© UNICEF/UNI284500/CHINYENZE

CHAPTER 1

# Introduction

## BACKGROUND OF THE STUDY

The COVID-19 pandemic has disrupted education in many ways. Across the world, schools have been partially or wholly closed, teachers and students have been forced to quarantine at home for short or extended periods of time, social learning opportunities have been cancelled and community interactions curtailed. This has added a further obstacle to the achievement of the Sustainable Development Goals (SDGs) related to education (United Nations Department of Economic and Social Affairs, 2020; UNESCO, 2020a).

Simulations on the impact of COVID-19 school closures on learning outcomes have suggested that school closures could result in significant learning loss, which could continue to accumulate even after schools re-open. School closures can reduce the number of effective years of basic schooling that students achieve in their lifetime with the consequence of reduced or lost future earnings (Azevedo et al., 2020; Kaffenberger, 2021).

Initial evidence from Africa has suggested that the pandemic had negative consequences for students, with fewer undertaking learning activities leading to considerable learning losses and compounding disadvantage for students who come from households with fewer resources available (Ardington et al., 2021; Dang et al., 2021). Other research has shown that students who were already behind in their learning are also at risk of falling further behind compared to those who already had mastery of skills (Tarricone et al., 2021). These findings highlighted the need for an investigation into the impact of COVID-19 on learning outcomes and what efforts countries are making to help mitigate any learning loss.

In order to adequately measure any changes in learning outcomes, there needs to be data collected at two or more points in time; one prior to the outbreak of COVID-19 (a baseline measure) and one after the disruption (to determine any change over time). While other research in Africa during the COVID-19 pandemic has relied on simulations, retrospective data or smaller non-representative data collections, the Monitoring the Impact of Learning Outcomes (MILO) project is unique in that it is able to use reliable and valid assessments of reading and mathematics at two points in time. Using sampling approaches that represent the target populations, the MILO project allows direct comparison of student performance before the pandemic to performance in 2021 following the period of disruption. As described further, the MILO project also provides a way for countries to measure progress towards Sustainable Development Goal (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)

This introduction provides information about the purpose of the MILO project, outlining its overarching goals. It provides an overview of the study design, including information about the assessment blueprint for the tests, the conceptual framework for the contextual questionnaires, the MILO instruments, the historical instruments and the MILO countries and samples. Lastly, this chapter provides an outline of the rest of this MILO report.

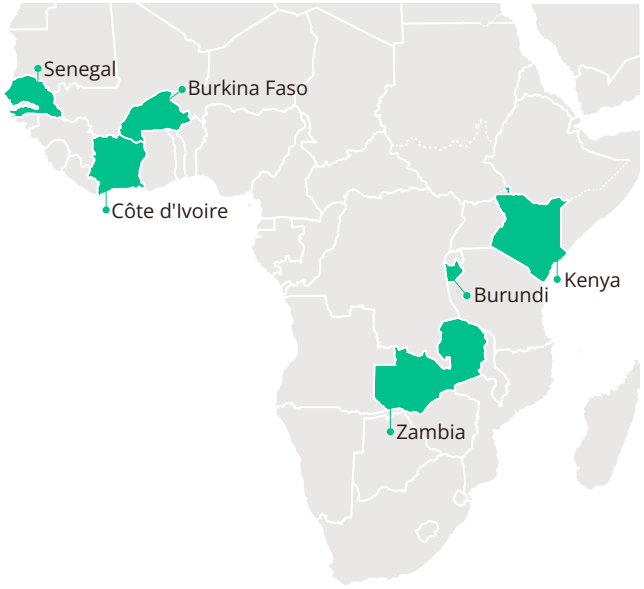
## PURPOSE OF MILO

The MILO study is a UNESCO Institute for Statistics (UIS) project and was funded by the Global Partnership for Education (GPE).

The four overarching goals of the project were to:

- evaluate the impact of COVID-19 on reading and mathematics learning outcomes by reporting against SDG indicator 4.1.1b.

