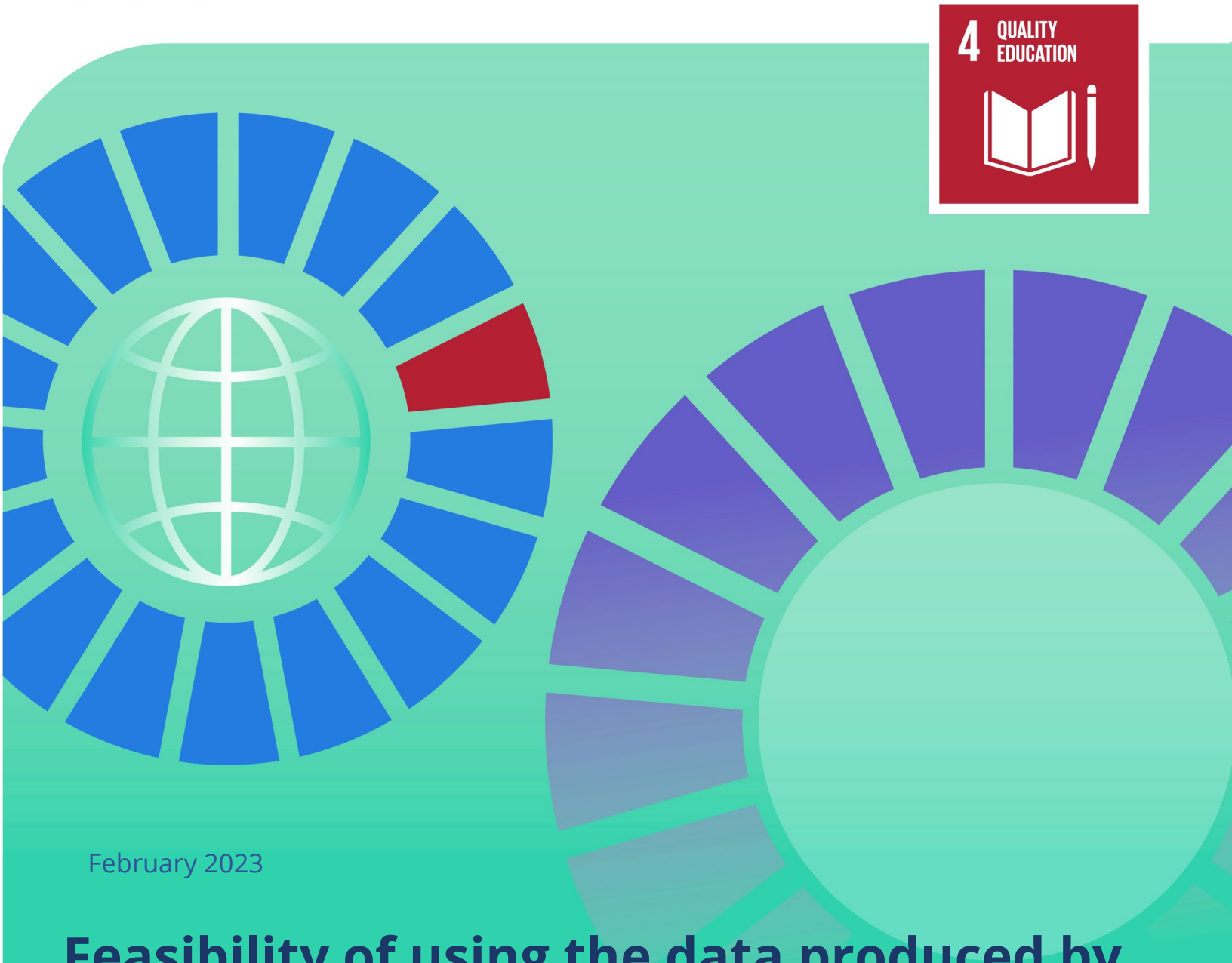
unesco
Institute for Statistics



February 2023

Feasibility of using the data produced by the Early Grade Reading (EGRA) and Early Grade Mathematics (EGMA) to measure and monitor SDG 4.1.1 by complementing it with other banks of items

Feasibility of using the data produced by the Early Grade Reading (EGRA) and Early Grade Mathematics (EGMA) to measure and monitor SDG 4.1.1 by complementing it with other banks of items

Paper prepared for the UNESCO Institute for Statistics by Artemio Arturo Cortez Ochoa (University of Bristol) and Andrés Sandoval Hernandez (University of Bach),

