

**COVID-19: Monitoring Impacts on Learning Outcomes
(MILO)**

Item Review Guidelines: Mathematics

March 2021

The Global Education Monitoring (GEM) Centre supports education stakeholders to collect, analyse and use high-quality data to improve learning outcomes. The GEM Centre is a long-term partnership between the Australian Council for Educational Research (ACER) and the Australian Government's Department of Foreign Affairs and Trade (DFAT).



Acknowledgments

This project, the Assessment and Study of COVID Impact on Learner Progress, is referred to as the COVID-19 MILO (Monitoring Impacts on Learning Outcomes) project. This UNESCO Institute for Statistics (UIS) project is funded by the Global Partnership for Education (GPE).

The Australian Council for Educational Research (ACER) is the technical partner for this project. Support is provided from the Global Education Monitoring (GEM) Centre, an ACER initiative in partnership with the Australian government's Department of Foreign Affairs and Trade. The GEM Centre is also contributing to the UIS Global Item Bank. Technical and implementation support, and contribution to the assessment item pool, is provided by CONFEMEN.

Contents

Acknowledgments	2
Notes for completing the Item Review Spreadsheet.....	4
Meaning of categories.....	4
Construct Validity.....	4
Translatability.....	4
Cultural issues	5
Technical Criteria	5
Other comments.....	7
Overall Rating	7

Notes for completing the Item Review Spreadsheet

An Excel spreadsheet has been prepared to assist you in recording your review of the items. This document explains the terms used in that spreadsheet and gives guidance on how the spreadsheet should be completed.

Meaning of categories

The meaning of each category is described in the following sections.

Construct Validity

Construct validity refers to the extent to which a test measures that which it claims to be measuring. A necessary step in ensuring construct validity is the definition of what is to be measured. The draft Minimum Proficiency Levels (MPLs) are accompanied by a glossary defining key terms for each of the constructs of reading and mathematics. In addition, the Global Proficiency Framework outlines the domain/ constructs/ sub-constructs to be used. Taken together, these documents provide the necessary guidance about how the constructs being measured have been defined. These documents (Revised Minimum Proficiency Levels and Global Proficiency Framework Overview) have been provided, and should be read prior the commencement of item review.

In the context of review of possible items for the Global Item Bank, it will be necessary to consider whether:

- each assessment task measures an element of the skills, knowledge and understanding contained in the construct, as represented in the MPLs (and if so, which element)
- each assessment task provides a piece of evidence about the extent to which students have developed the skills, knowledge and understanding defined in the MPLs and their associated documentation.

For each dot point above, please answer “yes” or “no” in the Excel spreadsheet. In the case of the first dot point, reviewers are asked to also identify the element of mathematics or reading that is addressed by the item in question.

Translatability

Material for review is in either English or French, but is likely that the material will be translated into other languages. Therefore an important part of the quality assurance procedure will involve reviewers considering whether the material appears to pose any difficulties in relation to translatability. For example, stimulus material for reading items

sometimes includes language-specific qualities that are difficult to capture in translation (e.g. word play, rhythm). Item format and issues of grammar and syntax also need to be considered. For example, sentence completion items can pose difficulties in translation because word order in sentences varies across languages (eg, Japanese= subject-object-verb, English= subject-verb-object, Tagalog= verb-subject-object). Use of the passive voice can also be an issue, as many languages do not have an equivalent. Please answer “Yes” or “No” in the “translation issues” column, and then use the adjacent comments column to elaborate if necessary.

Cultural issues

The Global Item Bank will contain items from multiple sources. Therefore reviewers should consider whether there are any cultural concerns about items or pieces of stimulus. That is, reviewers should reflect on whether the topic is appropriate for students in many different countries. This question is not about student familiarity with a context, as when items are drawn from multiple sources, it is not possible to ensure settings that are familiar to students in all countries. Rather, cultural concerns may arise when the contexts of references would not be acceptable or appropriate in a given country for reasons based on, for example, social norms or religious beliefs. Please answer “Yes” or “No” in the “cultural concerns” column, and then use the adjacent comments column to elaborate if necessary.

