VALIDITY AND RELIABILITY OF THE MICS FOUNDATIONAL LEARNING MODULE

MICS METHODOLOGICAL PAPERS

Paper No. 9, 2019

Data and Analytics Section
Division of Data, Research and Policy

United Nations Children's Fund
Data and Analytics Section, Division of Data, Research and Policy
3 United Nations Plaza, New York, NY 10017, USA
Tel: +1 (212) 326-7000
Email: mics@unicef.org

Suggested citation: Gochyyev P., Mizunoya S., and 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, UNICEF, New York.

VALIDITY AND RELIABILITY OF THE MICS FOUNDATIONAL LEARNING MODULE

MICS METHODOLOGICAL PAPERS

Paper No. 9, 2019

Data and Analytics Section
Division of Data, Research and Policy

About MICS

The Multiple Indicator Cluster Surveys, MICS, is one of the largest global sources of statistically sound and internationally comparable data on children and women. MICS data are gathered during face-to-face interviews in representative samples of households. The surveys are typically carried out by government organizations, with technical support from UNICEF.

Since the mid-1990s, MICS has supported more than 100 countries to produce data on a range of indicators in areas such as health, education, child protection and HIV/AIDS. MICS data can be disaggregated by numerous geographic, social and demographic characteristics.

As of 2019, five rounds of surveys have been conducted: MICS1 (1995-1999), MICS2 (1999-2004), MICS3 (2004–2009), MICS4 (2009–2012) and MICS5 (2012-2015). The sixth round of MICS (MICS6) is scheduled to take place in 2016–2019. Survey results, tools, reports, micro-data and information on the MICS programme are available at <*mics.unicef.org*>.

About the MICS Methodological Papers

MICS Methodological Papers are intended to facilitate exchange of knowledge and to stimulate discussion on the methodological issues related to the collection, analysis, and dissemination of MICS data; in particular, the papers document the background methodological work undertaken for the development of new MICS indicators, modules, and analyses. The findings, interpretation and conclusions do not necessarily reflect the policies or views of UNICEF.

Acknowledgements

The authors wish to thank Shane M Khan for review of the paper and Diogo Amaro for finalization of the document.

1

Executive summary

This paper describes the validation of the Foundational Learning Skills (FLS) module, a newly developed instrument in the Multiple Indicator Cluster Surveys (MICS). The focus of this study is the concurrent validity of the FLS instrument, which was administered to children aged 7 to 14 in Kenya, along with the Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) tools. Evidence on the reliability of the instrument – such as inter-rater reliability and Cronbach's alpha – was also obtained.

The FLS instrument of reading and mathematics skills, which focuses on children aged 7 to 14, is implemented as a self-contained module within a broader household survey. This has two important implications on equity: (1) It does not exclude those children who are not enrolled in schools, and (2) It provides substantial background information on the participants' households and families. The FLS module is designed to produce learning outcome data that can be compared across multiple languages and countries in an inclusive and efficient manner.

This study uses the data set collected in October 2016 from in-school children (average age 10) across four Kenyan villages (two schools per village). A total of 130 children (44 per cent female) were administered an FLS literacy instrument and 127 children (55 per cent female) were administered an FLS numeracy instrument. To gather evidence for the concurrent validity, same sets of students were administered EGRA and EGMA instruments.

Our analysis shows that the reliability estimates for the composite scores are 0.92 for reading and 0.76 for numeracy instruments. This indicates that both instruments are internally consistent, and the number of items used to measure the composite scores is sufficient. The reliability estimates for the two reading subtests are 0.83 and 0.91 and range from 0.54 to 0.62 for the four numeracy subtests (each having five to six binary items). We found that the number identification subtest has the lowest estimate of reliability (at 0.54). Due to the relatively small sample size of our study (n = 127), these findings are not definitive and need to be confirmed in a larger study.

Regarding reliability estimates for numeracy subsets, we further analysed data from a 2017 Sierra Leone MICS,[1] which includes responses from more than 6,000 children. The results were found to be substantively higher than the estimates obtained in this Kenya survey, with the reliability estimates for the four numeracy subtests ranging from 0.90 to 0.94.

Based on the interpretation suggested by McHugh,[2] inter-rater reliability results – averaged across items within each subtest – indicate strong (0.81-0.90) agreement for the reading and almost perfect (above 0.90) agreement for the numeracy subtests. These results are mainly based on the estimates of Cohen's kappa and the proportion of negative agreement (PNA). This implies that the interviewers were consistent in their scoring judgments in both reading and numeracy instruments.

