

Data Alignment for SDG 4 reporting

A process to support the monitoring of learning outcomes
for Sustainable Development Goal 4

Concept Note

Table of contents

Data Alignment for SDG 4 reporting	2
Concept Note	2
GAML’s approach to SDG 4 reporting	2
UIS Reporting Scales (UIS RS)	3
Principles of Good Practice in Learning Assessment (GP-LA)	3
UIS Catalogue of Learning Assessments (CLA).....	3
System-Wide Analysis of Assessment Practices (SWAAP)	4
Data Alignment process	5
Options for Data Alignment.....	6
Components of Data Alignment.....	7
Domain alignment	8
Strand alignment.....	8
Level alignment.....	9
Data Alignment Record Tool (DART)	11
Supporting procedural consistency	11
Data alignment quality concepts	13
Fitness for purpose	13
Clarity and consistency of purpose.....	13
Objectivity and independence	13
Transparency and accountability	13
Technical rigour	14
Ethicality and fairness	14
Next steps	15
Bibliography	0

GAML's approach to SDG 4 reporting

The UNESCO Institute for Statistics (UIS), through its Global Alliance to Monitor Learning (GAML), is working on an approach to monitoring learning outcomes for Indicator 4.1.1 of the UN's Sustainable Development Goal 4 (SDG 4): Quality Education:

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 SDG Indicators, 2016)

Monitoring progress against Indicator 4.1.1 will require measurement of student outcomes at several different stages of learning in a broadly consistent way across education systems, to enable meaningful international dialogue about learning progress and how it may be supported. This is a challenge, given that learning and how it is measured varies widely across local contexts. Education systems make independent interpretations and decisions about what learning means, how it is described in curriculum, and how it is assessed and reported.

This does not mean that international alignment of learning outcomes is out of reach, but nor does it imply the need for the imposition of universal measurement processes. Rather, GAML's approach is informed by the following principles:

- a fit-for-purpose approach to international monitoring must be achieved that supports consistency in reporting of outcomes, while being flexible enough to accommodate a variety of approaches.
- the approach must provide access to tools and methods informed by international technical expertise, to assist education systems in building capacity
- the approach must be driven by commitment to meaningful assessment of learning as a global public good, which is critical to the goal of quality education for all.

GAML's approach comprises three key components:

- an interpretive backbone for understanding the SDG 4 learning domains, with sufficiently broad and detailed descriptions to accommodate the variations in interpretation of learning domains that exist across education systems.
- an independent articulation of good practice in assessment, informed by (but not prescribed by) principles and practices from leading assessment programs.
- collection of robust information about the current state of assessment and evaluation systems around the world, to guide reporting and capacity development.

The main components of GAML's approach 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.

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 website¹. 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.

System-Wide Analysis of Assessment Practices (SWAAP)

The System-Wide Analysis of Assessment Practices (SWAAP) is an analytic process, which will be used to guide education system leaders and stakeholders through analysis of the quality of their evaluation and assessment systems and the use of results in policy and practice. The findings from the SWAAP will inform the identification of capacity development needs in educational evaluation and assessment, and the creation of improvement plans for individual education systems through which these needs may be addressed. As at August 2017, a draft outline of the SWAAP was under development.

Figure 1 shows how the components of GAML’s approach fit together, to support alignment in reporting against 4.1.1, and broader capacity development support.

