Australian Government
Department of Foreign Affairs and Trade

ACER

WG/GAML/5

International Standard Setting Exercise

Final Report
November 2022

The Global Education Monitoring (GEM) Centre drives improvements in learning by supporting the monitoring of educational outcomes worldwide. 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).

Global Alliance to Monitor Learning meeting
23 November 2022

unesco
Institute for Statistics

GLOBAL ALLIANCE TO MONITOR LEARNING

International Standard Setting Exercise: Final Report
Australian Council for Educational Research Limited © 2021

The Australian Council *for* Educational Research
19 Prospect Hill Road
Camberwell VIC 3124
Phone: (03) 9277 5555
ABN 19 004 398 145

www.acer.org
www.acer.org/au/gem

978-1-74286-688-8

Recommended citation
The Australian Council for Educational Research. (2022). International Standard Setting Exercise. https://doi.org/ 978-1-74286-688-8

Acknowledgements
This publication has been funded by the Australian Government's Department of Foreign Affairs and Trade, and the Australian Council for Educational Research Ltd. The views expressed in this publication are the author's alone and are not necessarily the views of the Australian Government.

Contents

Foreword

The GEM Centre provides technical support to the United Nations Educational, Scientific and Cultural Organization (UNESCO) Institute for Statistics (UIS), which has been mandated to monitor the progress of countries towards achieving the education goals of Sustainable Development Goal 4 (SDG 4) to "ensure inclusive and equitable quality education and to promote lifelong learning opportunities for all" (United Nations, 2021). The GEM Centre sponsors and contributes to public goods and activities that facilitate education systems reporting against SDG 4 in a globally consistent way. Consistent and high quality monitoring of student learning will help systems understand the strengths they have and the challenges they face. High quality monitoring provides evidence to inform the development of policies and practice to improve student learning. The International Standard Setting Exercise is a step closer to building a high quality global framework for countries around the world to use their own regional or national learning assessments for monitoring progress towards achieving SDG 4.

Introduction

Data on outcomes in the learning areas of reading and mathematics are central to monitoring and reporting countries' progress towards achieving SDG indicator 4.1.1, by 2030.

Large-scale assessments are widely recognised as a primary source for such data, however they vary in method and scope, posing major challenges for global monitoring. Also, not all countries participate in international or regional assessment programs. Therefore, ACER, with support from the GEM Centre, has developed approaches to harmonise quantitative data across assessment programs, and to provide substantive information about children's learning levels and progress benchmarked against international standards.

This International Standard Setting Exercise was undertaken as a contribution towards a common framework for countries around the world to use their own regional or national learning assessments to monitor progress towards achieving SDG 4.

The goal of the International Standard Setting Exercise (ISSE) was to place thresholds on empirical reading and mathematics Learning Progression Scales for:

- The Minimum Proficiency Level at the end of lower primary education
- The Minimum Proficiency Level at the end of primary education
- The Minimum Proficiency Level at the end of lower secondary education

Learning Progression Scales (LPS) in each of the key learning areas of reading and mathematics have been developed by ACER to provide a common reference point for

data from a range of different assessments, be they international, regional or national, and including assessments of learning in out-of-school children. The LPS build a bridge between statistical and conceptual approaches, provide meaningful descriptions of how learning progresses in reading and mathematics and can therefore be translated into targeted policy and interventions.

An outline of the SDG indicator 4.1.1 and the associated Minimum Proficiency Levels (MPL) is provided in the sub-section, 'SDG 4.1.1 and Minimum Proficiency Levels'. Brief descriptions of each Minimum Proficiency Level can be found in Appendix A.

The GEM Centre has been supporting the development of a Learning Progression Scale in each of reading and mathematics. 'Learning Progression' is a term used for a comprehensive description of what it typically looks like for learners to move from early through to advanced knowledge, skills and understandings within a learning area, such as reading or mathematics. The work undertaken by ACER to describe learning progressions in reading and mathematics is illustrated in the Learning Progression Explorer[1].

A Learning Progression Scale is an empirical measurement scale that is aligned with qualitative descriptions of a Learning Progression. Since 2016, ACER has been developing empirical Learning Progressions Scales in reading and mathematics (Turner, et. al 2018). A current description of the development of the Learning Progression Scales used in this exercise is provided in ACER (2022b).

The assessment items and data used to establish the empirical Learning Progression Scales were drawn from a wide variety of programs used in international, regional and national assessment programs. The panels assembled to determine the thresholds for the MPLs included a range of international partners and regional subject matter experts and practitioners (see 'Participants').

To establish the MPL thresholds on the Learning Progression Scales (LPS) the bookmark standard setting method was used, as described in the section: 'The Bookmark Method'. The ISSE was a collaborative virtually held exercise and the 'Procedure' sub-section summarises the activities undertaken.

This report describes the robust procedures and analyses to provide strong evidence that the standard setting exercise has produced reliable and valid global benchmarks of student proficiency. The standard setting consultations were organised to encourage and capture diversity of opinions. The proposed cuts scores for Minimum Proficiency Levels, provided as locations on the Learning Progression Scale, are presented in Table 10 of the recommendation section.

SDG 4.1.1 and the Minimum Proficiency Levels

[1] https://learning-progression-explorer.acer.org/

In 2015, the UN established the Sustainable Development Goals, 17 high level goals to be achieved by 2030:

> The 2030 Agenda for Sustainable Development… [and] …the 17 Sustainable Development Goals (SDGs), … recognize that ending poverty and other deprivations must go hand-in-hand with strategies that improve health and education, reduce inequality, and spur economic growth – all while tackling climate change and working to preserve our oceans and forests (UN DESA, n.d.a).