Coefficients of correlation between comparable tasks across the FLS and EGMA/EGMA instruments ranged mostly from .63 to .87, with one exception: the addition task, which was estimated at .52. Note that all three of these correlations are statistically significant. Considering some design-related differences between the FLS module and EGRA and EGMA instruments, results of the concurrent validity study were in line with expectations. Specifically, as EGRA and EGMA focus on fluency, children are timed on most of the tasks, whereas the FLS module is not timed. It is therefore reasonable to expect some dissimilarity in performances across the instruments.

In summary, results from psychometric analyses show that the new FLS module, with properties that pass scrutiny in terms of validity and reliability, has promise to serve as an efficient and inclusive measure of foundational learning – in terms of both reading and numeracy skills – among 7- to 14-year-old children in households across various countries.

2

Introduction

This paper builds on a previous methodological paper, 'Collecting Data on Foundational Learning Skills and Parental Involvement in Education',[3] that presented the rationale, development and initial validation process for the new FLS and Parental Involvement modules. They were designed as standalone modules for MICS using findings from field trials in four countries: Belize, Costa Rica, Ghana and Kenya, and were conducted in four languages (Spanish, Akwapem Tui, Kiswahili and English).

Results of the FLS module inform monitoring of Sustainable Development Goal 4 as the learning outcomes in reading and numeracy at the Grade 2 or 3 level are one of the goal's indicators. The design and practical nature of the module matters from an equity perspective in two important ways: First, this new module can reach children aged 7 to 14 who are excluded from school-based assessments. Second, it provides a wealth of information about household characteristics that reveal disparities among groups. The FLS module not only measures learning outcomes needed to monitor SDG4 indicators, but also collects indispensable data for policy discussion on how learning outcomes, children's socioeconomic status and schooling status interact.

Understanding learning outcomes of children in reading and comprehending texts is particularly important because these skills are crucial for children to continue to develop their academic competencies. The so-called 'Matthew Effect'[4] shows that as some children lag behind in learning to read, this gap grows over time, especially as their classmates transition to 'reading to learn'. Non-readers and children who can decode but not comprehend are overrepresented in socially disadvantaged groups. But such non-readers may not be captured by traditional school-based assessments because they tend to have never enrolled in school or have dropped out.

The FLS module's reading component is designed to be a short and practical instrument that can be attached to existing household surveys to assess reading comprehension while maintaining interlanguage comparability. This flexible design offers tremendous potential for cross-national comparability, while also meeting the requirement for both equity and efficiency. FLS data collected through a household survey inform the equity discussion in education policy by offering detailed information on children's socio-economic characteristics. The module is also efficient because it is administered as part of a household survey already being conducted and requires much fewer resources and relatively little training.

This paper focuses on a concurrent validity study using the data obtained from Kenya in order to examine the quality of learning data collected through a household survey. It provides an overview of each of the

FLS instruments' reliability and validity. Focus is on the concurrent validity, with EGRA being the external criterion. The next three sections focus on methods, results and conclusions.

3

Methods

This section consists of three subsections. We start with a discussion of the FLS instrument's reliability (both inter-rater reliability and Cronbach's alpha), followed by an assessment of concurrent validity with the rationale for the selection of comparators. Lastly, the study design used to collect data pertaining to both validity and reliability analyses is introduced.

Reliability

Reliability is a measure of a test's consistency. When a child answers an item, there will be factors that affect his/her response in addition to the true level of what is intended to be measured. This error is an unavoidable aspect of the measurement. The reliability coefficient summarizes this error, ranging from 0 to 1 (the higher the better). When an instrument is adequately reliable, measurement errors are sufficiently small to justify using the obtained score.[5]

Several types of reliability methods exist, e.g., test-retest reliability, parallel-forms reliability and Cronbach's alpha. Cronbach's alpha reflects the covariation among items of the test. Test-retest and parallel forms reliabilities do not apply to this instrument for the following reasons: (1) the small number of items, (2) the test is not repeatedly administered, and (3) the forms were identical in content and only differed in the ordering of the items.

When response data obtained from the instrument involve more than one rater, inter-rater reliability is of relevance.

Inter-rater reliability

Raters can themselves be sources of measurement uncertainty. Inter-rater reliability indicates the level of agreement between two (or more) raters rating performance of the same examinee (e.g., the child). Several methods can be used to obtain this inter-rater reliability score. Cohen's kappa[6] – the most popular statistic regardless of its limitations – readily applies to the FLS instrument as it is often used for situations involving two raters and a binary outcome (i.e., 'correct' or 'incorrect'). However, Cohen's kappa has flaws[7] and was not *specifically* designed to be used in testing situations like MICS or EGRA (see Appendix 2).