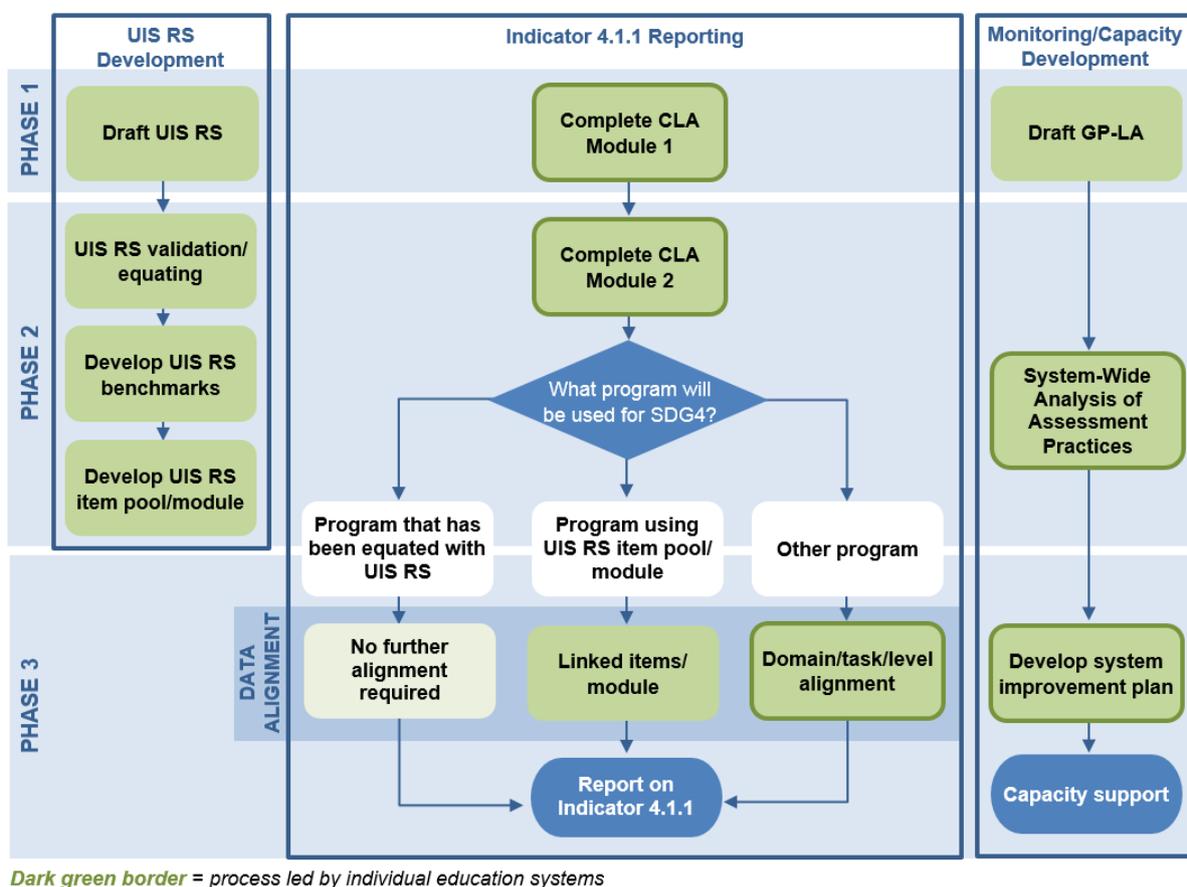


Figure 1: GAML’s approach to SDG 4 reporting

¹www.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

Data Alignment process

This document sets out the proposed Data Alignment process, which is a major component of GAML's approach to SDG 4 reporting. It has been developed by UIS and its technical partner, the Australian Council for Educational Research Centre for Global Education Monitoring (ACER-GEM), to guide further work on SDG 4 reporting and capacity development that will be undertaken through the GAML network.

The Data Alignment process will enable 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.

A longer-term aim of Data Alignment is to guide improvement in the alignment of assessment programs with the UIS RS, to enable more meaningful reporting against the SDG 4 indicators. Undergoing Data Alignment can help education systems to identify priorities for improving the quality of their reporting against Indicator 4.1.1, which may also inform the development of broader, system-level improvement plans for education evaluation and assessment. This longer-term aim reflects GAML's collaborative approach, which recognises knowledge-sharing and capacity-building opportunities as an integral part of global education monitoring.

Data Alignment is integrated with the other components of GAML's approach to SDG 4 reporting (outlined above) as follows:

- The UIS RS are used in Data Alignment as common scales against which the scales used in assessment programs can be referenced.
- The GP-LA is used in Data Alignment as principles of good practice to guide procedural consistency in reporting against Indicator 4.1.1.
- CLA Module 1 is used at the early stages of Data Alignment, to identify assessment programs that may be suitable for Indicator 4.1.1 reporting.
- CLA Module 2 will be complemented by the Data Alignment Reporting Tool, which will enable Data Alignment outcomes to be reported alongside Indicator 4.1.1 data.
- The Data Alignment process feeds into the SWAAP, by assisting education systems to analyse their assessment programs specifically in relation to SDG 4 reporting.

This document sets out a proposed process for Data Alignment, for consideration by GAML. It outlines the components of the process, and the tools that will be developed to

support and record them. It concludes by suggesting next steps for GAML and other stakeholders, to enable the process outlined in this document to be put into action.