SDG 4 addresses education: "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all". Each SDG has indicators. In this Study, the thresholds placed upon the Learning Progression Scales derive from SDG indicator 4.1.1:

> The proportion of children and young learners (a) in grade 2 or 3; (b) at the end of primary education; and (c) at the end of secondary education achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex. (UN DESE, n.d.b)

The wording of SDG 4.1.1 is not specific enough for devising appropriate measures of proficiency and there has been considerable effort put to the task of defining Minimum Proficiency Levels since 2016. The Global Proficiency Framework (GPF) is a significant example of such work (GAML, 2020a, 2020b).

The GEM Centre sponsored contributions to the GPF, for both reading and mathematics. However, the GPF specifies several levels of proficiency for each grade level from grades 1 – 9, whereas the GEM Centre took the position that a *single* set of MPLs should be applied globally at each of the 3 levels indicated in SDG 4.1.1 (ACER 2020).

ACER (2019) presents a concise set of definitions of MPLs, one for each of reading and mathematics, at the end of lower primary, end of primary and end of lower secondary levels. These definitions are consistent with the content described in the GPF. Moreover, the GEM Centre sought to describe and unpack the MPLs and provide exemplar assessment tasks. For the ISSE, slightly modified versions of these papers were used to operationalise the MPL definitions. An updated version of Minimum Proficiency Levels Unpacked is tabled at the GAML 2022 (ACER, 2022a). All of these documents produced by the GEM Centre are consistent with the Global Proficiency Framework.

Method

The overarching method of the ISSE was a series of facilitated live workshops, combined with participants working individually between workshops on exercises using tailored standard setting software. Workshops were conducted virtually; travel restrictions due to COVID-19, along with minimising participants' time commitment, ensured that a diverse range of participants were able to join from across international regions. Workshops were scheduled for times likely to be convenient for the greatest number of participants. More detail about participants is provided in the sub-section below –

'Participants'. In the workshops, participants were presented with content related to MPLs, the standard setting procedures, and technical demonstrations on how to undertake the standard setting. The workshops were also facilitated to ensure that participants could discuss issues and build consensus related to the standard setting. More detail is provided in the sub-section below – 'Procedure'– with specific detail about how assessment items for the exercises were selected and ordered in the subsequent sub-section – 'Instruments'.

Participants

Selecting participants

Essential to the success of the ISSE was the participation of a sufficient number and diversity of people with the appropriate experience, background and skill. The first step to achieving this was developing a matrix of the desired qualities of participants. This matrix included the following categories: gender, language, geographical region of relevant experience, role within organisation, educational learning area experience, practitioner experience and subject matter knowledge.

An executive team from ACER was convened to identify relevant stakeholders. This included ensuring representation from international education development organisations (including UN agencies), Policy and Assessment Networks, and Government Education Ministries. Drawing on GEM Centre networks, an email was sent to stakeholders explaining the ISSE and requesting that they nominate participants from either their organisation or their broader network.

Upon receiving the nominations, the nominees were vetted to ensure that each one satisfied the minimum requirements. First, the participants needed to satisfy one of these 2 requirements:

- Expert or master teachers of reading or mathematics, with experience in at least one of the 3 levels of schooling referenced in SDG 4.1.1
- Reading or mathematics subject matter experts, with experience in one of the following three educations learning areas: assessment development, curriculum development, or pedagogical training.

Second, participants were required to have at least a strong working proficiency in English, as the materials provided and the workshops were in English.

Subsequently, an invitation was sent to the nominees who satisfied these requirements, explaining the ISSE and requesting their participation. An honorarium of $US200 was offered to participants as a token of appreciation.

Participant profile

Number of participants

Individuals who attended at least one session and then undertook either the bookmarking or Angoff exercise are considered 'participants'. Engaging in discussion and asking questions throughout the 4 sessions was the ideal form of participation. However, the sessions were recorded which meant participants could engage in all the required material without synchronously attending all sessions, and therefore make a fully informed contribution.

Table 1: Number of participants and level of attendance

Level of attendance	Mathematics	Reading
Signed up	32	31
Attended **at least 1** session	29	31
Attended only 1 session	*2*	*2*
Attended only 2 sessions	*3*	*2*
Attended only 3 sessions	*5*	*7*
Attended 4 sessions	19	*20*

Level of schooling

As described in the section below on Procedure, for much of the ISSE workshop sessions, participants worked in 2 groups, with one focussing on early and late primary school, and the other on late primary and lower secondary school material. When participants had experience in both levels, they were allocated to the one with fewer participants.

Table 2: Group participation of participants

School level	Mathematics	Reading
Early-late primary school	13	17
Late primary - lower secondary school	16	14
Total	**29**	**31**

As shown in Table 2, there was a good split between the number of participants who participated in the early/late primary school and late primary/lower secondary school groups.

Gender

A mix of women and men participated in both mathematics and reading learning areas. There was a larger proportion of women in each group, especially in reading.

Table 3: Gender of participants

Gender	Mathematics	Reading
Female	18	24
Male	11	7
Total	**29**	**31**

Region

The ISSE had global representation. Africa and Oceania were particularly well represented. The totals come to more than the number of participants, as some participants had experience working in more than one region.

Table 4: Region of participants

Region	Mathematics	Reading
Africa	10	15
Asia	2	4
Europe	*0*	6
Oceania	14	*8*
Americas	*3*	*5*
Total	**31**	**38**

Educational learning area

Participants had experience across all 3 educational learning areas presented in the table below. Most commonly though, participants had experience in national or standardised assessment development. The totals come to more than the number of participants, as some participants had experience in more than one learning area.

Table 5: Educational learning area

Educational learning area	Mathematics	Reading
Curriculum development	19	15
National or standardised assessment development	25	23
Teacher training, pedagogical development	18	19
Total	**62**	**55**

Procedure

The standard setting exercise used the bookmark method to establish proposed MPL thresholds for SDG 4.1.1 indicators. This bookmark method is an item-centred standard setting method.