Technical Criteria

Items being considered for inclusion in the Global Item Bank should be evaluated against a set of technical criteria. Reviewers are asked to consider issues of Clarity and correctness, Centrality and Difficulty.

Clarity and correctness

Stimulus

When considering the clarity and correctness of stimulus materials, reviewers should consider the following:

- Is the stimulus cohesive and coherent? Are the main ideas clearly developed?
- Is the presentation/wording of stimulus as clear and succinct as possible?
- Is the stimulus self-contained and not reliant on irrelevant assumed knowledge?
- Does the stimulus require contextualisation and, if so, is this brief, concise and direct?
- Does the stimulus avoid unintended ambiguity or unnecessary distracting information?

- Does the stimulus support the assessment of key skills? Where appropriate, can it be read on multiple levels to assess deeper understanding?
- Is the complexity/length of the stimulus justifiable for the items required?

Items

The items included in the testlets will all be closed format: multiple choice, complex multiple choice or closed constructed response (where the answer can be defined narrowly – e.g. a number).

The following factors should be considered when rating the quality of multiple-choice items:

- Is the item unambiguous and clearly worded?
- Is the layout of the item clear?
- Is the Key the only correct, or only good answer?
- Is the Key of about the same length, level of language and complexity as the distractors?
- If the Key is the longest option, does the key blend in with the other options?
- Does the Key stand out because it is the only option that includes technical or complex language and symbols, and might be attractive to students for the wrong reasons? [This is to be avoided.]
- Are the distractors plausible (but clearly incorrect)?
- Are the distractors unique in their focus (i.e. content does not overlap)?

Please answer “Yes” or “No” in the “Concerns with clarity and correctness” column, and then use the adjacent comments column to elaborate if necessary.

Centrality

This field applies only to items – not to stimulus. (The cells that apply to stimulus material will be greyed out for this criterion.) Here, reviewers should consider whether the item assesses important/ central content, and whether any inbuilt complexity in the item is justifiable in relation to eliciting evidence about the construct as defined.

Specifically, consider whether an item focusing on a particular central concept contains some minor element or nuance (e.g. in mathematics, a difference in measurement units that is not made explicit; or a requirement in a constructed response marking guide to round an answer where rounding is not the central concept) that would mean a student could fail to gain any credit for their response, despite having a good understanding of the central concept. Please answer “Yes” or “No” in the “Concerns regarding centrality” column, and then use the adjacent comments column to elaborate if necessary.

Difficulty

In this field, consider whether the item is appropriate for presentation to upper primary students, remembering that the assessment will need to contain items representing a range of difficulty in order to obtain information about all students in the cohort, from the least to most able. It is important to consider curriculum exposure in this section. Specifically, is there any content in the item that it would be unreasonable to expect an upper primary student to have had the opportunity to learn as yet, and cannot be worked out from “first principles” using existing knowledge? Does the item contain specialised terminology or convention that is unlikely to have been encountered in the course of their schooling as yet? If in doubt, please refer to the guidelines provided as to the types of mathematical content associated with different grade levels. Please answer “Yes” or “No” in the “Appropriate difficulty” column, and then use the adjacent comments column to elaborate if necessary.

Other comments

If you have any comments that are not covered under the previous headings, please free feel to add any comments in the “other comments” column.

Overall Rating

In this section, please provide a rating (1 to 5) that reflects your overall opinion about the suitability of the unit/ items for inclusion in the assessment. In this rating scheme (1 to 5), 1 = low; 2 = moderately low; 3 = medium level, 4 = moderately high; 5 = high. Please do not use other “ratings” such as 0 or decimal fractions (e.g. 3.5): the excel spreadsheet has been formatted to disallow such ratings.