

# Study Design

## Assessment of Minimum Proficiency Level (AMPLab)

February 2023, version 1.0

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).

# Acknowledgments

This project, Development of an Assessment of Minimum Proficiency Level (AMPL) aligned to Global Proficiency Framework (GPF) Level 2, is referred to AMPLab. This UNESCO Institute for Statistics (UIS) project is funded by the Bill & Melinda Gates Foundation.

The Australian Council for Educational Research (ACER) is the technical partner for this project. Support is provided from ACER's Global Education Monitoring Centre (GEM Centre), an ACER initiative in partnership with the Australian government's Department of Foreign Affairs and Trade.

# Contents

Background.....	1
Introduction.....	1
Design overview.....	1
Roles and responsibilities.....	3
Method.....	4
The population of interest.....	4
Survey and sample design.....	4
Technical standards.....	5
Assessment of learning outcomes.....	5
AMPLab test design.....	8
Contextual information.....	10
Materials preparation.....	10
Data collection.....	11
Ordering and timing.....	11
Data management.....	11
Standard setting exercise.....	12
Psychometric analysis.....	12
Data analysis.....	12
Reporting.....	12
Capacity development.....	13
References.....	15

# Background

As part of Sustainable Development Goal (SDG) 4, Indicator 4.1.1 aims to measure the “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.” To meet this goal, UIS has coordinated efforts to establish common reading and mathematics scales for all three points of Indicator 4.1.1, building on existing cross-national and national assessments. As a result of these efforts, two important points of consensus have been reached: the definition of the Minimum Proficiency Level (MPL) and the Global Proficiency Framework (GPF).

The overarching objective of the AMPLab project is to measure and analyze the performance of students at the end of lower and upper primary using an assessment that aligns with the GPF. This will:

- enable the collection of more informative data about where students are performing in terms of the MPLs at the end of lower and upper primary in reading and mathematics,
- produce baseline measures to set targets and compare learning gains/losses over time
- facilitate reporting on SDG 4.1.1
- aid the tracking of learning progress over time
- complement tools that had been already developed in 2021 in the Monitoring the Impacts on Learning Outcomes (MILO) study.

## Introduction

This document provides an overview of the study design and an outline of the method that will be used to implement the project which will take place in 2023. Detailed information on procedures will be made available progressively as part of the deliverables for other product areas, including sampling, field operations and data management.

## Design overview

The benchmarks that will be used to quantify performance will be:

- the proportion of students in the end of lower primary school, in participating countries, that meet the Minimum Proficiency Levels (MPL) referred to in SDG indicator 4.1.1(a) and described in ACER-GEM (2022).
- the proportion of students in the end of upper primary school, in each country, that meet the Minimum Proficiency Levels (MPL) referred to in SDG indicator 4.1.1(b) and described in ACER-GEM (2022).

The links to these benchmarks will be established in the AMPLab assessments as part of this study.

In addition, to characterise reading and mathematics performance, contextual information will be gathered alongside the assessments at the student and school levels. This information will help in exploring learning outcomes for different subpopulations, especially by gender.

As of February 2023, India, Kenya, Lesotho and Zambia are Anglophone participants. There are currently no Francophone participants in the 2023 study. The following study design by country has been confirmed.

**Table 1: Study design by country**

Participating Country	SDG indicator	Target grade	Learning Outcomes	Contextual information
India (pilot only)	4.1.1(a)	3	AMPL-A booklets	AMPLab questionnaires
	4.1.1(b)	5	AMPL-B booklets	
Kenya	4.1.1(a) and 4.1.1(b)	6	AMPL-A+B booklets	
Lesotho	4.1.1(a)	4	AMPL-A booklets	
	4.1.1(a) and 4.1.1(b)	7	AMPL-A+B booklets	
Zambia	4.1.1(a)	4	AMPL-A booklets	
	4.1.1(a) and 4.1.1(b)	7	AMPL-A+B booklets	

## Roles and responsibilities

The AMPLab project is organised around three distinct roles, reflecting the entities that are involved and their different responsibilities. These roles are study manager, technical partner and participating country. An outline of the responsibilities of each entity in the different roles is provided in Table 2.

