

Strengthening MICS FLS Module to report on SDG 4.1.1a (FLM2.0)

December 2023

João Pedro Azevedo

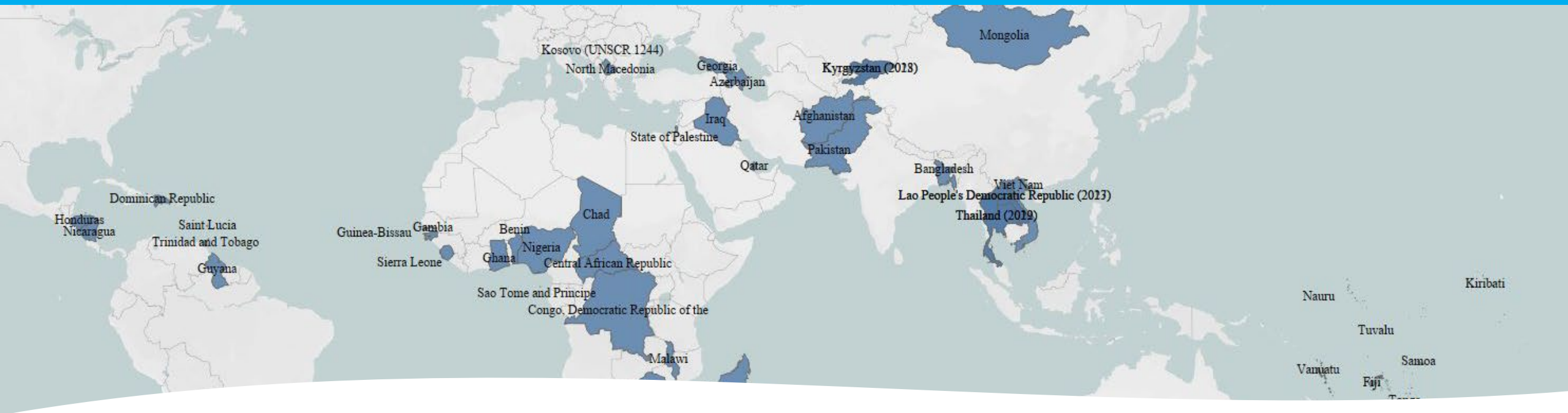
Eva Quintana

Sakshi Mishra



unicef 
for every child

1. Purpose



- To present to GAML, UNICEF's plan for improving the MICS Foundational Learning Skills (FLS) module to report on SDG4.1.1a
- To discuss with GAML participants, the proposed approach and outstanding questions
- To discuss with GAML participants approaches for using existing stock of assessments

2. Initial considerations

- SDG 4.1.1a assesses foundational learning skills which are building blocks for success in further education and life-long learning
- Preserving the indicator is critical to maintain policy momentum and drive investments and interventions to improve the quality of early-grade education
- Internationally comparable early-grade assessments are still comparatively nascent; and countries and the global community has significantly less experience in measuring learning at this level in a globally comparable way;
- the SDG 4 process led by UIS has created for the first time a global standard of minimum proficiency levels and we must build on that, while building on what exists following the precedent of several indicators in the space of SDG 1 and SDG 3.

2. Initial considerations (cont.)

- UNICEF is actively engaged in the support of countries implementing a range of learning assessments (i.e. SEA-PLM, PASEC, National Learning Assessments, and MICS-FLS) and our choice is always driven by country demand, fit for purpose, and analytical value to inform national policies.
- We are fully aligned with the importance of disrupting the assessment market, focusing on a minimum viable product and using all the information available.
- Household surveys can play an important role in filling existing data gaps and complement school assessment data:
 - Cost-effectiveness: the marginal cost of adding a module to a multi-topic household survey than implementing a standalone assessment
 - Linking education data to household factors: Ability to understand variance in learning by child, parental, household and school factors (MICS link is an example of how we can link household responses with EMIS data).
 - It is a flexible approach that can be instrumental in enabling the measurement of learning in humanitarian and emergency contexts.