Options for Data Alignment

The Data Alignment process is designed to be as efficient and cost-effective as possible, to maximise the number of education systems that can use the UIS RS to support consistency in their Indicator 4.1.1 reporting. Broadly, education systems may proceed through Data Alignment in three possible ways, depending on what kind of assessment program they are using:

1. Existing programs for which equating with UIS RS has occurred

Discussions are occurring through the GAML process to explore a potential equating method against the UIS RS, for selected existing assessment programs. This will enable education systems using an equated assessment program to easily understand the level of alignment between that program and the UIS RS and indicative standards, without any further Data Alignment process. The proposed process for equating involves item-based, rather than test-based, equating, allowing countries flexibility in incorporating linked items in existing assessment programs, rather than requiring the adoption of a program in its entirety.

2. Programs using items or modules mapped to the UIS RS

The equating process may involve the creation of a pool of items or modules, which have been mapped to the UIS RS. These items or modules would be designed to be incorporated into an assessment program that is not already equated to the UIS RS, to make it easier to show alignment with the UIS RS learning domains. Education systems incorporating UIS RS linked items or modules could examine the alignment of their assessment program with the UIS RS through empirical methods, using linked items to locate their assessments on the UIS RS scale.

3. Programs for which no Data Alignment has previously occurred

It is necessary to establish clear, consistent Data Alignment procedures for education systems using assessment programs that do not fall into either group identified above. This will ensure that such systems can still report against Indicator 4.1.1 in a way that maintains public confidence in the integrity of SDG 4 data, and will also guard against misinterpretation of differences in student learning that may in fact arise from differences in assessment processes or scales. This concept note sets out a proposal for these procedures, and identifies the tools that will be required to support them.

The details of the equating and validation process, and development of the UIS RS item pool or modules, will be set out in future UIS concept notes. The steps in the Data Alignment process set out below may be used in combination with these options for

empirical alignment, or as a stand-alone process to improve consistency in Indicator 4.1.1 reporting, where no UIS RS-linked empirical method is available. In the latter case, the Data Alignment process may provide valuable preparation for the subsequent pursuit of more rigorous empirical linking or equating of country assessment programs against the UIS RS, to achieve a further increase in confidence and international comparability.

For all pathways through Data Alignment, the end result is the same: reporting against the indicator, including clear articulation of the level of alignment between the country's assessment program and the UIS RS and GP-LA. For education systems whose pathway includes the Content or Process Alignment components of Data Alignment, the end results will also include valuable information that can be used to guide their evaluation and assessment improvement plans, and to provide a potential focus for capacity development support.

It is possible that a country may have multiple assessment programs suitable for Indicator 4.1.1 reporting; for example, separate assessment programs for different year levels. In this case, it is recommended that Data Alignment is repeated for each assessment, to capture any variations in how each program is structured. The structure of an assessment program is likely to prioritise different knowledge, skills, and understandings at different levels or for students participating at different ages.

Components of Data Alignment

The UIS RS has four key components with which an assessment program may align:

- *Domains*: the UIS RS provide a description of learning progression in the learning domains to be assessed (for example, reading and mathematics)
- *Strands*: within each domain, the UIS RS is organised into strands that reflect the knowledge, skills and understandings that are at the core of each domain
- *Levels*: the UIS RS identifies levels of learning progress for each domain, based on a continuum of development for each strand, and for the domain as a whole
- *Benchmarks*: benchmarks for minimum proficiency in reading and mathematics will be established on the UIS RS, for the purpose of SDG 4 reporting.²

Data Alignment examines the alignment of the country assessment program with each of these components, to yield an overall assessment of its alignment with the UIS RS. It may be that the assessment program aligns with some of these components but not others; for example, a program may align broadly with UIS RS at the domain level, but not in its coverage of specific strands, or levels of progress, within those domains. The Data Alignment process is therefore designed to arrive at a nuanced understanding of

² See June 2017 Discussion Paper: *Setting benchmarks on the UIS reporting scales*

alignment, which captures the specific ways in which the program is aligned with UIS RS, as well as any points of divergence.

The components of the Data Alignment process are as follows:

Domain alignment

The domain alignment process examines the overall conceptual alignment of the domains as defined in the assessment program, and as described in the UIS RS.

