to interpret the live mark scheme each examination series**

Reliable marking depends on markers' ability to interpret and apply the mark scheme accurately. If this is not achieved, marks will not reflect a student's true score and inter-marker reliability will be low. Training all markers to apply the live mark scheme using students' answers from the current examination series to illustrate how it would work in practice will improve marking accuracy. Training should focus on achieving a shared understanding of the requirements of the mark scheme and what constitute creditworthy and non-creditworthy answers. One-off training for new markers or general marking training for all markers will not result in the same improvement in marking quality.

The following are features of an effective marker training process that incorporates and develops some of the processes already in use in Lebanon:

a) **Producing the mark scheme**
This should be done face-to-face by the Chair of the Examinations Committee and Supervisors from each regional centre working together.

b) **Selecting examples for the training materials**
Once the mark scheme is produced, the Chair of the Examinations Committee and Supervisors from each regional centre should select students' answers for use in training materials. Answers should be selected to illustrate:

- common and uncommon responses
- the line between what is creditworthy and non-creditworthy, including plausible but incorrect answers

- accurate and inaccurate understanding of the requirements of the question
- answers deserving credit across the full mark range.

iii) Form of training:

Training should take place in each regional marking centre. It should take the form of a face-to-face meeting using a combination of formal instruction and practical activities. The practical activities are vital to the process as they allow the trainer and the marker to check that the objectives of training have been met and that all markers understand how to interpret the mark scheme.

a) Trainers

Depending on the total number of markers and the size of the venue available, the Chair of the Examinations Committee can either train all of the markers together, including the Supervisors, or use a system of cascade training. In this case the Chair of the Examinations Committee trains the Supervisors and in turn they train the markers in their team. Training all markers together makes more efficient use of time but cascade training ensures that Supervisors fully understand how the mark scheme should be applied and improves the trainer to marker ratio.

b) Timing and duration

Marker training should take place just before the start of marking. This will help ensure key messages from the training are fresh in markers' minds and they retain their understanding of how to apply the mark scheme when they start marking. To maintain security, nobody should be given access to the question paper or mark scheme before the examination has been sat by students. In addition, to allow the principal markers sufficient time to prepare, training should generally be scheduled a few days after the mark scheme has been produced. The duration of training will depend on the complexity of the questions and mark scheme and the total number of marks for the question paper. However, a day is usually sufficient to train the Supervisors with a further day allowed for training markers.

c) Resources for training

These should be budgeted for carefully and administrative support should be available to help with organisation. Both material and time resources are required including:
- sufficient time for all parts of the process
- the availability of all markers at the right time
- sufficient copies of the marker training materials so that everyone has their own copy
- a suitably sized and equipped venue.

iv) *Harmonize markers using a practical marking exercise*

Training on how to apply the mark scheme should be followed directly by practical harmonization of markers. During practical harmonization markers mark a sample of students' answers to questions. After this their marks should be compared with the marks awarded earlier by the Chair of the Examinations Committee and Supervisors from each regional centre working together. To pass practical harmonization, the marks awarded by the marker must be within tolerance of the marks awarded by the Chair and Supervisors.

The purpose of practical harmonization is to improve inter-marker reliability by identifying:

- markers who can be approved to mark because they passed harmonization
- markers who require retraining or who should be withdrawn from the marking process because they are unable to pass harmonization and are not applying the mark scheme correctly.

Here, 'not applying the mark scheme correctly' refers to a marker showing undue leniency or severity when applying the mark scheme to a student's answer. This is particularly important for questions that require a degree of judgment over how many marks should be awarded. Objective questions worth one mark or where there is only one answer, do not usually require marker judgment. Open questions worth more than one mark and where extended answers are needed are more likely to require judgment. Some markers will be lenient, awarding more marks than a different marker would have, and some will be severe, awarding fewer marks. Practical harmonization improves inter-marker reliability by ensuring that the student would be awarded the same mark no matter who the marker is.

These are the features of an effective practical harmonization process:

a) Process of practical harmonization

Set the harmonization tolerance: a 'harmonization tolerance' is set for the question paper. The tolerance is the allowable difference between the marks awarded by a senior marker and those awarded by a marker. The use of a tolerance recognizes that where judgment is required, it is unlikely that a senior marker and a marker would make exactly the same judgment, but that they should be close. The size of the tolerance will depend on the type of questions in the question paper and therefore the amount of judgment required to mark students' answers.

Question papers with a large proportion of objective questions	Set a small tolerance
Question papers with a mix of objective questions and open questions	Set a larger tolerance
Question papers only comprised of open questions	Set the largest tolerance

- **Select students' answers:** the Chair of the Examinations Committee and Supervisors from each regional centre working together should select students' answers for the harmonization exercises and mark them together. They should select enough answers to create two separate harmonization exercises. They can do this when they produce the mark scheme (microbarème).
- **Complete the harmonization exercise:** markers should complete the harmonization exercise independently of each other. When a marker completes the exercise, the Supervisor should review the marking identifying any answers where the marker is outside the tolerance. The Supervisor then decides whether to approve the marker. If they are unsure they can ask the marker to complete the second harmonization exercise. After reviewing this second exercise the Supervisor should decide whether to approve the marker or remove them from the team of markers.

b) Criteria for selection of students' answers for the practical harmonization exercises

Use similar criteria to those used to select answers for training. The number of answers selected for each question depends on the degree of judgment required to mark the question. For open-ended questions requiring a lot of judgment 10 answers to each question might be sufficient.

c) Resources for harmonization

These are similar to those required for training. Care should be taken that sufficient numbers of markers are available to allow marking to be completed on schedule regardless of whether any of them are not approved to mark.

v) *Increase monitoring of markers at the start of marking*

Early monitoring is vital because markers will be inexperienced in applying the mark scheme and some will make enough errors to classify them as outside tolerance or 'divergent'. Early monitoring relies on having a sample of each marker's answer papers go through the double marking process as quickly as possible after they start marking. In practice, this may mean that double marking focuses on a marker as soon as they have been approved to start marking and once early monitoring indicates they are within general tolerance, the focus switches to the next new marker. This means that each marker will receive early monitoring and feedback, followed by a gap until they are next monitored.

vi) *Monitor markers during live marking using statistical data*

The system of double marking provides opportunities to give feedback and training for markers if they are divergent, as it allows the supervisors to be aware of who is marking well and who is not. Using statistical data for monitoring the outcomes of double marking would provide easily understood information for the Chair of the

Examinations Committee, Supervisors and other stakeholders who have an interest in monitoring general changes in marking quality. However, using statistical data in this way is reliant on improvements to the current data entry process as data cannot be generated until marks are recorded on the system.

a) Statistical monitoring using tolerance

Cambridge International recommends that divergent markers are identified by monitoring the percentage of each marker's marks that are significantly greater than agreed. In other words, the percentage of the marker's question papers that are referred to a third marker.

In the short-term the current tolerances could be used, but tolerances should be reviewed at the end of each examination to decide if there is capacity to a smaller tolerance for any subjects. Where the number of answer papers that require a third mark is reduced, the tolerance for the subject can also be reduced. This will provide consistency throughout the whole double marking process and a method of recognising that tolerances should vary according to the type of questions in the paper. If data entry systems permit, tolerances could be set at question level.

Markers should be classified as divergent if the percentage of their answer papers referred to a third marker is significantly greater than the mean percentage for all markers. This recognizes that it is very unlikely that none of a marker's answer papers will be outside tolerance, but that if the percentage rises above the average then the marker is not applying the mark scheme correctly.

b) Using monitoring data

Once a marker is identified as divergent they should be retrained in the application of the mark scheme, using answer papers that they have already marked as a basis for the training. If they continue to be identified as divergent once they have started marking again, following consultation with the Chair of the Examinations Committee, they should be removed from the marking team and not allowed to return to mark in following series.

Monitoring the total percentage of answer papers outside tolerance over examination series will provide an indicator of how effective practices for improving the quality of marking are. This information can be used to identify subjects for investigation because they are difficult to mark reliably.