A more apt metric in some situations is the proportion of positive agreement or its corresponding metric, in this case, the PNA, because an 'incorrect' or negative response by the subject is the only mark made by the raters. We obtained averages of Cohen's kappa statistics for each of the subtests. In cases when Cohen's kappa is misleading (due to its limitations),[8] we have obtained an alternative computation of this statistic that addresses these limitations, proposed by Brennan and Prediger,[9] which is also known as prevalence-adjusted and bias-adjusted kappa.[10]

Cronbach's alpha

For MICS FLS subtests, none are timed, and thus Cronbach's alpha is appropriate as a measure of reliability. Both overall test (i.e., the composite score for reading or numeracy) and subtests can be assessed for their reliability. Note, however, that Cronbach's alpha tends to be lower in tests with a small number of items. While the FLS module's 'Oral Reading Accuracy' task has 43 items, the FLS subtests have only five or six items. This implies that the scale generally used for outcomes on Cronbach's alpha should serve as a guideline rather than a set of strict cut-offs for determining internal consistency.

To evaluate the test's internal consistency, we also estimated item-test correlations. Item-test correlation, also known as point biserial correlation when items are binary, allows evaluation of whether an item is a good contributor to what is measured by the test. The total score includes an item itself, and hence the correlation of the item with the total score is an overestimate of the relationship. Therefore, we obtained a corrected item-test correlation, also known as item-rest correlation, which represents the correlation of the item with the rest of the total score. Ebel suggested that an item with a correlation of less than 0.20 may require some revision and items exceeding a correlation of 0.40 are judged as 'good'.[11]

Concurrent validity

One way to gather evidence for the validity of a new test like FLS is to compare it to another widely accepted and established instrument that is intended to measure the same set of skills. Therefore, FLS results were compared to EGRA/EGMA results to obtain evidence for the concurrent validity as a measure for early grade reading and mathematics skills.

Due to equity concerns around learning outcomes and their implications for education policies, the focus of a MICS results analysis will be not only the average scores or improving averages for a portion of the population, but also the proportion of children that have achieved (or not achieved) a set of foundational learning skills.

The MICS achievement groups were compared to similar achievement groups in the EGRA and EGMA. RTI International, the creators of EGRA and EGMA, have outlined cut-offs for levels of proficiency for these two tests.[12] In EGRA, for example, the recommended cut-offs are:

- Non-/Beginning reader: Reading fluency greater than or equal to 0 words per minute and reading comprehension equal to 0 per cent
- Emergent reader: Reading fluency greater than 0 words per minute and reading comprehension greater than 0 per cent but less than 80 per cent
- Reader: Reading fluency greater than 0 words per minute and reading comprehension of at least 80 per cent

Similar cut-offs can be applied to the MICS reading subtests, and a direct comparison between the cut-offs of the two tests can be obtained for each child to evaluate consistency across two different instruments.

RTI's recommended skill groups for EGMA are:

- Non-/Early mathematician: Either missing number and/or addition and subtraction level 2 below 30 per cent
- Emergent mathematician: Missing number and addition and subtraction level 2 both above 30 per cent
- Mathematician: Missing number and addition and subtraction level 2 both above 80 per cent

Unfortunately, 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. The revised cut-offs are:

- Non-/Early mathematician: Either missing number or addition (both level 1 and level 2 combined) below 30 per cent
- Emergent mathematician: Missing number and addition both above 30 per cent
- Mathematician: Missing number and addition both above 80 per cent

Study design

This subsection describes the study design used to collect data pertaining to both validity and reliability.

Field operations and data collector backgrounds

The field team was structured into two teams (A and B) with each team having a field coordinator, a supervisor and 20 research assistants. An overall team leader was in charge of the whole research team. The 20 research assistants (10 female and 10 male) were divided as follows: eight assessors, eight raters and four solo assessors. The research assistants were divided equally between numeracy and reading.

A total of eight schools were visited and 48 children were assessed in each of the schools. Only one of the schools did not have a sufficient number of learners in some of its classes. For those classes, a census was drawn.

The team leader was in charge of quality control and ensuring that the two teams followed the established protocol and met the objectives set for the assignment.

The field coordinator was charged with: (1) ensuring that the overall team was well-organized; (2) addressing any issues with the school administration; (3) ensuring that the research assistants were well-facilitated; and (4) advising on sampling.

The supervisors were mainly tasked with checking that the process of sampling the pupils was undertaken according to agreed procedures for random sampling. They were also in charge of making sure that the children were informed of the process and that the children's initial consent was obtained after explaining the research process in a child-friendly way.

School selection

Eight schools were identified from two Kenyan counties (Nairobi and Kiambu) and selected considering the following characteristics: (1) urban versus peri-urban areas; (2) informal versus formal settlements of

the lower-income levels in both counties; and (3) public and alternative basic education and training centres (APBET, previously referred to in Kenya as non-formal schools).