This step involves examining how the relevant domain is defined and operationalised in the assessment program, drawing on assessment frameworks, curriculum documents, and other available information as required. It is likely that the broad area of academic achievement being considered within the program will extend beyond the scope of the UIS RS; for example, a program capturing reading achievement may also capture writing, spelling, grammar, or punctuation. In this case, this step will identify the components of the program that align with UIS RS.

If the broad definition of the domain in an assessment program is aligned with the UIS RS, it is possible to proceed to the next step, and examine the strands. If the broad definition of the domain is markedly different from the UIS RS, then further alignment of the assessment program against the UIS RS is not recommended. The results of the domain alignment can still be recorded in the Data Alignment Record Tool (described below), to enable these differences to be taken into consideration in Indicator 4.1.1 reporting.

Strand alignment

The strand alignment process aims to answer two questions:

- *What degree of coverage of the UIS RS strands is evident in the assessment program?*

The second step requires the items used in the assessment to be categorised according to the strands of the UIS RS. This process will be made easier if the strands used in the assessment program are already broadly aligned with the UIS RS descriptors. Individual items may fit within multiple strands; in which case, the items should be categorised according to the *main* strand that they represent. Reference to curriculum documents and assessment frameworks may assist in this classification process.

- *What is the breadth of items for each strand?*

This step involves detailed examination of the items used in the assessment. For a strand to be well-aligned to the UIS RS, the assessment should include a range of item types and texts (for example, multiple choice and constructed response items, and personal and technical texts). Using different kinds of texts helps create tasks of varying complexity, and also represents the breadth of the domain as it is operationalised in that country.

Level alignment

The level alignment process aims to answer two further questions:

- *What range of learning progress is covered for each strand?*

This step involves categorising the items in the assessment for each strand, according to the different levels of learning progress they represent. For some assessment programs, it may be possible to arrange items on a continuum of learning progress, whereas others might only allow broad categories of learning progress to be defined (such as low, medium and high). Evidence should be collated to support the categorisation of items, such as the percentage of students answering each item correctly, or descriptions from the assessment framework or curriculum about the skills assessed. This serves a similar purpose to the step above, in ensuring that each strand covers a breadth of levels of learning progress (breadth of learning progress is distinct from breadth of item types and texts, as it may be possible to create a lower-level item using a complex text, and *vice versa*). The categorisation of items at this step will also be used in the next step, in aligning levels.

- *How are levels of proficiency defined in the assessment?*

This process will map items at different levels of learning progress from within the assessment program against the proficiency levels of the UIS RS. This is an important step for enabling education systems to report against SDG 4, because it is essential that the assessment have coverage of at least one of the SDG 4.1.1 benchmarks.

Aligning levels against the UIS RS basically involves three steps. All of these steps may be applied, or any combination of them, depending on what information is available for the assessment program:

1. *Scale mapping* involves aligning the overall scale of achievement between the assessment program and the UIS RS. This step will occur if the country assessment defines a scale or continuum that describes the knowledge, skills, and understandings that are observed at certain points. Such a scale is likely to be defined qualitatively (such as in descriptive standards in a curriculum document or assessment framework), and may also be based on statistical information (such as proficiency standards yielded from psychometric studies or reporting).
2. *Item mapping* involves a qualitative rating of the items compiled to illustrate different levels of learning progress on the assessment program, relative to the UIS RS illustrative items at similar levels. This is an iterative process, which may require identification of additional examples of items at certain levels, if the initial selection is not adequate to cover the entire range. Where the location of items from the assessment on the UIS RS scale is not obvious, supporting information may be used for clarification, including descriptions of proficiencies or curriculum, or statistical information, such as proportion of students who get that item correct.

Data Alignment Record Tool (DART)

A tool will be created to record the outcomes of the Data Alignment process: the Data Alignment Record Tool (DART). The DART will enable the information from each of the steps above to be recorded, and used to generate a series of statements about the alignment of the content of the assessment program with the UIS RS. It is recommended that these statements are incorporated into SDG 4 reporting, and may also be used to inform the country's assessment improvement plan, if desired.

For example, DART will include a facility for entering the results of the analysis of items for each strand, in terms of their coverage across types of items. If the range of item types or texts is limited (for example, if all questions are multiple choice), DART will generate a statement such as:

[Assessment] has a limited variety of task types in [strand].