3. Multiple Indicator Cluster Surveys (MICS): Overview

- Largest source of internationally comparable data on the situation of children, with nearly 360 surveys across 120 countries in the last 30 years
-
- Strong country ownership. Carried out by National Statistical Offices with UNICEF support
 - Typically, every 4-5 years, although a higher frequency is possible
 - MICS data:
 - Representative samples of households
 - Individual face-to-face interviews
 - Covers a wide range of topics (e.g. child survival, nutrition, early childhood development, education and learning, and child protection)
 - Generates nearly 200 indicators (including 40 SDG indicators)
 - MICS as a platform:
 - Scalability: Network of global experts
 - MICS Plus: phone-based follow-up surveys
 - Population-based mental health module statistically validated



4. Multiple Indicator Cluster Surveys (MICS): Sampling

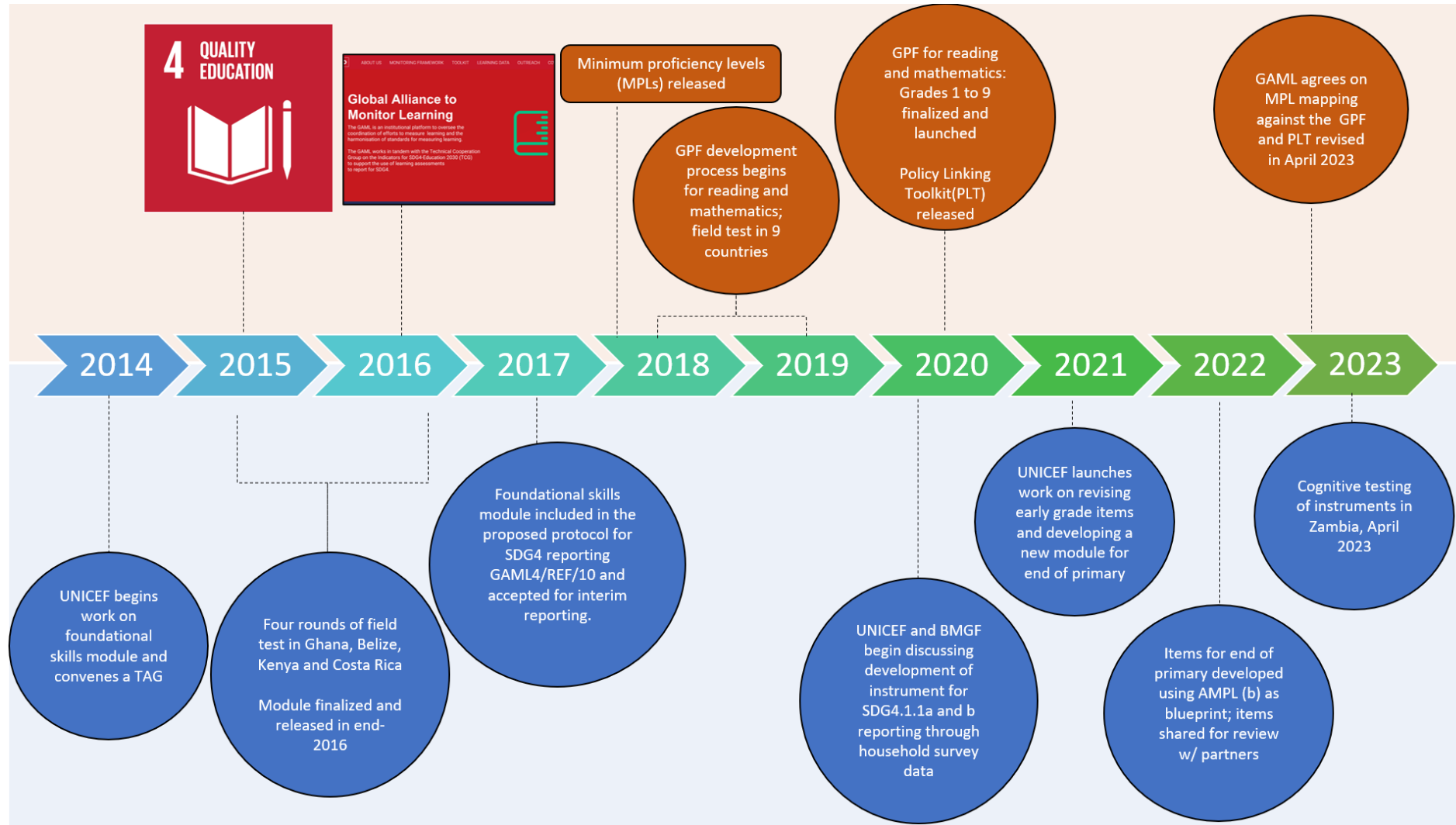
- MICS sample is based on probability sampling methods to generate representative and unbiased results.
- Typically, a two-stage design:
 - Enumeration Areas (EAs) selected from appropriate sampling frame (e.g., national census) using probability proportional to size.
 - Random systematic sampling used for household selection within listed EAs
- Every household and member has a known probability of selection
- Valid inferences to the population or subgroups are possible through weighted data
- Sampling errors can be estimated
- Oversampling of specific population subgroups or age subsets (e.g., under 5 children) for reliable indicators if needed