**FIGURE 1.1** Map of MILO countries




- identify the impact of different distance learning mechanisms put in place to remediate the learning disruption generated by COVID-19
- expand the UIS bank of items for primary education
- generate a toolkit to scale assessment results to international benchmarks, reporting against SDG 4.1.1b.

The MILO project focused on six African Anglophone and Francophone countries, chosen by the UIS as they had existing pre-pandemic national or regional assessment data. The countries were Burkina Faso, Burundi, Côte d'Ivoire, Kenya, Senegal and Zambia, as shown in Figure 1.1. A National Centre in each country was responsible for implementing the project within their country. The Australian Council for Educational Research (ACER) was the technical partner for this project and technical and implementation support was provided by The Conference of Ministers of Education of French-Speaking Countries (CONFEMEN).

This report will focus on the first two goals. The report addresses the first goal by evaluating the degree to which learning outcomes for students at the end of primary schooling change between two time points: one pre-pandemic and the other

in mid-2021 after the pandemic had inflicted substantial disruption upon education contexts. The report addresses the second goal by examining contextual factors at the student, family, school and system levels for their response to the pandemic disruption. How these contextual factors relate to any change in outcomes over time is also explored. Discussion on the variety of educational responses to COVID-19 and recommendations for building more resilient education systems are included in this report. For details on how the MILO project addressed the third and fourth goals, see Box 1.

**BOX 1**

## The Global Item Bank and the Assessments for Minimum Proficiency Levels

The UIS's Global Item Bank provides a global public repository of items which can be used to generate assessment data to measure reading and mathematics and report against SDG 4.1.1. As part of the MILO project, items were added to the Global Item Bank expanding the pool of high quality items available to countries. Quality assurance guidelines were also developed to enhance future contributions to the item bank.

The Global Item Bank can be used to develop assessments efficiently using high-quality material that enables reporting against SDG 4.1.1. In the MILO project, the Assessments for Minimum Proficiency Levels (AMPL) were created using English and French-source items from the UIS's Global Item Bank (see Chapters 2 and 3).

Additionally, the MILO project generated a set of tools for the UIS that can be used by countries to measure and report leaning outcomes against SDG 4.1.1b. The toolkit includes the AMPL-b, along with supporting documentation used in the MILO project to support the implementation – technical standards, the assessment blueprint, contextual framework, field operations guidelines and a description of the analysis methods used in the study (see Appendix B).

The AMPL-b is a robust and efficient assessment tool that measures the proportion of students meeting SDG 4.1.1b. Beyond 2021, the AMPL-b are resources provided by the UIS that can be used by countries and assessment programs to monitor progress against SDG 4.1.1b. Should a country, region or system want to report against SDG 4.1.1b in the future, the AMPL-b can be implemented as a standalone assessment. The AMPL-b targets the Minimum Proficiency Levels (MPLs). However, should a country, region or system want to measure and describe the broad range of abilities that children at the end of primary school may exhibit in reading and mathematics, in addition to reporting against SDG 4.1.1b, the AMPL-b can be integrated into existing national or regional assessments. For example, this can be done by administering the AMPL-b forms alongside existing assessments, as was done in the MILO project.

The development of the AMPL-b is a significant step forward and has the potential to align national and cross-national assessment programs to a single set of global standards in mathematics and reading as articulated in SDG 4.1.1, and elaborated by the definitions of the Minimum Proficiency Levels (ACER-GEM, 2019, 2020) and the Global Proficiency Frameworks (USAID et al., 2020a, 2020b). The AMPL-b is currently available in English and French but can readily be adapted and translated, and could include additional items set above or below the MPLs.

Currently, the AMPL-b covers the end of primary schooling outcomes, SDG 4.1.1b. However, the same methods could be applied if further assessments are developed to measure learning outcomes at the end of lower secondary to address SDG 4.1.1c (AMPL-c) or the end of lower primary, SDG 4.1.1a (AMPL-a).