Where the input shows that the assessment includes a broad range of items and texts, DART will indicate that no limitations need be reported in this area. This information will help to understand where any limitations in alignment with the UIS RS might occur.

The Data Alignment process will generate a series of statements that clearly identify any limitations in alignment with the UIS RS, including:

- any domains or strands in which coverage is limited;
- any differences in the way proficiency levels are defined; and
- any differences in the points at which minimum proficiency benchmarks are set.

In addition to these automatically-generated statements, DART will include a facility for recording free-text comments in relation to all the areas above, recognising that Data Alignment will often involve a level of subjective expert judgement. The tool will also include a facility for recording the evidence used in any step in the process. This will help education systems to record and communicate the rationale for their Data Alignment outcomes to UIS, and to guide UIS in using this information to improve transparency in SDG 4 reporting.

Importantly, limitations in alignment with the UIS RS do not necessarily imply limitations in the quality of the assessment program, relative to its purpose and context. In many cases, valid reasons for discrepancies may exist. These will require time and negotiation to resolve, to support longer-term consistency in Indicator 4.1.1 reporting.

Supporting procedural consistency

Procedural consistency is a highly desirable component of any program that aims to measure learning outcomes consistently across diverse education systems (Lietz et al 2017). However, the Data Alignment process is unusual in that it is intended to be

applied to a wide variety of assessment programs, each of which will have their own procedural standards designed to respond to local challenges and needs. It is therefore not appropriate to specify rigid standards to apply in the Data Alignment process, in the way that might be possible for standardised assessment programs.

It nevertheless remains desirable for UIS to collect some basic procedural information from education systems about the data that is provided against Indicator 4.1.1, to understand the level of confidence with which results can be reported. Three technical aspects of the data collection process are considered essential for this purpose:

1. A sample that is representative of the national population

This includes representation of all major population groups for which educational outcomes may vary (such as socio-economic, language or disability groups), and geographic coverage across metropolitan, rural/regional and remote locations.

2. A response rate that is appropriate for the chosen sampling method

UIS will not set a threshold response rate for Indicator 4.1.1 reporting, but will aim to obtain information about the response rate from each education system reporting against Indicator 4.1.1. Education systems are therefore encouraged to keep robust documentation about response rates, and to make every effort to maximise them.

3. Translation procedures that ensure consistency across language groups

Where a program is administered in multiple languages (for example, in an education system with students from multiple language groups), it is important to ensure that variation in test results across language groups does not result from inconsistencies in the test instrument arising from translation. This may be ensured through rigorous translation methods (such as forward-and-back translation by skilled translators), and examining item-level data for variation by language groups.

UIS will work to incorporate efficient mechanisms for reporting on these key aspects of procedural consistency into existing GAML processes and tools. A facility for recording this information may be incorporated into the DART tool described in this document, or into further development of Module 2 of the CLA. If information about these procedures is not available for an assessment program, this does not preclude reporting against Indicator 4.1.1, but provides insight into potential issues that may affect the consistency of the data. This opens the way for dialogue between the education system and GAML about how the data will be reported, and help to identify any capacity building opportunities that may strengthen the assessment program in the future.

There are many other aspects of the data collection process that are also important to the quality and rigour of large-scale assessments. These include the consistency of field operations, the expertise and resourcing of the assessment team, and the security measures taken to protect the integrity of the data. While UIS will not be collecting information in these areas – as they are likely to vary too widely across programs to

enable specific procedural standards to be applied – the GP-LA provides valuable guidance for countries in all of these areas, to maximise the quality of their data. The system analysis and capacity development components of GAML’s broader approach will help education systems to examine their practices against the GP-LA, and develop improvement plans that are achievable and relevant to their context.

Data alignment quality concepts

Many of the principles that guide best practice in conducting assessments of student learning also apply to the conduct of Data Alignment itself. This section outlines how the six key quality concepts for assessment outlined in the draft GP-LA apply to the Data Alignment process.

Fitness for purpose

Data Alignment is unlikely to involve a uniform pathway for all types of large-scale assessment programs. In some instances, Data Alignment may be a mere formality, because the assessment program’s alignment with the UIS RS may have already been assured through another process. In other education systems, Data Alignment will involve a comprehensive examination of a program that has not previously undergone international alignment of any kind. The Data Alignment approach adopted in each country should represent the most efficient, cost-effective and timely option available.

