



Eligibility criteria for reporting **SDG 4.1.1**

Colin Watson

7 December 2023

Australian Council *for* Educational Research

Criteria overview

The criteria were selected to ensure consideration of the quality of the assessment instrument and its implementation.

- **Criterion 1** – is the assessment sufficiently aligned to the MPL?
- **Criterion 2** – is there evidence that the items in the assessment have been reviewed qualitatively and quantitatively to determine their suitability for inclusion in the assessment?
- **Criterion 3** – is the sample of learners that took the assessment representative of the population against which the results will be reporting?
- **Criterion 4** – is there evidence that the assessment was administered in a standardised way?
- **Criterion 5** – are the outcomes of the assessment sufficiently reliable?

Criterion 1 – Alignment (MPLa Option 1)

All assessments must contain a minimum of 20 items

- **Reading** – minimum of 10 score-points assessing *decoding*, 5 score-points assessing *listening comprehension/comprehension of spoken or signed language* and 5 score-points assessing *reading comprehension*. The assessment must also cover 5 of the 9 subconstructs at grade 2 in the GPF.
- **Mathematics** –minimum of 10 score-points assessing *number and operations*, 5 score points assessing *measurement and geometry* and 2 score-points assessing *statistics and probability* and *algebra*. The assessment must also cover 7 of the 14 subconstructs at grade 2 in the GPF.

Criterion 1 – Alignment (MPLa Option 2)

All assessments must contain a minimum of 20 items

- **Reading** – minimum 10 score-points assessing *reading comprehension* and the assessment must cover both reading comprehension subconstructs at grade 2 in the GPF. The remaining items can be drawn from any of the domains (*decoding, listening comprehension or reading comprehension*).
- **Mathematics** – minimum 10 score-points assessing *number and operations* and the assessment must cover all four *number and operations* subconstructs at grade 2 in the GPF. The remaining items can be drawn from any of the domains (*number and operations, measurement, geometry, statistics and probability or algebra*).

Criterion 1 – Alignment (MPLa)

	Advantages	Disadvantages
Option 1: All domains	<ul style="list-style-type: none">• The domain requirements align with the description in the MPL• Encourages countries where learners are still focused on foundational skills in reading to include all domains in their assessments	<ul style="list-style-type: none">• Countries that assess with an assessment of reading comprehension only would not be able to report against SDG 4.1.1a• Countries that are focused only on foundational skills in mathematics (number and operations) would not be able to report against SDG 4.1.1a
Option 2: Minimal domains	<ul style="list-style-type: none">• Maximises the number of assessments that can be used for reporting	<ul style="list-style-type: none">• Assessments being used for reporting are likely to contain different domains outside the core elements required

Criterion 1 – Alignment (MPLb)

All assessments must contain a minimum of 20 items

- **Reading** – the minimum 20 items must all relate to the *reading comprehension* domain. There should be 5 score-points assessing the *retrieve information* construct and 5 score-points assessing the *interpret information* construct from the GPF. The assessment should also cover 4 of the 8 reading comprehension subconstructs at grade 5 in the GPF.
- **Mathematics** – minimum of 10 score-points assessing *number and operations*, 5 score-points assessing *measurement and geometry* and 5 score-points assessing *statistics and probability* and *algebra*. The assessment must also cover 12 of the 21 subconstructs at grade 5 in the GPF.

Criterion 1 – Alignment (MPLc)

All assessments must contain a minimum of 20 items

- **Reading** – the minimum 20 items must all relate to the *reading comprehension* domain. There should be 5 score-points assessing the *retrieve information* construct, 5 score-points assessing the *interpret information* construct and 5 score-points assessing the *reflect on information* construct from the GPF. The assessment should also cover 5 of the 10 *reading comprehension* subconstructs at grade 8 in the GPF.
- **Mathematics** – there should be a minimum of 10 score-points assessing *number and operations*, 5 score points assessing *measurement and geometry* and 5 score-points assessing *statistics and probability* and *algebra*. The assessment must also cover 12 of the 21 subconstructs at grade 8 in the GPF.

Criterion 2 – Item review

- To be suitable for reporting against SDG 4.1.1, there must be evidence that the items in the assessment have followed an appropriate test development process.
- Items must have been reviewed quantitatively and qualitatively to determine their suitability for inclusion in the assessment.
- Details of the test development process followed, and evidence that suitable qualitative and quantitative reviews have been carried out should be in the public domain as part of a technical report.