The standard setting exercise was conducted virtually and consisted of group and individual work sessions. Participants were allocated to a panel that would set thresholds for 2 of the 3 SDG 4.1.1 indicators for mathematics and reading. The 2 panels for each learning area met at the end of the process to discuss, evaluate, and if needed, adjust the proposed thresholds within each learning area.

An overview of the activities conducted and the time commitment is provided in Table 6.

The ISSE process is summarised for participants in the 'International Standard Setting Overview' document, with slight variation between the mathematics and reading versions. The same process was followed with separate panels for each of reading and mathematics. The information in this section should be assumed to apply to both learning areas, unless otherwise stated. Of the 7 activities comprising the exercise, Activities 2, 5 and 7 were completed individually, while activities 1, 3, 4 and 6 were group activities.

Table 6: Overview of ISSE activities

Activity number	Duration	Activity	Notes
1	1 hour	Plenary session	• Introduction • Exercise briefing • Materials and access dissemination
2	4 hours	Individual deskwork	• MPL descriptors review • Sample MPL items test completion
3	2 hours	Group session	• Unpacking the MPL descriptors and discussion
4	2 hours	Group session	• Bookmark standard setting introduction • Item maps introduction
5	3 hours	Individual deskwork	• Item map completion • Preliminary bookmark placement
6a	2 hours	Group session	• Preliminary bookmark review • Panel bookmark placement
6b	2 hours	Group session	• Panel bookmarks review and discussion • Build bookmark placement consensus
7	1 hour	Individual deskwork	• Participation questionnaire and feedback completion

Prior to the commencement of the exercise, participants received 2 documents:

1. *International Standard Setting Reading Overview*: This document described the activities to be undertaken as part of the ISSE.
2. *Minimum proficiency levels unpacked*: This document set out and explained the Minimum Proficiency Levels that are referred to in one of the indicators for judging achievement of Sustainable Development Goal 4, Indicator 4.1.1.

In Activity 1, the ISSE was introduced. The first part of the session focussed on giving background and context for the exercise. At present, we have a single definition of Minimum Proficiency Level for each of reading and maths at each level of schooling in SDG 4.1.1 to apply globally. The overall purpose of the ISSE was to derive a single standard at each Minimum Proficiency Level to apply globally. In general terms, standard setting is the process used to distinguish between a student who meets the MPL and a student who does not. This is achieved by determining the location on an empirical measurement scale associated with the minimal level of skill or knowledge required to reach the MPL. In the second part of the Activity 1 session the target student

was defined, and the MPLs explained, first in general terms, and then by giving participants the opportunity to review a series of items and consider whether the target student, who was just meeting the MPL would be more likely than not to respond successfully to a given item.

In the final part of Activity 1, a demonstration of ACER Signum, the software used for the exercise was given. This training covered how to access ACER Signum, the appearance and key features of the interface within ACER Signum, system requirements for using ACER Signum and the support available to participants during the exercise, in preparation for Activity 2.

For most of the remaining ISSE Activities, participants were assigned into smaller groups:

- 'Reading AB' investigated items associated with the MPLa and MPLb thresholds, in reading.
- 'Reading BC' investigated items associated with the MPLb and MPLc thresholds, in reading.
- 'Mathematics' AB investigated items associated with the MPLa and MPLb thresholds, in mathematics.
- 'Mathematics BC' investigated items associated with the MPLb and MPLc thresholds, in mathematics.

This grouping meant that individuals were able to focus on their area of most experience (primary / secondary education) but facilitated a strong link between the groups as every participant interacted with MPLb in their respective learning area.

The goal of Activity 2 was for the participants to familiarise themselves with the items within the item pool and their relationship to the Minimum Proficiency Levels using a version of the Angoff standard setting task. Participants were presented with items from the pool in a random order in the Signum software and asked to make Yes/No judgements about whether the target student would be more likely than not to successfully answer the item correctly. Participants in the AB groups were asked to evaluate each item twice, once each for MPLa and MPLb. The BC groups followed the same procedure for target students at MPLb and MPLc. Activity 2 was not a standard setting exercise, its purpose was twofold: to familiarise participants with the items in terms of construct coverage, design and range of difficulty and to assist them in developing and better understanding of the target student profiles for each of the MPL standards.

In Activity 3, participants convened in smaller groups (the AB groups focussing on MPLa and MPLb, and the BC groups focussing on MPLb and MPLc). The ACER facilitators had reviewed all Angoff judgements prior to this session. While for some items there was (near) uniformity in the judgement made about an item (that is all, or almost all participants gave the same judgement about whether the target student

would be able to successfully answer the item) for other items this pattern was less clear, with participants evenly split between "yes" and "no" judgements. ACER facilitators identified the items in this latter category, and it was these items that formed the basis for the discussion in Activity 3. Participants reviewed and discussed a set of items where there was a lack of consensus with reference to the MPLs and were encouraged to change their judgement if the group discussion convinced them that this was warranted.

Activity 4 comprised an introduction to bookmarking, including the central notion of the 'ordered item booklet', a booklet in which items appear in order of their difficulty, with the easiest item appearing first, and the most difficult, last. A demonstration of how to use the ACER Signum Software to place bookmarks was also given. The training provided in Activity 4 allowed participants to complete Activity 5: individually placing their bookmarks. Participants worked with the same MPLs they had worked with for the Angoff exercise: the AB group placed one bookmark for MPLa, and one bookmark for MPLb, while the BC group placed one bookmark for MPLb, and another for MPLc.

Activity 6 had 2 parts. In the first, all participants (the AB group and the BC group) reviewed and discussed their placement of the bookmark common to both groups: the bookmark for MPLb. In the second part of the session, participants again worked in their smaller groups to review placement of the MPLa bookmark (AB group) and the MPLc bookmark (BC group). In both parts of the session, participants were encouraged to change their bookmark placement within the ACER Signum system if they were sufficiently convinced by the discussion to do so. They could change these placements either during the session, or for several days afterwards.