Clarity and consistency of purpose

While the Data Alignment pathway may vary across education systems, its purpose will remain clear and consistent across all education systems: to enable robust international reporting against Indicator 4.1.1. This well-defined goal will help maintain focus in the Data Alignment process as a contained and manageable process, and prevent undue diversion or expansion to unrelated goals.

Objectivity and independence

Objectivity and independence will aid in maintaining accuracy and integrity in the Data Alignment process. The organisation undertaking the process is encouraged to declare any actual or potential conflicts of interest at the commencement of the Data Alignment process, and resolve these to the satisfaction of all parties involved. Oversight of SDG 4 reporting from GAML and UIS will also help to ensure that education systems are supported to make objective judgements about Data Alignment.

Transparency and accountability

Transparency in the Data Alignment process will be supported by documentation of the outcomes of the process, supported by the tool described above. The outcome of Data Alignment is also an opportunity for education systems to demonstrate transparency and accountability in their assessment programs. This will help to foster confidence in

the SDG 4 reporting program, as well as helping to identify potential efficiencies in addressing capacity-building needs, where issues are shared across multiple systems.

Technical rigour

Data Alignment is driven by GAML's commitment to achieving the greatest possible technical rigour in SDG 4 reporting, within a fit-for-purpose approach that continues to accommodate variability between systems and assessment programs. The resources that guide Data Alignment (UIS RS and GP-LA) are informed by technical considerations of how learning progressions may be measured, and good practices used around the world. The Data Alignment process helps education systems to consider the technical rigour of their assessment programs, and provides a basis for dialogue with international assessment experts about how rigour can best be assured.

Ethicality and fairness

The principle of ethicality and fairness requires that an assessment program minimises as much as possible the risk of harm to assessment participants. Undergoing Data Alignment may be perceived as creating a risk of harm to participating education systems, especially if the process is regarded as imposed or critical in nature. The process proposed in this concept note addresses this risk by locating leadership of the Data Alignment process within each education system, with the GAML network providing expert oversight and support. This places the interests of each education system at the heart of the Data Alignment process, and ensures that SDG 4 reporting remains a collaborative, ethical shared enterprise to improve the quality of student learning around the world.

Next steps

It is recommended that the GAML network discuss and agree on the final set of tools and methods for Data Alignment; taking into account the points at which the goals of Data Alignment differ from those of standardised regional or international assessments.

It is also recommended that the GAML network consider appropriate governance protocols for Data Alignment, including procedures to review information provided by education systems through the Data Alignment Record Tool, and to make decisions about the consequences of this information for SDG 4 reporting.

GAML may consider establishing a panel or board to oversee Data Alignment, and working parties to examine the information collected through the proposed tools. The Assessment Implementation Task Force may be an appropriate body to lead this work.

Another important consideration for GAML is how information collected through the Data Alignment process may be used in public reporting against Indicator 4.1.1 (if at all); recognising the potential sensitivity of information about any limitations in the alignment of assessment programs, relative to an internationally-validated scale.

Bibliography

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards in Educational and Psychological Testing. (2014). *Standards for Educational and Psychological Testing*. Washington DC: AERA.
- Australian Bureau of Statistics. The ABS Data Quality Framework. Retrieved 30 May, 2016, from <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Quality:+The+ABS+Data+Quality+Framework>
- Australian Bureau of Statistics, & National Statistical Service. Data Fitness: A guide to keeping your data in good shape.
- Caldwell, N., Foy, P., Martin, M. O., Mullis, I. V. S., & Sibberns, H. (1999). Technical Standards for IEA Studies. M. O. Martin, K. Rust & R. J. Adams (Eds.), Retrieved from http://www.iea.nl/publication_list.html?&no_cache=1
- Clarke, M. (2012). What Matters Most for Student Assessment Systems: A framework paper SABER Working Paper Series, Retrieved from <http://saber.worldbank.org/index.cfm?indx=8&pd=5&sub=4>
- Council of Europe: Language Policy Unit. (2011). Common European Framework of Reference for Languages: Learning, teaching, assessment. Retrieved from http://www.coe.int/t/dg4/linguistic/cadre1_en.asp
- Educational Testing Service. (2014). ETS Standards for Quality and Fairness. Retrieved from <https://www.ets.org/about/fairness/>
- Ehling, M., Körner, T., Bergdahl, M., Elvers, E., Földesi, E., Kron, A., *et al.* (2007). EuroStat Handbook on Data Quality Assessment Methods and Tools. Manfred Ehling & T. Körner (Eds.), Retrieved from <http://unstats.un.org/unsd/dnss/docs-nqaf/Eurostat-HANDBOOK%20ON%20DATA%20QUALITY%20ASSESSMENT%20METHODS%20AND%20TOOLS%20%20I.pdf>
- European Statistical System. (2011). European Statistics Code of Practice. Retrieved from <http://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice>
- European Statistical System. (2011). Quality Assurance Framework for the European Statistical System. Retrieved from <http://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice>
- Gregory, K. D., & Martin, M. O. (2001). Technical Standards for IEA Studies: An Annotated Bibliography. Retrieved from http://www.iea.nl/publication_list.html?&no_cache=1
- Independent Expert Advisory Group on a Data Revolution for Sustainable Development. (2014). A World that Counts: Mobilising the Data Revolution for Sustainable Development.
- Inter-Agency and Expert Group on Sustainable Development Goal Indicators. (2016). Report of the Inter-Agency and Expert Group on Sustainable Development Goal Indicators.