Introduction

This document evaluates the feasibility of using data produced by the early grade reading assessment (EGRA) and early grade mathematics assessment (EGMA) to measure and monitor SDG 4.1.1 by complementing it with other banks of items, such as the Foundational Learning Skills Module of UNICEF's Multiple Indicators Cluster Survey (MICS). The SDG indicator 4.1.1 is concerned with 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" (UN, n.d.). Minimum proficiency level (MPL) definitions for benchmarking knowledge of mathematics and reading, as measured via assessments, were agreed upon in September 2018, see Table 1:

Table 1. Minimum proficiency levels defined by each learning assessment

Reading	
Educational Level	**Descriptor**
Grade 2	They read and comprehend most of written words, particularly familiar ones, and extract explicit information from sentences.
Grade 3	Students read aloud written words accurately and fluently. They understand the overall meaning of sentences and short texts. Students identify the texts' topic.
Grades 4 & 6	Students interpret and give some explanations about the main and secondary ideas in different types of texts. They establish connections between main ideas on a text and their personal experiences as well as general knowledge.
Grades 8 & 9	Students establish connections between main ideas on different text types and the author's intentions. They reflect and draw conclusions based on the text.

Mathematics	
Educational Level	**Descriptor**
Grades 2-3	Students demonstrate skills in number sense and computation, shape recognition and spatial orientation.
Grades 4-6	Students demonstrate skills in number sense and computation, basic measurement, reading, interpreting, and constructing graphs, spatial orientation, and number patterns.
Grades 8 & 9	Students demonstrate skills in computation, application problems, matching tables and graphs, and making use of algebraic representations.

Source: (UN, 2022)

EGRA was developed in 2006 by the Research Triangle Institute, also known as RTI International (Platas et al., 2014). It has received support from the World Bank and the United States Agency for International Development (USAID), among other donors. The test seeks to measure early grades of primary students' progress in learning to read, and it is administered orally one-to-one for about 15 minutes per child. In contrast to paper-based assessments, which require prior reading achievement, EGRA's oral assessments are expected to uncover what the student already knows in relation to reading, which is considered critical for learning other content (RTI International, 2016). More than 30 organisations in 70 countries, mainly low-income contexts, have employed EGRA in 120 languages (Gove & Wetterberg, 2011; RTI International, 2016). It is important to note that "EGRA does not measure literacy behaviours, background knowledge or attitudes about reading" (Dubeck & Gove, 2015, p. 316). For example, the test does not measure other aspects related to reading skills, such as "motivation, attention, memory, reading strategies, productive vocabulary, comprehension of multiple text genres, retell fluency, etc." (RTI International, 2016, p. 40), which would lead to a lengthy assessment, potentially impacting the students' performance negatively.

EGMA was also developed by RTI International in 2008 and has been endorsed by USAID. This early-grade primary mathematics oral test focuses on numbers and operations (Platas et al., 2014). The number of countries where EGMA has been applied is smaller, with 14 nations partaking, including "the Democratic Republic of Congo, Dominican Republic, Ghana, Iraq, Jordan, Kenya, Liberia, Malawi, Mali, Morocco, Nicaragua, Nigeria, Rwanda, and Zambia" (Platas et al., 2014, p. 1). Unlike EGRA, the mathematics assessment is not meant for cross-country comparison. The main reason is that EGMA aims to evaluate students' performance against the local curriculum (Australian Council for Educational Research, 2020; Platas et al., 2014). However, programme evaluation and informing teacher training initiatives are not discarded (Platas et al., 2014).

In order to evaluate whether EGRA and EGMA assessments, alongside other banks of items, can inform SDG 4.1.1 advancement at a global scale, this report analyses the aforementioned assessments considering their conceptual frameworks and the interpretations associated with scores arising from their implementation. After this, the Multiple Indicator Cluster Survey (MICS) is examined, and a comparison between the UN's MPLs and these assessments is presented.

This report concludes with an analysis of EGRA, EGMA and MICS regarding the feasibility of complementing each other to measure and monitor SDG 4.1.1.

Conceptual frameworks

Early Grade Reading Assessment

A tenet of EGRA is that “Children first need to “learn to read” so that they can “read to learn.” (RTI International, 2016, p. 2). In that sense, this assessment focuses on low-order skills in relation to reading, including phonological awareness, letter sounds, and decoding (RTI International, 2016). Its creators maintain that “Children who do not learn to read in the early grades risk falling further and further behind in later ones, as they cannot absorb printed information, follow written instructions, or communicate well in writing” (Gove & Wetterberg, 2011, p. 1). Furthermore, the RTI considers reading skills critical for a nation’s economic development.