A major finding of the MILO project was that, in general, the learning outcomes across time were stable (see Chapter 4). That is, the student population assessed after the pandemic did not perform worse than the population assessed before the pandemic. This report details key information about the design of the MILO study that allows that measurement to take place, the contexts in which the stability of outcomes took place (at the student, home, school and educational system level), the possible reasons why no changes were observed, and the implications of the findings for students in the region.

## STUDY DESIGN

The main aim of this study was to determine the impact of COVID-19 on learning outcomes at the end of primary schooling. To quantify current learning outcomes, an assessment of reading and mathematics was administered to students at the end of primary school in mid-2021. These assessment data also provided a means of comparison against assessment data from previous years. The performance for the target population was compared against an equivalent cohort prior to the outbreak of the COVID-19 pandemic in 2019.[2] For further information about the methods used to compare the results pre-pandemic to the results in 2021, see Appendix B.

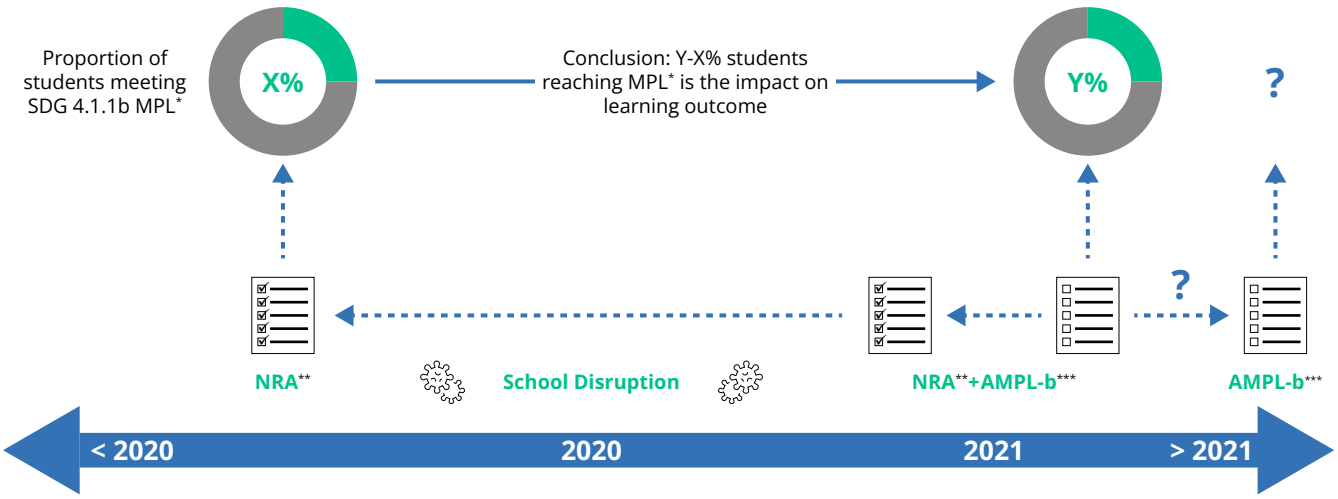
The design of the study is outlined in Figure 1.2. The figure shows that prior to the onset of the pandemic, students in the MILO countries participated in a national or regional learning assessment (NRA). This assessment is referred to as 'the historical assessment' in this report. A sub-set of the same historical assessment was administered to an equivalent cohort in 2021 alongside the *Assessments for Minimum Proficiency Levels for SDG 4.1.1b* (AMPL-b) tests, with the period of COVID-19 disruption somewhere in-between.

### Assessment blueprint for AMPL-b tests

The MILO Assessment Blueprint outlines details of the two learning areas that are assessed in the MILO project: reading and mathematics. In line with the Global Proficiency Frameworks (GPFs) (USAID et al., 2020a, 2020b), reading and mathematics are referred to as 'learning areas', which are then broken down into domains, constructs, and sub-constructs.

The MILO project focuses on students at the end of primary schooling. However, the definition of 'the end of primary schooling' differs across systems and countries. In this Study, the benchmark used to indicate learning outcomes is aligned with SDG indicator 4.1.1b.

