System-Wide Analysis of Assessment Practices (SWAAP)
Concept note

GAML4/REF/11

The ACER Centre for Global Education Monitoring supports the monitoring of educational outcomes worldwide, holding the view that the systematic and strategic collection of data on education outcomes, and factors related to those outcomes, is required to inform high quality policy aimed at improving educational progress for all learners.

Table of Contents

3

GAML4/REF/11
System-Wide Analysis of Assessment Practices (SWAAP)
Concept note

UNESCO
INSTITUTE
FOR
STATISTICS

GLOBAL
ALLIANCE
TO MONITOR
LEARNING

Background

This note is concerned with the development of a process to analyse how assessment practices are integrated into education systems in order for assessment outcomes to be useful for evidence-based decision making throughout the system. It examines the quality of learning assessment and the extent to which it is embedded into an education system and acts as an effective contributor to the improvement of learning outcomes. The process is referred to as the System-Wide Analysis of Assessment Practices (SWAAP).[1] The primary objective of SWAAP is to support education systems to develop national assessment strategies that generate meaningful data and assist the systems with the improvement of learning outcomes.

SWAAP is part of an initiative of the Global Alliance to Monitor Learning (GAML), spearheaded by the UNESCO Institute for Statistics (UIS). GAML was established to develop the standards and methodologies needed to measure learning globally, while supporting countries to produce and use the information to achieve UN Sustainable Development Goal 4 (SDG 4): *Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all* (UNESCO Institute for Statistics, 2016). SWAAP has been developed by UIS and its technical partner, the Australian Council for Educational Research Centre for Global Education Monitoring (ACER-GEM), to assist GAML's work on documenting system-wide assessment practices along with other related work on reporting against Indicator 4.1.1 of SDG 4. The indicator is set out as:

> *Proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex (Inter-Agency and Expert Group on Sustainable Development Goal Indiators, 2016).*

One of the key objectives of GAML is to support countries to implement learning assessment initiatives and use the results to improve the learning outcomes and skills of children and adults (UNESCO Institute for Statistics, 2016). SWAAP plays an important role in GAML's effort to achieve the objective. It takes the view that the implementation of assessment should be deeply integrated with all other elements of an education system in order for the system to be effective in improving learning outcomes. Therefore, as a first step, it is imperative to identify the extent to which assessment practices are embedded and integrated into education systems. SWAAP provides a process and tool to undertake this step.

Findings from SWAAP can be useful for education systems to identify strengths of their assessment systems and room for improvement, and the information may facilitate the development of new national assessment strategies (or the improvement of existing strategies). The findings from the analysis can also inform the process of identifying needs for capacity development in educational assessment – a process that can be valued by education systems wishing to develop their own capacity, and by related initiatives supporting capacity development efforts.

With the envisaged use of SWAAP as outlined above, education systems may choose to use the process offered through SWAAP for the following related reasons:

- The education system may wish to undertake a robust review of its assessment practices and their integration into other systems, drawing on international best practice, to develop its own plan for system improvement.

---

[1] In other contexts the SWAAP has been referred to as UIS Catalogue of Learning Assessment (CLA) Module 3.

- The education system may wish to access capacity-building opportunities through GAML or other networks. The SWAAP provides an analytical process on which a strong rationale for access to capacity-building support can be based, and helps to ensure that capacity-building efforts are targeted to areas of need.

SWAAP and other GAML initiatives

SWAAP is not a stand-alone initiative, but an essential part of the concerted effort undertaken by GAML to build global capacity in learning assessment, and monitor progress towards SDG 4. UIS is developing an integrated process to assist GAML's work for supporting countries in particular to report against SDG Indicator 4.1.1. The process consists of three work areas.

The first area is documenting system-wide assessment practices for improving learning outcomes. SWAAP and the UIS Catalogue of Learning Assessments (CLA) (UNESCO Institute for Statistics, 2017) both are designed to contribute to this area. The recently upgraded CLA provides a robust starting point for gathering essential information about the current state of learning assessment programs worldwide. SWAAP goes beyond data gathering about assessments and endeavours to assure that assessment practices are firmly embedded in national education systems.

The second area is reporting against SDG 4.1.1. This includes the development of the UIS Reporting Scales (UIS RS), and equating or Data Alignment processes to align existing assessment programs with the UIS RS, for the purposes of internationally consistent reporting against the indicator. While the Data Alignment process as well as UIS RS may highlight some specific aspects of assessment programs that could be improved, SWAAP reviews assessment practices in the wider context of education systems. Thus, SWAAP provides important system-level background on the assessment programs that are used for SDG 4 reporting purposes.

The third area is developing capacity in learning assessment practices. Findings from the Data Alignment process can offer an insight into countries' existing assessment capacity at program level. On the other hand, outcomes from SWAAP should inform the identification of opportunities for capacity development at system level. The Principles of Good Practice in Learning Assessment (GP-LA) may provide a reference point for this identification of opportunities. SWAAP and the Data Alignment process complement each other to assist GAML's efforts to build countries' assessment capacity.

SWAAP along with this suite of initiatives will support education systems to pursue effectiveness in their assessment practices to drive improvement in learning outcomes and to maximise the quality of their reporting against SDG 4.

Appendix 1 has more details of each of the GAML initiatives: UIS Reporting Scales, Data Alignment process, Principles of Good Practice in Learning Assessment and UIS Catalogue of Learning Assessments.

Conceptual Framework

The development of an analytical process and tool for SWAAP is underpinned by a conceptual framework that views learning assessment practices as a key element that should be integrated into the overall education system. Best assessment practice is an information gathering activity that is central to evaluating the effectiveness of an education system and informing action that is intended to improve educational outcomes.

The framework for SWAAP is based on the framework developed for the OECD's *Reviews of Evaluation and Assessment in Education* (OECD, 2013).[2] Like the OECD's framework, the conceptual framework locates assessment within the overall education system, and suggests the need for coherence between *assessment policy and practice* and the *overall goals of an education system* in supporting learning outcomes.