**Table 2: Roles and responsibilities of participating entities**

Role	Entities	Responsibilities
<b>Study Manager</b>	UIS	Determine project parameters and funding Liaise with participating countries Liaise with technical partners
<b>Technical Partner</b>	ACER	Design and provide all documentation and materials for procedures and production Liaise with countries to design and implement school samples Provide data management systems and services Organise and conduct standard setting exercise Undertake analysis Reporting Quality assurance Capacity development
<b>Participating Countries</b>	India Kenya Lesotho Zambia	Liaise with UIS Liaise with technical partners, as appropriate Provision of school sample frame Recruitment of schools Materials production and distribution Data collection, including organising and training test administrators Data entry and data management Documentation of all technical processes Quality assurance Participation in standard setting exercise Participation in capacity building modules

# Method

## The population of interest

In line with SDG indicators 4.1.1(a) and 4.1.1(b) this study aims to investigate learning outcomes at the end of lower primary schooling and at the end of upper primary schooling. However, the definition of the end of lower and upper primary schooling differs across systems and countries. In reporting against SDG indicators 4.1.1(a) and 4.1.1(b) the UIS also allows the flexibility for countries to report at grade levels close to the end of lower and upper primary schooling if the quality and appropriateness of the outcomes data is more suitable for purpose.

In determining the population of interest in each participating country, two main factors will be considered:

1) the structural definition of the end of primary schooling within the country, aligned to international standards (ISCED); and

ACER and the UIS will work with each participating country to determine the most appropriate target population within that country. To assist this, the *AMPLab Sampling Framework* and *AMPLab School Sampling Preparation Guide* documents will be provided and decisions recorded using *AMPLab Sampling Forms*.

## Survey and sample design

To support the goals of this program to measure and analyse the performance of students at the end of lower and upper primary using an assessment that aligns with the GPF, a survey design framework will be developed. The survey design framework will outline the overall approach to obtaining reliable estimates of outcomes and recording survey outcomes, drawing on the methods used in established large-scale assessments.

The sample will be designed to yield an effective sample size of 400 students. As the sample design involves a first stage of sampling schools and a second stage of sampling classes, it is expected that a sample of at least 150 schools and between 3500 and 5000 students will be drawn. More precise estimates of school and sample size will be made on a country by country basis once the effects of clustering students within schools are explored during the sample design phase.

ACER will work with countries to document and quantify key sample design decisions including the target population, coverage and exclusions, explicit and implicit stratification, and school and within-school level exclusions. Negotiation and formal documentation of key sample design decisions is expected. These discussions will be supported by the *AMPLab Sampling Framework* document and a *AMPLab School Sampling Preparation Guide*. Sample design decisions are recorded in the *AMPLab Sampling Forms*.

## Technical standards

Technical standards will help ensure the precision and validity of the sample, the quality of the operations and resulting data and the quality of the psychometric analysis.

Technical standards will be provided for an in-school assessment administration of the assessments and questionnaires.

The standards will include the following key areas:

- Sampling standards relating to the level of precision and validity of the sample, including sample size, response rates and sample coverage of the target population.
- Linguistic quality control in the production of the French and English versions of the assessments and questionnaires.
- Data standards relating to: assessment and questionnaire construction and adaptation; test administration, quality monitoring and security of assessments, questionnaires and data; data management, data cleaning and data protection and security.
- Psychometric standards relating to the methods used to scale the results and to align the results to the SDG 4.1.1.a and SDG 4.1.1.b Minimum Proficiency Levels to ensure that the data is valid and reliable.

The standards will be expected to be adhered to by participating countries and technical partners. ACER will document technical partners' adherence to the standard.

Participating countries will document their own adherence to the standards. ACER will critique documented deviations to the standards and will judge the impact of any deviations on the quality of the resulting data.

## Assessment of learning outcomes

Participating countries will be provided with assessment instruments that will measure the attainment of SDG 4.1.1(a) and SDG 4.1.1(b) Minimum Proficiency Levels (MPL) in reading and mathematics in students at the end of lower and upper primary education.

- The AMPL-A assessment for lower primary education will be audio- and paper-based and available in English and French. The listening comprehension portion of the assessment will be provided in high quality audio files which countries can choose to adapt. The paper-based assessment will be provided in print-ready format.
- The AMPL-A+B assessments for upper primary education will be audio- and paper-based and available in English and French. The listening comprehension portion of the assessment will be provided in high quality audio files which countries can choose to adapt. The paper-based assessment will be provided in print-ready format.