An important aspect to bear in mind about EGRA is that it has been intentionally devised to predict students’ reading skills acquisition by gaining insight into their learning processes of symbols, sounds and reading more broadly (RTI International, 2016). This aspect can be valuable in anticipating future performance, particularly as the assessments are targeted at the early years of primary education. Moreover, their supporters maintain that over several rounds of data collection, findings suggest that EGRA can indicate the direction of reading skills development (RTI International, 2016).

Concerning the above, the second edition of EGRA’s toolkit recommends that when benchmarks of minimum proficiency are decided, these must be sufficiently justified and reviewed in line with school or district expectations of reading development and contextual limitations and enablers (RTI International, 2016). This point is essential concerning cross-context comparisons because assessment data should be able to be contrasted as much like-with-like as possible. This is a debate Dubeck and Gove (2015) have been grappling with following the publication of a Brookings Institute’s Centre for Universal Education-led Learning Metrics Task Force report (LMTF, 2013). The report puts forward a series of criteria that would discard EGRA as suitable for international comparisons, including global indicators of educational achievements, such as SDG 4.1.1. It is

worth reminding that the matter of concern of EGRA is language, which necessitates adaptations and considerations on a case-by-case basis, according to context.

EGRA was conceived to be a large-scale standardised assessment of children's reading skills, aspiring to support the development of reading comprehension and measure it (RTI International, 2016). More generally, the test measures accuracy, fluency and comprehension. The six primary components of EGRA shown in Table 2 will be revisited later. Given the framework's flexibility, additional subtasks can be employed to gain information on the students' early reading skills; these optional components can be consulted in Dubeck and Gove (2015). The application of EGRA generates quantitative data for deciding on minimum levels of reading competency, and the results can be communicated using descriptive and inferential statistics.

Table 2. EGRA's components, early reading skills and indicators

Component	Early reading skill	Skill demonstrated by students' ability to:
1. Listening comprehension	Listening comprehension; oral language	Respond correctly to different types of questions, including literal and inferential questions about the text the assessor reads to them
2. Letter identification: Letter names and/or letter sounds	Alphabet knowledge	Provide the name and/or sound of letters presented in both upper case and lower case in a random order
3. Nonword reading	Decoding	Make letter–sound (grapheme–phoneme correspondences, or GPCs) through the reading of simple nonsense words
4. Oral reading fluency with comprehension	Oral reading fluency Reading comprehension	Read a text with accuracy, with little effort, and at a sufficient rate Respond correctly to different types of questions, including literal and inferential questions about the text they have read
5. Initial or final sound identification, or letter sound discrimination, or phoneme segmentation, identification of onset/rime sounds	Phonological awareness	Identify/differentiate the onset/rime sounds of words or the initial or final sounds of words, or segment words into phonemes by having the assessor and then the student read the phonemes aloud
6. Familiar word reading	Word recognition	Read words which are randomly ordered and drawn from a list of frequent words

Source: (RTI International, 2016, p. 41)

Early Grade Mathematics Assessment

EGMA considers that children develop mathematical thinking before entering school, and they do so in their playing and other context-specific situations (Platas et al., 2014). Pedagogically speaking, it is assumed that children learn about mathematics vocabulary with fellow children and adults around; in other words, that understanding is socially constructed (Platas et al., 2014). Similarly to EGRA, the mathematics version is meant to predict later achievement, and the role of knowledge of mathematics is also recognised as key in the future economic life of individuals (Platas et al., 2014). The lack of knowledge regarding reading and writing is considered not to limit early-year primary students' demonstration of their mathematics knowledge and competencies. The oral application of EGMA seeks to avoid the confounding influence of low or null knowledge of the written language (Platas et al., 2014).