Criterion 2 – Item review (qualitative review)

The qualitative review should consider whether:

- Each assessment item is considered appropriate by relevant experts for inclusion in the assessment
- The scoring guides are consistent with what the item is intended to measure.

Criterion 2 – Item review (quantitative review)

The quantitative review should consider whether:

- Item difficulty (e.g., item facility (CTT) or item location on the scale (IRT)) is appropriate for the grade level
- Item discrimination (e.g., Discrimination Index for each item is generally greater than 0.2, with any exceptions rationalized or the distractors in a multiple-choice item should be negatively correlated with ability).

Criterion 3 – Sample

- To be suitable for reporting against SDG 4.1.1, there must be evidence that the group of learners who took the assessment is representative of the population against which the results will be reported.
- Details of the target population definition, population coverage, design effect, sampling frame development and the post sampling treatment of data to account for any issues identified in the achieved sample (for example weightings used to account for sampling bias) should be described in a technical report. This report must be made publicly available.

Criterion 3 – Whole cohort vs. Sample

- Where the assessment is administered to the whole cohort, the project team should consider whether there are any subgroups of the population that have been systematically excluded. For example, learners not in school, learners in conflict-affected areas, learners with special educational needs.
- Any systematic exclusions should be noted for reporting along with an estimate of the number of exclusions, and the exclusions as a proportion of the population.
- Where the assessment is administered to a sample of the population, evidence must be provided to demonstrate the representativeness of the sample. The margin of error should be 5 percent or less at the 95 percent confidence level.

Criterion 4 – Administration

- To be suitable for reporting against SDG 4.1.1, there must be evidence that the assessment was administered in an appropriate and standardised way (for example, administration conditions were consistent, or length of time to administer the assessment was adhered to).
- Documentation relating to administration should be in the public domain. Details of administrator training, quality assurance procedures and quality assurance outcomes should also be made available publicly.

Criterion 4 – Administration quality assurance

- Administration guides must be reviewed for clarity and monitoring of the implementation must be undertaken.
- Any incidents of inappropriate administration, identified through monitoring or reporting of concerns, should be recorded.
- Where significant incidents of inappropriate administration are recorded, relevant results should be excluded from the outcomes.
- This will require additional checks to confirm that this does not affect the representativeness of the sample.

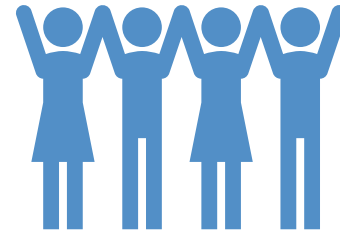
Criterion 5 – Reliability

- To be suitable for reporting against SDG 4.1.1, the value of coefficient alpha/Cronbach's alpha (or equivalent reliability statistic) for the assessment must be greater than or equal to 0.7.
- There must be evidence of appropriate quality assurance arrangements for any human-scored items.
- The approach to quality assurance must be documented and provided to UIS as a minimum, though publication is advised. UIS must also be provided with statistical outcomes from the quality assurance arrangements, for example agreement rates between scorers or with pre-agreed scores.

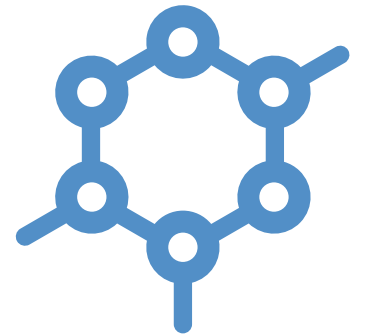
Criterion 5 – Reliability of scoring

- As a minimum, this quality assurance should take place during the training for those responsible for scoring the items.
- Ideally, however, such quality assurance should take place during the live administration.
- The method of quality assurance may be determined locally, though common procedures include scoring of items with a pre-agreed score to check that the scorer assigns the same score or double scoring of a sample of responses to check levels of agreement.

ACER



Thank you



References

- Eligibility criteria for reporting against SDG 4.1.1 – Global Alliance to Monitor Learning (GAML) paper, December 2023. https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2023/12/Eligibility-criteria-for-reporting-against-SDG-4.1.1_final.pdf