The conceptual framework for SWAAP differs from that of the OECD in two aspects. First, it requires a move from a focus on school education, to providing a framework for examining assessment practices across all stages of lifelong learning, from early childhood through to post-school education and training. Second, it requires a move beyond general recommendations for improvement, to a specific analytical process that will lead to the development of an improvement plan relevant to the context of each education system. This shift from general recommendations to specific action plans is supported by the Principles of Good Practice for Learning Assessment (GP-LA), which sets out specific principles for good practice in learning assessment. It is also supported by GAML, which brings together assessment expertise spanning from system level into the classroom.

The improvement of learning outcomes is a central goal of education systems at all levels. SWAAP views assessment of learning as the key mechanism for monitoring how well this goal is being achieved, how information from different assessment practices is used for evidence-based decision-making and to inform teaching. Assessment of learning outcomes informs a variety of monitoring and evaluation activities throughout the education system, including activities that focus on systems, policies, programs, providers or curriculum; professional learning and reflection for educators or leaders; and certifying learners. These monitoring, reflection and evaluation activities in turn inform actions that are taken to improve learning throughout the system, including goal setting, policy/program development, resource allocation, curriculum development, instructional practice and workforce development. For example, in school education, teachers and school leaders use assessment of student learning to reflect on the impact of their practice, which in turn informs action around needs-based resource allocation and teacher professional development.

For this reason, the whole-of-system analytical approach of SWAAP is not solely aimed at improving assessment practices, but also at using learning assessment to improve education systems as a whole. The ultimate objective of learning assessment is to generate data that are useful for evidence-based decision making. The decisions that are made should enhance the effectiveness of the education system and, consequently, lead to the improvement of learning outcomes. Figure 1 illustrates the conceptual framework of SWAAP, which reflects this objective.

---

[2] The OECD review provides an international comparative analysis and policy advice on how evaluation and assessment arrangements can be embedded within a consistent framework to improve the quality, equity and efficiency of school education. The review builds upon a three-year review of evaluation and assessment policies in 28 countries, the *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes*. As well as analysing strengths and weaknesses of different approaches, the report provides recommendations for improvement including how results should be incorporated into policy and practice.

Figure 1: Conceptual framework of SWAAP
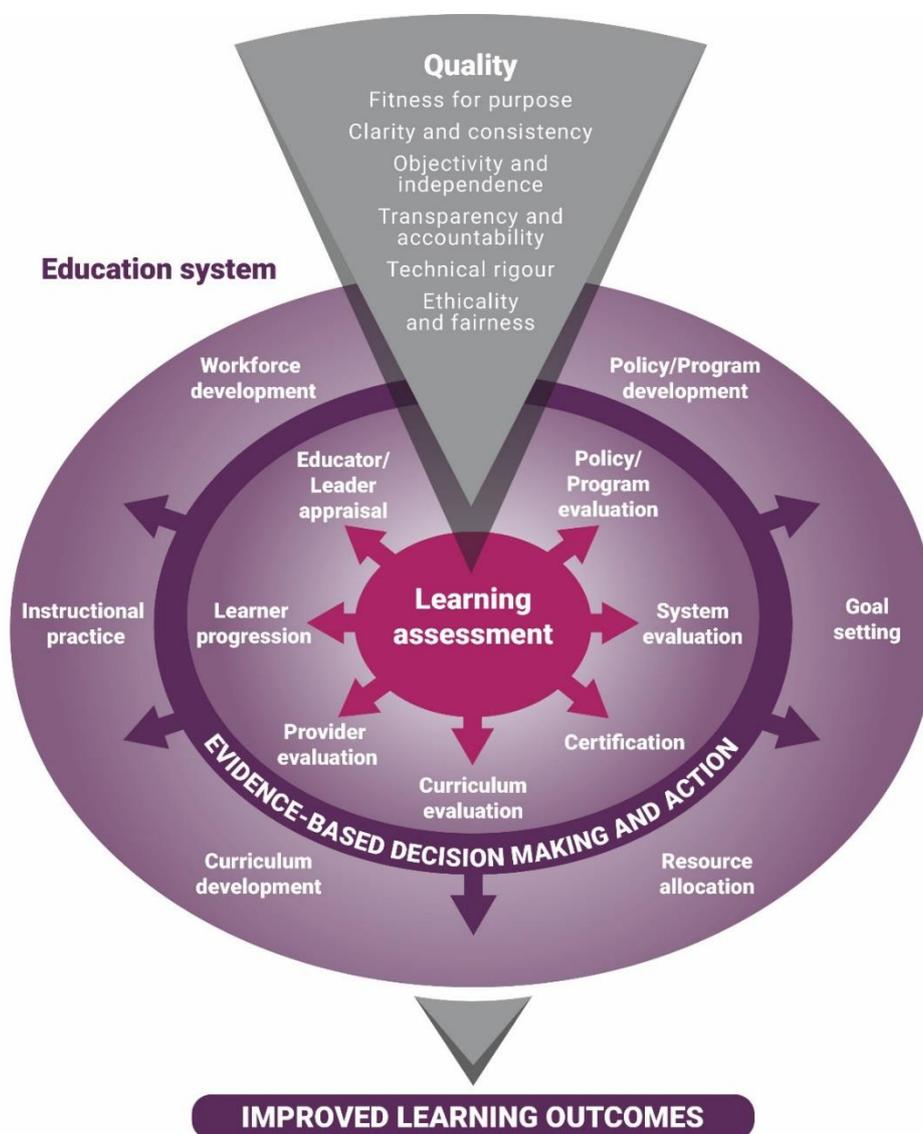
Key concepts

The conceptual framework for SWAAP (see Figure 1) focuses on learning assessment as the central source of information for evidence-based decision making and action.