The final activity in the process (Activity 7) was the completion of a feedback form, enabling participants to express their views on all aspects of the exercise. As the ISSE was undertaken by ACER for the first time, using a novel remote facilitation method, ACER wanted to evaluate the process through participant feedback.

The feedback on the ISSE from both reading and mathematics participants was overwhelmingly positive and summarised in Appendix B.

Instruments

Items included in the standard setting exercise were carefully selected to cover the whole range of the MPL standards from lower primary to the end of lower secondary levels. For reading and mathematics, the items were drawn from three different assessment programs as well as the pool of MPL items written by ACER with the purpose of exemplifying different levels of growth in reading and mathematics respectively. The item sets for each of reading and mathematics comprised of 100 items in a range of response formats (e.g. multiple-choice and constructed-response) and, as sets, the items represented the range of constructs for reading and mathematics as set out in the *Minimum Proficiency Levels Unpacked* (ACER, 2022a) document provided to participants.

The bookmark method

The bookmark standard setting method uses an Ordered Item Booklet (OIB), which consists of items ordered by difficulty. The easiest item is presented first, and the most difficult item is presented last. In the ISSE, participants were presented with the same set of items that they worked with for the Angoff exercise, but this time, the items were ordered according to their level of difficulty. In working through all the items, participants were asked to consider the question – 'Is the target student, working just at the MPL, more likely than not to answer this question correctly?' Then in response to this question, participants were asked to place the bookmark on the first item that the target student, who just meets the MPL, is more likely to answer incorrectly than correctly. In making this judgement, participants were advised to make holistic, on-balance judgements about where to place each bookmark Subsequently, participants needed to examine at least the next 5 items after the bookmark, thereby either confirming the placement of their bookmark, or informing adjustments to the placement of their bookmark. In general, the appropriate point for the bookmark is the point at which all or almost all of items before the bookmark have a 'Yes' judgement and all or almost all of the items after the bookmark have a 'No' judgement, as illustrated in Figure 1.

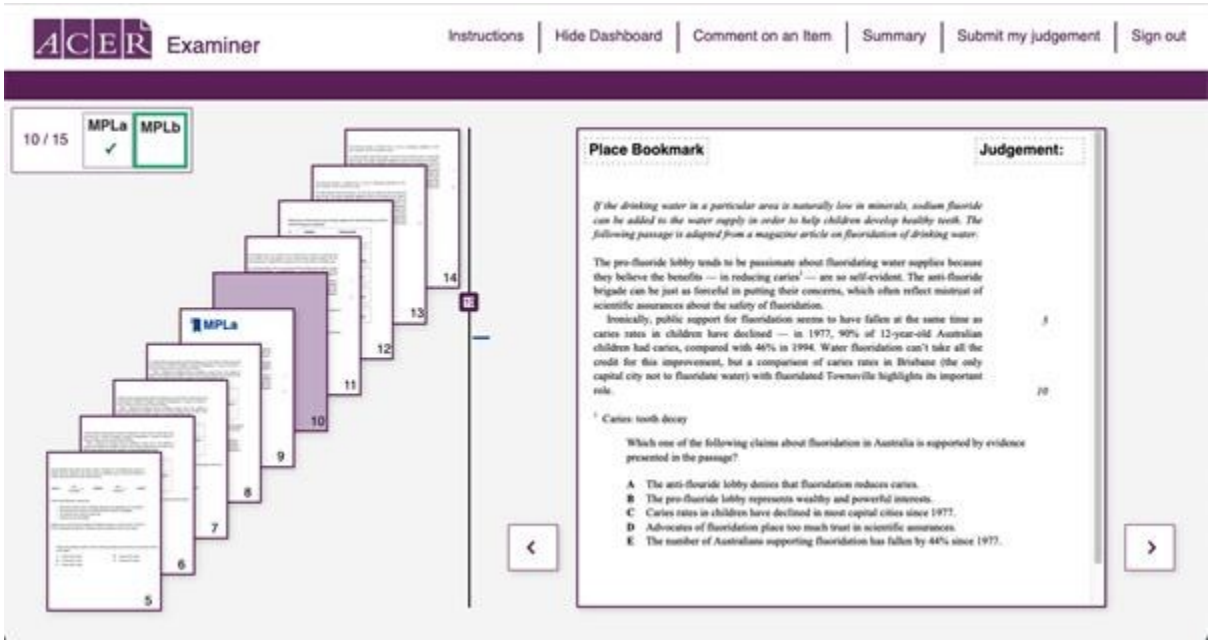

Figure 1: Bookmarking standard setting method

Results and Analysis

As described in the Method section, plenary and group discussion sessions were conducted for each learning area. No attempt was made to enforce consensus. Participants were able to change the position of their initial bookmarks for any of the standards during and after the consensus building activities and for this reason only the final cut-score locations for each participant are reported and analysed.

In the bookmark method the cut-score for a standard is set at the item preceding the item on which participants have placed the bookmark. This is the last item that the target student is more likely than not to answer correctly and thus it represents demarcation line for students that meet the minimum requirements of the respective standard.

Given that OIB for the 2 groups (AB and BC) overlapped, instead of the OIB location the cut-scores are reported against the location of corresponding items on the Learning Progression Scale. Table 7 provides summary statistics for the cut-score location for the three standards and 2 learning areas.

