CHAPTER 3

The mathematics assessment

HIGHLIGHTS

- The AMPL for mathematics assesses the following five key aspects of mathematics at upper primary level: number and operations, measurement, geometry, statistics and probability, and algebra (i.e. number patterns and missing number problems) (Table 3.1).
- The AMPL for mathematics is strongly aligned to the Global Proficiency Framework, enabling the reporting against SDG 4.1.1b.
- The assessment material included in the AMPL was selected from the UIS's Global Item Bank using a set of quality assurance guidelines. The 29 selected mathematics items came from nine different sources, with some originating in French, and others in English.
- The AMPL booklets contained a set of mathematics material and a set of reading material. The booklets were provided to students in their language of instruction (French or English) and students had one hour to complete the booklet.

INTRODUCTION

As outlined in Chapter 1, a main goal of the MILO study was to determine the impact of COVID-19 on learning outcomes for students at the end of primary school. In order to achieve this aim, assessments for minimum proficiency levels (AMPL) were designed to measure proficiency in reading and mathematics at the end of primary school in 2021. The construct validity of these assessments is addressed in this chapter. The development process led to highly reliable instrumentation (for details see Appendix B).

The performance of the 2021 population was compared to that of an equivalent cohort from a period prior to the COVID-19 outbreak. For a technical description of the analysis methods used to link the MILO data with the past historical assessment results see Appendix B. The focus of this chapter is on the features of the AMPL for mathematics.

ASSESSMENT OF MATHEMATICS PERFORMANCE IN MILO

The MPL for upper primary for mathematics provided the overarching conceptualisation of mathematics in the AMPL. The parts of mathematics referred to in the MPL (described in detail later in this chapter) parallel those in the Global Proficiency Framework (GPF) (USAID et al., 2020b) and the GPF was used to provide additional detail for the item selection. The AMPL included five domains of mathematics in line with the GPF. As outlined in the MILO assessment blueprint, content targets were set at the domain level.

These targets were:

- Number and operations items make up 35–45% of the assessment
- Measurement items make up 15–20% of the assessment
- Geometry items make up 15–20% of the assessment
- Statistics and probability items make up 10–15% of the assessment
- Algebra items make up 10-15% of the assessment.

The mathematics targets were developed with reference to existing large-scale and regional assessments, and the work of the GPF advisory group on alignment.[4] For example, the content breakdown is analogous to that used in the large-scale international assessment Trends in International Mathematics and Science Study (TIMSS) (Mullis & Martin, 2017). In TIMSS Grade 4, the equivalent breakdown is 50% Number (including Algebra as defined in the GPF), 30% Measurement and geometry, and 20% Data.

Table 3.1 shows the classification of the final items in the assessment against the specified targets.

TABLE 3.1 Final AMPL mathematics items and targets by construct

Construct	Items in AMPL (no.)	Items in AMPL (%)	Target percentage (%)
Number and operations	12	41	35–45
Measurement	4	14	15–20
Geometry	5	17	15–20
Statistics and probability	4	14	10–15
Algebra	4	14	10–15

The GPF advisory group on alignment specified that in order to be considered 'strongly aligned' with the GPF, an assessment needs to include:

- at least five items from the Number and operations domain
- at least five items from the Measurement and Geometry domains
- at least five items from the Statistics and probability, and Algebra domains
- a total of 50% of all the sub-constructs in the mathematics GPF that are relevant to the target grade level. For example, if there are 20 sub-constructs at Grade 5, at least 10 of the sub-constructs should be included in the assessment.

The upper primary MPL has been defined by the Global Alliance for Monitoring Learning and the Technical Cooperation Group (TCG) on the Indicators for SDG 4 as spanning grade levels 4, 5 and 6. There are 23 sub-constructs in the GPF for mathematics that are relevant to either or all of these three grade levels. Of the 23 sub-constructs relevant to Grades 4, 5 and/or 6, there were 16 included in the AMPL. The AMPL was therefore strongly aligned to the GPF, with all mathematics targets met. Appendix C provides further detail about the coverage of GPF constructs and sub-constructs within the AMPL.