Learning assessment

The term *learning assessment* is used to refer to judgements concerning the learning outcomes and achievement of individual learners (OECD, 2013). It may include:

- international large-scale assessment
- regional large-scale assessment
- national large-scale assessment
- household-based assessment

- examinations
- classroom-based assessment.

The types of learning assessments listed represent typical assessment activities of education systems around the world.

Appendix 2 has a brief description of each type of learning assessment.

Evaluation and appraisal

Learning assessment can be used for different purposes, which can be subsumed under the terms *evaluation* and *appraisal*, according to the definitions of the OECD (2013):

- The term *evaluation* is used to refer to judgements about the effectiveness of curriculum, schools and other educational organisations, education systems, policies and programs.

- The term *appraisal* is used to refer to judgements about the performance of educators and educational leaders, e.g. teachers, school leaders.

As shown in Figure 1, these purposes include policy/program evaluation, system evaluation, certification, curriculum evaluation, provider evaluation, establishing and monitoring learner progression, or educator/leader appraisal.

Evidence-based decision-making and action

The data gathered can then be used for evidence-based decision making and to inform action oriented towards improved learning outcomes. Major areas for evidence-based decision-making and action are policy/program development, goal setting, resource allocation, curriculum development, instructional practice, and workforce development. An outline of how assessment practice can inform evidence-based decision-making and action in each of these areas is provided below.

Policy/program development

Education policy is a broad term covering the wide range of decisions and actions that governments take to shape education systems, including legislation, guidelines and other policy initiatives. Education programs are often developed to operationalise specific education policies. Assessment of learning is an important resource for evidence-based policy-making and action. For example, results from a national large-scale assessment may show a low level of reading achievement of students with a particular socio-economic status. The assessment results can provide useful information for developing an intervention program that targets the particular group.

Goal setting

In broad terms, educational goals are the priorities that education systems set for improvement, to ensure that all learners are achieving the competences, skills, and attributes with which the system aims to equip them. Goal setting may be informed by assessment in a variety of ways. For example, student assessment may be used to evaluate education providers, leading to specific, evidence-based goals for school improvement in particular areas. Similarly, results from an international large-scale assessment are often analysed with the purpose of system monitoring, and may provide information about where an education system is situated in comparison with other systems. It may also prompt education ministry personnel to review the current education goals, and to use the findings to inform the new goal setting.

Resource allocation

The purposeful and practical allocation of resources to support equitable access to high-quality learning opportunities is a major element of education policy at system, sub-system, local and school levels. Leaders at all levels of the education are charged with making decisions about how to effectively distribute and leverage resources to support teaching and learning. Data from a learning assessment may be used when decisions are made regarding the change of funding allocation between schooling sectors (e.g., primary and secondary), or targeted funding for an at-risk student group.

Curriculum development

Curriculum development is a planned, purposeful, progressive, and systematic process in order to create positive improvements in the educational system. Every time there are changes or developments happening around the world, the school curricula are affected. There is a need to update them in order to address the society's needs. Results from learning assessment, particularly large-scale assessment, can inform evaluation of curriculum, and resulting changes to curriculum development. For example, Best *et al.* (2013) found that the use of data from national and international large-scale assessments in developing countries quite often resulted in informing curriculum standards and reform.

Instructional practice

In order to improve learning outcomes, a strong focus can be placed on guiding and improving school- and classroom-level factors such as instructional practice – or teaching and learning practice. Results and findings from learning assessment can provide educators and teachers with evidence that informs their pedagogical actions. When a pedagogical action is taken in response to observations drawn from assessment findings, it creates a process of 'assessment for learning' to guide future teaching and learning.

Learning assessment can play a role in improving instructional practice, and relevant issues can be explored such as:

- Leaders'/Principals' responsibility for establishing and managing assessment policies, including national/sub-national assessments held at institution/school,
- Use of learning assessment results to develop institution's/school's educational goals and programs
- Educators'/Teachers' use of assessments to improve teaching and learning.

Workforce development

Workforce development includes any professional development activities that are designed for those who work, or intend to work, in education. If learning assessment should be embedded in an education system, and if assessment practice should be integrated into instructional practice at school- and classroom-levels, it is important to ensure that educators and leaders receive adequate professional development opportunities in the area of assessment practice. It is also important that the findings from learning assessment are used, through educator and leader appraisal systems, to identify areas in which teachers and leaders may benefit most in their professional learning. For example, assessments of learning may help to identify learning domains in which instructional practice could be improved.

Quality

In order to be effective, learning assessment, and the interaction of different assessment practices, need to be guided by key quality principles. SWAAP draws upon the six key quality concepts for learning assessment outlined in the Principles of Good Practice in Learning Assessment (GP-LA). While the quality concepts in the GP-LA are described specifically in relation to large-scale learning assessments, their principles represent broader concepts of quality in educational assessment and are applicable to SWAAP.

Fitness for purpose

'Fitness for purpose' describes the concept that the ultimate goal of an assessment is to generate data that are appropriate for their designated purposes. In the context of SWAAP, fitness for purpose asks whether the outcomes from various learning assessment practices are useful to contribute collectively to evidence-based decision making, which should result in improving the effectiveness of the education system and, hence, learning outcomes.

Clarity and consistency

Clearly stating the purpose(s) of an assessment is essential because a statement of the reason for the collection of data and its intended uses will serve as a guide for all future decisions made in relation to the assessment practice. Such decisions should be consistent with its stated purpose. This will help maintain consistency across all areas of the assessment and ensure that the final results are relevant and useful for education policy makers and other stakeholders. SWAAP draws on this quality concept to examine the clarity of the purpose of each learning assessment and the consistency of the decisions made in relation to the assessment practice with the purpose.

Objectivity and independence

In order for stakeholders to trust the results of an assessment, the collection and the interpretation of the data must be objective and independent. Objectivity and independence of the interpretation of data is particularly relevant to SWAAP as it looks closely into how assessment outcomes are used for evidence-based decision making at system-level.