Table 7: Summary statistics for the cut-score location

Learning area	MPL	Mean	SD	N	Median	Mode	Min	Max
Reading	MPLa	73	4.5	11	72	72	72	87
Reading	MPLb	90	4.7	22	88.5	88	83	106
Reading	MPLc	119	3.4	11	120	120	112	124
Mathematics	MPLa	87	1.3	10	88	88	85	88
Mathematics	MPLb	104	7.2	24	103	103	95	127
Mathematics	MPLc	129	6.9	14	132	132	107	133

Table 7 shows that across all learning areas and standards the 3 measures of central tendency, Mean, Median and Mode are extremely close. Such an outcome indicates that a relatively high level of overall agreement on the position of cut-scores was achieved for all three standards

For each learning area, the number of individual judgements is highest for the MPLb standard because this was the task for participants in both the AB and BC groups. Consequently, it is important to investigate whether there was any systematic difference in how MPLb cut-score were distributed in these groups. Figure 2 presents a histogram of cut-score location across standards and learning areas for the AB an BC groups.

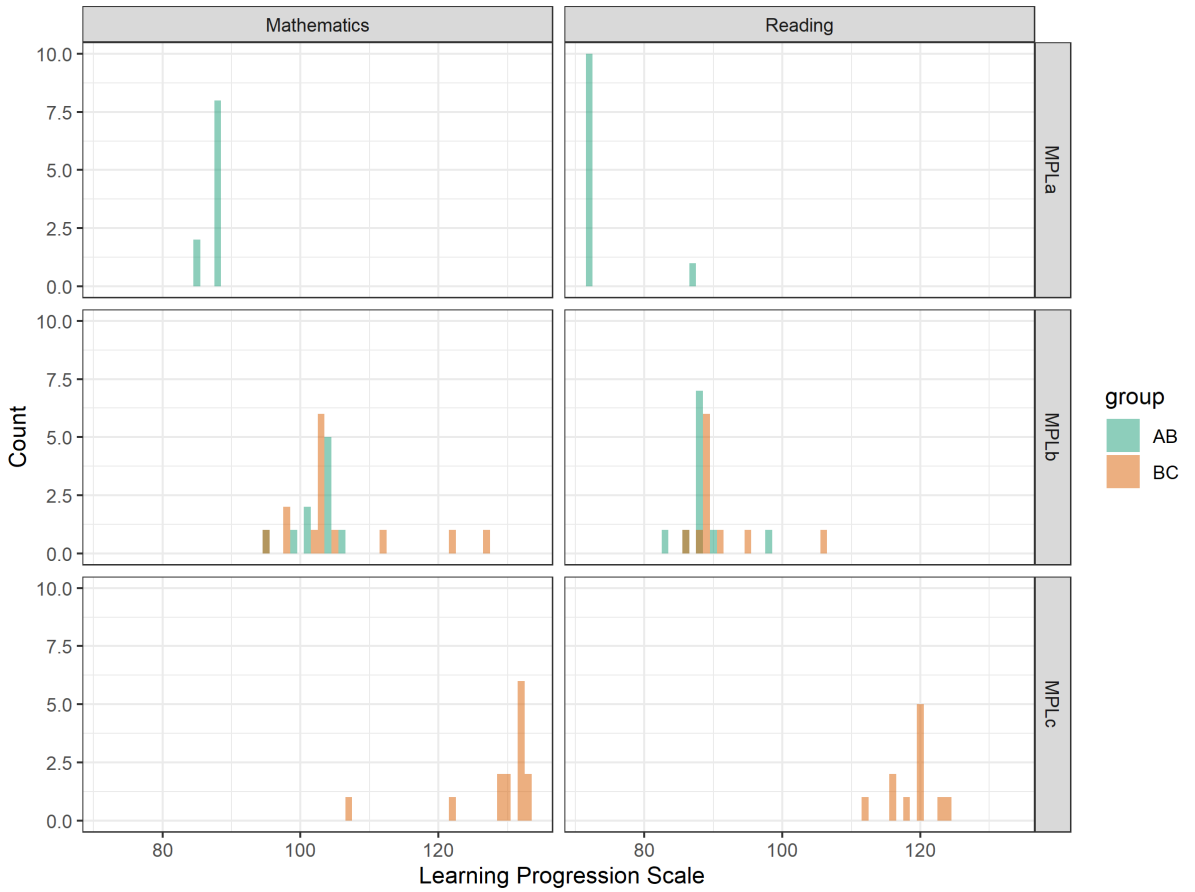

Figure 2: Cut-score location across standards and learning areas

As can be seen in Table 2 there are no systematic difference in the location of the MPLb cut-score between the 2 groups. The Mode locations for both groups are placed next to each other in both learning areas. Disregarding outliers in the BC group, the range of the cut-score locations are similar for the 2 groups.

The method employed in the bookmark exercise was novel: with remote participation, and MPL and target student familiarisation training conducted through a modified Angoff exercise. As a validation exercise, several of ACER's mathematics and reading assessment experts also participated in the ISSE. These experts have sound understanding of MPLs, MPL descriptors and the Learning Progression underpinning the descriptors development. Consequently, their cut-scores were used as a validation set to compare participant outcomes.

Table 8 shows these comparisons between ACER experts and external participants for the key cut-score location statistics across the standards and learning areas.

Table 8: Cut-score location statistics across the standards and learning areas

Learning area	Experts	MPL	Mean	N	Median	Mode
Reading	ACER	MPLa	72	1	72	72
Reading	Participants	MPLa	73	11	72	72
Reading	ACER	MPLb	91	4	89.5	88
Reading	Participants	MPLb	90	22	88.5	88
Reading	ACER	MPLc	116	3	116	112
Reading	Participants	MPLc	119	11	120	120
Mathematics	ACER	MPLa	86	3	85	85
Mathematics	Participants	MPLa	87	10	88	88
Mathematics	ACER	MPLb	102	5	104	104
Mathematics	Participants	MPLb	104	24	103	103
Mathematics	ACER	MPLc	133	2	133	132
Mathematics	Participants	MPLc	129	14	132	132

There is very little difference between the ACER validation set cut-scores and those produced by external participants and where some difference was observed these are well within the standard deviation of mean cut-scores produced by participants reported in Table 7. Such an outcome provides evidence that MPL training and the bookmarks standard exercise managed to produce a set of valid MPL standards' cut-scores.