Comparisons between regional centers can be made using these data, although it is important to take care that pressure to reduce the number of divergent markers does not lead to a relaxation of marking standards to avoid markers being outside tolerance.

c) Improve data entry

Monitoring marking in real time is reliant on timely and accurate entry of students' marks into the data management system. Traditionally entry is

carried out by clerical staff who transcribe the marks from the students' answer papers into the data system. Even with a sophisticated system of checks this can lead to transcription errors with students being awarded the right mark on their answer paper and the wrong mark in the data management system.

Currently a process where markers enter marks directly into the system is being piloted in Lebanon. If successful, this process will reduce transcription errors, ensure that marking data are available immediately for monitoring purposes and capture data at question level.

However, this system will incur additional costs. Instead of using clerical staff time, markers will have to be paid to carry out data entry. The process of entry will slow markers down and it is likely that the number of answer booklets allocated to them would need to be reduced, leading to a requirement for more markers and therefore more Supervisors. Using more markers will also lead to greater time costs in terms of carrying out practical harmonisation and monitoring markers. Increasing the number of markers also increases the risk of lower inter-marker reliability.

vii) *Revise the tolerances in the double marking system*

Currently two different tolerances are used to trigger remarking of answer booklets. One tolerance triggers remarking by an auditor, the other triggers remarking by the Chair of the Examinations Committee. This lacks consistency. In addition, the size of both tolerances is quite large. A larger tolerance implicitly accepts that inter-marker reliability will be low and that the mark awarded to a student will, to some extent, depend on who has marked their answer booklet.

The recommendation for practical harmonisation sets out criteria for smaller harmonisation tolerances tailored to the type of questions in the paper so that tolerances will vary across subjects. Using a similar system of tolerance for double marking would improve inter-marker reliability by triggering the remarking of answer booklets when there are smaller discrepancies between the marks awarded by the markers and/ or auditors. However, reducing the size of the tolerance should be done in stages, as outlined in statistical monitoring using tolerance above. Additionally, reducing the size of the tolerance before implementing the other recommendations in this section will lead to more remarking than is practical. Taking care when deciding how much to reduce the tolerance and only reducing the tolerance once the recommendations are effective will make the change more manageable.

As electronic marks data become available at question level setting tolerances at question level, rather than question paper level could be investigated. This would make more efficient use of time, particularly on question papers where there is a mixture of objective questions and open-ended questions.

6. Communication strategy

6.1 Introduction

The NSLAF will address every form of student assessment in the primary and secondary school sectors from monitoring primary readiness, through ongoing teacher assessment to national examinations and international surveys. The NSLAF reforms are ambitious and wide ranging and the stakeholders (parents, teachers, assessment professionals and the wider educational community) will have widely differing experiences and understanding of assessment systems and the processes that drive them. To build confidence in the NSLAF and make it understandable to such a diverse audience, a coherent and stratified communications and engagement strategy is needed. Below are the principles, processes and interventions needed to achieve this.

i.   Principles
The core principles governing the communications are:

- It is an iterative and ongoing process that relies on partnership and shared ownership
- It is jointly owned, with World Learning, MEHE-GDE and CRDP acting as active engaged participants
- MEHE has strategic oversight of the communications plan and its deliverables.
- The strategy is about communicating change AND what is not changing
- The strategy provides clear messages about assessment
- From its inception, MEHE-GDE/CRDP will lead coordination and implementation of the communications and engagement and implementation plans
- It incorporates a system to review the communications implementation to support the sustainability of implementation.

ii.   Assumptions
The communications strategy is based on the following assumptions gathered from stakeholder consultations, the Assessment Dialogue workshops and a seminar on communications and engagement. The key assumptions are:

- **Knowledge:** understanding of the NSLAF and its development is limited amongst key audience groups. Improved knowledge and understanding of assessment, of the NSLAF and its principles is essential to successful implementation
- **Application:** the NSLAF applies across all school grades and in both the private and public sectors, so securing engagement across all target groups is vital
- **Ownership:** the communications strategy is a shared responsibility between key educational stakeholders especially MEHE-GDE and CRDP
- **Implementation:** the strategy will be implemented by MEHE-GDE and CRDP as lead participants
- **Interventions:** the strategy should be reviewed during implementation to assess feedback, ensure alignment with stakeholder needs and resonance with key audience groups.

- **Resourcing:** staff time, funding and opportunity have to be identified to ensure effective implementation. This is the responsibility of MEHE-GDE.
- **Activity plan and matrix:** the implementation plan requires an activity matrix to ensure that all key audience groups and districts have been engaged.
- **50 percent engagement:** at least 50per cent of core audience groups require direct engagement with and understanding of the NSLAF in order to create a tipping point for understanding of the framework.
- **Coherence:** the implementation of the communications strategy should be aligned with the NSLAF workplan.

The strategic oversight of the communication strategy is the responsibility of MEHE who are responsible for budget planning, quality assurance and ensuring the plans are delivered through partner agencies such as GDE and CRDP.

6.2 Aims and outcomes

The aims of the communications strategy are to:

- provide an integrated communications strategy to support the implementation of the NSLAF
- secure widespread engagement and understanding of the NSLAF
- achieve a minimum baseline of 70per cent understanding and knowledge of the NSLAF
- support the delivery of the NSLAF.

To be successfully implemented the communications strategy needs to acknowledge the diversity of audiences, limited assessment literacy and complex stakeholder relationships. To build consensus for, and engagement with, the NSLAF, the strategy should focus on the following key outcomes:

- Outcome 1: *Knowledge* – ensuring that audiences are aware of the NSLAF, its context, its outcomes, why it matters and how it will improve educational outcomes
- Outcome 2: *Awareness* – building understanding and awareness of the need for assessment reform amongst key stakeholders
- Outcome 3: *Advocacy* – building audience participation and advocacy for assessment reform
- Outcome 4: *Shared understanding* – securing a common understanding of the NSLAF, its role and impact for students and schools
- Outcome 5: *Materials and delivery* – use of targeted materials and delivery channels, including social media, for communication
- Outcome 6: *Metrics* – agreed metrics on audience penetration, engagement and understanding.

Stakeholders and audiences
The key stakeholders and audiences for the communications strategy are identified below.

Segment	Need	Delivery channel
National examinations administration	Updated handbook on national examinations, the NSLAF and impact	Newsletter, updated handbook, F2F policy briefing and monthly reports
MEHE-GDE /CRDP/inspectorate **Educational administrators**	Briefing paper on recommendations and impact Briefing paper on timelines and implementation	Information booklet on NSLAF, monthly update and F2F policy briefing Information booklet on NSLAF, monthly update and policy briefing
Elementary schoolteachers **Intermediate schoolteachers**	Information on purpose, impact and timelines Information on purpose, impact and timelines	Information booklet on NSLAF, monthly update and school briefing pack Information booklet on NSLAF, monthly update and school briefing pack
Secondary schoolteachers	Information on purpose, impact and timelines	Information booklet on NSLAF, monthly update and school briefing pack
Special needs teachers	Information on special needs considerations	SEND information booklet on the NSLAF, school briefing pack and monthly update

School principals and administrators	Information on administration, timescales and accountabilities	Newsletter, updated handbook, F2F policy briefing and monthly reports
Parents – primary sector	Information on what the NSLAF is, timelines and impact	Information booklet on NSLAF, monthly update and F2F policy briefing Information booklet on NSLAF, monthly update and policy briefing
Parents – secondary sector **Higher education stakeholders**	Information on what the NSLAF is, timelines and impact Information on what the NSLAF is, timelines and impact	Information pack and briefing by school, access to MEHE-GDE website with NSLAF Q and A Information pack on NSLAF, Q and A, and timelines
Education donors **Media**	Information on what the NSLAF is, timelines and impact Purpose and context of NSLAF and impact	Information pack on NSLAF, Q and A, and timelines Information pack on NSLAF, policy briefing and media updates

Interventions

The interventions in the communications strategy for the NSLAF consist of three levels:

- Level 1: Knowledge exchange – the basic details of the NSLAF and who it affects
- Level 2: Planning and implementation – the implementation of the NSLAF and timelines
- Level 3: Policy alignment and educational outcomes – the educational benefits of the NSLAF.