Transparency and accountability

Assessment practice should be transparent and held accountable to its stakeholders. Therefore, all aspects of assessment practice should be open to outside scrutiny. In an education system where transparency and accountability are in place, assessment purposes, methods to achieve the purposes and intended uses of the assessment outcomes should be clearly documented and publicly available as a national assessment strategy.

Technical rigour

It is essential that assessment methodology, analysis and interpretation of data follow scientific principles so that inferences drawn are valid and their level of certainty can be determined. Technical rigour is particularly important in order for assessments to be technically robust to produce data that are meaningful for evidence-based decision making.

Ethicality and fairness.

The broad goal of assessment ethics and fairness is to ensure that no harm is done to individuals or groups as a results of an assessment, and that no participants are unfairly disadvantaged. The concepts are also related to the importance of ensuring confidentiality, well-being and inclusiveness of assessment participants. In SWAAP, these quality concepts can be useful to

examine whether the welfare of assessment participants is well taken into consideration across all types of learning assessment.

Process

The conceptual framework serves as a foundation for setting a process of carrying out SWAAP. As aforementioned, the analysis undertaken through SWAAP will examine how assessment practices are used to establish a broad evidence-base for decision making and action throughout an education system. Such an analysis will require careful planning, systematic gathering of evidence, and effective use of findings. Therefore, the process of SWAAP is likely to involve the following major steps:

1. setting an implementation plan,
2. collecting and analysing data (evidence),
3. disseminating findings to identified audiences,
4. developing an improvement plan based on findings.

1. Setting an implementation plan

Setting an implementation plan of SWAAP entails consideration and clarification of the following issues:

1) Responsible body and key stakeholders: Who is conducting SWAAP: An education ministry, a donor agency or fund manager, an external partner? Who are the key stakeholders in the process of conducting SWAAP and to whom will the findings of the SWAAP be directed?

2) Purpose, priorities and focus of conducting SWAAP: What is the main reason for carrying out SWAAP? What are the priorities? Which assessment practices should be included and focussed on in the analysis?

- Resources: What are the available budget, timeframe, human and material resources to undertake SWAAP? How does the SWAAP process need to be scaled to meet the available resources?

- Data collection and analysis strategy: Based on the purpose, priorities and focus of the SWAAP, a data collection and analysis strategy needs to be developed.

- Dissemination of findings: A dissemination strategy needs to be developed. Who will be the audience? How will the findings and conclusions arising from the analysis be disseminated?

- Improvement plan: Based on the findings of the SWAAP, an improvement plan needs to be developed. Who will be responsible for developing an improvement plan and monitoring its implementation? Who are the key stakeholders to be included in this process?

2. Collecting and analysing data (evidence)

A data collection and analysis toolkit will need to be developed to assist education systems in conducting the SWAAP process. The toolkit will reflect the understanding of how assessment practice is embedded in an education system articulated in the conceptual framework, and will translate this into a concrete process for gathering and analysing data. The toolkit will be designed to provide evidence that learning assessment is an integral part of the education system, by

identifying the coherence between learning assessment and various elements of the education system outlined in the conceptual framework.

The purpose of the toolkit is to break the data collection process into manageable steps, centred around a set of clearly designated priorities and processes. This recognises that examining the relationship between key elements of an entire education system can appear to be an overwhelming task, and system leaders may struggle to identify where to begin their search for opportunities for improvement. By listing a clear set of potential priorities, and the steps that may be taken to examine them, the toolkit will support a focused approach to system-wide improvement.

For each priority area, the toolkit will help to identify:

- The questions that education systems need to ask

- The evidence that might be used to answer these questions

- Improvement actions that system leaders may take, depending on their findings.

The priority areas identified in the toolkit will reflect the type of functions and activities that may be supported by learning assessment outlined in the 'Key concepts' section. The questions will be designed to enable system leaders to examine how learning assessment is used to inform each of these areas in detail, or to examine the points of intersection between them, using learning assessment as a common point of focus. Examples of the kinds of priority areas and questions to be listed in the toolkit are shown in Table 1.

Table 1: Example priority areas and questions in the data collection tool

Example priority area	Example questions
Single priority areas	
Curriculum evaluation	How are the findings from learning assessment used for the revision of the curriculum?
Education system evaluation	How is learning assessment used to inform system evaluation at the sub-national and national level?
Teacher/educator appraisal	How is learning assessment used to inform teacher/educator appraisal, and in turn inform evidence-based decision-making about teachers' and educators' professional development?
Intersections	
Relationship between: curriculum evaluation; and classroom-based assessment	How do teachers' insights from classroom-based assessment inform curriculum evaluation? How are student assessments designed to ensure that they respond rapidly to curriculum reform?
Relationship between: provider evaluation; and teacher/educator/leader appraisal	How is appraisal of teachers/leaders embedded in an improvement agenda at school level, and how is learning assessment used in their alignment? How are individual professional development priorities aligned with whole-school priorities?
Relationship between: policy evaluation; and large-scale learning assessment	How are results from large-scale assessments of learning used to inform policy evaluation? How is policy designed to maximise its visible impact on measures of student learning?

The detailed examination of any one of these priority areas could constitute a substantial analytic exercise in itself. It is therefore recommended that systems undertaking the SWAAP focus on a small number of priorities (say, one or two) at a time, building over time to a complete examination of the entire system.

Improvement in one or two areas can also open up pathways for improvement in related activities, meaning that the process is likely to gain momentum over time. While the first priorities selected for examination may require the most intensive evidence-gathering and analytic effort, subsequent priorities will become easier, as they build on insights from these first analyses. This reinforces the desirability of beginning with a small number of priorities, and building up gradually to system-wide analysis.