An additional validation investigation was undertaken by comparing the MPLb cut-scores for reading and mathematics between the ISSE with the COVID-19: Monitoring the Impacts on Learning Outcomes (MILO) study. The assessment tool developed for MILO can be used to measure and report on SDG Indicator 4.1.1b. MILO items were incorporated in the development of the Learning Progression Scale and thus it was possible to equate the location of the MPLb standards developed for MILO assessments on to the Learning Progression Scale. This in turn enabled direct comparison of 2 sets of MPLb cut-scores. Table 9 provides a comparison of the MILO MPLb cut-scores to the mean MPLb cut-scores produced from all bookmark participants.

Table 9: MPLb cut-scores comparisons

Learning area	MILO MPLb on LPS	Mean MPLb	SD MPLb
Reading	93	90	4.5
Mathematics	100	104	6.7

As can be seen, the 2 set of MPLb cut-scores are very close and well within the standard deviation of the bookmark exercise cut-scores. This is a remarkable outcome given that different item sets, standard setting procedures and different participants were used to produce these cut-scores for the MPLb standard. These results provide strong evidence for the reliability and validity of the MPLb standards and of the international standard setting exercise overall.

Recommendations of cut-scores

The analyses provide strong evidence that the standard setting exercise has produced reliable and valid outcomes. As described earlier the bookmark standard setting group work and consultations were organised to encourage and capture diversity of opinions and thus it is appropriate that the Mean cut-scores for each standard and learning areas are used as the final MPL standards cut-scores proposal. The proposed cuts scores as locations on the Learning Progression Scale are presented in Table 10. The Learning Progression Scales were transformed to have a mean of 120 and standard deviation of 10 scale score points.

Table 10: Proposed cuts scores as locations on LPS

Learning area	MPL	cut-score	cut-score SD
Reading	MPLa	73	4.3
Reading	MPLb	90	4.5
Reading	MPLc	118	3.6
Mathematics	MPLa	87	1.4
Mathematics	MPLb	104	6.7
Mathematics	MPLc	129	6.6

Conclusion and applications

The International Standard Setting Exercise successfully placed thresholds for Minimum Proficiency Levels (MPL) in reading and mathematics on empirically based Learning Progression Scales. For each of reading and mathematics, these Learning Progression Scales contained items used to measure skills and knowledge from lower primary to beyond the end of lower secondary, on a single scale. The items originated from a range of large scale assessments or were specifically authored to target the relevant MPL.

The outcomes of the ISSE demonstrates proof of concept that 'long-span' Learning Progression Scales that are empirically constructed using a range of methods with items from a variety of sources can serve as the basis for locating Minimum Proficiency Levels on an empirical continuum.

The next steps towards supporting countries and regions to report against SDG 4.1.1 using there own instrument involve 2 potential applications:

1. The process of constructing Learning Progression Scales and locating MPLs in language other than English. A bilingual exercise could also be considered where items are translated into rather than sourced from other languages.

2. The use of Pairwise Comparison Method incorporating items from a national or regional assessment alongside items from the Learning Progression Scales used in the ISSE would facilitate the placement of the MPLs onto the national or regional scale.

The results of the ISSE and the overwhelmingly positive feedback of participants support the first application/expansion. The second application/expansion is outlined further in ACER (2022b) 'Pairwise Comparison Method: Concept note'. The ISSE and both of the potential applications are steps towards a common framework for countries around the world to use their own regional or national learning assessments for monitoring progress towards achieving SDG 4.

References

Australian Council for Educational Research (ACER). (2019). *Minimum Proficiency Levels: Described, unpacked and illustrated*. Version 2. (GAML6/REF/2). UNESCO Institute for Statistics. https://research.acer.edu.au/monitoring_learning/47/

Australian Council for Educational Research (ACER). (2020). *Minimum proficiency levels: Revisions proposed by ACER*. (WG/GAML/6). http://tcg.uis.unesco.org/wp-content/uploads/sites/4/2020/10/WG-GAML-6-Minimum-Proficiency-Levels_revised.pdf

Australian Council for Educational Research (ACER). (2022). *Minimum Proficiency Levels: Described, unpacked and illustrated*. Version 3.

The Australian Council for Educational Research (ACER). (2022b). *Pairwise Comparison Method: Concept note.*

Global Alliance to Monitor Learning (GAML) (2020a). *Global Proficiency Framework for Reading: Grades 1 to 9*, https://gaml.uis.unesco.org/#

Global Alliance to Monitor Learning (GAML) (2020b). *Global Proficiency Framework for Mathematics: Grades 1 to 9*, https://gaml.uis.unesco.org/#

Turner, R. et al. (2018). *Development of Reporting Scales for Reading and Mathematics: A report describing the process for building the UIS Reporting Scales*. Melbourne: Australian Council for Educational Research. https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso&db=cat03974a&AN=ACER.365885&site=eds-liv

United Nations Department of Economic and Social Affairs (UN DESA). (n.d.a). *Do you know all 17 SDG Goals*. Sustainable Development. https://sdgs.un.org/goals

United Nations Department of Economic and Social Affairs (UN DESA). (n.d.b). *Goals 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all*. https://sdgs.un.org/goals/goal4

Appendix A: Brief descriptions of the Minimum Proficiency Levels

This appendix contains brief descriptions of reading and mathematics Minimum Proficiency Levels (MPL) at each of lower primary, end of primary and end of lower secondary. For each MPL a concise 'nutshell statement' is provided. An expanded description is also provided. More detail unpacking and illustrating the MPL with assessment items can be found in ACER 2022b.