**FIGURE 1.2** Study design




* MPL: Minimum Proficiency Level ** NRA: National or regional learning assessment
*** AMPL - b: Assessments for Minimum Proficiency Levels for SDG 4.1.1b at the end of primary

The assessments used in the MILO project are labelled Assessments for Minimum Proficiency Levels for SDG 4.1.1b (AMPL-b). The Minimum Proficiency Level (MPL) in reading for end of primary schooling is:

> 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. (ACER-GEM, 2020, p. 6)

The MPL in mathematics for end of primary schooling is:

> 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 2D 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. (ACER-GEM, 2020, p. 4)

The items in the AMPL-b tests were chosen to match the constructs expressed through the GPF (USAID et al., 2020a, 2020b). A participatory standard setting exercise involving experts from all six MILO countries was used to set a single cut-point for reading and a single cut-point for mathematics in the tests. The cut-point is the MPL at the end of primary schooling, as referred to in SDG 4.1.1b These cut-points were used to determine the proportion of students above and below the SDG 4.1.1.b MPLs in 2021 and in the historical assessments.

Note that AMPL-b focused on a single cut-point of the MPL for efficiency. An assessment that would be used to more deeply describe the entire range of reading or mathematics outcomes in a population requires more items, more development time, and they are usually more complex to implement and analyse. The urgent need for information on the impact of the pandemic precluded such a lengthy process. The AMPL-b were designed to be efficiently developed and implemented. AMPL-b are fit for the purpose of providing estimates against a single global indicator of learning outcomes. For further information about the standard setting exercise, see Appendix A, and full details about the AMPL-b tests can be found in Chapters 2 and 3.

## Conceptual Framework for contextual questionnaires

A Conceptual Framework underpins the design of the MILO questionnaires. It includes the types of data needed in order to achieve the MILO objectives, which were:

- to understand how the COVID-19 disruption affected learning
- to quantify any learning loss
- to identify how to support student learning.

The Framework is organised into six themes, and the impact of the COVID-19 disruption organised into three layers (see Figure 1.3):

- student characteristics
- the home environment
- the school environment, which includes two sub-themes, teaching and learning, and assessment and monitoring.

Six themes were used to provide a comprehensive picture of the way the pandemic affected each of these three levels:

- **UNDERSTANDING THE COVID-19 DISRUPTION:** Data were collected on how the pandemic disruption impacted different school systems, schools and students. This was the foundation on which other data were collected, in that the other data related to the effects of COVID-19.

- **STUDENT CHARACTERISTICS:** Student characteristics data were collected based on demographic categories that included, gender, students with disability or special needs, and students from ethnic, linguistic, refugee or internally displaced backgrounds. These categories were derived from the literature, where evidence shows that students with particular qualities or from certain backgrounds are more vulnerable to learning loss during emergencies in education.

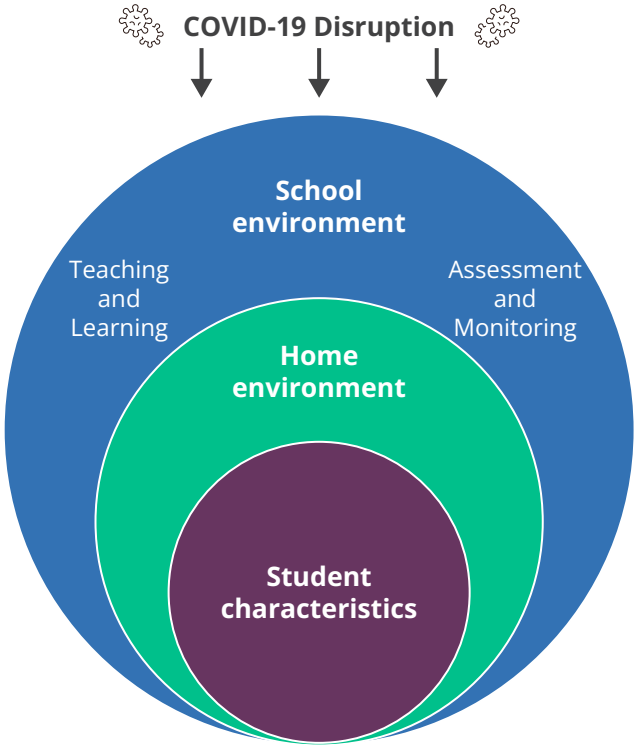
- **HOME ENVIRONMENT:** Data about the home environment focused on the home circumstances of students that might enable or inhibit learning during the COVID-19 disruption, regardless of more enduring personal characteristics. The contexts of the home environment are expected to have a profound influence on the degree in which the COVID-19 disruption enables or inhibits learning, regardless of more enduring personal characteristics.