- The AMPL-B assessments for upper primary education will be paper-based, provided in print-ready format, and available in English and French.

The assessments will follow an *Assessment Blueprint* that defines the coverage of the learning areas, domains and constructs as referenced in documentation of the MPLs. The assessment blueprint will specify the coverage of learning areas and the relative proportion of domains. The assessment items will be drawn from the UIS Global Item Bank and will include French and English source items.

The Minimum Proficiency Level in mathematics for end of lower primary schooling is described in ACER (2022) as:

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.

The Minimum Proficiency Level in reading for end of lower primary schooling is described in ACER (2022) as:

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.

The Minimum Proficiency Level in mathematics for end of primary schooling is described in ACER (2022) as:

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. The Minimum Proficiency Level in reading for end of primary schooling is described in ACER (2022) as:

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.

An outline of the selected domains and constructs covered by the AMPLab assessments appears in Table 3 and Table 4.

**Table 3: Domains and constructs in the AMPLab assessments for SDG 4.1.1(a)**

Learning areas	Reading	Mathematics
<b>Domains</b>	Listening Comprehension Decoding Reading Comprehension	Number and Operations Measurement and Geometry Statistics, Probability and Algebra
<b>Constructs</b>	Retrieving information Interpreting information Precision	Whole numbers Length, weight, capacity, volume, area, and perimeter Time Properties of shapes and figures Spatial visualisations Position and direction Data management Patterns

**Table 4: Domains and constructs in the AMPLab assessments for SDG 4.1.1(b)**

Learning areas	Reading	Mathematics
<b>Domains</b>	Reading comprehension	Number and operations Measurement Geometry Statistics and probability Algebra
<b>Constructs</b>	Retrieving information Interpreting information Reflecting on information	Whole numbers Fractions Decimals Integers Exponents and roots Operations across number Length, weight, capacity, volume, area and perimeter Time Currency Spatial visualizations Properties of shapes and figures Position and direction Data Management Chance and probability Patterns Expressions Relations and functions

## AMPLab test design

Countries have the choice of administering one or more of three test designs.

### AMPL-A at the end of lower primary stage

AMPL-A assesses students at the end of lower primary stage on the Minimum Proficiency Levels corresponding to 4.1.1(a). The clusters of items include:

- One cluster of 10 listening comprehension and 5 decoding items delivered via audio with students answering in their booklets
- One cluster of 25 reading items and 5 decoding items, paper-based
- One cluster of 30 mathematics items, paper-based

There are two AMPL-A booklets, to be rotated across learners. The listening comprehension/decoding cluster appears at the beginning of both booklets. Each booklet contains both the reading/decoding and mathematics clusters, but the ordering of the clusters is reversed across the two booklets.

**Table 5: AMPL-A Assessment: test design**

Booklet	Part 1	Part 2	Part 3
AMPL-A Booklet 7	Listening Comp Decoding	Mathematics	Reading Decoding
AMPL-A Booklet 8	Listening Comp Decoding	Reading Decoding	Mathematics

### AMPL-B at the end of upper primary stage

AMPL-B assesses students at the end of upper primary stage on the Minimum Proficiency Levels corresponding to 4.1.1(b). The clusters of items include:

- One cluster of 32 reading items, paper-based
- One cluster of 30 mathematics items, paper-based

There are two AMPL-B booklets, to be rotated across learners. Each booklet contains both the reading and mathematics clusters, but the ordering of the clusters is reversed across the two booklets.

**Table 6: AMPL-B Assessment: test design**

Booklet	Part 1	Part 2
AMPL-B Booklet 1	Mathematics	Reading
AMPL-B Booklet 2	Reading	Mathematics

### AMPL-A+B at the end of upper primary stage

AMPL-A+B assesses students at the end of upper primary stage on the Minimum Proficiency Levels corresponding to both 4.1.1(a) and 4.1.1(b).