- Inter-Agency and Expert Group on Sustainable Development Goal Indicators. (2016). Update on the work to finalize the proposals for the global indicators for the Sustainable Development Goals.
- International Institute for Educational Planning. (2016). Financing Education 2030 - The IIEP Letter (Vol. 1): United Nations Educational, Scientific and Cultural Organisation.
- International Monetary Fund. Data Quality Reference Site, from <http://dsbb.imf.org/Pages/DQRS/EDQ.aspx>
- International Monetary Fund. (2006). Data Quality Assessment Framework: A Factsheet.
- International Monetary Fund. (2012). Data Quality Assessment Framework - Generic Framework.
- International Monetary Fund. (2013). The General Data Dissemination System - Factsheet. Washington DC: International Monetary Fund.
- International Monetary Fund. (2013). The General Data Dissemination System : guide for participants and users. Retrieved from <http://dsbb.imf.org/pages/gdds/home.aspx>
- Izard, J. (2005). Module 6: Overview of test construction. In K. N. Ross (Ed.), *Quantitative Research Methods in Educational Planning*. Paris, France: UNESCO International Institute of Educational Planning. Retrieved from <http://www.iiep.unesco.org/en/library-resources/briefs-papers-tools>.
- Izard, J. (2005). Module 7: Trial testing and item analysis in test construction. In K. N. Ross (Ed.), *Quantitative Research Methods in Educational Planning*. Paris, France: UNESCO International Institute of Educational Planning. Retrieved from <http://www.iiep.unesco.org/en/library-resources/briefs-papers-tools>.
- Johansone, I. (2012). Operations and Quality Assurance. In Michael O Martin & Eva V.S. Mullis (Eds.), *Methods and Procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from http://timssandpirls.bc.edu/methods/pdf/TP_Operations_Quality_Assurance.pdf.
- Kenneth N. Ross, T. Neville Postlethwaite, Marlaine Lockheed, Aletta Grisay, & Gabriel Carceles Breis. (1990). Improving data collection, preparation and analysis procedures: a review of technical issues. In K. N. Ross & L. Mählck (Eds.), *Planning the Quality of Education: The collection and use of data for informed decision-making*. Paris & Oxford: UNESCO International Institute for Educational Planning & Pergamon Press.
- Lietz, P., Cresswell, J. C., Rust, K. F., & Adams, R. J. (2017). *Implementation of Large-Scale Education Assessments*. Somerset : John Wiley & Sons, Incorporated, 2017.
- Livingstone, I. D. (2005). Module 2: From educational policy issues to specific research questions and the basic elements of research design. In K. N. Ross (Ed.), *Quantitative Research Methods in Educational Planning*. Paris, France: UNESCO International Institute of Educational Planning. Retrieved from <http://www.iiep.unesco.org/en/library-resources/briefs-papers-tools>.
- National Statistical Service. Data Quality Online. Retrieved 30 May, 2016, from <https://www.nss.gov.au/dataquality/index.jsp>
- Organisation for Economic Cooperation and Development. (2012). Quality Framework