- **SCHOOL ENVIRONMENT:** School environment data related to the resources and actions of schools within their national systems and which could either exaggerate or insulate children from the COVID-19 disruption. While policies and procedures at the school to combat disruption were often dictated at the system level, individual school characteristics and individual school responses to address these policies may play a role in reducing the extent of the disruption on students. This included factors such as school leadership, school characteristics, resources and location as well as national policies and plans that impact schools.

- **TEACHING AND LEARNING:** Teaching and learning practices fall within the broader school environment and data were collected about pre- and post-disruption classroom and school practices as well as about student experiences in their school work.

- **ASSESSMENT AND MONITORING:** There is greater risk of unequal learning progress during periods of disruption. The MILO questionnaires captured information on assessments conducted on students and the degree to which both students and staff were monitored throughout the disruption in relation to wellbeing and for students, academic progression.

**FIGURE 1.3** Conceptual Framework for MILO contextual data




## The MILO instruments

There was a suite of MILO instruments:

- The AMPL-b **test of reading performance** consisted of 29 items within the reading comprehension domain that covered the constructs *retrieve information*, *interpret information* and *reflect on information*. The same items were used in two AMPL-b test booklets with the items presented by domain in set order; Booklet 1 contained reading and then maths items, and Booklet 2 contained maths and then reading items. The AMPL was created using items from the UIS's Global Item Bank and is used to estimate the proportions of students who meet the MPL referred to in SDG 4.1.1b. See Chapter 2 for more information.

- The AMPL-b **test of numeracy performance** consisted of 29 items within the mathematics domain relating to number and operations, measurement, geometry, statistics and probability and basic concepts of algebra. The same items were used in two booklets. As with reading, the AMPL was created using items from the UIS's Global Item Bank and is used to estimate the proportions of students who meet the MPL referred to in SDG 4.1.1b. See Chapter 3 for more information.
- A **Student Questionnaire** was given to the same students who completed each of the two AMPL-b tests. The questionnaire consisted of 91 items grouped into 27 questions, with topics about demographic and home characteristics of the students, their experiences using technologies as a result of the pandemic and information about the nature of their schooling during the pandemic.
- A **School Questionnaire was completed by school principals or their delegates** and consisted of 177 items grouped into 27 questions. The questionnaire collected information about how COVID-19 impacted each school's ability to deliver teaching and learning activities, as well as any ongoing consequences of the pandemic. In addition, the questionnaire elicited information about schools in general to aid the interpretation of the reading and numeracy performance and Student Questionnaire responses.
- **A System Questionnaire of 13 questions was completed by respondents at the national level** who were asked to provide responses about the education system of the whole country with specific regard to the impacts of COVID-19. They were asked how the period of disrupted schooling could be characterised, how responsibility for the pandemic response was distributed in the school sector, and what plans and policies had been implemented to respond to the COVID-19 disruption.

### *The historical instruments*

Historical assessment data were used for each of the six participating MILO countries to compare the performance of the MILO target population against equivalent populations before the outbreak of the pandemic. Only a subset of historical items was readministered as part of MILO in order to minimise the testing time required for students. The historical assessments used for comparison in MILO were:

- Programme for Analysis of Educational Systems (PASEC) 2019 (CONFEMEN, 2020) (used for Burkina Faso, Burundi, Côte d'Ivoire, Senegal)
- National Assessment Survey (NAS) Grade 5, 2016 (used for Zambia)
- National Assessment System for Monitoring Learner Achievement (NASMLA) Grade 7, 2019 (Karogo et al., 2020) (used for Kenya - only a link to mathematics is available; the 2019 assessment of English in Kenya did not contain a sufficient number of reading comprehension items to align with the reading constructs within the GPF.)