In an earlier version of the EGMA's conceptual framework, the test was concerned with mathematics learning "with an emphasis on numbers and operations and on geometry through second grade or, in developing countries, perhaps through third grade" (Reubens, 2009, p. 1). Later, the test focused on knowledge and six competencies and subtests, as shown in Table 3:

Table 3. EGMA Subdomains and Subtests

Subdomains	Subtests	Description
Number identification	Number identification	It consists of 20 items that increase in difficulty. It includes three single-digit numbers, 12 two-digit numbers and five three-digit numbers. Students are asked to say each number aloud (timed for 1 minute).
Number discrimination	Number discrimination	It consists of ten items. Each item consists of a set of two numbers, one of which is greater than the other. The subtest includes one set of one-digit numbers, five sets of two-digit numbers, and four sets of three-digit numbers. Students are asked to state the higher of each set of two numbers (not timed).
Number pattern identification	Missing number	It consists of ten items. The items are presented as four horizontally aligned boxes, three of which contain numbers and one of which is empty (the target missing number). Eight of the items increase in number from left to right; two of the items decrease in number from left to right. Items 1, 2, and 6 increase by one (in a set of one-, two-, and three-digit numbers, respectively). Items 3, 4, 5, and 8 increase by tens, hundreds, twos, and fives, respectively. Items 7 and 9 decrease by two and tens, respectively. The last item with numerals within the range of 1–20 increases by fives, but does not begin with a multiple of five. Students are asked to state the number that belongs in the empty box (not timed).
Addition and subtraction	Addition Level 1	It consists of 20 items that increase in difficulty. No addends are greater than 10, and no sums are greater than 19 (timed for 1 minute).
	Addition Level 2	It consists of five items that increase in difficulty. No sums are greater than 70. Addition Level 2 is not given to students who receive a score of zero for Addition Level 1 (not timed).
	Subtraction Level 1	It consists of 20 items that increase in difficulty. The Subtraction Level 1 problems are the inverse of the Addition Level 1 problems (timed for 1 minute).
	Subtraction Level 2	It consists of five items that increase in difficulty. Subtraction Level 2 is not given to students who receive a score of zero for Subtraction Level 1. The Subtraction Level 2 problems are the inverse of the Addition Level 2 problems (not timed).
	Word problems	It consists of six items that increase in difficulty. Three of these items use numbers that match three items from the Addition and Subtraction Level 1 subtest. Assessors also keep track of whether the student used one of three problem-solving strategies: finger/tick marks, paper and pencil calculation, or solved problems in his or her head (not timed).

Source: (Australian Council for Educational Research, 2020, p. 3)

Intended interpretations

Early Grade Reading Assessment

As mentioned earlier, EGRA has been used in several contexts. Furthermore, studies using EGRA have followed different research designs, and the aims may also vary. One example is a multi-year mixed-methods study conducted with more than five thousand students from first,

second and third grades in Nicaragua (Víctor et al., 2021). In this case, the researchers aimed to design a pedagogical intervention based on the reading test results and inform teachers' professional development. Remarkably, average results from the Nicaraguan students regarding the number of words per minute children in this age group should be able to read differed from international benchmarks. Generally, this study's participants read between 10 and 20 words less than global indicators. The opposite was found regarding comparative indicators of reading fluency, with Nicaraguan students typically performing like or above international standards, leading the researchers to adapt such thresholds for local interpretation and analysis (Víctor et al., 2021).

Another example by Stern et al., (2018) explored language acquisition in Bahasa Indonesia, drawing on a representative sample (n=4812) of grade 2 students. In this case, eight sub-tasks of EGRA were employed in the research. The purpose of the research was to determine "five learning profiles: 1) Grade 3 Ready; 2) Fluent; 3) Instructional; 4) Beginner; and 5) Nonreader" (Stern et al., 2018, p. 67). Researchers acknowledge the role of orthography in learning to read, mainly as Bahasa Indonesia is a phonetic language. In other words, it is read as it is written. This situation might lead to different interpretations of learning profiles in languages that do not share this characteristic. Nevertheless, the study was meant to contribute a framework to identify the skills that might help teachers and students support progress across the learning profiles.

The interpretations regarding reading and literacy of EGRA have been subject to academic debate, including the seemingly fragmented approach to language acquisition, which is in stark contrast with views of reading development as integral and interdependent among its various components (Bartlett et al., 2015). It is argued that the relationship between the umbrella areas of EGRA, i.e., fluency, accuracy and comprehension, needs to be well-established, and because results vary across contexts, the test has limitations as a global benchmark and for comparative purposes (Bartlett et al., 2015).

Early Grade Mathematics Assessment

The creators of EGMA explicitly state the situations in which the tests "should ***not*** be used: (1) for cross-country comparisons, (2) for high-stakes testing, (3) as input for student report cards, and (4) for the simultaneous program and country-level diagnostics" (Platas et al., 2014, p. 4 emphasis and bold from original).