The section below outlines the likely interventions that will be needed in any activity plan:

Interventions	L1: Knowledge exchange	L2: Planning – implementation	L3: Policy and outcomes
Face-to-face (F2F) briefings	MEHE-GDE/CRDP staff	MEHE-GDE/CRDP staff School principals	
Information packs – schools	School principals Educational stakeholders School generic information packs	Educational stakeholders	School information packs
Briefing notes – teachers	Targeted information sheets by subject	Targeted information and updates by subject specialism	School information packs
Briefing packs – SEND Guide for parents	Targeted information packs for SEND teachers Short information brochure for parents	Short information brochure for parents	SEND packs Information brochure for parents
Media packs & briefings	Media pack on NSLAF targeted at, for example, journalists and NGOs		Impact of NSLAF and policy
Frequently asked question (FAQ) guides – user guides	FAQs to capture key issues and answers	FAQs to capture key issues and answers	

Online web pack	Online materials on NSLAF – open access	Online updates on implementation plan and next stages	Online materials
Stakeholder conference			NSLAF and its impact – F2F briefing and update
Seminars – MEHEGDE/CRDP		Review of planning and implementation of NSLAF and implementation	
School leaders' F2F briefing	Regional meeting of school leaders for F2F briefing		
School leader updates	Targeted materials from MEHE for school leaders	Targeted materials from MEHE-GDE on implementation	
School workshops	School-based events led by principals on NSLAF		School-based events led by principals using MEHE-GDE materials
Regional workshops	Teacher cluster groups by subject to outline NSLAF	Teacher cluster groups by subject on implementation	

The final shape of the activity will be confirmed as soon as the NSLAF is officially accepted and adopted.

Delivery channels

The evidence from dialogues with stakeholders is that a blended approach to communications will be most effective using a variety of delivery channels. This approach should include the following:

- **Passive Information:** information packs, briefing updates and contextual information
- **Personalised briefings:** targeted briefings and workshops for F2F dialogue and exchange
- **Group briefings:** workshops and group meetings to explain developments and share knowledge
- **Action information:** information requiring a response or action (timelines and action plans)
- **Online web presence:** static web presence providing access to all key information
- **Online and social media interactive:** including presence on Twitter, Instagram and other social media sites that support active engagement
- **Media:** radio and television broadcasts and updates to share information on reform and its import.

The balance of these will be dependent on the activity plan and the resources available.

Metrics

There are three levels to assess the successful implementation of the communications strategy of the NSLAF. These are:

- **Core:** this is the base level where most people (over 50per cent) have heard about the NSLAF and accept its changes
- **Engaged:** this is the tipping point where most education stakeholders have moved from acceptance to engagement with the NSLAF and its implementation
- **Advocacy:** this represents the highest level of engagement with the NSLAF where stakeholders are now advocates for implementation creating a multiplier effect. The specific dimensions of this are contained in the table overleaf:

Core	Engaged	Advocacy
• All teachers, school leaders and administrators are aware of the NSLAF • 50% of teachers accept • the changes in the • NSLAF 50% of school leaders feel that they are prepared for • changes within the • NSLAF 50% of special needs teachers are aware of the changes • and the impact they will • have on their work • At least 50% of parents have heard about the changes	• 70% of teachers, school leaders and administrators are engaged with the communications strategy and NSLAF • 70% of parents are aware and understand the changes proposed • 70% of special needs teachers are engaged with the NSLAF • 65% of parents endorse the changes proposed • 80% of public and private school leaders endorse the change	• 80% of teachers and school leaders have a clear understanding of reform and what it means • 80% of special needs teachers are aware and engaged with the NSLAF preparation • 70% of schools are aware and understand the changes proposed • Classroom pedagogy demonstrates early signs of positive engagement • Service departments actively collaborate and reinforce benefits from the NSLAF
• General acceptance of change • General understanding of reforms • Service agencies accept eform	• 80% of school leaders feel they are prepared for the changes and are positive about the • NSLAF 70% have a clear understanding of the drivers for reform Key • service agencies – MEHE-GDE, CRDP-PITB, • MEHE-GDE-DOPS, • CRDP, the inspectorate • – are aligned with the reform	• Parents, media and stakeholders talk positively about NSLAF • NSLAF is viewed as a • gateway to further development • Public and private • sector provision is balanced in advocacy • Stakeholders expect • pupil performance to improve

Timelines
An indicative timeline for the communications strategy would be as follows:

Timeline	Activity
Baseline	• NSLAF approved
Within 1 month	• Communications strategy, activity plan and resources agreed
Within 3 months	• National Seminar launches NSLAF and communications strategy • Website established and populated • Information packs prepared • Media preparations • First updates for schools and teachers
Within 4 months	• F2F briefings scheduled • School conferences scheduled • Information pack for schools approved for publication • Information brochure for parents approved for publication
Within 6 months	• Information packs and brochures distributed to schools and parents • FAQs online go live • Social media strategy fully engaged • First baseline metrics and review of engagement
6-month review	• Review and adjust the communications plan and strategy in the light of feedback

Budgets
The indicative costs of the communications strategy are contingent on the form the agreed plan. This are the responsibility of MEHE as the lead agency. The following chart provides an outline of indicative costs based on the following assumptions

- Staff time from MEHE and partner agencies is provided free of charge
- MEHE is able to use its purchasing power to provide value for money in material supply
- Materials are available for the state and private sector
- Materials are designed to be subject specific
- Not all materials are needed in the first phase
- The budget should be phased over a two-year implementation period and over a follow up period.

The total cost of an integrated campaign over two years would be $439,550 or $219,775 per year. This allocation is similar to the costs for other national campaigns in assessment. Costs are indicative and the final resource requirements would be subject to a procurement analysis once the assessment strategy has been agreed to ensure value for money and impact. There are opportunities to achieve savings and rationalisation within the final strategy.

Activity	Volime	Cost (US$)
Specialist consultancy for materials development	Dedicated consultancy support for material development for the communication strategy – 15 days at $750 per day	$11,250
Materials development	Development of a suite of materials for use by schools, parents and the media, including graphics	$50,000
Subject specific materials	Specialist materials for key subject areas. Specialist time and materials	$20,000
Face-to-face briefings	Venue hire where needed, travel expenses for participants, refreshments and materials support. Twenty meeting meetings @ $250 per meeting	$5,000
Information packs for schools	3200 packs for all school types. The pack includes 'how to' guides and basic information. Cost per pack $4	$12,800
Briefing notes for teachers	105,000 briefing notes for teachers. Cost per unit 0.50 cents	$52,500
Briefing packs for SEND	Specialist SEND packs for schools and SEND teachers. Allow for 1500 school packs and 1500 teacher sheets at unit cost of $4.	$12,000
Guide for parents	450,000 booklets for parents outlining the proposed assessment reforms @ 0.25 cents	$106,500
Media packs and briefings	Targeted media briefing packs 400 packs @ $10 per pack	$4,000
Frequently asked question (FAQ) guides – user guides	FAQ user guides 500,000 for later campaign usage @ 0.05 cents per guide	$25,000

Online web pack	Dedicated website for queries, FAQs and core information, with active management. Cost over two years $100,000	$75,000
Stakeholder **conference** **Seminars – MEHEGDE/CRDP**	Venue hire, payments for speakers, travel expenses, meals and logistics. 10 seminars and workshops – costs for venue, refreshments and some logistics @ $300 per event	$10,000 $3,000
School leaders' face-toface briefing **School leader updates**	Twenty face-to-face briefings for school leaders including travel costs, venue hire and refreshments @ $300 per event	$6,000
	Materials and web updates for school leaders including email messaging	$10,000
School workshops	School-based workshops with an additional briefing materials pack provided for each workshop 3200 packs @$ 10 per pack	$32,000
Regional workshops	Five regional workshops for educational administrators and stakeholders. Venue hire and some logistics only, Material already available. Cost per event $1000	$5,000
Total Costs		$439,550

6.3 The next stages

The key next stages for CRDP / MEHE-GDE and the implementation of the communications strategy are:

- agreement on NSLAF informs the communications plan
- CRDP / MEHE-GDE agrees the communications and engagement strategy leading to the activity plan
- CRDP / MEHE-GDE operationalises the activity plan aligned with the NSLAF.