5. MICS Foundational Learning Skills (FLS) module

MICS FLS module features:

- Administered to one randomly selected child (age 7-14) per sampled household
- Requires informed consent from primary caregiver and assent from the child
- Individually administered by an interviewer specially trained

Implemented in more than 50 countries, including local languages, since 2017



6. MICS FLS Scoring protocol: Scoring and benchmarking criteria

Reading:

- Child has foundational reading skills if they:
 - Accurately read 90% of the words in a Primary Grade 2 level story (word count varies by language/country customization)
 - Correctly answer 3 simple literal comprehension questions based on the story
 - Correctly answer 2 simple inferential comprehension questions based on the story
- Prior to reading story, children aged 7 to 9 and out of school children are given practice items consisting of a short paragraph and simple questions

Numeracy:

- Child has foundational numeracy skills if they respond correctly to:
 - Reading numbers (6 questions)
 - Number discrimination (5 questions)
 - Addition (5 questions)
 - Number pattern recognition (5 questions)



7. MICS FLS module: Localization protocol



Rigorous protocol to customize the reading assessment to the local context:

- MoE specialists lead the translation and adaptation of the assessment using early-grade curriculum knowledge, and in collaboration with MICS and UNICEF experts.

Step 1: Assessment Language Selection - Follows MoE's early-grade medium of instruction policy; complex decisions in multi-language settings.