Of the eight schools (shown in Table 1), five were public and three were APBET schools. All three of the APBET schools were from informal settlements in Nairobi (Kibera, Korogocho and Kayole), while the public schools were in the same locations and therefore had the same catchment areas as the APBET schools. The two schools in Kiambu county were from Thika sub-county, with one (Kianjau) located in the Kiandutu informal settlement, while Kiganjo primary was in a peri-urban and more formal environment.

Table 1: Selected schools

County	Public schools	APBET schools	Locality
Nairobi	Mwangaza Primary	LOGEF	Kayole area
	Ngunyumu Primary	Makao Junior	Korogocho area
	Ayany Primary	Oloo Children Home	Kibera area
Kiambu	Kianjau Primary		Thika - Kiandutu
	Kiganjo Primary		Thika

Sampling of children

A total of 48 children (shown in Table 2) were sampled from each school. The children were aged 7 to 14. It should be noted that a decision was made to exclude Class 8 from the sampling, primarily due to the fact that those pupils were preparing for the national end-of-year Certificate of Primary Education examinations at the time of the assessment.

Table 2. Number of pupils assessed per class

Class	No. of students	Reading		Numeracy		Reading solo assessor	Numeracy solo assessor
		Assessor 1 Rater 1	Assessor 2 Rater 2	Assessor 1 Rater 1	Assessor 2 Rater 2		
Class 1	8	2	1	2	1	1	1
Class 2	8	1	2	1	2	1	1
Class 3	8	2	1	1	1	1	2
Class 4	8	1	1	1	2	2	1
Class 5	6	1	1	1	1	1	1
Class 6*	5	1	1	1		1	1
Class 7*	5		1	1	1	1	1
Total assessed	48	8	8	8	8	8	8

*Class 6 and Class 7: Sampling by gender was done as follows: two boys, three girls in Class 6 and three boys, two girls in Class 7.

Determining the ages of Class 1 and some Class 2 pupils presented a general challenge. The supervisors sampled the pupils and then had to confirm their ages from their teachers. In instances where a child contradicted the age given by the class teacher, the research team opted to use the class teacher's information. Furthermore, most of the lower grade learners (classes 1 to 3) did not know their date of birth.

Identification of out-of-school children was also a challenge due to the fact that most areas employ multiple strategies to mobilize children to go back to school. In some areas, the type of mobilizers (e.g., community mobilizers) posed a challenge as they often do mobilizations (i.e., rounding-up children) together with local administrations. All of the identified and assessed out-of-school children were from informal settlement localities.

Data entry processes

The data was collected on paper and later transferred by trained data entry clerks to an electronic form on tablets using Tangerine Software assessors/raters.

Admittedly, this process can be problematic for reliability purposes. Specifically, it is difficult to determine if any disagreements between the two raters are due to: (1) complexities of the testing situation; (2) personal differences in scoring (i.e., the reasons inter-rater reliability testing is done); or (3) issues with transcription during data entry.

4

Results

Reliability results

Internal consistency: Cronbach's alpha

Results on the reading assessment

There are two reading subtests in MICS – the Oral Reading Accuracy task and the Reading Comprehension questions. For the Oral Reading Accuracy task (consisting of 43 items), Cronbach's alpha was estimated at 0.91, denoting an excellent degree of internal consistency. In the five-item Reading Comprehension subtest, Cronbach's alpha is 0.83, which represents satisfactory reliability, especially given the small number of items. Full item-level results can be found in the appendices. Table 3 below also shows average item-rest correlations for the items from the MICS Reading subtests, all of which are above the threshold of 0.4 suggested by Ebel.[13]

Table 3. Measures of reliability and item-rest correlation for the MICS Reading subtests

MICS Reading subtest	Number of items	Average item-rest correlation	Cronbach's alpha
Oral Reading Accuracy	43	0.53	0.91
Reading Comprehension	5	0.64	0.83
Both subtests' individual items	48	0.53	0.92

Results on the numeracy assessment

The Cronbach's alpha for the overall numeracy instrument consisting of 21 items was estimated at 0.76. The four numeracy subtests had fairly similar reliabilities. The Number Identification component had the lowest alpha, estimated at 0.54 while the reliabilities for the three remaining subtests ranged between 0.58 and 0.62. While these levels of reliability are generally seen as below acceptable, they are expected considering the number of items for each subtest.