It is also recommend that MEHE-GDE organises a national seminar on the NSLAF as a means of showcasing the assessment reform, alerting education stakeholders and using it as the initial platform for the communications and engagement strategy. An audience of 100+ would be invited to attend including:

- 30 school principals (public and private sector)
- 30 schoolteachers (public and private from primary and secondary)
- 10 special needs teachers
- 10 representatives from higher education.

It is important that these are drawn from across the country and include both urban and rural schools.

- five representatives from the donor community (USAID, DfID, British Council etc.)
- ten representatives from MEHE-GDE/CRDP
- two representatives from the inspectorate
- two media representatives.

The core aims of the seminar would be to:

- share the vision and structure of NSLAF
- commence the national dialogue on assessment reform
- align key educational stakeholders with the drivers for reform
- model partnership and engagement in reform
- create a multiplier impact through participants
- initiate the communications strategy.

An indicative structure for the seminar could be:

Time	Activity	Who
13.15 13.30	Arrival Context setting by MEHE-GDE	Education Minister or DG Schools
13.30–13.50	NSLAF and the Lebanese context • Vision for NSLAF • Why now and impact Presentation	MEHE-GDE /CRDP
14.00–14.20	Assessment Reform and Best Practice – International Lessons presentation	World Learning or Cambridge International expert
14.20–14.45	Break	

14.45–15.30	Panel discussion and open questions: Four panel members review the NSLAF and respond to audience questions	Chaired by Director of Lebanese University with one international participant, one each from CRDP, LU and MEHE-GDE
15.20–16.00	Refreshments and informal networking	
16.00–1800	Break	
18.00–20.00	Networking event with key stakeholders hosted by MEHE-GDE with additional invited guests	

7. Implementation of the NSLAF

This section provides proposed models and outline process maps for the implantation of the NSLAF

7.1 Initial screening

The time taken to introduce the grade 1 developmental assessment screening programme will depend on the source of assessment materials available and the availability of suitable software for assessment delivery, data capture and management and analysis. Communication with all stakeholders is a vital part of capacity building in order to ensure that the programme is successful. This capacity development plan sets out the phases of work that should be considered when planning to implement the programme.

Phase 1: agreeing the form of the programme

a. Agree a schedule: a schedule of implementation with all stakeholders should be completed early in the planning stage. Regular review and updating of the schedule will be required, although schedule changes will be limited once a date for introduction of screening into schools is announced. Scheduling should include development of assessment instruments including trialling and consultation with schools and stakeholders. Scheduling should also allow time for a pilot with review of decisions made being carried out after analysis.

b. Agree how assessment materials will be delivered into schools: materials can be delivered via paper or electronically. Paper based delivery requires a system for schools to order and receive materials, a system for return of materials and for double entry of the marks into a data capture system. Electronic delivery requires the development, piloting and testing of application software. Schools should have appropriate hardware available in school and the software should be installed and tested prior to application. Schools should have their own suitably trained IT personnel or access to a person from outside the school. A system for reviewing and regularly updating the software will be required, as well as a system to support schools that require help to use the software. Consultation with schools should be used to assist with this decision and consideration should be given to offering schools both options.

c. Agree on details of the assessment: this should include a decision on when the assessment should be administered in grade 1, the form of the assessment and its duration and who should administer the assessment in schools.

d. Agree on what data the assessment should deliver:

e. Agree on sources of assessment materials. Literacy and numeracy assessment tools could be developed relatively quickly and at a lower cost if the resources for this are available. Use of existing materials should be carefully investigated to ensure that the materials are suitable or could be made suitable for grade 1 screening and that they are of an appropriate level of technical quality.

f. Agree how follow-up student support will be provided: The form of support and resourcing should be agreed.

Phase 2: developing and acquiring resources

g. Develop technical process documents: to inform personnel involved in producing assessment materials, supporting schools, collecting and analysing data, producing reports etc. It is important that all processes are documented and disseminated with appropriate version control in place.

h. Develop communication materials: communications with stakeholders should be carried out as part of the communications strategy.

i. Develop process documents for schools: similar to the technical process documents, but for use in schools by school principals, teachers, administrators etc. These documents should provide details of responsibilities, timetabling of assessments, resource requirements, training programmes, sources of support etc.

j. Develop follow-up support processes for students who are identified as needing additional learning support.

k. Develop software: if assessments are being delivered electronically into schools the system will need to link to the data capture and reporting system. Software development, testing and piloting is a lengthy work stage and should be started early in the process

l. Develop assessment materials: these should include stimulus sheets, recording sheets, guidance for schools and teachers.

m. Identify and recruit personnel: while it is feasible that some personnel may already be available significant numbers of additional people will be required to implement the programme. These will consist of managers, technical staff, administrators etc.

Phase 3: developing training programmes

n. Agree and plan a training programme for personnel: this should include induction training for everyone as the programme will be new to all of them, and job specific training.

o. Agree and plan a training programme for schools: it is likely that the approach outlined for providing assessment literacy training for teachers would be a practical model for training schools to administer the assessments. However, the scale of training will be smaller.

Phase 4: developing a pilot

p. Develop a pilot: piloting should be carried out with a sample of schools that are representative of elementary schools. Piloting should include mechanisms for feeding back the outcomes and amending the programme where required, from its form to resources and training programmes.

7.2 Classroom assessment

Effective classroom assessment lies at the heart of a successful national assessment strategy, especially during the early stages of education in Cycles 1 and 2. Classroom teachers undertake assessment in all its forms, from diagnostic assessment in children's early school years, through formative assessment, which is a regular ongoing part of everyday classroom teaching and learning practice, to summative assessments at the end of semesters and school years. The NSLAF has indicated the shape, form and content of a comprehensive programme of assessment literacy CPD. The programme will be organised and coordinated by CRDP and will consist of a mixture of national and regional tutor-led sessions, local school clusters sharing their CPD resources, and in every school in form of self-reflexive community of professional practice.

Phase 1: agreeing the form of the programme

a. Agree a schedule: a schedule of implementation with all stakeholders should be completed early in the planning stage. Regular review and updating of the schedule will be required, although schedule changes will be limited once a date for introduction of screening into schools is announced. Scheduling should include development of assessment instruments including trialling and consultation with schools and stakeholders. Scheduling should also allow time for a pilot with review of decisions made being carried out after analysis.

b. Agree how assessment materials will be developed in schools: both paper or electronically. Schools should have appropriate hardware available in school and the software should be installed and tested prior to application. Schools should have their own suitably trained IT personnel or access to a person from outside the school. A system for reviewing and regularly updating the software will be required, as well as a system to support schools that require help to use the software. Consultation with schools should be used to assist with this decision and consideration should be given to offering schools both options.

c. Agree on details of the assessment: this should include a decision on when and how the school-based assessments should be administered; the form of the assessments and their duration.

d. Agree on sources of assessment materials. Literacy and numeracy assessment tools could be developed relatively quickly and at a lower cost if the resources for this are available.

e. Agree how follow-up student support will be provided: The form of support and resourcing should be agreed.

Phase 2: developing and acquiring resources

f. Develop technical process documents: to inform personnel involved in producing assessment materials, supporting schools, collecting and analysing data, producing reports etc. It is important that all processes are documented and disseminated with appropriate version control in place.

g. Develop communication materials: communications with stakeholders should be carried out as part of the communications strategy.

h. Develop process documents for schools: similar to the technical process documents, but for use in schools by school principals, teachers, administrators etc. These documents should provide details of responsibilities, timetabling of assessments, resource requirements, training programmes, sources of support etc.

i. Develop follow-up support processes for students who are identified as needing additional learning support.

j. Develop assessment materials: these should include sample questions sheets, recording sheets, guidance for schools and teachers.