Given that EGMA is concerned with assessing mathematical knowledge in line with the curriculum, it might be less pertinent for international comparisons (Australian Council for Educational Research, 2020; Platas et al., 2014).

The Multiple Indicator Cluster Survey

Also known as MICS, this is an international programme supported by UNICEF, aiming at monitoring households' circumstances, such as nutrition, vaccination status, access to running water, sexual health, and education, among others, mainly regarding children and women, via cross-sectional questionnaires (UNICEF-MICS, 2023). The survey has been rolled out since the 1990s focusing on various aspects, including the Millenium Development Goals and the Sustainable Development Goals in its last rounds (Khan & Hancioglu, 2019). Although participants are not the same every time, trends can be studied based on MICS data. Sampling is meant to establish a statistical representation of the population from where data is drawn.

MICS has collected data from 118 countries (UNICEF-MICS, 2023) and is meant to provide insight into local and global matters, which is possible given that the indicators are agreed upon among participant countries and regions (Khan & Hancioglu, 2019). Data is freely accessible for research purposes in different fields.

MICS includes a Foundational Learning Skills (FLS) module in which "children aged 7–14 years also participate in a short literacy and numeracy assessment" (Khan & Hancioglu, 2019, p. 280), and background information is collected. The test takes 15 minutes and has been applied to early-grade children 2/3 in MICS participating countries since 2017 (Cardoso, 2020). Table 4 presents the areas the FLS module assesses.

Table 4. Foundational Learning Skills tasks and subtests.

	Task/Subtest	Number of items
Reading	Oral Reading Accuracy	43
	Reading Comprehension	5
Numeracy	Number reading	6
	Number discrimination	5
	Addition	5
	Pattern recognition and completion (missing number)	5

Source: adapted from Cardoso (2020).

A study conducted in 2016 in Kenya (Gochyyev et al., 2019) indicates that the reading and numeracy instruments in the MICS-FLS module present appropriate reliability and inter-rater reliability estimates. These aspects are related to the capacity of the tests to obtain consistent results when applied for the same purpose repeatedly and by different test administrators (Fink, 2010). The MICS-FLS module is meant to provide data for SDG 4.1.1 across languages and contexts of application (Gochyyev et al., 2019). There needs to be more literature supporting the cross-context comparability of results from the application of MICS-FLS; this dearth of research may be because of the recency of the module; still, longitudinal studies could provide further support on this matter (Bartlett et al., 2015).

MICS-FLS has also been studied regarding its concurrent validity with EGRA and EGMA, concluding that the MICS-FLS module is valid as its results correlate well with the RTI assessments. However, critical differences should be considered; for instance, while reading in EGRA includes timed tasks, the reading section of MICS-FLS is not time restricted. Furthermore, mathematics in MICS-FLS focuses on numbers and currently does not include items on geometry, measurement, data and algebra (Cardoso, 2020). These limitations have led to the conclusion that identical cut-offs cannot be used for FLM results due to: (1) the lack of a subtraction subtest; and (2) the fact that the addition subtest has only one question compared

to five level 2 addition questions in EGMA. Therefore, cut-offs proposed for EGMA have been revised, and additional validity analysis of these tasks rely on visual scatterplots to evaluate the similarity of scores in these numeracy tasks (Gochyyev et al., 2019, p. 7).

Table 5. A comparison between UN's Minimum Proficiency Level (MPL) for reading and mathematics, and the assessments' components better aligned with it.

	Educational level	UN – Minimum Proficiency Level	EGRA – Early Reading Skills	EGMA – Early Numeracy Skills	MICS-FLS module
Reading	Grade 2	They read and comprehend most of written words, particularly familiar ones, and extract explicit information from sentences.	Decoding Oral reading fluency Reading comprehension	Not applicable	Oral reading accuracy Reading Comprehension
	Grade 3	Students read aloud written words accurately and fluently. They understand the overall meaning of sentences and short texts. Students identify the texts' topics.	Word recognition	Not applicable	
	Grades 4 & 6	Students interpret and give some explanations about the main and secondary ideas in different types of texts. They establish connections between main ideas on a text and their personal experiences as well as general knowledge.	Not applicable	Not applicable	Not yet
	Grades 8 & 9	Students establish connections between main ideas on different text types and the author's intentions. They reflect and draw conclusions based on the text.	Not applicable	Not applicable	Not yet
Mathematics	Grades 2 - 3	Students demonstrate skills in number sense and computation, shape recognition and spatial orientation.	Not applicable	Number identification, discrimination, pattern, addition and subtraction	Number reading Number discrimination Addition Pattern recognition
	Grades 4 - 6	Students demonstrate skills in number sense and computation, basic measurement, reading, interpreting, and constructing graphs, spatial orientation, and number patterns.	Not applicable	Not applicable	Not yet
	Grades 8 & 9	Students demonstrate skills in computation, application problems, matching tables and graphs, and making use of algebraic representations.	Not applicable	Not applicable	Not yet