It is expected that data will be collected from a full range of activities that support and constitute assessments in order to undertake a comprehensive review of a country's assessment practice. To gain comprehensive information from various data sources, data collection methods could include:

- Reviewing relevant documents
- Administering questionnaires
- Conducting interviews
- Observing assessment implementation

Depending on the desired scale and focus of the process, data could be collected from a range of stakeholders, including:

- Education ministry personnel
- National-level assessment authority
- National-level assessment implementation team
- Regional/provincial-level education authority
- School personnel/educational organisation personnel
- Students/children/young people/adults as learners

As SWAAP takes a system-wide approach, and is concerned with the full range of activities that support and constitute assessment practices, it is foreseen that it will take substantial time for an education system to complete SWAAP in its entirety. Therefore, it is essential to make the best use of existing data sources available where relevant, so that SWAAP can be implemented efficiently. The use of existing data sources also enable this analytic exercise to be more comprehensive by identifying and locating them to see where they fit within their overall assessment practices.

For example, UIS' Catalogue of Learning Assessment would be a good starting point to obtain essential data regarding learning assessment programs (UNESCO Institute for Statistics, 2017c). OECD's Teaching and Learning International Survey (TALIS) provides in-depth cross-country analysis about effective teaching and learning (OECD, 2014). World Bank's Systems Approach for Better Education Results (SABER), specifically the Student Assessment component of the initiative, produces comparative data and knowledge on education policies and institutions (World Bank, 2017).

The SWAAP toolkit will include suggested approaches for analysis, recognising that the data collected will vary for each education system, depending on their current assessment practices and priorities. Analytic tools may include mapping exercises to follow the influence of assessment data through different parts of the system (as listed in the conceptual framework), or gap analysis

tools to identify any areas in which assessment practice is not being used well in specific areas of the education system. The aim is for the analysis to result in a clear indication of how the education system is using assessment well, and opportunities for improvement and further action.

3. Disseminating findings to identified audiences

The dissemination of findings is an essential step to reach key stakeholders. To successfully disseminate the findings it will be important to develop a dissemination strategy. The strategy may involve: identifying key audience groups (e.g. important stakeholders, the public/media), the different information needs of the key audience groups about relevant aspects of findings; and the kinds of dissemination products to be developed (e.g. a report, policy briefs, infographic, etc.).

4. Developing an improvement plan based on findings

Based on what is documented, findings from SWAAP should address how assessment practice is integrated into a country's education system, and the coherence between all parts of the system. If the findings identify incoherence and inconsistencies within assessment practice or between assessment practice and key elements of the education system, then the analysis will assist the education system in the formulation of an improvement plan for assessment practices. Ideally, the improvement plan should contribute to the development of new, or the improvement of current, national assessment strategies – but may also benefit education systems through more incremental approaches to improvement. By choosing one or two priorities at a time, systems can achieve incremental improvement with more visible results, compared to attempts to reform whole systems simultaneously.

Other issues education systems might wish to include in their improvement plan are: funding arrangements for required improvements; needs for capacity development to implement the plan; and a monitoring and evaluation plan to assess the improvement.

Relationship to other initiatives

It is important to ensure that SWAAP complements rather than duplicates the work of other initiatives. In this section, the main features of some of the relevant initiatives are described, and how SWAAP can complement and add value over them is discussed.

CapED

Created in 2003 as the Capacity Development for Education for All (CapEFA) Programme, the Programme was modified in 2016 to better align with the SDG4-Education 2030 agenda and rebranded as Capacity Development for Education. Through CapED, UNESCO provides targeted assistance and reinforces national capacities to undertake evidence-based national education reforms in least developed countries (UNESCO, n.d.). To support the CapED SDG 4 Pilot Initiative, UIS has developed a wide range of tools to assist the participating countries in collecting data to monitor progress towards the SDG 4 (UNESCO Institute for Statistics, 2017a).

SWAAP will provide the initiative with another useful tool to assist the countries to monitor progress towards SDG 4, specifically in the area of monitoring learning outcomes. The CLA data collection tool and GP-LA have been already made available to support the countries to carry on the initiative. The CapED SDG 4 pilot initiative demonstrates a practical example of UIS' collaborative work with national and international actors on data gathering and analysis efforts. SWAAP along with other GAML initiatives can contribute to the success of the collaborative work.

SABER

World Bank's Systems Approach for Better Education Results (SABER) evaluates the quality of education system policies against evidence-based global standards, with the aim of helping countries systematically strengthen their education systems. (World Bank, 2012). The Student Assessment component of SABER (SABER-SA) is particularly relevant to SWAAP as it is designed to support countries in strengthening their student assessment systems.

SWAAP will draw on the data collected through SABER-SA, and complement it by analysing not only the quality of assessment practices, but also the quality of relationships between assessment practices and other elements existing in an education system. It should assist education systems to gain better understanding of how assessment practices can be integrated to the education system and contribute to the improvement of learning outcomes. The findings from SWAAP will also produce further information that could help identify where capacity development support might be most needed.

Assessment for Learning

Assessment for Learning (A4L) is an international platform to assist national learning assessment systems, which is supported by the Global Partnership for Education (GPE). The initiative aims to: develop and strengthen national learning assessment systems; and improve the use of learning data to inform educational policy and teacher training. The A4L platform is said to have a plan to channel financial and technical assistance to developing countries according to their individual needs and priorities, in order to build the capacity of national learning assessment systems with the aim of improving learning and equity (Brookings Institution, 2015).

SWAAP shares with A4L the objective of supporting the strengthening of learning assessment systems, and will provide a process and tool to realise the objective. Findings from SWAAP can inform the identification of needs for capacity development in assessment areas, which may be highly valued by initiatives supporting capacity development efforts like A4L.

PISA for Development Capacity Needs Analysis

PISA for Development (PISA-D) Capacity Needs Analysis is another initiative that is closely related to SWAAP. As a diagnostic tool, PISA-D Capacity Needs Analysis evaluates the capacity of participating countries to implement PISA-D (OECD, 2016a). The Capacity Needs Analysis framework provides a set of project requirements for implementing PISA-D, accompanied by corresponding description of capacity entailed to fulfil the requirements. Then, the current capacity of the participating country is evaluated using the description.