Phase 3: developing training programmes

k. Agree and plan a training programme for teachers to include training in assessment for learning (formative assessment) and the development of new objective, standards- related and comprehensive classroom assessments (summative assessment) and job specific training.

Agree and plan a training programme for schools which includes national and regional training sessions as well as local and school-based ones.

7.3 Official examinations

In order to translate the professional development programmes, outlined in the earlier sections into concrete action and implementation plans it is useful to created flow diagrams that indicate the responsibilities, the time frames, the dependencies and the outcomes -

Where appropriate flow diagrams have been included to indicate work flows. Colour coding in each flow diagram to indicate timeframes is shown in the key.

Processes are shown with this symbol. The labelling of processes matches the labelling of work stages.
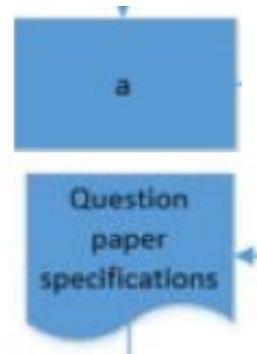
Documents are shown as outputs of processes with this symbol:

7.3.1 Improving the quality of the setting of the national examinations

Develop a schedule

It would be possible but very challenging to improve the quality of all Brevet and Baccalaureate question papers and mark schemes in one phase. Phased improvement would allow for piloting of processes and avoid overstretching resources. If improvements are phased, priority should be given to the Baccalaureate.

Develop generic processes

a. Review and update question production and banking processes: identify weak points in the current process and agree on the process changes required to strengthen these points. Develop written procedures and personnel guidance documentation for the whole question production and banking process.

Develop subject specific materials

b. Review the question paper specifications for each subject: identify key information to be included in all subject question paper specifications. This should be followed by review and revision of the existing question paper specifications for each subject.
c. Review and update guidelines for question writers for each subject: this written guidance should be based on the revised question paper specifications.

Develop assessment professionals

d. It is recommended that all personnel involved in producing question papers and mark schemes receive a form of generic training and where appropriate subject specific training. It is likely that subject specialists could carry out this training, but if not then

suitably experience personnel who are both assessment specialists and subject experts should be identified and recruited to carry out this training.

e. Identify assessment professionals to produce the question papers and mark schemes: once 'c' has been carried out it will be possible to identify the assessment professional roles to be filled. It is likely that most of these roles can be filled by existing personnel involved in writing questions and producing question papers but some roles may be vacant and in need of filling.

f. Plan training for assessment professionals: plan both the generic training and subject specific training and produce the training materials.

g. Carry out training of assessment professionals according to the agreed schedule of improvement. Evaluation of trainees should be carried out during training.

Produce question papers

h. Produce question papers: begin the process of writing questions and constructing question papers according to the new processes. Gather data on compliance with processes during production and evaluate the quality of materials produced.

i. Review: analyse the data and information gathered and review all processes and documentation materials in the light of this analysis. Revise processes and documentation as appropriate. This review and revision of processes should be carried out after the first examination and every subsequent series. Where processes are revised it is important to ensure that all documentation is updated and disseminated. Where changes to processes are extensive further training of assessment professionals should be carried out.

These steps are summarised in the flow chart below:
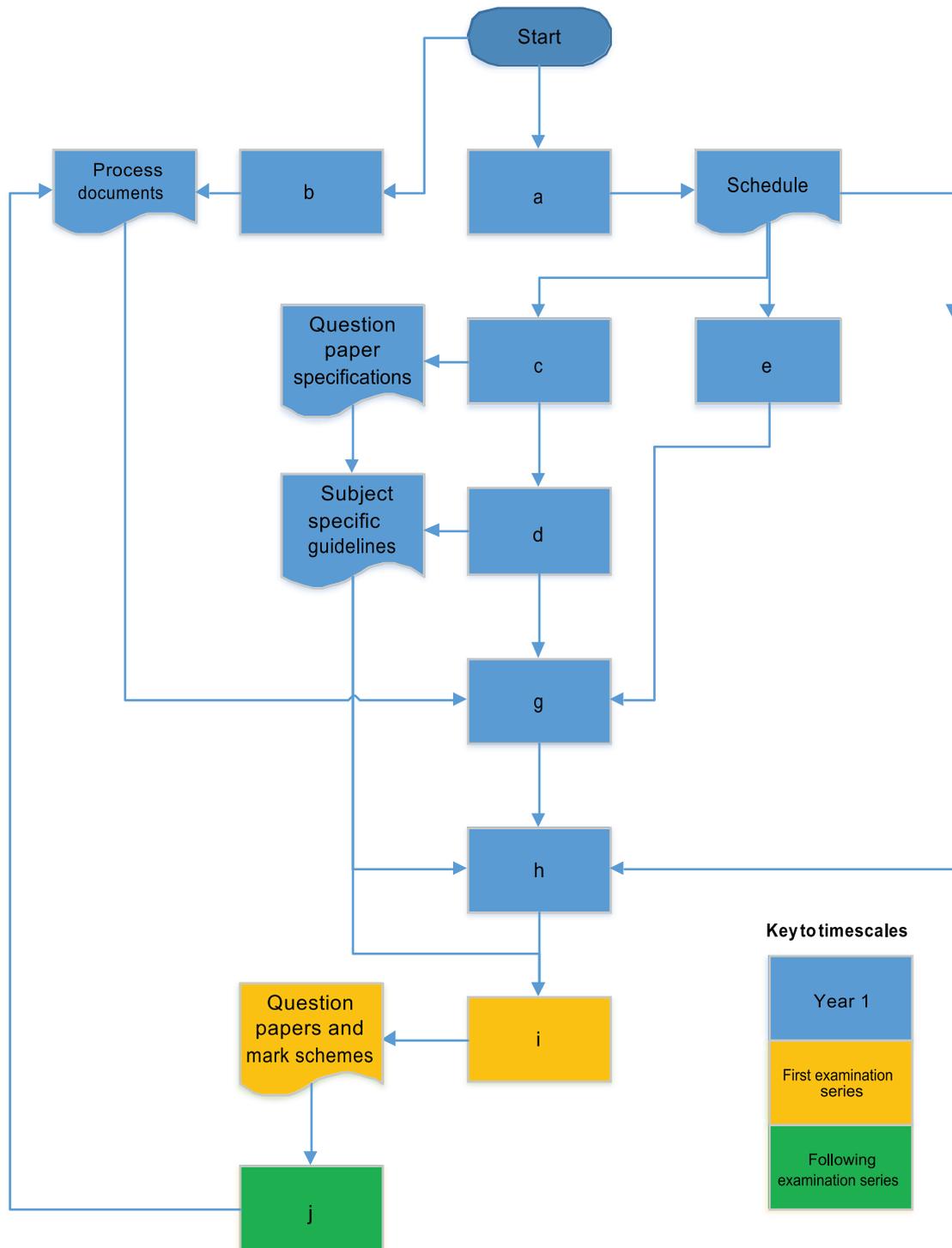
7.3.2 Improving the quality of the marking of the national examinations

Improving marking process can create a challenging additional workload for all personnel involved in marking until the new processes are refined and embedded. If unexpected problems arise during marking as a result of the changed processes there is a risk that marking will not be

completed reliably and to schedule. For this reason Cambridge International has recommended that measures to improve the quality of marking are phased in. Batches of question papers should be selected for piloting procedures before implementation of the particular improvement for those question papers. This phased approach to improvement ensures that the scale of change in each examination series is manageable and question papers that may prove to be problematic are not introduced until processes are suitably refined. Question papers that fail piloting should repeat the process until they pass.

Piloting of the group of question papers in the first examination series should focus on the organisation of marking, training of markers to interpret the live mark scheme, harmonising markers using a practical marking exercise and increasing the monitoring of markers at the start of marking . In the second examination series the tolerance of question papers that were successfully piloted in the first examination series should be revised if monitoring shows that this is appropriate. These question papers should then be piloted for implementing a system of monitoring markers during live marking using statistical data. This second pilot should only be carried out if data entry has been improved. If it is not possible to enter marks data into the data management system in a timely and accurate way, piloting should be postponed.