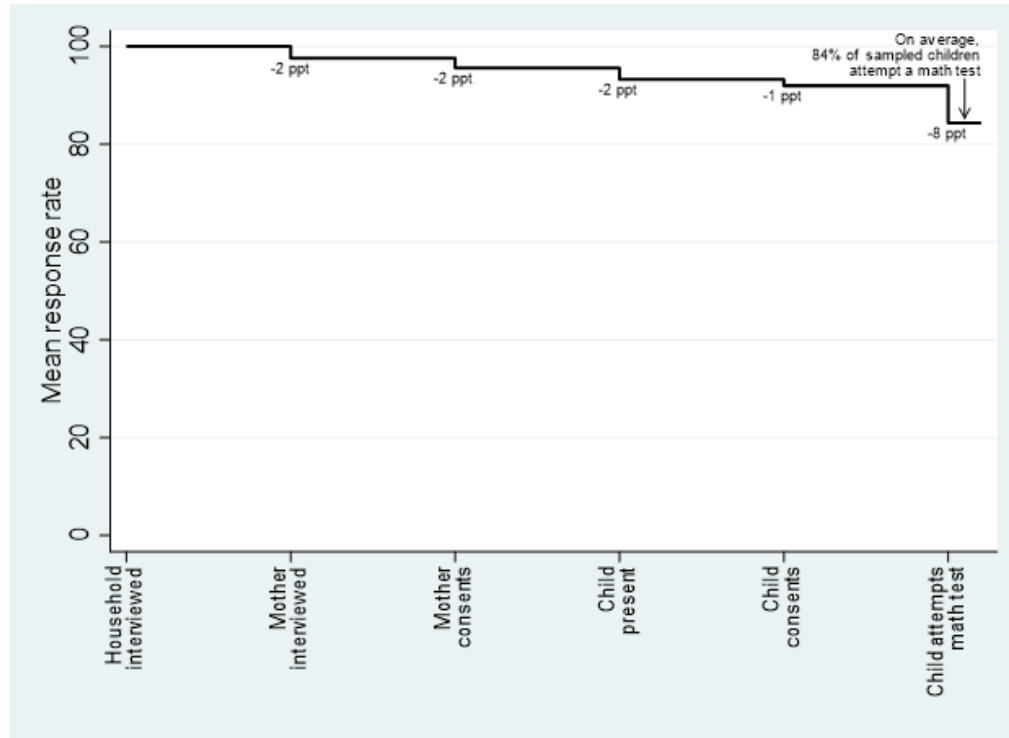
Step 2: Translation and Adaptation - Translated from standards available in 5 languages; adapted for vocabulary and cultural relevance; aligns with grade 2 textbooks.

Step 3: Text Analysis - Backtranslation and rigorous analysis to maintain tool quality and difficulty level.

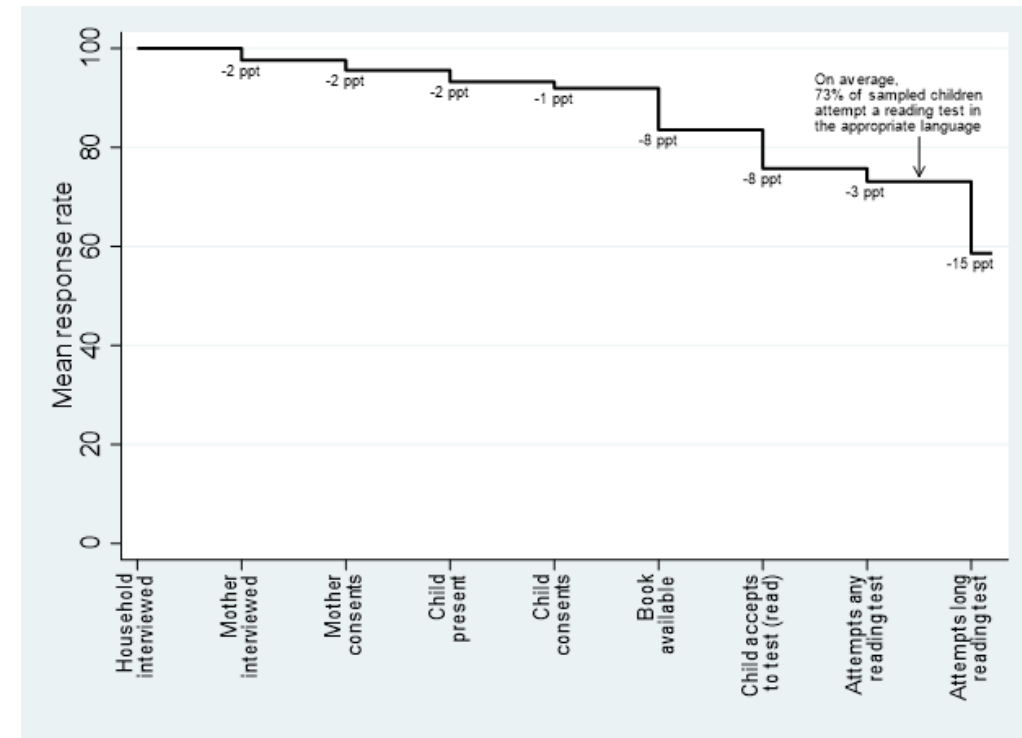
8. MICS FLS module: Consent and response rates