PSIA-D Capacity Needs Analysis demonstrates how specific project requirements can be incorporated into a capacity needs analysis tool. This is a good example of an analytical tool to gather information relevant to the identification of capacity development needs, with reference to a specific assessment program. SWAAP will draw on the example and develop data collection tools that can be useful not only for a particular project, but also for a system-wide purpose.

Next steps

This note presents the framework and associated key concepts of SWAAP along with a preliminary outline of the process that assists to translate the concept into action.

The next step to move forward would be for the GAML network and relevant task force members to review, discuss and reach a consensus on the concepts addressed in this note. This is expected to be an iterative process. Issues and questions raised during the review would help shape the concepts further.

Once the concepts are agreed upon, further details of the process and accompanying data collection toolkit will need to be developed. Development of the toolkit will involve a substantial research process, to identify questions for each priority that are informed by international best practice. This may include compilation of practice examples from leading education systems, to help countries understand what they are looking for, and guide development of suggested improvement actions in the tool. A pilot will also be conducted in a small number of countries, in which the questions are tested in a variety of contexts, for a variety of priorities, to ensure that they yield worthwhile information for system improvement. This will result in a high-quality, evidence-based analytic toolkit applicable in a range of contexts.

As the development phase proceeds, the relationship between SWAAP and other SDG 4-related initiatives led by GAML would become clearer, and greater coherence between these initiatives could be achieved. This would include clearer identification of opportunities for UIS and GAML partners to support the SWAAP process, and to consolidate findings from SWAAP implementation to improve understanding of the uses of learning assessment in system-wide improvement around the world.

Timeline

We propose that the development of concept, process and toolkit to implement SWAAP should be undertaken over a period of 12-months starting from September 2017, based on the tasks and estimated time required as summarised in Table 2.

Table 2: Timeline for the development of SWAAP

Tasks	Estimated time required (month)	Indicative timeline to complete
Finalise concept note	2	November 2017
Develop the process and toolkit	4	March 2018
Undertake a pilot	4	July 2018
Finalise the process and toolkit based on the results from the pilot	2	September 2018
Total	12	

Budget

There are two phases that are associated with SWAAP: development phase and implementation phase.

The budget required in the development phase includes technical assistance costs related to the tasks aforementioned in the 'timeline' section. In addition, implementation costs in relation to the pilot need to be taken into account.

Part of the technical assistance costs could be possibly funded by the Australian Department of Foreign Affairs and Trade's Australian Aid Program and ACER through ACER-GEM. Additional funding, especially funding to implement a pilot will need to be sought from donor-supported funding allocations.

Costs required in the implementation phase are more difficult to estimate as the specifics of the process and tool that will be operationalised in this phase are yet to be determined.

Appendix 1: GAML initiatives

The main components of GAML's initiatives are outlined below. Together, this suite of initiatives will equip education systems to maximise the consistency of their reporting against Indicator 4.1.1, as well as to pursue broader improvements to assessment and evaluation systems to drive improvement in student learning.

UIS Reporting Scales (UIS RS)

The UIS Reporting Scales (UIS RS) are numerical scales and associated substantive descriptions, which explain developing proficiency in the learning domains that feature in the SDG 4 targets. The substantive descriptions on the UIS RS will provide a backbone for interpreting the words 'reading' and 'mathematics' in Indicator 4.1.1. Particular locations on the UIS reporting scales will be established as benchmarks, also providing a backbone for interpreting the expression 'at least minimum proficiency' for each of grades 2/3, the end of primary and the end of lower secondary.

As at August 2017, draft scales for reading and mathematics have been constructed, and are undergoing review by education, assessment, and domain experts globally. The draft UIS RS include theoretically-driven descriptions of the kind of observables that indicate the knowledge, skills, and understandings within each strand and domain at each level. A proposal is also being prepared for validating the draft scales in the field. Further information about the draft scales and their development can be found in the UIS RS Concept Note, and Phase I technical report. In addition, a draft proposal for setting benchmarks on the UIS RS has been developed for GAML's consideration.

Data Alignment process

The Data Alignment process enables education systems to examine and report on the current level of alignment of their (national) assessment programs with the UIS RS clearly, efficiently and consistently. It is designed to capitalise on the potential of the UIS RS to improve consistency in reporting from diverse assessment programs against Indicator 4.1.1, and enhance international dialogue about the measurement of learning. It will support greater transparency and integrity in SDG 4 reporting, and create a common point of reference so that all stakeholders can better understand what Indicator 4.1.1 data is saying about student learning around the world. As at August 2017, a draft concept note of the Data Alignment process was under development.

Principles of Good Practice in Learning Assessment (GP-LA)

The Principles of Good Practice in Learning Assessment (GP-LA) is a statement of principles to guide the development and implementation of robust assessment programs. The GP-LA is an independent articulation of good practices that accommodates the diversity of large-scale learning assessment activities being undertaken throughout the world. It describes what is involved in efficiently developing and implementing a robust large-scale assessment program, with the aim to effectively use the data for education system monitoring and evidence-based education policy. At August 2017, a draft of the GP-LA had been developed, and was under review by relevant GAML Task Forces.