Stages of work
The first examination series

a. Review and revise marking processes for organisation of marking, training of markers and harmonisation: identify areas of strength and weaknesses in the existing marking processes and plan to improve as required. Update existing process documentation or produce new documentation as appropriate.

b. Revise current monitoring processes to increase monitoring at the start of marking: this should build upon the existing marking monitoring processes.

c. Identify question papers for the first phase: identify subjects where early marks data entry is relatively simple. This is likely to be subjects where the total number of marks for the question paper is low or question papers where there are fewer questions on the paper.

d. Carry out training: the relevant Chair of the Examinations Committee and supervisors for each question paper in the first phase will require training to successfully implement the new procedures. Training should be based around the procedures and could either be face-to-face on take place online. It is possible that subject specialists for question paper production could carry out the training.

e. Implement processes in the first examination series: for the first phase of question papers. Monitor implementation of the processes, collecting evidence to contribute to an end of examination series review.

f. Revise processes: analyse the monitoring data collected and revise and refine the processes as appropriate. If changes are extensive it is possible that further training of personnel will be required.

The second examination series

g. Analyse marking data from the first examination series: analyse the marking data from the first examinations series to decide whether the tolerance for each paper that passed piloting can be reduced. If there is evidence that the tolerance should not be reduced, that question paper should be withheld from piloting of statistical monitoring procedures for the examination series to allow the quality improvement processes to become properly embedded.

h. Revise tolerances of question papers selected for continuation into the second pilot.

i. Write statistical monitoring procedures if data entry is sufficiently improved to allow this first pilot to go ahead.

j. Carry out training: the relevant Chair of the Examinations Committee and supervisors for each question paper in the second examination series pilot will require training to implement the new procedures effectively. Training should be based around the procedures and could either be face-to-face on take place online.

k. Implement processes in the second examination series for the question papers selected as a pilot. Monitor implementation of the processes, collecting evidence to contribute to an end of examination series review.

l Revise processes: analyse the monitoring data collected and revise and refine the processes as appropriate.

Below is a flow chart that indicates the order of work for improving the quality of marking

Start

Process documents

a

Process documents

b

d

e

f

If 'no'

g

If 'yes'

h

c

i

Process documents

j

k

l

Key to timescales

First examination series

Second examination series
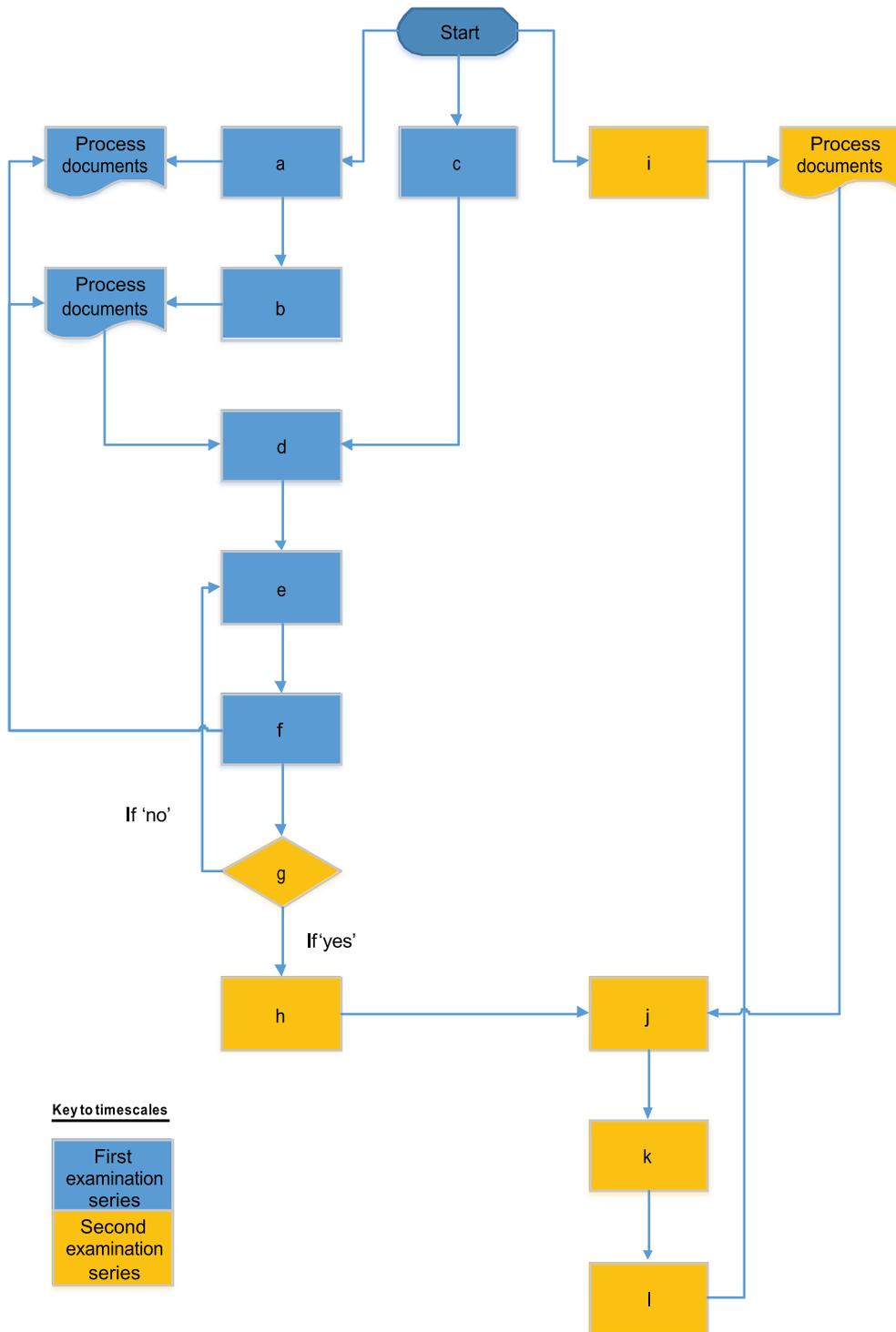
7.3.3 Improving the maintenance of the standard of the national examinations

Cambridge International has recommended that in the shorter term the improvements to question papers, mark schemes and marking should be allowed to become embedded before consideration is given to introducing a post-examination awarding process. While it is possible, although challenging, that improvement in the quality of all question papers could be carried out over two examination series if phasing is not used, improvement in the quality of marking will take considerably longer depending on how implementation is phased.

If only three phases of implementation are used it will take a minimum of four years for all papers to complete the first two pilot stages but providing that piloting of the first phase of at least some of the question papers is successful it would be possible to consider the introduction of a post-examination awarding process after two years for some subjects.

However, to avoid significant mark adjustments resulting from changes to the demand of the examination it is recommended that post-examination awarding is delayed for at least another two examination series to allow improvements to become embedded. This delay will also allow adequate preparation time for the introduction of post-examination awarding.

Once the decision to introduce a post-examination awarding process has been taken consideration has to be given as to whether to introduce the process to the Brevet and Baccalaureate simultaneously or to take a staggered approach. Implementing the recommendations to improve maintenance of the standard of both the Brevet and the Baccalaureate simultaneously would be challenging to undertake. The Baccalaureate is a higher stakes examination for students and therefore it is recommended that priority is given to stabilising the standard of the Baccalaureate before the standard of the Brevet. This will also allow any improvements to the quality and reliability of the end of semester classroom assessments to become embedded before post-examination awarding of the Brevet is attempted.

7.4 Prepare to take part in TIMSS grade 4

The following are stages of work that could be undertaken to carry out the recommendations to take part in TIMSS at grade 4. Timing of implementation should be based around the timing of TIMMS but communications should begin at least two years before the first participation in grade 4 TIMSS and training should begin approximately a year in advance. Acquiring resources should begin as early as possible in the process.