Items were selected from the UIS's Global Item Bank to meet the assessment blueprint after an extensive review process. Two expert reviewers for each of English and French independently reviewed a set of material. They were provided with item review guidelines and asked to consider issues such as construct validity (whether the item assesses a part of mathematics as described in the GPF), translatability (whether there are features of the material that might make it difficult to translate), cultural issues and technical criteria (clarity and correctness, centrality, appropriate level of difficulty). Only items that attained a high overall rating and for which no significant concerns were identified were considered for inclusion in the AMPL. From the set of suitable items, a selection was made that:

- met the requirements of the assessment blueprint
- contained items that originated in both English and French
- represented a range of sources (nine different sources for the 29 items were included)
- represented a range of materials (e.g. context-free items, real-world problems)
- represented a range of difficulty levels that were appropriate for the target population and for measuring the minimum proficiency levels at the end of primary.

As described in Chapter 2, there were two AMPL booklets, each containing a set of mathematics material and a set of reading material. The testing time was one hour (30 minutes each for reading and mathematics.

As described in Chapter 1, a key goal of the MILO project was to evaluate the impact of COVID-19 on learning outcomes by reporting against SDG indicator 4.1.1b '... the proportion of children and young learners ... at the end of primary ... achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex' (United Nations, 2015).

The MPL for the end of primary for mathematics is discussed and illustrated in the section that follows. This information is taken from two papers that were presented and endorsed in 2019 and 2020 meetings of the Global Alliance to Monitor Learning entitled 'Minimum Proficiency Levels: Described, unpacked and illustrated.' (ACER-GEM, 2019) and 'Minimum Proficiency Levels Revisions Proposed by ACER' (ACER-GEM, 2020).

The MPL is described and elaborated in the following four ways, providing different levels of detail for different audiences:

1. A nutshell statement: provides brief information for all readers about each learning area, by educational level.
2. An expanded statement: provides information for those working in the field of education.
3. Descriptors by construct: these elaborations use more technical language and are suitable for educators and researchers.
4. Sample items: a small set of sample items, one below, one at, and one above the MPL.

MATHEMATICS: END OF PRIMARY (SDG 4.1.1B)

Nutshell statement

Students recognise, read, write, order, compare and calculate with whole numbers, simple fractions and decimals. Students can measure length and weight using standard units, calculate the perimeter of simple two-dimensional shapes and area of rectangles. 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 1000 and demonstrate fluency with multiplication facts up to 10 × 10 and related division facts; solve simple real-world problems with whole numbers using the four operations (consistent with the grade and performance level) and identify simple equivalent fractions; select and use a variety of tools to measure and compare length, weight and capacity/volume; understand the relationships between different units of time, e.g. seconds, minutes, hours, days, weeks, months, and years; retrieve multiple pieces of information from data displays to solve problems; recognise and name two-dimensional shapes by their simple attributes; and apply the concept of equivalence by finding a missing value in a number sentence.

Constructs and Descriptors

Number knowledge

NUMBER SENSE (COUNTING, READING, WRITING, COMPARING, AND ORDERING)

Read, write, compare, and order whole numbers up to 10,000.

Skip count forwards and backwards using twos, fives, tens, hundreds, and thousands.

NUMBER SENSE (USING PLACE VALUE AND ROUNDING)

Round numbers up to the nearest hundred and thousand.

OPERATIONS (ADDING AND SUBTRACTING)

Add and subtract whole numbers within 1000.

OPERATIONS (MULTIPLYING AND DIVIDING)

Demonstrate fluency with multiplication facts up to 10 × 10, and related division facts.

REAL-WORD PROBLEMS

Solve simple real-world problems using the four operations, with the unknown in different positions.

FRACTIONS

Identify simple equivalent fractions where one denominator is a multiple of another (e.g. $\frac{1}{3} = \frac{2}{6}$).

Compare and order unit fractions (e.g. $\frac{1}{4}, \frac{1}{3}, \frac{1}{2}$) or fractions with the same denominator ($\frac{1}{8}, \frac{3}{8}, \frac{5}{8}$).

DECIMALS

Identify and represent decimal numbers up to the tenths place (e.g. identify that 0.8 is 8 tenths).

Compare and order decimal numbers up to the tenths place (e.g. sort the following decimals from high to low: 0.8, 0.3, 0.1).

Measurement

MEASUREMENT UNITS (STANDARD AND NON-STANDARD)

Select and use a variety of tools to measure and compare length, weight, and capacity/volume.

AREA, PERIMETER, AND VOLUME

Solve problems, including real-world problems, involving the perimeter of a rectangle using concrete or pictorial representations of units (e.g. grid squares).

TIME

Tell time using an analogue clock to the nearest quarter hour.