Reading: End of lower primary (SDG 4.1.1a)

Nutshell statement

Students accurately read aloud and understand written words from familiar contexts. They retrieve explicit information from very short texts. When listening to slightly longer texts, they make simple inferences.

Expanded statement

In a short simple text of one or 2 sentences, students read aloud most words – including some unfamiliar words – accurately but slowly and often word by word. They identify the meaning of familiar words, including when they have common morphological changes, and also some unfamiliar words. They retrieve explicit information from a single sentence. When listening to longer texts, and looking at the illustrations, students retrieve explicit information about main events, ideas or characters and use that information to draw simple inferences.

Reading: End of primary (SDG 4.1.1b)

Nutshell statement

Students independently and fluently read simple, short narrative and expository texts. They retrieve explicitly stated information. They interpret and give some explanation about the main and secondary ideas in different types of texts, and establish connections between main ideas in a text and their personal experiences.

Expanded statement

In a short, simple narrative or expository text, students read aloud at a pace and a level of accuracy and expression (prosody) that demonstrate understanding. They use previously taught morphological (word-level) and contextual (sentence- or text-level) clues to understand the meaning of familiar and unfamiliar words and to distinguish between the meanings of closely related words. When reading silently or aloud, they locate explicit information in a paragraph. They use that information to make inferences about behaviours, events or feelings. They identify the main and some secondary ideas

in a text if they are prominently stated, and recognise common text types when the content and structure are obvious. They make basic connections between the text and their personal experience or knowledge.

Reading: End of lower secondary (SDG 4.1.1c)

Nutshell statement

Students retrieve and connect multiple pieces of related information across sections of texts to understand key ideas. They make straightforward inferences when there is some competing information. They reflect and draw conclusions in a variety of text types.

Expanded statement

In a range of continuous and non-continuous texts, including narrative, expository, descriptive, argumentative, instructional, and transactional texts, students locate multiple pieces of information across a text, including information in paratextual elements. They make straightforward inferences by drawing on prominent explicit and implicit information to summarise key ideas, and select evidence to support an interpretation. They reflect on texts in relation to personal experience and draw on general knowledge to identify if there is an obvious flaw in a text-based idea.

Mathematics: End of lower primary (SDG 4.1.1a)

Nutshell statement

Students recognise, read, write, order and compare whole numbers up to 100. They demonstrate computational skills involving the processes of addition, subtraction, doubling and halving for whole numbers within 20. They recognise and name familiar shapes and describe their basic attributes. They recognise time in days, weeks and months. They describe location in a space using simple language.

Expanded statement

Students can read, write and compare whole numbers up to 100. They can add and subtract numbers within 20, double and halve whole numbers within 20, and solve application problems involving numbers within 20. Students can recognise simple shapes and their attributes and use these shapes to make other shapes. They can also measure and compare lengths of shapes and lines using non-standard units. They use calendars and recognise days in a week and months in a year. They can read simple data displays. They possess foundational knowledge of spatial orientation, and can appraise the relative size of real-world objects.

Mathematics: End of primary (SDG 4.1.1b)

Nutshell statement

Students recognise, read, write, order and compare whole numbers within 100,000, unit fractions and their multiples. They add/subtract with whole numbers within 1,000 and multiply/divide with whole numbers within 100. Students can measure length, weight and capacity using standard units; read time on an analogue clock; calculate the perimeter of simple 2D shapes and the area of rectangles; and describe the attributes of familiar 2D and 3D shapes. They read, interpret and construct different types of data displays such as tables, column graphs and pictographs, and recognise, describe and extend number patterns. They can solve simple application problems.

Expanded statement

Students can add and subtract whole numbers within 1,000 and demonstrate fluency with multiplication facts up to 10 x 10 and related division facts; solve simple application problems with whole numbers using the four operations; identify simple equivalent fractions; compare and order unit fractions and fractions with related denominators; identify and represent quantities using decimal notation up to the tenths place; select and use a variety of tools to measure and compare length, weight and capacity/volume; read time to the minute on an analogue clock and calculate elapsed time in minutes within and across the hour; construct data displays with data arranged into categories and single or multi-unit scales; retrieve multiple pieces of information from data displays to solve problems; recognise and name 2D shapes and familiar 3D objects by their simple attributes such as number of faces, edges and vertices for 3D shapes and number of sides and corners for 2D shapes; describe and continue number patterns that increase or decrease by a constant value from any starting point; or that increase or decrease by a constant multiplier; and apply the concept of equivalence by finding a missing value in a number sentence.

Mathematics: End of lower secondary (SDG 4.1.1c)

Nutshell statement

Students demonstrate skills in computation with fractions, decimals, rates, ratios, percentages and integers. They apply geometric relationships and formulae such as area, volume, Pythagoras' theorem, and the angle sum of a triangle. They interpret and construct a variety of data displays and calculate measures of central tendency. They make use of algebraic representations of linear relationships. They can use their mathematics knowledge to solve application problems.

Expanded statement

Students can apply the order of operations and solve simple problems involving fractions, decimals and whole numbers. They can apply geometric relationships and formulae (namely, area of a triangle, circumference and area of a circle, volume of a rectangular prism, Pythagoras' theorem, and angle sum of a triangle) to solve straightforward problems in simple contexts. They can interpret and construct a variety of data displays and calculate measures of central tendency. They can graph linear

equations on a coordinate grid. They can solve equations in one variable and model context-based situations using simple algebraic representations. They can evaluate and calculate with simple algebraic expressions. They can use proportional reasoning to solve problems.

Appendix B: Process evaluation

As the ISSE was undertaken by ACER for the first time, using a novel remote facilitation method, ACER wanted to evaluate the process through participant feedback. Upon completion of the workshops, participants were asked to complete a short survey. The questionnaire asked respondents to rate their level of agreement with a series of statements. Each question also provide opportunity for respondents to comment. Eighteen respondents in the Mathematics learning area and 17 respondents in Reading learning area provided feedback, as seen in the table below. The feedback on the ISSE from both reading and mathematics participants is overwhelmingly positive. Almost all responses were in agreement or strong agreement with the positive statements.