Stages of work

Plan communication

a. Develop a communication schedule: communications with stakeholders should be carried out as part of the communications strategy. Specifically for TIMSS, communication should start two years in advance of the grade 4 TIMSS survey that Lebanon will take part in. Communications to key stakeholders such as school principals, teachers and parents should communicate the benefits of introducing TIMSS in schools at grade 4. This will also enable them to discuss this with other teachers, parents and students.

b. Develop communications resources: communications documents and resources will be required to inform all stakeholders of the benefits and purposes of Lebanon's participation in grade 4 TIMSS.

Plan to train school administrators

c. Agree and plan on a training programme for schools: identify who will administer TIMSS in schools and plan a training programme for them. Training could be based on existing training provided for grade 8 TIMMS administrators. Training should be planned to start at least one year in advance of administering the first TIMSS surveys at grade 4.

Plan to acquire resources

d. Review the availability of existing resources and plan for acquisition of additional resources: as significant additional resources will be required to take part in TIMSS grade 4. Wherever feasible the same sources of resources used for grade 8 TIMSS should be utilised.

e. Adapt grade 8 process documentation: for both technical personnel and schools. Much of this could be adapted from the documentation provided for grade 8 TIMSS. Ensure that all documentation is disseminated in good time.

Annexes

Annex 1) Glossary of terms and acronyms

Acronym	Full term
AUB	American University of Beirut
Baccaluareate	The diploma awarded to students who are successful in the official examinations at the of Third Secondary
Barème	Marking scheme
Brevet	The diploma awarded to students who are successful in the official examinations at the end of grade 9
BP	Brevet Professionel – the national award for students who have successfully completed an initial technical education programme
Cambridge International	Cambridge Assessment International Education
CBA	Curriculum Based Assessment – a national survey of cohort attainment in the core subjects at grade 3 and grade 6
CAP	Certificat d'Aptitude Professionel
CRDP	Centre de Recherche et de Développement Pédagogiques
DfID	Department for International Development, UK

Diagnostic assessment	An instrument used as form of pre-assessment to evaluate students' strengths, weaknesses, knowledge and skills prior to instruction
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessments
Formative assessment	Assessment conducted during the learning process which evaluations of student comprehension and progress and, learning and which informs future learning
GDE	The General Directorate of Education
IB	International Baccalaureate
ICT	Information and communications technology
IEA	International Association for the Evaluation of Educational Achievement
IT	Information technology
LU	Lebanese University
MEHE	Ministry of Education and Higher Education
MEHE-GDE	General Directorate of the Ministry of Education and Higher Education
MEHE-GDE-DOPS	Department of Guidance and Counselling within the General Directorate of Education
NSLAF	The National Student Learning Assessment Framework

OECD	Organisation for Economic Co-operation and Development
PISA	Programme for International Student Assessment
PITB	Pre-service and in-service training bureau
QITABI	Quality Instruction Towards Access and Basic education Improvement
SDG	Sustainable Development Goal
SEND	Special Educational Needs and Disabilities
TIMSS	Trends in International Mathematics and Science Study
Summative Assessment	Assessment that is conducted at the end of an instructional unit or learning episode which is usually compared against some standard or benchmark
UNESCO - UIS	United Nations Educational, Scientific and Cultural Organisation Institute for Statistic
USAID	United States Agency for International Development
WBSR	World Bank's S2R2 initiative

Annex 2) Record of stakeholder conversations

The NSLAF and its recommendations were informed by a wide range of stakeholder consultations undertaken by Cambridge Assessment in Lebanon between February and October 2019. The major consultation events and their contributions to the development of the NSLAF are listed below in chronological order.

Stage 1 – January to April 2019

Inception Report presentation and review (February 2019)

Senior members of the Cambridge International project team visited Lebanon between 6 and 9 February to make a presentation based on the Inception Report (produced by Cambridge International) and undertake planning and baseline discussions with key stakeholders, including USAID, World Learning/QITABI, MEHE-GDE and CRDP.

Key stakeholder strategy discussions (March 2019)

This initial visit was followed in mid-March by further meetings and focus group discussions with stakeholders in Lebanon. The main aim was to provide the Cambridge International team with a key source of information and opinions from partner organisations and stakeholders. This ensured that the team had an accurate picture of both the current situation regarding curriculum and assessment as well as more recent developments (particularly since 1997).

The meetings also gave key stakeholders the chance to find out more about the NSLAF and take part in the development process. The meetings and discussions were held over two weeks, at CRDP during the first week (12–15 March) and MEHE-GDE during the second (19–21 March). Cambridge International team members also visited two schools for discussions with principals and teachers during this period.

The meetings in the first week included discussions with CRDP Academic Departments, the Pre-service and In-service Training Bureau and the Educational Research Bureau and addressed the following key issues:
- curriculum and classroom assessment
- application of data to support learning
- understanding the summative assessment measures
- Official Examination
- International Assessment
- Data management system
- Channeling within the General Education System.

The same areas were covered in the second week of meetings at MEHE-GDE, with representatives from MEHE-GDE-DOPS, the Secondary and Primary Directorates, Information

Technology Department and the Department for National Examinations taking part. However, the exact focus was slightly different reflecting the responsibilities and roles of the organisations taking part and the different departments in each.

Relevant comments from the participants, the Cambridge International team's key observations and findings and possible implications for the project from across the activities were assimilated and summarised and have been used to inform the development of the NSLAF.

School visits

In order to understand and gain a better understanding of issues and current practices relevant to the NSLAF, Cambridge International visited two public schools during Stage 1 (as facilitated by MEHE-GDE) and met with teachers and principals.

Other stakeholder consultations

During the February visit and a subsequent visit from 2 to 4 March, Cambridge International team members, together with QITABI/World Learning, met with representatives from donor organisations, including USAID and DfID, to discuss the wider context of reform and international interventions in Lebanon and in the region.

Stage 1 Report Literature Review and Situational Analysis presentation and review (May 2019)

The main deliverable and focus of activities for Stage 1, as stated in the initial project proposal and project Inception Report, was a 'Literature Review'. Its purpose was to outline international best practice in assessment of student learning and to present data and information on the national enabling context, policy framework and current practices of the assessment of student learning. It was compiled in Cambridge and was based on a desktop review of documents in the public domain and findings from the Stage 1 consultations and activities. Both in-country team members and team members based outside the region contributed to the report.

The Literature Review and Situational Analysis was well received by the key stakeholders and their feedback was recorded by Cambridge International and used to amend the report, the final version of which was delivered at the end of May 2019.

Stage 2 – May to September 2019

As described in the project Inception Report, the theme of Stage 2 of the NSLAF project was the analysis of the national assessment landscape and proposed national vision. The in- country consultations and activities listed below were carried out to inform the main Stage 2 deliverable, the National Vision document.

Assessment dialogues (May 2019)

Cambridge International ran a series of workshops for teachers and principals between 21 and 24 May 2019. The main aims were to:

- gauge the relative levels of assessment literacy among teachers and principals
- understand and validate the modes of assessment used by teachers
- evaluate the challenges arising from use of formative assessment
- gather feedback and critical reflections on the alignment of curriculum and assessment
- understand current assessment practice in the classroom
- listen to the views and concerns of participants.

The assessment dialogues were a listening and validation exercise, designed to encourage the participants to reflect on their experiences and practice. They were a valuable way to ensure authentic voices from the classroom were heard and could inform the development of the NSLAF.

The dialogues were framed around 'formative assessment' and consisted of four workshops. Each workshop had similar content but differing audience segments to enable the emerging data to be tested and validated. Each workshop had two themes: one on continuous assessment and the other on the principles and practice of formative assessment.

The audiences for the dialogues were structured to ensure

- a mix of public and private sector teachers and principals
- a spread between all key stages
- inclusive schools
- A total of 68 teachers and principals attended the four workshops. In addition, colleagues from CRDP, MEHE-GDE-DOPS and MEHE-GDE attended each day as observers throughout the process. Several attended all four sessions
- Meeting with Dean of Faculty of Education, Lebanese University
  Cambridge International team members met with the Dean of Education from Lebanese University who provided an overview of the initial teacher training in Lebanon over the last few decades
- Observation of national examinations processes (June 2019).