Solve problems involving elapsed time in half hour increments within an hour (e.g. difference between 3.00 and 3.30).

Understand the relationships between different units of time (e.g. seconds, minutes, hours, days, weeks, months, and years).

Statistics and probability

DATA MANAGEMENT

Complete missing information in simple data displays using data arranged into categories, with some support provided (e.g. labelled horizontal and/or vertical axes).

Retrieve multiple pieces of information from data displays to solve problems (e.g. calculate a total represented by multiple bars on a graph).

Geometry

CONSTRUCTIONS

Compose a larger two-dimensional shape from a small number of shapes in more than one way (if possible).

Decompose a larger two-dimensional shape into a small number of shapes in more than one way (if possible).

Recognise parallel and perpendicular lines.

PROPERTIES

Recognise and name two-dimensional shapes by their attributes (e.g. their lines and informal angle properties).

Recognise the congruence and similarity of two-dimensional shapes (e.g. shapes that have been reflected, translated, rotated, enlarged, or reduced).

POSITION AND DIRECTION

Follow more complex directions and/or give simple directions to a given location (e.g. go straight, turn right at the corner with the tree, turn left at the next corner, keep going to the green house).

Algebra

PATTERNS

Describe numerical patterns as increasing by a constant value but starting at a number that is not a multiple of the value of the pattern (e.g, the pattern 5, 8, 11, 14 starts at 5 and goes up by 3).

RELATIONS AND FUNCTIONS

Demonstrate understanding of equivalence by finding a missing value in a number sentence using addition or subtraction of numbers within 100 (e.g. 23 + ____ = 29).

SAMPLE ITEMS

Three sample items are included, one below, one at, and one above the MPL. Two English items and one French item are included. Sample items are released items from the PASEC 2014 assessments (CONFEMEN, 2015) and were included in the AMPL.[9]

EXAMPLE 1: An item below the MPL

What units do you use to measure the length of a classroom?

A. metres
B. kilograms
C. litres
D. hours

Source: PASEC (CONFEMEN, 2015)

Domain	Construct	Descriptor	International percentage correct
Measurement	Length, weight, capacity, volume, area and perimeter	Select appropriate standard units to measure length and distinguish from mass, capacity and time units.	65%

Task solution and commentary

Option A (metres) is selected. This task invites students to consider a familiar measurement task (length of a classroom); and then to identify the relevant unit to use for the measurement from those presented. Given that, of the units offered, only the first option is a length unit, this should be a very straightforward task for end of primary students. The relevant MPL statement at the end of primary states 'Select and use a variety of tools to measure and compare length, weight, and capacity/volume'. As this item requires students to recognise a common unit, it is below the end of primary MPL.

EXAMPLE 2: An item at the MPL

Le tableau suivant donne le nombre de filles et de garçons dans les classes d'une école

	CP1	CP2	CE1	CE2	CM1	CM2
Filles	16	15	18	16	20	1
Garçons	20	18	15	12	16	14

Quel est le nombre total de filles de CP1 et CP2?

A. 15
B. 16
C. 31
D. 38

Source: PASEC (CONFEMEN, 2015)

Domain	Construct	Descriptor	International percentage correct
Statistics and probability	Data management	Interpret a two-way table to calculate a total running across two categories.	45%

Task solution and commentary

Option C (31) is selected. This task requires students to read a two-way table and select multiple pieces of information to solve a simple word problem. The MPL at the end of primary refers to the ability to 'retrieve multiple pieces of information from data displays to solve problems'. Therefore, this item is clearly at the end of primary MPL. The skill of adding two-digit numbers, which is also a feature of this item, is well below the descriptor at the end of primary MPL, which refers to the ability to 'add and subtract whole numbers within 1000.' Therefore, the major cognitive load in this item is on the statistical skill of retrieving data from a table, rather than the skill of adding two-digit numbers.

EXAMPLE 3: An item above the MPL

The length of a rectangle is 50 m, and its area 500 m^2.
What is the width of the rectangle?