UIS Catalogue of Learning Assessments (CLA)

The UIS Catalogue of Learning Assessments (CLA) collects information about learning assessments used in different education systems in a standardised way. It uses a questionnaire to collect information from education systems about the types of assessment programs they have in place, the levels of education that are addressed, and the years in which the assessments occurred. To

date, Module 1 of the CLA has collected information on 80 assessment programs, with results of these studies available on the UIS website[3]. A second module for the CLA is currently under development by UIS, which will collect data from assessment programs to use in SDG 4 reporting. Figure 2 shows GAML's overall approach to SDG 4 reporting.
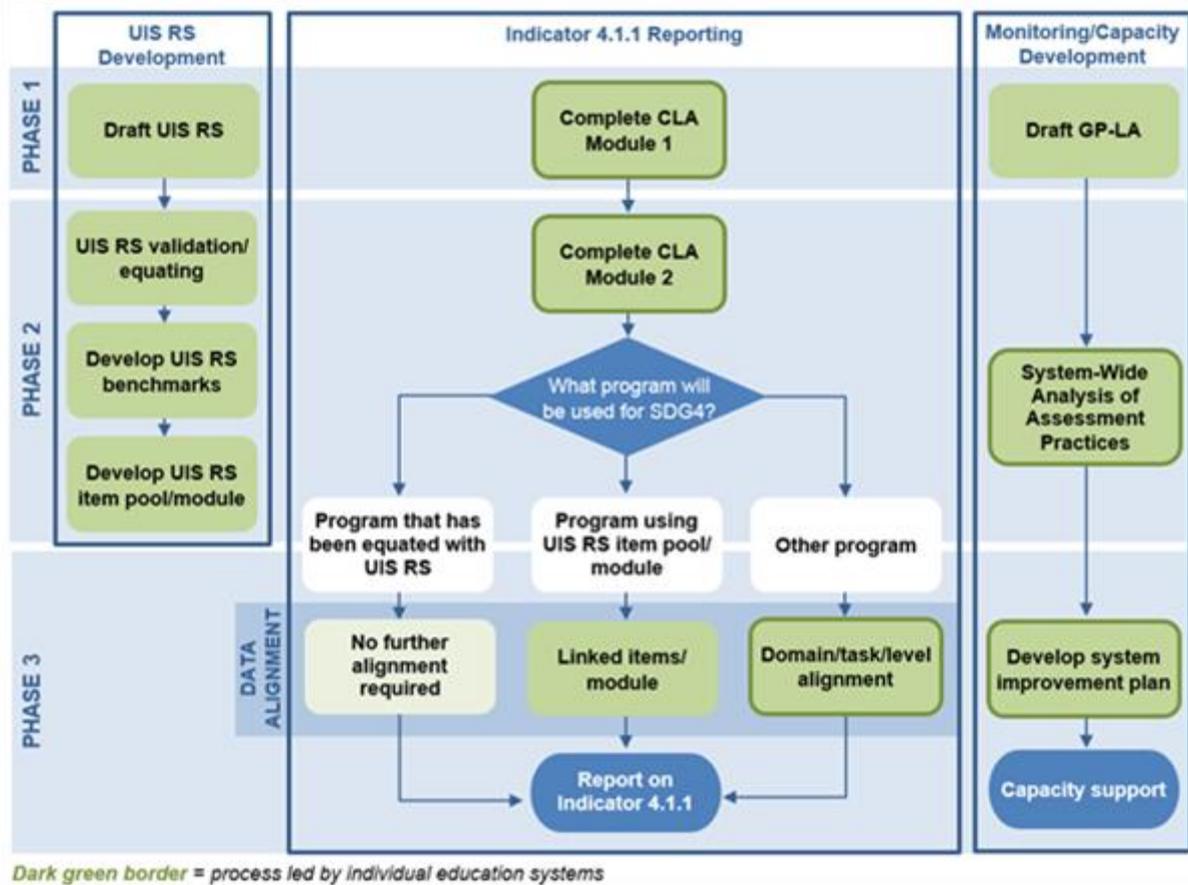

Figure 2: GAML's approach to SDG 4 reporting

[3]www.uis.unesco.org/nada/en/index.php/catalogue/learning_assessments#_r=&collection=&country=&dtype=& from=1990&page=1&ps=100&sid=&sk=&sort_by=nation&sort_order=&to=2015&topic=&view=s

Appendix 2: Type of learning assessment

This appendix provides a brief description of each of the assessment types that are covered by the SWAAP.

International, regional and national large-scale assessments

Large-scale assessments are designed to provide information on system performance levels and the contributing factors in order to inform education policy and practice (Lietz, Cresswell, Rust, & Adams, 2017).

According to Tobin, Lietz, Nugroho, Vivekanandan, and Nyamkhuu (2015), large-scale assessments of students' learning:

- are standardised to enable comparability across students, schools and in some cases, countries

- are intended to be representative of an education system either at the sub-national (i.e., state, province) or national levels

- are equally likely to be conducted in centralised or decentralised education systems

- do not have as their main purpose to certify individual student achievement, and

- do not refer to assessment used by teachers in classrooms, or to selective assessments such as graduation examinations or university entrance examinations.

In addition, large-scale assessments in some instances can compare education systems across countries in the same region[4] or internationally[5]. The former is referred to as regional large-scale assessment, and the latter is called international large-scale assessment. Those with national focus are referred to as national large-scale assessments.

Examinations

An examination is a formal test administered to assess a student's knowledge, skill or ability in a particular subject, or to obtain a qualification or certification. Examinations often take the form of standardised tests and usually have significant consequences for individuals (Thissen & Wainer, 2001). Examinations are widely used to provide information for high-stakes decision making about individual students, such as graduation certificates or high school/university entrance examinations in many education systems (World Bank, 2012).

Some countries require all their secondary school students to take an examination on individual subjects as a requirement for graduation. Examples include General Certificate of Secondary Education in the United Kingdom (Office of the Qualifications and Examinations Regulator, n.d.) and Baccalauréat in France (Sayare, 2013). In contrasts, high school students in other countries may not be required to take an examination to graduate. Rather, students in these countries

---

[4] Examples include: the Latin American Laboratory for Assessment of the Quality of Education (LLECE); Pacific Islands Literacy and Numeracy Assessment (PILNA); Conference of the Ministers of Education of French Speaking Countries' (CONFEMEN) Programme for the Analysis of Education Systems (PASEC); and the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ).
[5] Examples include: International Civic and Citizenship Education Study (ICCS); Programme for International Student Assessment (PISA); Progress in International Reading Literacy Study (PIRLS); and Trends in International Mathematics and Science Study (TIMSS).

usually take examinations, for example, the Scholastic Assessment Test (SAT) in the United States (The College Board, 2017), to apply for a position in a university program.