Table 4. Measures of reliability and item-rest correlation for the MICS Numeracy subtests

MICS Numeracy subtest	Number of items	Average item-rest correlation	Cronbach's alpha
Number Identification	6	0.48	0.54
Quantitative Comparisons	5	0.43	0.61
Addition	5	0.35	0.58
Missing Number	5	0.41	0.62
Overall numeracy	21	0.35	0.76

In Table 5, we provide reliability estimates for the same subscales in the EGMA instrument. Notice that the number of items in each of the EGMA subscales differs from the number of items in the corresponding MICS Numeracy subscales. This difference in subtest lengths does not allow comparison of

the same subscales in their reliability indices across two different instruments. Using the Spearman-Brown prediction formula, we can obtain predicted reliabilities for the MICS Numeracy subtests by matching the number of items in FLS to the corresponding subtest of the EGMA instrument, shown in the last column of Table 5.

Table 5. Predicted estimates of the reliability using Spearman-Brown prediction formula

Subtest (EGMA/FLS)	Number of items in EGMA	EGMA Cronbach's alpha	FLS Numeracy Cronbach's alpha (corrected for test length)
Number Identification	20	0.83	0.80 (if 20 items)
Quantitative Comparisons	10	0.72	0.76 (if 10 items)
Addition	25	0.52	0.87 (if 25 items)
Missing Number	10	0.65	0.77 (if 10 items)
Overall EGMA	65	0.65	0.91 (if 65 items)

Inter-rater reliability results

As mentioned earlier, Cohen's kappa works well for tests where there is no default answer (i.e., an answer that is automatically recorded when an assessor does not enter anything on the paper). The Reading Comprehension, Number Identification, Quantitative Comparisons, Addition and Missing Number subtests are examples of such tests. Table 6 below presents the mean kappa and PNA values for each test. Please note that full distributions can be found in the appendices.

Reading results

The MICS Reading subtests have very high inter-rater reliability scores obtained from the data collected in Kenya. The Oral Reading Accuracy task shows substantial agreement for PNA, and the Reading Comprehension subtest shows the higher 'almost perfect agreement' rating between assessors on average.

Table 6. Summary of average kappa and PNA estimates for the MICS Reading subtests

Reading (n = 130 pairs)	Average kappa	Average PNA
Oral Reading Accuracy	0.81	0.76
Reading Comprehension	0.90	0.88

Another way to look at the data is by examining which of the individual items were subject to high levels of disagreement between the two raters. This could be either a situation where one rater marked an item incorrect while the other marked it correct, or where one rater marked the item as not attempted and the other marked it as attempted and either correct or incorrect.

Overall, the level of agreement between the two raters on each individual item was high. Items with lower levels of agreement are examined here to determine if such disagreements can be avoided or training can be improved (to decrease the rater effects) in future MICS assessments.

In the Oral Reading Accuracy task, only two items fell below 91 per cent agreement. They were the words 'Nyanyake' and 'Mkoba'. The level of agreement for each item can be seen in Figure 1 below. Please note that the vertical scale begins at 85 per cent, so even dips in the graph represent high levels of agreement.