The audio-based cluster of 10 listening comprehension and 5 decoding items appears at the beginning of each booklet. Booklets 3 and 4 rotate the AMPL-B reading and mathematics clusters. Booklets 7 and 8 rotate the AMPL-A reading/decoding clusters and the mathematics clusters. Booklets 5 and 6 contain reading and mathematics clusters which are comprised of a mix of AMPL-A and AMPL-B items.

**Table 7: AMPL-A+B Assessment test design**

	Part 1 Audio	Part 2	Part 3
<b>AMPL-A+B Booklet 3</b>	Listen Comp(a) 10 items Decoding(a) 5 items	Maths(b) 30 items	Reading(b) 32 items
<b>AMPL-A+B Booklet 4</b>	Listen Comp(a) 10 items Decoding(a) 5 items	Read(b) 32 items	Maths(b) 30 items
<b>AMPL-A+B Booklet 5</b>	Listen Comp(a) 10 items Decoding(a) 5 items	Maths(a) 15 items Maths(b) 15 items	Read(a) 15 items Read(b) 15 items
<b>AMPL-A+B Booklet 6</b>	Listen Comp(a) 10 items Decoding(a) 5 items	Read(a) 15 items Read(b) 15 items	Maths(a) 15 items Maths(b) 15 items
<b>AMPL-A+B Booklet 7</b>	Listen Comp(a) 10 items Decoding(a) 5 items	Maths(a) 30 items	Decode(a) 5 items Read(a) 25 items
<b>AMPL-A+B Booklet 8</b>	Listen Comp(a) 10 items Decoding(a) 5 items	Decode(a) 5 items Read(a) 25 items	Maths(a) 30 items

## Contextual information

A *Contextual framework* will provide a rationale for the contextual data collected as part of this study. Two questionnaires will be constructed, each focusing on a different level: student and school. The student-level questionnaire will be completed by the students undertaking the assessments and the school-level questionnaire will be completed by school principals. The questionnaires will be paper-based, provided in print-ready format, and available in English and French.

The student questionnaire will gather information on student characteristics, household resources, home support and resources, and student nutrition and sanitation.. The school questionnaire will gather information on the characteristics of the school principal, school characteristics, school facilities and resources, and teachers and students.

## Materials preparation

Printing of the booklets, questionnaires and administration manuals will be handled by the institutions within participating countries. ACER will provide print ready assessments and questionnaires, and test construction guidelines.

Some adaptation of the international source version of the questionnaires will be required and will be the responsibility of the participating institution and the technical partner.

ACER will also provide high quality international source audio files for the listening comprehension portion of the assessment. If countries choose to re-record the audio in their own accents, participating institutions will be responsible for producing the localised version. Guidance will be provided by ACER in the test construction guidelines.

The procurement of audio playback devices and speakers to facilitate the listening comprehension assessment will be the responsibility of the institutions within participating countries.

## Data collection

Institutions within participating countries will be responsible for the data collection. Data on student learning outcomes and contextual information will be collected through administering traditional pencil and paper tests and a questionnaire to children at the end of lower and/or upper primary school<sup>1</sup>, in school settings. Students participating in AMPL-A or AMPL-A+B will also complete an audio-based component of the test. ACER will provide supporting materials such as field operations manuals.

## Ordering and timing

For countries administering AMPL-B, the assessment is administered in one day and the ordering and timing is as follows:

Session 1. First hour AMPL-B (+introduction/practice)

Session 2. Second hour AMPL-B

Session 3: 20-25 min Questionnaire

For countries administering AMPL-A or AMPL-A+B, the assessment should be administered over two days and the ordering and timing is as follows:

Day 1

- Session 1. 25 minutes: AMPL-A/AMPL-A+B audio (+introduction/practice)
- Session 2. 65 mins: AMPL-A/AMPL-A+B (+introduction/practice)

Day 2

- Session 3: 60 minutes AMPL-A/AMPL-A+B
- Session 4: 20-25 minutes Questionnaire

## Data management

ACER will provide a data management tool called ACER Maple to each participating country. It is essential that this tool is used to:

---

<sup>1</sup> The definition of end of lower and upper primary schooling will likely vary across countries. Defining the end of lower and upper primary school for the purpose of reporting against SDG 4.1.1(a) and SDG 4.1.1(b) will be determined on a case by case basis in consultation with each country and the UIS.