A. 10m
B. 50m
C. 450m
D. 550m

Source: PASEC (CONFEMEN, 2015)

Domain	Construct	Descriptor	International percentage correct
Measurement	Length, weight, capacity, volume, area and perimeter	Given an area and the length of a rectangle, identify the width of the rectangle.	38%

Task solution and commentary

Option A (10 m) is selected. This task requires students to have a sound conceptual understanding of the method for calculating the area of rectangle, and also knowledge of how to apply this to a slightly more complex problem. The task is not just a straightforward application of area where students multiply side lengths of the rectangle to calculate area. Rather this item requires students to use a given area and one side length to find an unknown width. The MPL at the end of primary refers to the ability to 'calculate the perimeter of simple 2D shapes and area of rectangles'. This question goes beyond the simple calculation of the area of a rectangle, as it requires students to use a strategy such as transposing the relationship length multiplied by width to divide area by length and find the unknown width, or to set up a 'missing number' problem such as 50 × ? = 500 and then identify which of the options provided would be the required missing number. Therefore, this item is above the end of primary MPL.

Endnotes

1 The proportion of children and young learners … at the end of primary … achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex (United Nations, 2015).

2 In 2016 for Zambia

3 Contextual data from the historical population for Zambia was not available in a format suitable for direct comparisons of populations. Some contextual data was not available from the Kenyan historical assessment.

4 The GPF advisory group on alignment was a working group comprised of psychometricians and subject matter experts who contributed to the development of the Global Proficiency Framework in 2020. The group was convened to formulate a set of alignment criteria to allow assessments to be compared to the GPF in order to determine their suitability for evaluating and reporting against SDG 4.1.1. The alignment criteria are outlined in detail in: USAID, UIS, UK Aid et al. (2020) *Policy Linking Toolkit for Measuring Global Learning Outcomes – Linking assessments to the Global Proficiency Framework.*

5 From SDG 4.1.1 Review Panel: March 2021.

6 These items were reproduced with permission from CONFEMEN.

7 For the purposes of AMPL, this item was classified as "Retrieve information" rather than "Decoding" as consistent with the GPF for reading (USAID et al, 2020a) which lists matching a given word to an illustration as an example of retrieving information.

8 The four French-speaking countries were Burkina Faso, Burundi, Côte D'Ivoire and Senegal.

9 These items are used with permission from CONFEMEN.

10 Zambia's historical assessment was conducted in 2016. All other countries' historical assessments were conducted in 2019.

11 Historical results are not reported for Kenya since the 2019 assessment of English in Kenya did not contain a sufficient number of reading comprehension item to align with the reading constructs within the GPF.

12 In the MILO project, students were the primary sampled unit. All results from the School Questionnaire are reported using student weights that are representative of the population. Therefore all results from school principals need to be interpreted in numbers of students.

13 There is no consensus among researchers and practitioners on which are the best indicators to operationalise SES. Typical children SES indicators are parents' occupation and education level, household income and home possessions. For a review of SES indicators used in educational research and other disciplines such as health, economics and sociology see Osses et al. (forthcoming).

14 Results for Kenya have been excluded based on data validation issues

15 The population chosen by countries to report against varied from Grade 5 to Grade 7.

16 A wealth index for Kenyan students was computed based on common items from the historical assessment and the AMPL. Comparisons for boys over time revealed higher scores on the wealth index in the 2021 population in comparison to the historical population.

17 For further information on different learning approaches and the benefits, considerations and enabling conditions, see for example Dabrowski et al. (2020).

18 For further recommendations relating to education in emergencies, see the Policy Monitoring tool developed for building resilient education systems (Tarricone et al., 2021).

19 Magnitude of item by gender interaction estimates from a facet model. See PISA 2006 Technical Report (OECD, 2009a).

20 'Not reached' items were defined as all consecutive missing values at the end of the test, except the first missing value of the missing series which was coded as 'embedded missing' i.e. coded the same as other items that were presented to the student but which did not receive a response. Omitting the 'not reached' items from the item calibration ensures the item difficulties not to be over-estimated.

21 The psychometric properties of the reading items administered in Burundi was unexpectedly inconsistent with those of the other countries. In particular, the response patterns in nearly all of the reading items was consistent with high rates of guessing and resulted in very low discrimination. It was therefore decided to exclude Burundi from the international reading item calibration. Burundi student reading proficiency estimations were subsequently based on the international calibration.

22 Expected a-posteriori/plausible value (EAP/PV) reliability (Adams, 2005).

23 A two-dimensional model with Quadrature estimation with 40 nodes was used.

24 So-called weighted likelihood estimates (WLEs) were used as ability estimates in this case (Warm, 1989).

25 Conceptual background and application of macros with examples are described in the PISA Data Analysis Manual SPSS®, 2nd edn (OECD, 2009b).