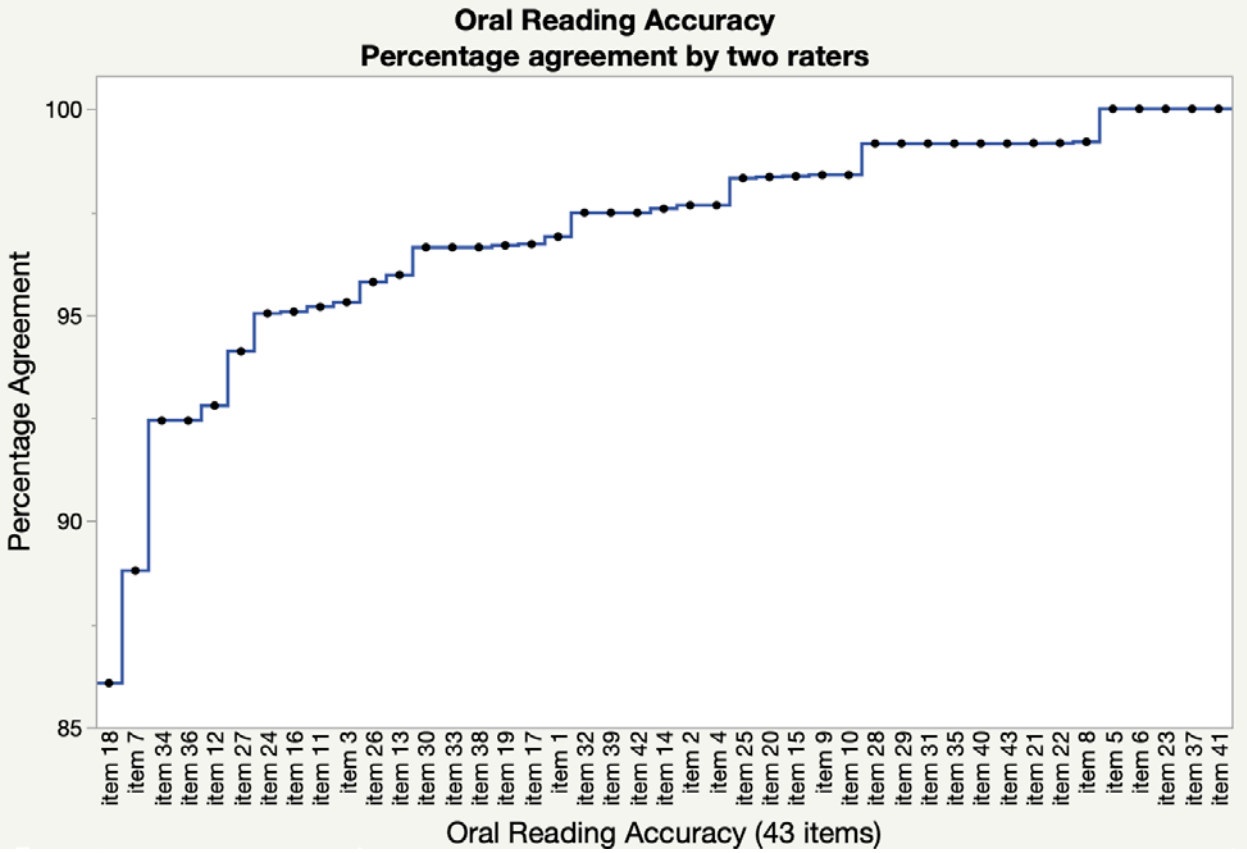


Figure 1. **Item-level rater agreement for the Oral Reading Accuracy subtest**

A high degree of agreement is visible in the Reading Comprehension subtest as well, as shown in Figure 2. Question 4: “Kwa nini Maria alipoteza pesa?” (Why did Maria lose the money?) had the potential answer: “Kwa sababu mkoba wake ulikuwa na shimo kubwa.” (Because it fell through the hole in the bag/The bag had a hole). The inferential nature of this question may have contributed to the rater disagreement. However, it is unclear as to why the agreement for Question 5, an inferential question as well, was substantively higher.