Source: original, based on (Australian Council for Educational Research, 2020; RTI International, 2016; UN, 2022)

Table 5 shows four areas of EGRA which conceptually align better with the UN's MPL in Grades 2/3. These are decoding, oral reading fluency, reading comprehension, and word recognition. Other reading assessment components appear less helpful in evaluating and monitoring SDG 4.1.1a and/or are definitely unrelated to SDG 4.1.1b and 4.1.1c.

Decoding is a low-order skill in the acquisition of written language that, in EGRA, aligns with the UN's MPL concerning reading, particularly aloud. However, in EGRA assessments, this section relates to nonsense words, which contradicts reading for comprehension, as in MPL. This has been criticised in the literature because of the primacy given to grapheme-phoneme correspondence over the actual meaning of words (Bartlett et al., 2015). Yet, EGRA claims to explore familiar **word reading**, in line with MPL, and factual and inferential information in their **reading comprehension** assessment. EGRA's **oral reading fluency** relies heavily on the number of words a child can read per minute. While MPL does not explicitly state it, accuracy and fluency are also part of the UN's approach to reading proficiency. Still, it can be considered that understanding the overall meaning of sentences and short texts suggests that the MPL is more concerned with reading skills beyond decoding and recognising sounds and letters.

The subdomains of EGMA that could inform the evaluation and monitoring of SDG 4.1.1a are **number identification, discrimination, pattern, addition and subtraction** (Table 5). However, EGMA is also limited beyond grades 2/3, meaning that SDG 4.1.1b and 4.1.1c cannot be assessed using this mathematics assessment.

EGMA's approach to mathematics learning is heavily oriented on numbers, unlike previous versions mentioning geometry explicitly (Reubens, 2009). This is important because, for the UN, children in 2 and 3 grades should demonstrate shape recognition and spatial orientation. In addition, it is difficult to determine whether EGMA's **number identification** and **discrimination,** including **addition** and **subtraction,** may inform UN's MPL regarding number sense and computation. Furthermore, while EGMA assesses **number patterns**, the UN's MPL mentions this from Grades 4 -6. Yet, EGMA was not designed for such grades. Finally, as mentioned earlier, EGMA is concerned with assessing mathematics learning according to the curriculum, which make cross-national comparisons difficult. In fact, cross-national comparisons are explicitly discouraged by RTI (Platas et al., 2014).

Table 5 also presents MICS-FLS's components that align with the UN's MPLs. **Oral reading accuracy** and **reading comprehension** are directly relevant to MPL for Grades 2/3 as these indicators look at the reading of words in a story, accuracy, and comprehension in the literal and inferential sense (Cardoso et al., 2020). Cardoso et al., (2020) indicate that although the goal is that the reading MICS-FLS measures skills regardless of language, it remains a challenge to decide what to do in places where there is more than one language of instruction to determine what is feasible while maintaining statistical representativeness. The validity of the reading component of MICS-FLS has been established based on concurrent validity studies with EGRA (Cardoso, 2020). In this regard, it is critical to reflect on whether MICS-FLS also inherits similar limitations in serving as a global indicator of achievement of SDG 4.1.1 (Bartlett et al., 2015). It can be argued that given the specificities of language acquisition and development, exact comparisons between different education systems and contexts will prove challenging and misinterpretations regarding reading skills achievement might arise. In the context of the Programme for International Student Assessment (PISA), researchers have been dealing with whether language plays a role in comparative results, and the findings have shown that it does (Asil & Brown, 2015; Pamei et al., 2022; Soyler et al., 2021).