### *Sampling approach*

The MILO Sampling Framework sets out the standards of participation in terms of sampling, and these are aimed at maximising the comparability of survey outcomes across countries before and after the onset of the pandemic. The target grade in the MILO project was the grade closest to the end of primary schooling within each country for which historical assessment data were available to use as the pre-pandemic baseline. All students enrolled in the target grade in each participating country were included in the target population. This included students from schools across all educational sub-systems and types within a country where the language of instruction was English or French. Some school and student-level exclusions applied, consistent with other large-scale surveys.

### *The MILO countries and samples*

The language of administration as well as the sampling characteristics of all six countries are shown in Table 1.1. The school participation rate in all countries was extremely high (with replacements where necessary). All countries were able to achieve a very high level of student response rate (the proportion of students who were sampled to participate who actually participated). The grade level of the students assessed in 2021 in each of these countries was selected to replicate the grade levels assessed in the historical assessment and enable comparisons between the populations.

One of the overarching goals of the MILO project is to identify whether learning loss took place from the time of the historical assessments (held in 2016 for Zambia and 2019 for other countries) to the time of the MILO data collection in 2021. Therefore, it is important to compare characteristics of the two populations. This allows any differences observed in achievement over time to be taken into context (given the established relationship between student and home background characteristics and achievement, see Chapter 7 for further details).

**TABLE 1.1** Numbers of students and schools participating in MILO with participation rates

| Country | Language of administration | Grade assessed | Participating schools (no.) | School response rate (%)* | Participating students (no.) | Student response rate (%)* |
|---|---|---|---|---|---|---|
| Burkina Faso | French | 6 | 289 | 100 | 5684 | 84 |
| Burundi | French | 6 | 252 | 100 | 4993 | 95 |
| Côte d'Ivoire | French | 6 | 250 | 100 | 4867 | 96 |
| Kenya | English | 7 | 265 | 100 | 6417 | 98 |
| Senegal | French | 6 | 247 | 99 | 4675 | 98 |
| Zambia | English | 5 | 252 | 99 | 4954 | 93 |

* Unweighted response rate including substitutes

Table 1.2 presents comparative data on wealth of students, gender, age, maternal and paternal literacy and school type for populations of students from the historical assessments and the AMPL.[3] In many ways, the characteristics of populations are similar across the two points in time, although some differences can be observed. For instance, students in the 2021 population were comparatively wealthier in Kenya and Senegal, but comparatively less wealthy in Burkina Faso in comparison to the historical population. The literacy rate of parents in Burundi and Côte d'Ivoire, and the proportions of students who attended public schools in Burkina Faso were also comparatively higher for the MILO population.