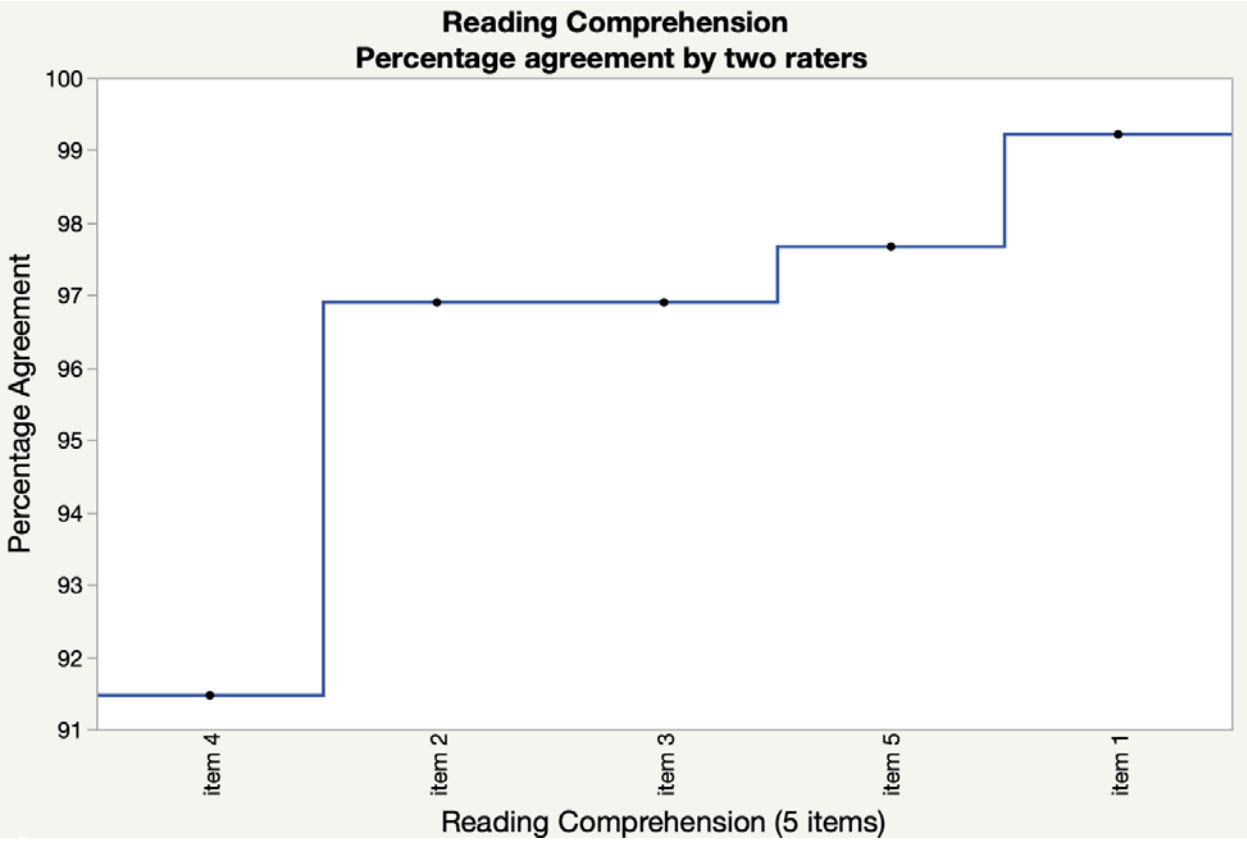


Figure 2. **Item-level rater agreement for the Reading Comprehension subtest**

Numeracy results

The numeracy subtests of the FLS show even higher scores for the kappa, all estimated at or above 0.89. There is a minimal disagreement between raters on the MICS numeracy subtest questions

Table 7. **Summary of average kappa and PNA estimates for the MICS Numeracy subtests**

Numeracy (n = 127 pairs)	Average kappa	Average PNA
Number Identification	0.98	0.94
Quantitative Comparison	0.96	0.94
Addition	0.93	0.90
Missing Number	0.92	0.90

Individual item analysis revealed that some items have relatively higher levels of disagreement among the raters. Additional emphasis during the data collector training may result in further improvements in agreement among raters.