Training materials

All respondents either strongly agreed or agreed that the training materials were helpful. Some comments included: "The presentations in the session were very clear and well explained – I also found the MPL unpacked document particularly useful"; and, "If it were not for the training materials, I would not been able to do it so thoroughly".

Table 11: The training materials were helpful

Learning area	Strongly disagree	Disagree	Agree	Strongly agree
Maths	0	0	5	13
Reading	0	0	4	13

MPL definitions

All respondents, except one, either strongly agreed or agreed that the MPL definitions were clearly communicated. One respondent commented that:

> they were very well communicated, and made so much sense. I was happy to get some insight as to the amount of preparatory work that goes into the Standard Setting process. Such wide consultation with English Language practitioners was great. I feel empowered to have been part of the process

Table 12: The MPL definitions were clearly communicated

Learning area	Strongly disagree	Disagree	Agree	Strongly agree
Maths	0	1	4	12
Reading	0	0	6	11

Defining the target students

All respondents either strongly agreed or agreed that defining the target students helped them to make judgements. Some comments include: "My ideas of proficiency levels of the target student have changed and I will be consulting with these documents when I set items"; and, "Indeed, it was also a reminder that one should not necessarily only consider one's own country's context."

Table 13: Defining the target students helped me to make my judgements

Learning area	Strongly disagree	Disagree	Agree	Strongly agree
Maths	0	0	9	9
Reading	0	0	7	10

Angoff-style task

All respondents either strongly agreed or agreed that they understood the Angoff-style task and how to judge an item. One respondent commented that:

> This was the most insightful part of the workshop, from which I learned enormously. The Angoff-style of judging items is a new concept in my professional life as an item developer, which I find fascinating, and very helpful in this regard.

Table 14: I understood the Angoff-style task and how to judge an item

Learning area	Strongly disagree	Disagree	Agree	Strongly agree
Maths	0	0	6	12
Reading	0	0	1	16

Bookmarking

Almost all respondents either strongly agreed or agreed that they understood the Bookmarking task and how to place the bookmarks. One person each in Reading and Mathematics disagreed. Some comments include: "This was slightly challenging for me. However, the subsequent discussions gave some clarity"; and, "The platform made it much easier to apply the task. It also cuts the time spent on this process in half."

Table 15: I understood the Bookmarking task and how to place the bookmarks

Learning area	Strongly disagree	Disagree	Agree	Strongly agree
Maths	0	1	6	10
Reading	0	1	4	12

Enough time to make judgements

Almost all respondents either strongly agreed or agreed that they had enough time to make judgements during the days between meetings. However, 2 people from the mathematics learning area disagreed. A participant commented that:

> I would've liked a little more time – extra couple of days so that I could fit it in with my own work commitments and importantly, give sufficient thinking time so I can participate fully in the discussions

Table 16: I had enough time to make my judgements during the days between meetings

Learning area	Strongly disagree	Disagree	Agree	Strongly agree
Maths	0	2	2	13
Reading	0	0	2	15

Common understanding of MPLs

All respondents, except one, either strongly agreed or agreed that their group reached a common understanding of its two MPLs. One person commented that: "We more or less came to consensus on our judgements, but I am not sure if everyone had a same understanding of the two MPLs".

Table 17: My group (group ab or group bc) reached a common understanding of its two MPLs

Learning area	Strongly disagree	Disagree	Agree	Strongly agree
Maths	0	1	9	7
Reading	0	0	6	11

The full panel reached a common understanding

A strong majority of participants either strongly agreed or agreed that the full panel (Groups ab and bc together) reached a common understanding of the shared MPL, MPLb. However, this was the most contentious of all the statements, with 3 respondents from mathematics and 2 respondents from reading, disagreeing. One person commented that:

> I think there was some confusion at first stemming from the confusion around the two sets of ordered booklets that the different groups received. However, I feel that the final plenary cleared up any confusion and allowed us to arrive at a shared understanding.

Table 18: The full panel (Groups ab and bc together) reached a common understanding of the shared MPL, MPLb

Learning area	Strongly disagree	Disagree	Agree	Strongly agree
Maths	0	3	8	6
Reading	0	2	8	7

Enough time to reconsider judgements

All respondents, except one, either strongly agreed or agreed that they had enough time to reconsider their judgements during the discussions. One respondent commented that they "had time to reconsider during the discussions but did not have time to study again and then make changes."

Table 19: I had enough time to reconsider my judgements during the discussions

Learning area	Strongly disagree	Disagree	Agree	Strongly agree
Maths	0	1	4	13
Reading	0	0	3	14

Cut-scores reflected thinking

All respondents either strongly agreed or agreed that they felt that the procedure allowed them to recommend cut-scores that reflected their thinking. Some comments include: "personally would've liked a little more time; that being said the procedure does allow one to reflect and express ones thinking regarding recommending cut-scores"; and, "Definitely – the discussion allowed me to both clarify my own understanding and to discuss my thoughts on particular items".

Table 20: I felt that this procedure allowed me to recommend cut-scores that reflected my thinking

Learning area	Strongly disagree	Disagree	Agree	Strongly agree
Maths	0	0	6	10
Reading	0	0	5	11

The standard setting website

There was overwhelming strong agreement that the standard setting website was easy to use. One person commented:

> The software was brilliant – very intuitive, easy to use and it ran smoothly. I was very impressed with the ease of use of this software having only run standard setting meetings with physical materials previously.

Table 21: The standard setting website was easy to use

Learning area	Strongly disagree	Disagree	Agree	Strongly agree
Maths	0	0	3	15
Reading	0	0	0	17