**TABLE 1.2** Student and home background characteristics of historical assessment and AMPL

| | AMPL-NATIONAL ASSESSMENT WEALTH INDEX (LOGITS) | | | GENDER (% GIRLS) | | | AGE (YRS) | | |
|---|---|---|---|---|---|---|---|---|---|
| **Country** | **AMPL** | **Historical** | **Difference (AMPL-Historical)** | **AMPL** | **Historical** | **Difference (AMPL-Historical)** | **AMPL** | **Historical** | **Difference (AMPL-Historical)** |
| Burkina Faso | -0.79 | -0.36 | -0.42 | 56 | 53 | 3 | 13.1 | 13.5 | -0.4 |
| Burundi | -1.73 | -1.82 | 0.09 | 56 | 56 | 0 | 14.4 | 14.7 | -0.3 |
| Côte d'Ivoire | -0.30 | -0.41 | 0.11 | 48 | 46 | 2 | 12.0 | 12.1 | -0.1 |
| Kenya | 0.58 | 0.19 | 0.39 | 51 | 55 | -4 | 12.6 | 12.5 | 0.1 |
| Senegal | 0.28 | -0.20 | 0.48 | 54 | 46 | 8 | 14.8 | 14.5 | 0.3 |
| Zambia | N/A | N/A | N/A | 51 | N/A | - | 12.3 | N/A | - |
| | **MATERNAL LITERACY (%)** | | | **PATERNAL LITERACY (%)** | | | **SCHOOL TYPE (% PUBLIC)** | | |
| **Country** | **AMPL** | **Historical** | **Difference (AMPL-Historical)** | **AMPL** | **Historical** | **Difference (AMPL-Historical)** | **AMPL** | **Historical** | **Difference (AMPL-Historical)** |
| Burkina Faso | 46 | 50 | -4 | 57 | 66 | -9 | 87 | 68 | 0 |
| Burundi | 78 | 67 | 11 | 83 | 73 | 10 | 98 | 94 | 0 |
| Côte d'Ivoire | 49 | 37 | 12 | 70 | 56 | 14 | 81 | 83 | 0 |
| Kenya | 63 | N/A | - | 79 | N/A | - | 77 | 82 | 0 |
| Senegal | 56 | 60 | -4 | 78 | 77 | 1 | 87 | 91 | -4 |
| Zambia | 79 | N/A | - | 86 | N/A | - | 83 | N/A | - |

## OUTLINE OF THE REPORT

This report explores the findings from the MILO project with a focus on comparing learning outcomes over time, as well as the contexts for teaching and learning during the COVID-19 pandemic.

Chapters 2 and 3 explore the AMPL-b reading and mathematics assessments, with respect to the framework, construction and contents. Chapter 4 provides details on the proportions of students across countries who met the SDG 4.1.1b MPLs in reading and mathematics.

Chapters 5, 6 and 7 explore the second overarching goal of the MILO project by investigating the impact of different distance learning mechanisms to remediate the pandemic learning disruption. Contexts at the system, school and student level help explain how learning loss was not recorded and these explanations are discussed in Chapter 8.

Chapter 5 describes the national contexts of each MILO country. It describes how each participating country was impacted by the COVID-19 disruption, details the aspects of national policies of the educational systems in response to the pandemic, discusses national communication and outreach during the pandemic and summarises assessment and data collection used by each country. Chapter 5 draws primarily on information from the MILO System Questionnaire.

Chapter 6 focuses on the school and classroom contexts of the MILO countries during the pandemic. It presents information on how schools in each country were impacted during the period of COVID-19 disruption, provides details on school infrastructure and teaching and learning resources available during the disruption, and discusses assessment and monitoring during and after the period of disruption. These findings draw on data from the MILO School Questionnaire.

Chapter 7 focuses on the contexts for students in the MILO countries. It explores student performance in the AMPL reading and mathematics by subgroups of students based on home background characteristics and those students who may be considered vulnerable. The chapter will also present various supports available to students during periods of COVID-19 disruption to their studies and detail various impacts of the disruption on students. The findings in the chapter will largely draw on data from the MILO Student Questionnaire.

Chapter 8 provides a summary of the main findings about the national contexts, the school and classroom contexts, and the student contexts and how they address the four overarching goals of the project. This chapter discusses possible reasons why no learning loss was observed as a whole across the MILO countries. The findings are presented in the context of other similar research, where mixed findings about the impact of the pandemic on learning outcomes were found. Possible implications for policy and practice of the findings of the study are presented, and recommendations are made for building more resilient education systems as well as future considerations for evaluating the impact of the pandemic on educational systems.

Appendix A provides further details about the standard setting exercise to determine the level of performance that corresponds to students meeting the MPLs at the end of Primary School, as referred to in SDG 4.1.1b.

Appendix B includes technical descriptions of data analyses used to link the MILO data with past historical assessment results.

Appendix C provides additional supplementary tables including information about the GPF reading and mathematics domains, constructs and sub-constructs and tables containing the standard errors for the proportion of students meeting the MPLs for reading and mathematics and information about contributors to the MILO project.

# 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).