- Field procedures in place to minimize non-response, including rescheduling interviews and revisiting (3 attempts)
- Strong ethical protocols including primary caregiver's consent to interview child and child's assent to participate in survey
- Preliminary analysis across 40 surveys (2017-2022) with publicly available data shows:
 - On average 92% of sampled children are interviewed

Math



Reading



Note figures and findings are preliminary for illustration and discussion. Forthcoming MICS methodological paper will have the final results

9. Improvement and validation plan: qualitative and quantitative review and reliability

- FLM2.0: Content alignment -Instrument design to match MPL-GPF mapping
- Field test 1:
 - **Purpose:** Field testing the instrument under realistic conditions in Zambia focusing on cognitive testing of items in April-May 2023 in collaboration with ZamStats and Examinations Council of Zambia (ECZ)
 - **Sample:** About 25 students in grades 4 and 5 in urban and rural Lusaka clusters for literacy and numeracy each, purposive sampling for maximum variation
 - **Tools:** Two different sets of reading and math items tested
 - **Qualitative approaches:** Cognitive interviewing, retrospective think-aloud, and coding of interviewer-respondent interaction
 - **Observer:** ECZ observer paired with each ZamStat interviewer
 - **Training and Fieldwork:** Classroom-based training and practice for 5 days, followed by data collection in sample households for about three weeks
- Field test 2, planned in 2024 focusing on internal consistency (reliability) of the instrument as well as inter-rater reliability

10. Improvement and validation plan: Proposed concurrent validity studies/ Non-statistical and statistical linking

- **Concurrent Validity Studies:**
 - Quantitative testing to compare UNICEF's final instrument with similar validated tools (e.g., AMPL-a, SEA-PLM, EGRA, EGMA, iCARE)
 - Aims to gather evidence on how well the new instrument aligns with existing scales
- **Policy Linking or Pairwise Comparison:**
 - Evaluate how revised FLM2.0 instrument align with a common scale
- **Statistical Linking of FLS and FLM2.0:**
 - Establishing a link between revised early-grade items and MICS6 FLS module items

11. FLM2.0 against the 5 criteria: Criteria 1 and 2

Criterion	Description	Minimum requirement	FLM2.0	FLS
1	is the assessment sufficiently aligned to the GPF?	3 alignment: Minimal, Additional and strong	<p>Content is currently being revised to strongly align with Grade 2 competences, following June/2023 guidance.</p> <p>Mathematics: Number and operations; Measurement and Geometry; Statistics/Probability and Algebra Reading: Listening comprehension, Reading comprehension, Decoding</p>	<p>UNICEF FLS was launched before the GPF and MPL and was designed to be aligned with Grade 3. And therefore, it is weakly aligned</p> <p>Mathematics: Number and operations; Measurement and Geometry; Statistics/Probability and Algebra Reading: Accuracy/decoding and reading comprehension</p>
2	is there evidence that the items in the assessment have been reviewed qualitatively and quantitatively to determine their suitability for inclusion in the assessment?	<p>The qualitative review should consider whether:</p> <ul style="list-style-type: none"> • Each assessment item is considered appropriate by relevant experts for inclusion in the assessment • The scoring guides are consistent with what the item is intended to measure. <p>The quantitative review should consider whether:</p> <ul style="list-style-type: none"> • Item difficulty (e.g., item facility (CTT) or item location on the scale (IRT)) is appropriate for the grade level • Item discrimination (e.g., Discrimination Index for each item is generally greater than 0.2, with any exceptions rationalized or the distractors in a multiple-choice item should be negatively correlated with ability). 	<p>Qualitative review; cognitive and quantitative testing (ongoing) including:</p> <ul style="list-style-type: none"> • Face validity by respondents (through cognitive interviewers) and household survey experts/end users • Field-pretesting under realistic conditions (in Zambia and Field Test 2) • Future psychometric analysis/techniques (e.g., IRT to reduce item pool, reliability testing to evaluate the scale) • Future criterion validity (to evaluate the extent of the relationship between test score/performance and another test(s) considered a 'gold standard') • Second field test will focus on quantitative review of items 	<p>A TAG supported the development of MICS FLS and the instrument was field tested in Belize, Kenya, Ghana and Costa Rica.</p> <p>Quantitative review can be conducted ex-post</p>