Classroom-based assessment

Assessment in classroom can be undertaken using a wide variety of methods from recording anecdotal notes while observing a student to administering standardised tests. These methods are often divided into two broad categories based on the way it is used – formative assessments and summative assessments. Where summative assessment primarily focuses on assessing learning outcomes, formative assessment aims to gain insights into learning processes that can be used to support learning through tailored instruction and targeted feedback (Heitink, Ven der Kleij, Veldkamp, Schildkamp, & Kippers, 2016). Formative assessment is often called classroom-based assessment (Western and Northern Canadian Protocol for Collaboration in Education, 2006).

Classroom-based assessment, or sometimes referred to as 'Assessment for Learning' (Western and Northern Canadian Protocol for Collaboration in Education, 2006), occurs as part of ongoing classroom practices (Klenowski, 2009), and focuses on the quality of the learning process (Stobart, 2008). Feedback is continually incorporated in this process to guide future learning, and is aimed at the class or individual levels. Students play a vital role in classroom-based assessment and are expected to engage in assessing their own learning (Western and Northern Canadian Protocol for Collaboration in Education, 2006).

Household-based assessment

The major defining characteristic of household-based assessments is the location of the assessments – it is household-based, not school-based. Within the category of household-based assessments, there are different types of surveys depending on who organises it.

Citizen-led assessments are led by citizens group, and are often conducted in households. According to ACER (2015) and Plaut and Jamierson-Eberhardt (2015), citizen-led assessments share the following defining characteristics. Citizen-led assessments are:

- an assessment of basic reading and math competencies,
- conducted in households,
- conducted orally and one-on-one,
- statistically representative,
- independent of government as they are organized by civil society (p.11).

In particular, they highlight the importance of conducting the assessment in households, not schools, so as to include all children–not just those enrolled in and present in recognised schools on testing day. This means that out-of-school children are included as well as children in unrecognised schools.

However, not all household-based assessments are led by civil society. Some are organised by international organisations. For example, the Programme for the International Assessment of Adult Competencies (PIAAC) is a household-based survey of adult skills conducted by OECD (2016).

References

ACER. (2015). Report on the concurrent validity and inter-rater reliability studies of Uwezo. Washington, DC: Results for Development Institute (R4D).

Best, M., Knight, P., Lietz, P., Lockwood, C., Nugroho, D., & Tobin, M. (2013). The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries: Final report London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Brookings Institution. (2015). Assessment for Learning (A4L): An international platform to support national learning assessment systems: Discussion document.

Heitink, M. C., Ven der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review, 17*, 50-62.

Inter-Agency and Expert Group on Sustainable Development Goal Indiators. (2016). Update on the work to finalize the proposals for the global indicators for the Sustainable Development Goals.

Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice, 16*(3), 263-268.

Lietz, P., Cresswell, J., Rust, K., & Adams, R. (2017). Large-scale education assessments. In P. Lietz, J. Cresswell, R. Adams, & K. Rust (Eds.), *Implementation of large-scale education assessments*. New York: Wiley.

OECD. (2013). Synergies for better learning: An international perspective on evlauation and assessment *OECD reviews of evaluation and assessment in education*. Paris: OECD.

OECD. (2014). TALIS 2013 results: An international perspective on teaching and learning. Paris: OECD.

OECD. (2016). Skills matter: Further results from the Survey of Adult Skills. Paris: OECD.

Office of the Qualifications and Examinations Regulator. (n.d.). GCSEs: The official student guide to the system. United Kingdom: Ofqual.

Plaut, D., & Jamierson-Eberhardt, M. (2015). Bringing learning to light: The role of citizen-led assessments in shifting the educaiton agenda. Washington, DC: Results for Development.

Sayare, S. (2013, 27 June, 2013). Rite of passage for French students reeives poor grade. *The New York Times*. Retrieved from http://www.nytimes.com/2013/06/28/world/europe/a-rite-of-passage-for-french-students-receives-a-poor-grade.html?src=me&ref=general&_r=2&

Stobart, G. (2008). *Testing Times: The Uses and Abuses of Assessment*: Taylor & Francis.

The College Board. (2017). SAT Suite of Assessments. Retrieved 31 January, 2017, from https://collegereadiness.collegeboard.org/sat?navId=gf-sat

Thissen, D., & Wainer, H. (2001). Test scoring. Mahwah, NJ: Erlbaum.

Tobin, M., Lietz, P., Nugroho, D., Vivekanandan, R., & Nyamkhuu, T. (2015). Using large-scale assessments of students' learning to inform education policy: Insights from the Asia-Pacific region. Melbourne, Australia.

UNESCO. (n.d.). The CapED Programme. Retrieved August 22, 2017, from http://en.unesco.org/themes/education-21st-century/caped

UNESCO Institute for Statistics. (2016). Global Alliance to Monitor Learning. Retrieved 8th August, 2017, from http://uis.openplus.ca/gaml/

UNESCO Institute for Statistics. (2017a). Capacity Development Tools. Retrieved August 22, 2017, from http://uis.unesco.org/en/node/561619

UNESCO Institute for Statistics. (2017c). UIS Catalogue of Learning Assessments 2.0. Retrieved April 24, 2017, from http://uis.unesco.org/en/cla-survey

Western and Northern Canadian Protocol for Collaboration in Education. (2006). *Rethinking*

classroom assessment with purpose in mind: Assessment for learning, assessment as learning, assessment of learning.

World Bank. (2012). Systems Approach for Better Education Results: What matters most for student assessment systems: A framework paper. *SABER Working Paper Series 1.*

World Bank. (2017). SABER: Student Assessment. Retrieved April 24, 2017, from http://saber.worldbank.org/index.cfm?indx=8&pd=5