As shown in Table 5, MICS-FLS emphasises **number reading, number discrimination, addition,** and **pattern recognition.** The heavy reliance on numbers partially addresses the UN's MPL in mathematics, leaving other areas untested in Grades 2/3, including geometry (Cardoso, 2020). Similarly to EGRA, the numeracy component of MICS-FLS has been validated by comparing it with EGMA, finding that UNICEF's test works in line with its purpose (Cardoso, 2020; Cardoso et al., 2020; Gochyyev et al., 2019). As examined previously, EGMA is meant to assess mathematical skills in line with the curriculum (Platas et al., 2014). This aspect should be considered in pondering the relevance and feasibility of using MICS-FLS for comparative numeracy achievement. In addition, as different approaches to mathematics teaching and learning might undergo in different contexts, including problem-solving and other traditions, such as the Contextual Teaching and Learning approach (Intaros et al., 2014; Selvianiresa & Prabawanto, 2017), evaluators using the MICS-FLS should consider whether the pedagogical underpinnings of the instrument are relevant to the context of application and what sort of misinterpretations might arise.

The evidence suggests that EGRA, EGMA and the MICS-FLS module have limitations in their capacity to serve as a global indicator for SDG 4.1.1. Combining their different elements would

be difficult as it would require a central quality assurance procedure that guarantees equivalency in language and learning approaches and conceptualisations in reading and mathematics across contexts.

Final thoughts

Based on the arguments presented here, the following considerations regarding the feasibility of using data from EGRA and EGMA in conjunction with other banks of items, such as MICS-FLS, to measure and monitor SDG 4.1.1 are derived.

- EGRA and EGMA have been designed to assess early reading and mathematics skills related to SD 4.1.1.a, but not b) or c). Similarly, although MICS-FLS has been conceived for using it with 7–14-year-olds, it has been tested mainly with early-grade children 2/3. This situation makes these assessments unsuitable for monitoring grades 4 and above.
- EGRA might be limited in assessing language acquisition by omitting to acknowledge the interconnectedness of the various components that the test assesses separately. Also, given the specific characteristics of languages, empirical research using EGRA has found different results, again lessening the comparability of outputs (Bartlett et al., 2015).
- EGMA was explicitly developed to understand mathematics learning in line with local or national curricula. Therefore, their creators recommend against using it for cross-national comparisons (Platas et al., 2014).
- EGMA has been subjected to concurrent validity with MICS-FLS, and while EGMA does not aim to be a global indicator of mathematics learning by design, the numeracy component of MICS-FLS has such an aspiration; however, this intention omits to recognise the limitations it might bring with it, as EGMA creators might have noticed when they developed their test.
- Given the flexibility of EGRA and EGMA, each context may decide on different components to assess their students' early reading skills, which may lead to findings that are not directly comparable to inform achievement of SDG 4.1.1 (LMTF, 2013). In addition, the tests have been employed for different purposes and following various research methodologies.

- MICS-FLS research is still limited, and longitudinal studies demonstrating its capacity to produce consistent results regardless of the language of the application and local curriculum specificities are needed (Bartlett et al., 2015; LMTF, 2013).
- If it were decided to combine elements from EGRA and EGMA with those from the MICS-FLS module, the lack of centralised quality assurance mechanisms could pose important risks to the validity of international comparisons, mainly because of the individual limitations of the assessments reviewed in this paper (Stern et al., 2018; Víctor et al., 2021),

References

Asil, M., & Brown, G. T. L. (2015). Comparing OECD PISA Reading in English to Other Languages: Identifying Potential Sources of Non-Invariance. *Http://Dx.Doi.Org/10.1080/15305058.2015.1064431*, *16*(1), 71–93. https://doi.org/10.1080/15305058.2015.1064431

Australian Council for Educational Research. (2020). *The Early Grade Mathematics Assessment: Assessing children's acquisition of basic numeracy skills in developing countries*. http://www.acer.edu.

Bartlett, L., Dowd, A. J., & Jonason, C. (2015). Problematizing early grade reading: Should the post-2015 agenda treasure what is measured? *International Journal of Educational Development*, *40*, 308–314. https://doi.org/10.1016/J.IJEDUDEV.2014.10.002

Cardoso, M. (2020). *School-age learning assessment tools: Frequently Asked Questions on the Foundational Learning Skills module*. UNICEF-MICS. https://www.oecd.org/pisa/pisa-for-development/PISAD-Workshop-2020-Manuel-Cardoso.pdf

Cardoso, M., Quintana, E., & Mizunoya, S. (2020). School-age learning assessment tools: The Foundational Learning Skills module Children left behind in learning-measuring foundational learning skills in Multiple Indicator Cluster Surveys (MICS). In *PowerPoint presentation*. http://uis.unesco.org/apps/visualisations/laci/