12. FLM2.0 against the 5 criteria: Criteria 3,4 and 5

Criterion	Description	Minimum requirement	FLM 2.0	FLS
3	is the sample of learners that took the assessment representative of the population against which the results will be reporting?	To report against SDG 4.1.1, there must be evidence that the sample of learners who took the assessment is representative of the population against which the results will be reported. Where the assessment is administered to a sample of the population, the margin of error should be 5 percent or less at the 95 percent confidence level.	MICS is a nationally representative household survey. MICS uses probability sampling which allows to make valid inferences about the population because the probability of selection is known, and weights can be applied to the data. Each Survey findings report provides details about sampling Based on ongoing work, oversampling may be explored	MICS is a nationally representative household survey. MICS uses probability sampling which allows to make valid inferences about the population because the probability of selection is known, and weights can be applied to the data. Each Survey findings report provides details about sampling
4	is there evidence that the assessment was administered in a standardized way?	To report against SDG 4.1.1, there must be evidence that the assessment was administered in an appropriate and standardized way (for example, administration conditions are consistent, or length of time to administer the assessment is adhered to).	FLM 2.0 will build on and expand the FLS protocols, building results from our field validation, ex post analysis of the FLS data, and guidance from the TAG.	FLS Standard detailed protocols, guidelines and survey tools are available on customization, training and administration. → https://mics.unicef.org/tools
5	are the outcomes of the assessment sufficiently reliable?	To report against SDG 4.1.1, the value of coefficient alpha(or equivalent reliability statistic) for the assessment must be greater than or equal to 0.7. In addition, there must be evidence of appropriate quality assurance arrangements for any human-scored items. This could include scoring of items with a pre-agreed score or double scoring of a sample of responses.	As done in the FLS, the FLM 2.0 will also follow stringent cross-country quality assurance with the support of the Global and Regional MICS team; we also plan to report on reliability statistics calculated from the assessment as part of accompanying methodological papers	Gochyyev P., Mizunoya S., and Cardoso M. (2019). <i>Validity and reliability of the MICS foundational learning module</i> . MICS Methodological Papers, No. 9, Data and Analytics Section, Division of Data, Research and Policy, UNICEF, New York. The Cronbach's alpha for reading subtest' individual items was estimated at 0.92. The Cronbach's alpha for the overall numeracy instrument consisting of 21 items was estimated at 0.76.

14. Recommendations for GAML

- As a principle, we must find a technical solution to preserve SDG 4.1.1a, which should include improving both the flow of new early-grade assessments, but also build on the stock of existing assessments.
 - UNICEF has been working on the improvement of our tool to measure learning and appreciate the increasing clarity that has been provided
 - UNICEF stands ready to collaborate on methodological work to build on the stock of 56 assessments
- Flexibility in implementing the standards so we create the much-needed market contestability in the learning assessment space;
- Member states, with the support of partners, should work towards a list of recommended languages of assessment (including the information required on how to weight these children to create population estimates);
- PAL, USAID, and UNICEF collectively recommend constituting a TAG to support assessment design decisions for reporting on SDG 4.1.1a
 - Probabilistic sampling for unbiased and representative estimates
 - Aligned to MPL standards and guidance for reporting
 - Guidance on how to set the benchmark protocol
 - Temporal comparability to allow to monitor progress towards SDG indicator