- and Guidelines for OECD Statistical Activities. Retrieved from <http://www.oecd.org/std/qualityframeworkforoecdstatisticalactivities.htm>
- Organisation for Economic Cooperation and Development. (2013). Technical Report of the Survey of Adult Skills (PIAAC) - pre-publication copy: OECD.
- Organisation for Economic Cooperation and Development. (2014). PIAAC Technical Standards and Guidelines.: OECD.
- Organisation for Economic Cooperation and Development. (2014). PISA 2012 Technical Report Retrieved from <https://www.oecd.org/pisa/pisaproducts/pisa2012technicalreport.htm>
- Organisation for Economic Cooperation and Development. (2015). *PISA 2015 Technical Standards.*: OECD.
- Postlethwaite, T. N. (2005). Module 1: Educational research: some basic concepts and terminology. In K. N. Ross (Ed.), *Quantitative Research Methods in Educational Planning*. Paris, France: UNESCO International Institute of Educational Planning. Retrieved from <http://www.iiep.unesco.org/en/library-resources/briefs-papers-tools>.
- Rashtriya Madhyamik Shiksha Abhiyan Technical Cooperation Agency. (2016). *Large-Scale Learning Assessments: A Handbook for the Indian Context*. Retrieved from http://rmsaindia.org/en/?option=com_pdf&view=__recentdocs&catid=all&type=1&main=0&Itemid=224
- Ross, K. N. (2005). Module 3: Sample design for educational survey research. In K. N. Ross (Ed.), *Quantitative Research Methods in Educational Planning*. Paris, France: UNESCO International Institute of Educational Planning. Retrieved from <http://www.iiep.unesco.org/en/library-resources/briefs-papers-tools>.
- Schleicher, A., & Saito, M. (2005). Module 10: Data preparation and management. In K. N. Ross (Ed.), *Quantitative Research Methods in Educational Planning*. Paris, France: UNESCO International Institute of Educational Planning. Retrieved from <http://www.iiep.unesco.org/en/library-resources/briefs-papers-tools>.
- Siniscalco, M. T., & Auriat, N. (2005). Module 8: Questionnaire design. In K. N. Ross (Ed.), *Quantitative Research Methods in Educational Planning*. Paris, France: UNESCO International Institute of Educational Planning. Retrieved from <http://www.iiep.unesco.org/en/library-resources/briefs-papers-tools>.
- Statistics Canada. (2002). *Statistics Canada's Quality Assurance Framework*. Ottawa: Ministry of Industry.
- UNESCO Institute for Statistics. (2016). *Meet the Education 2030 Data: A guide to the indicators to monitor SDG 4 - Education 2030 (Brief about indicator 4.1.1)* Retrieved 10 June, 2016, from <http://www.uis.unesco.org/Education/Pages/meet-the-education-2030-data-indicator-4-1-1.aspx>
- UNESCO Institute for Statistics. (2016). *Meet the Education 2030 Data: A guide to the indicators to monitor SDG 4 - Education 2030 (Brief about indicator 4.1.2)* Retrieved 10 June, 2016, from <http://www.uis.unesco.org/Education/Pages/meet-the-education-2030-data-indicator-4-1-2.aspx>
- UNESCO Institute for Statistics, & World Bank. (22 October 2012). *Ed-EDQ wiki*

- Retrieved 30 May, 2016, from
http://EDQ.uis.unesco.org/index.php?title=Main_Page
- United Nations General Assembly resolution 68/261. Fundamental Principles of Official Statistics. *A/RES/68/261 (from 29 January 2014)*.
- United Nations Statistics Division. (2015). United Nations Fundamental Principles of Official Statistics: Implementation Guidelines. Retrieved from
<http://unstats.un.org/unsd/dnss/gp/impguide.aspx>
- Withers, G. (2005). Module 5: Item writing for tests and examinations. In K. N. Ross (Ed.), *Quantitative Research Methods in Educational Planning*. Paris, France: UNESCO International Institute of Educational Planning. Retrieved from
<http://www.iiep.unesco.org/en/library-resources/briefs-papers-tools>.
- Wolf, R. M. (2005). Module 4: Judging educational research based on experiments and surveys. In K. N. Ross (Ed.), *Quantitative Research Methods in Educational Planning*. Paris, France: UNESCO International Institute of Educational Planning. Retrieved from <http://www.iiep.unesco.org/en/library-resources/briefs-papers-tools>.
- World Bank. (n.d.). SABER - Student Assessment Questionnaires and Rubric: World Bank.