Dubeck, M. M., & Gove, A. (2015). The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development*, *40*, 315–322. https://doi.org/10.1016/J.IJEDUDEV.2014.11.004

Fink, A. (2010). Survey research methods. *International Encyclopedia of Education*, 152–160. https://doi.org/10.1016/B978-0-08-044894-7.00296-7

Gochyyev, P., Mizunoya, S., & Cardoso, M. (2019). *Validity and Reliability of the MICS Foundational Learning Module*. MICS Methodological Papers, No. 9, Data and Analytics Section, Division of Data, Research and Policy,. https://mics.unicef.org/publications/methodological-papers#:~:text=MICS%20Methodological%20Paper%20No.%209%2C%202019%2C%20Validity%20and,instrument%20in%20the%20Multiple%20Indicator%20Cluster%20Surveys%20%28MICS%29.

Gove, A., & Wetterberg, A. (2011). The Early Grade Reading Assessment: An Introduction. In *The Early Grade Reading Assessment: Applications and Interventions to Improve Basic Literacy*. https://doi.org/10.3768/rtipress.2011.bk.0007.1109

Intaros, P., Inprasitha, M., & Srisawadi, N. (2014). Students' Problem Solving Strategies in Problem Solving-mathematics Classroom. *Procedia - Social and Behavioral Sciences*, *116*, 4119–4123. https://doi.org/10.1016/J.SBSPRO.2014.01.901

Khan, S., & Hancioglu, A. (2019). Multiple Indicator Cluster Surveys: Delivering Robust Data on Children and Women across the Globe. *Studies in Family Planning*, *50*(3), 279–286. https://doi.org/10.1111/sifp.12103

LMTF. (2013). *Toward universal learning: a global framework for measuring learning: executive summary*. https://unesdoc.unesco.org/ark:/48223/pf0000225210

Pamei, G., Cheah, Z. R. E., & McBride, C. (2022). Construct validity of international literacy measures: implications for dyslexia across cultures. *Journal of Cultural Cognitive Science*. https://doi.org/10.1007/S41809-022-00115-X

Platas, L., Ketterlin-Gellar, L., Brombacher, A., & Sitabkhan, Y. (2014). *Early Grade Mathematics Assessment (EGMA) Toolkit*. https://shared.rti.org/content/early-grade-mathematics-assessment-egma-toolkit#

Reubens, A. (2009). *Early Grade Mathematics Assessment (EGMA): A Conceptual Framework Based on Mathematics Skills Development in Children*.

RTI International. (2016). *Early Grade Reading Assessment (EGRA) Toolkit Second Edition*. United States Agency for International Development.

Selvianiresa, D., & Prabawanto, S. (2017). Contextual Teaching and Learning Approach of Mathematics in Primary Schools. *Journal of Physics: Conference Series*, *895*(1), 012171. https://doi.org/10.1088/1742-6596/895/1/012171

Soyler, P. B., Aydin, B., & Atilgan, H. (2021). PISA 2015 Reading Test Item Parameters Across Language Groups: A measurement Invariance Study with Binary Variables. *Journal of Measurement and Evaluation in Education and Psychology*, *12*(2), 112–128. https://doi.org/10.21031/EPOD.800697

Stern, J. M. B., Dubeck, M. M., & Dick, A. (2018). Using Early Grade Reading Assessment (EGRA) data for targeted instructional support: Learning profiles and instructional needs in Indonesia. *International Journal of Educational Development*, *61*, 64–71. https://doi.org/10.1016/J.IJEDUDEV.2017.12.003

UN. (n.d.). *Goal 4 - Sustainable Development Goals*. Retrieved February 4, 2023, from https://sdgs.un.org/goals/goal4

UN. (2022, March 31). *SDG indicator metadata*. Metadata. https://unstats.un.org/sdgs/metadata/files/Metadata-04-01-01.pdf

UNICEF-MICS. (2023). *COVID-19 - MICS Guidance - UNICEF MICS*. https://mics.unicef.org/covid-19/mics-guidance

Víctor, E., Urcuyo, A., & Zamora, N. (2021). Comparative analysis of the results of the EGRA tests applied in a network of educational centers in Fe y Alegría, Nicaragua (2016-2018). *Revista Innovaciones Educativas*, *23*(35), 18–34. https://doi.org/10.22458/IE.V23I35.3476