- Track school participation
- Draw the within-school sample
- Produce student tracking forms
- Enter the assessment data and the contextual data from the questionnaires
- Record the participation of students as per the tracking form
- Submit the data to ACER.

## Standard setting exercise

To enable robust and valid reporting of student achievement against the MPL requirements, a systematic approach will be taken to establish cut-scores that correspond to the end of the lower primary MPL requirements for each AMPL domain (reading and mathematics). Note that the cut-scores that correspond to the end of primary MPL requirements have already been established.

A standard setting exercise will be conducted in cooperation with officials and subject matter experts from all AMPL participating countries. A modified Yes/No Angoff method will be used to determine a single MPL(a) cut-score for mathematics and a single MPL(a) cut-score for reading. The Angoff method is based on the concept of the borderline or minimally competent student– target student.

## Psychometric analysis

The assessment data will be psychometrically scaled, using a two-dimensional model to produce estimates for mathematics and reading learning outcomes. Using the benchmarks set in the standards setting exercise, the proportion of students meeting or exceeding the MPLs for SDG 4.1.1a and SDG 4.1.1b will be estimated.

## Data analysis

Analysis of the data will focus on the following research questions:

- What proportion of each target grade are reaching SDG4.1.1a Minimum Proficiency Levels in reading and mathematics, by gender?
- What proportion of each target grade are reaching SDG4.1.1b Minimum Proficiency Levels in reading and mathematics, by gender?

## Reporting

The reporting will clearly articulate the outcomes of the study, in a way that is clear to policy and practitioner stakeholders.

ACER will provide the following study reports a final report on the performance of students at the end of lower and upper primary school, with a focus on gender

The final report will include:

- A brief description of the study purpose and design, the target population, country samples and response rates
- A statistical summary of the scale score distribution by country and by gender
- A statistical summary of the proportion of students reaching/exceeding SDG 4.1.1.a Minimum Proficiency Level cut-point by country and by gender
- A statistical summary of the proportion of students reaching/exceeding SDG 4.1.1.b Minimum Proficiency Level cut-point by country and by gender
- Descriptive statistics of contextual factors at the student and school levels
- Inferential statistics of associations of contextual factors and achievement, aggregated at different levels.
- Individual country summaries with a focus on gender

The final report will be provided by ACER to the UIS for dissemination.

ACER will compile a comprehensive technical note that summarises and clearly describes all data and statistical conventions and approaches applied in the AMPLab project. The technical report will be written as a reference text.

## Capacity development

Capacity development will be an integral part of the project implementation. The overall aim of the capacity development is to build sustainable capacity of national teams in developing, implementing and using data from large-scale learning assessments for education system monitoring.

The capacity development activities will be based on the Principles of Good Practice in Learning Assessment (ACER & UIS, 2017) and aligned to the different products described in this proposal.

For the areas of Sampling, Test Administration, and Data Management, capacity building activities have been integrated into the project. A focus is thereby on skill development through working directly with, and in consultation with, ACER experts on key tasks. The documentation of technical aspects such as key sample design decisions, quality assurance of standardised administration procedures, and data validation procedures, is important to ensure the technical rigour of the assessment. Enhanced skills in using software tools for sampling and data management that have in-built quality assurance mechanisms, are essential to increase and ensure the accuracy and quality of the data.

In order to design additional capacity building activities that will support the national teams, ACER will provide 2 capacity building modules:

- Module 1



- Plausible Values (PVs)
- PV reliability (reliability/measurement error as a design effect)
- Population models
- Law of total variance (Eve's law)
- Module 2
  - Complex survey design (sampling weights, and 2-stage sampling)
  - Sampling error
  - Accounting for sampling error using replicate weights
  - Putting it all together – an example using large scale assessment data

These courses are delivered online through ACER's learning management system (Moodle), are self-paced, and include facilitated live sessions delivered via virtual conferencing software (Microsoft Teams). A learning assurance task (LAT) is included at the end of each course, which participants will be asked to complete in order to receive receive grading and feedback.

## References

ACER & UIS. (2017). *Principles of good practice in learning assessment*. <http://uis.unesco.org/sites/default/files/documents/principles-goodpractice-learning-assessments-2017-en.pdf>

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