This was an additional activity not listed in the project Inception Report and was arranged in consultation with, and support of, key stakeholders.

With the cooperation of MEHE-GDE, the Technical Lead from Cambridge International travelled to Beirut during the administration of the Brevet (first session) to observe the administration, production and marking processes. During this time, the following activities took place:

- meeting with the DGE, who summarised some of the processes and outlined issues around the national examinations. This meeting was also attended by members from World Learning/QITABI and USAID
- meeting with key members at MEHE-GDE involved with the administration and operations for the national examinations
- visit to a marking centre in Beirut to observe marking processes and get information from operational staff, subject experts and marking team members.

Curriculum and assessment alignment workshop (July 2019)

This activity was also not included in the project Inception Report and was arranged in consultation with, and support of, key stakeholders.

The Curriculum Expert from Cambridge International, supported by the In-Country Team Leader, ran a one-day session with colleagues from CRDP on 23 July 2019. The aim of this collaborative workshop was to build capacity by exchanging information on curriculum- related issues in the context of developing the NSLAF. Specific points covered, at the request of CRDP, included

- exemplar assessment and curriculum frameworks, the relationship between the curriculum and assessment and sharing Cambridge International approaches
- discussion on the content and scope of the Baccalaureate.

NSLAF National Vision – stakeholder meetings (July 2019)

Members of the Cambridge International team travelled to Lebanon from 23 to 26 July 2019 to hold a series of meetings with stakeholders as part of the consensus-building process and production of the National Vision document.

The meetings produced feedback on the content of the NSLAF National Vision document, with key stakeholders agreeing on the format and general design principles. In addition, decisions were taken around the finalisation and implementation of the project Communications Plan and preparations for the Stage 3 activities were started.

NSLAF National Vision – stakeholder meetings (August 2019)

Members of the Cambridge International team travelled to Lebanon from 20–23 August to present the initial draft of the National Vision document for review and comment by key stakeholders of MEHE-GDE and CRDP. The second version, revised to take account of the feedback received, was delivered the same week, on 23 August. The feedback focused mainly on adding more emphasis to building the foundation of the vision and on postponing until the next stage the development of more detailed examples of connection to the capacity building, communications and application to the assessment landscape in Lebanon.

NSLAF vision feedback review and vision finalisation (September 2019)

After the release of the second version of the National Vision document, further feedback was provided to Cambridge International from CRDP, MEHE-GDE and World Learning. This was discussed and clarified in three conference calls with World Learning on 11, 19 and 23 September. The majority of the feedback related to providing more details about the development of a communications plan, capacity building and the current landscape of assessment in Lebanon. The feedback was addressed in two further versions of the Vision, issued on 16 and 27 September. The latter was accepted as the final version by World Learning.

Stage 3 – September to October 2019

As described in the project Inception Report, the goal of Stage 3 was to create a unified national student learning assessing framework (NSLAF or Framework) based on the content of the Vision document. It would provide practical guidance on designing and implementing improved practices in learning assessment and education governance and engagement strategy.

NSLAF National Vision – stakeholder meetings (late September – early October 2019)

Members of the Cambridge International team travelled to Lebanon from 25 September to 7 October 2019 to have a series of meetings, workshops and education activities with World Learning, MEHEGDE and CRDP. The objective was to develop the structure of the Framework shared in September with World Learning.

During the initial period of the visit, the project meetings with World Learning, MEHE-GDE and CRDP executive management focused on re-purposing the previously planned Communication Workshop into a Framework Planning and Implementation meeting. The Reliability Workshop was postponed to the following week, due to the Technical Lead, who had to suddenly leave the project for health reasons.

During the last days on the visit, with the arrival of the Project Executive, the meetings focused on reviewing the structure and current status of the Framework, the progress check points and the future education activities in Lebanon.

During the two-week visit, the activities involved the analysis of possible classroom assessment enhancements and workshops on Quality Review of 2019 Brevet and Baccalaureate Question Papers and Test Specifications.

NSLAF National Vision – Analysis on capacity development (late September – early October 2019)

Cambridge International held stakeholder meetings from 30 September to 4 October. These took the form of interviews conducted by the Cambridge International Policy and Communications expert. The meetings supported the following key themes that the NSLAF would address:

- Capacity for implementation needs to be strengthened
- Professional Development requires an organising framework and structure
- Assessment literacy needs to be strengthened amongst teachers
- National assessment (Brevet/Baccalaureate) quality
- Broad support for an enriched assessment landscape
- Multiple request for access to the Vision document and NSLAF, including from the Inspectorate.

NSLAF National Vision – Quality review of national examinations (early October 2019)

Cambridge International held two sessions on the quality review of national examinations, one at MEHE-GDE on 3 October, the other at CRDP on 4 October. The main purpose of the sessions was to share the findings of the Cambridge review into the quality of Lebanese Brevet and Baccalaureate exams, as well as the test specifications they are based on. The presentation covered the concept of validity at question and question paper level, the importance of test specifications and specific issues that were identified in the 2019 question papers. It ended with a number of recommendations. The sessions were also an opportunity to discuss the findings with the attendees and to gain further clarification on some issues helping Cambridge to develop their understanding of the assessment processes used in Lebanon.

NSLAF National Vision – stakeholder meetings (mid-October 2019)

Members of the Cambridge International team travelled to Lebanon from 14 to 18 October 2019 for a series of meetings with World Learning, MEHE-GDE and CRDP. These meetings covered:

- recommendations for Framework implementation
- actions relating to providing training and coaching for teachers to develop their assessment literacy
- issues relating to the quality of the Brevet and Baccalaureate examinations.

The discussions on recommendations for the Framework Implementation focused on the review of the final version of the Vision which was agreed to form the basis of the NSLAF. The stakeholders gave feedback on some sections of the Vision which informed the subsequent development of the NSLAF. This included points related to curriculum, classroom assessment, comparison of the Lebanese Baccalaureate with International and French Baccalaureates, the grade 1 screening test, data use and student learning pathways.

The discussion with MEHE-GDE-DOPS and CRDP-PITB on ways to implement extended training and coaching to develop teachers' assessment literacy clarified some of the processes and practices of assessment in Lebanon and the training currently provided and provided some guidance on issues the NSLAF should address. These included:

- the assessment systems and tools already in place (school information management system programme, EGRA/EGMA tool for early grade screening) that the NSLAF recommendations can built on
- how to strengthen capacity building activities
- connecting curriculum reform and assessment through well-defined progression models
- current training provision and areas for future improvements (for example sample assessment materials for all grades)
- examples of current assessment documentation, steps and practices as a basis for proposed future improvements
- the need for clear and concise guides at each grade on how to construct classroom assessments including sample assessment materials

- the importance consistent messages on producing assessments across CRDP-PITB, MEHE-DGE and DOPS
- the inclusion of examples of innovative assessment practice in the NSLAF.

NSLAF National Vision – Discussion relating to the quality of the national examinations (Brevet/Baccalaureate) (mid-October 2019)

The discussion relating to the review of the Brevet and Baccalaureate national examinations was completed by holding the postponed reliability workshop. Again, this meeting provided some clarifications, recommendations and points to analyse more in depth related to the National Examinations for the NSLAF. These included:

- clarifications on marking centres, marker monitoring, data entry, standardisation and typesetting processes
- recommendations for consistency in test specifications and of different tolerance
- the need for more research on classification of difficulty of items and feedback loops to item writers.

NSLAF National Vision – Stakeholders other consultations (mid to end October 2019)

While a further meeting with stakeholders in Beirut was agreed during the meeting on the review of the national examinations meeting, this could not go ahead due to difficulties to travel to Beirut during the end of October. Some communication by phone went ahead instead.

On 18 October, 2019 Cambridge International sent the draft version of the NSLAF to World Learning. World Learning provided feedback that Cambridge International has incorporated as appropriate in the final version sent to MEHE-GDE and CRDP on 31 October.