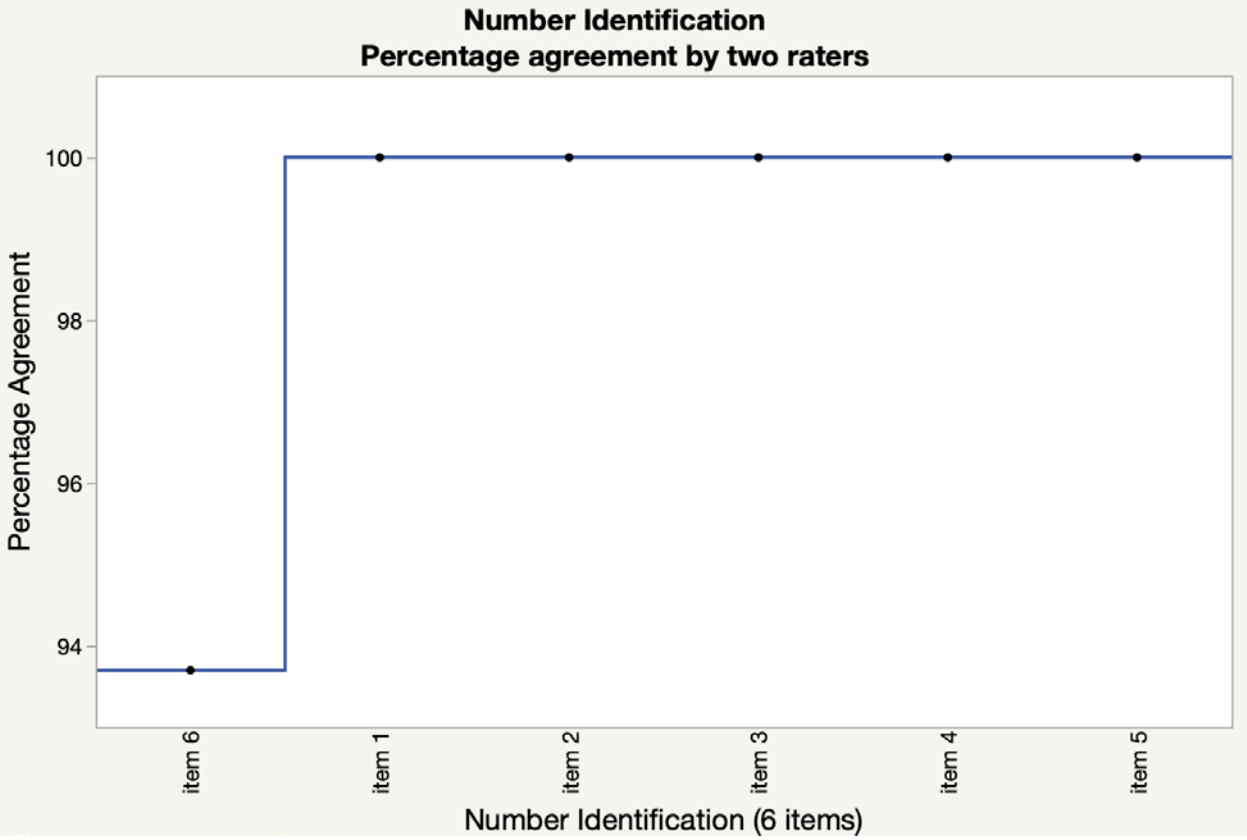


Figure 3. **Item-level rater agreement for the Number Identification subtest**

Agreement levels for the Number Identification subtest are shown in Figure 3. Only the item focusing on the three-digit number, 731, had imperfect level of disagreement between the raters. This is likely due to the administration of two different assessments by the data collectors. In EGMA, only the answer '731' was marked as correct. There was more leniency in the MICS scoring. For example, the word 'and' could have been left out and the answer would still have been marked correct in MICS. Trying to remember when a particular type of answer is correct or incorrect in this double administration likely caused these disagreements. There will likely be less confusion when only one instrument is administered.

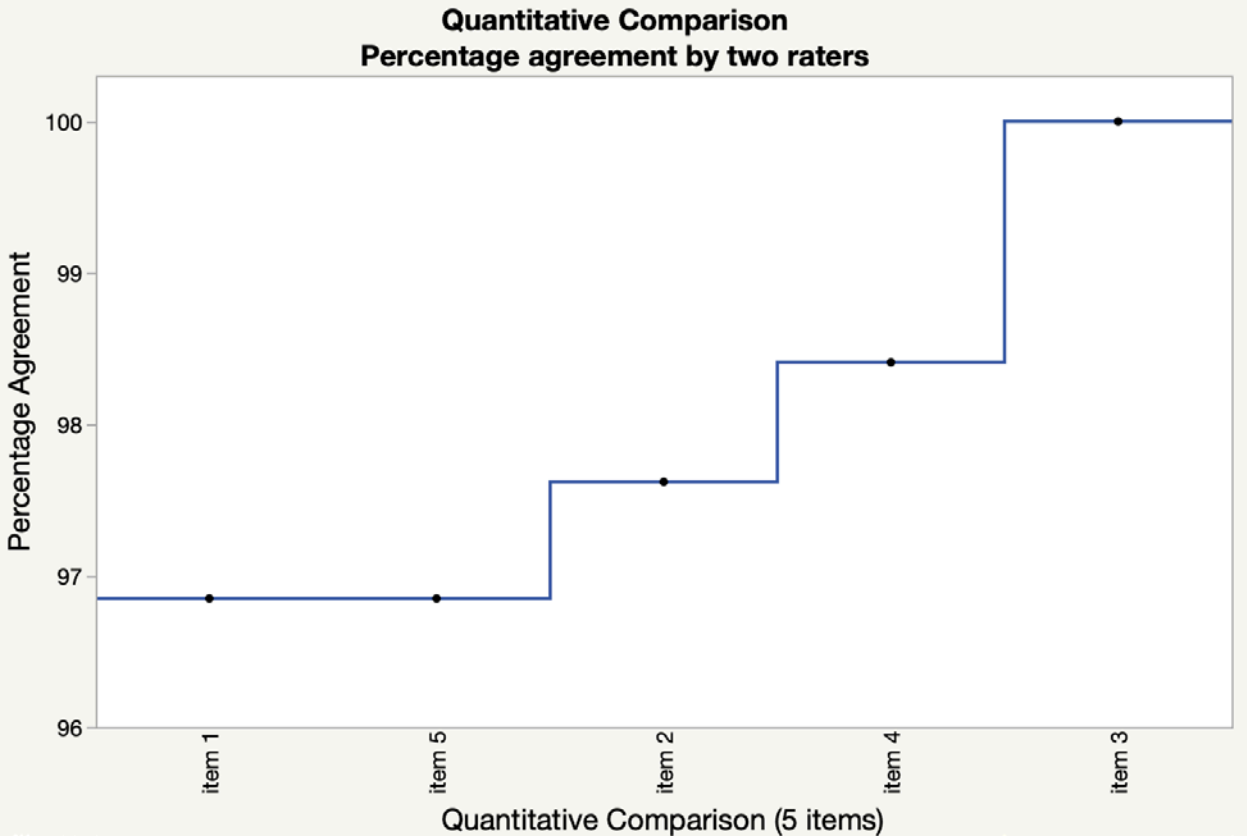


Figure 4. Item-level rater agreement for the Quantitative Comparison subtest

There was minimal disagreement among raters on most of the items in the Quantitative Comparison subtest. Again, the rules for administering EGMA and MICS were different. In EGMA, a child needs to merely point at the correct item to be marked as correct, whereas in MICS the child must also say the correct name. Confusion surrounding which instrument these rules correspond to might account for some of the disagreements we see in Figure 4.

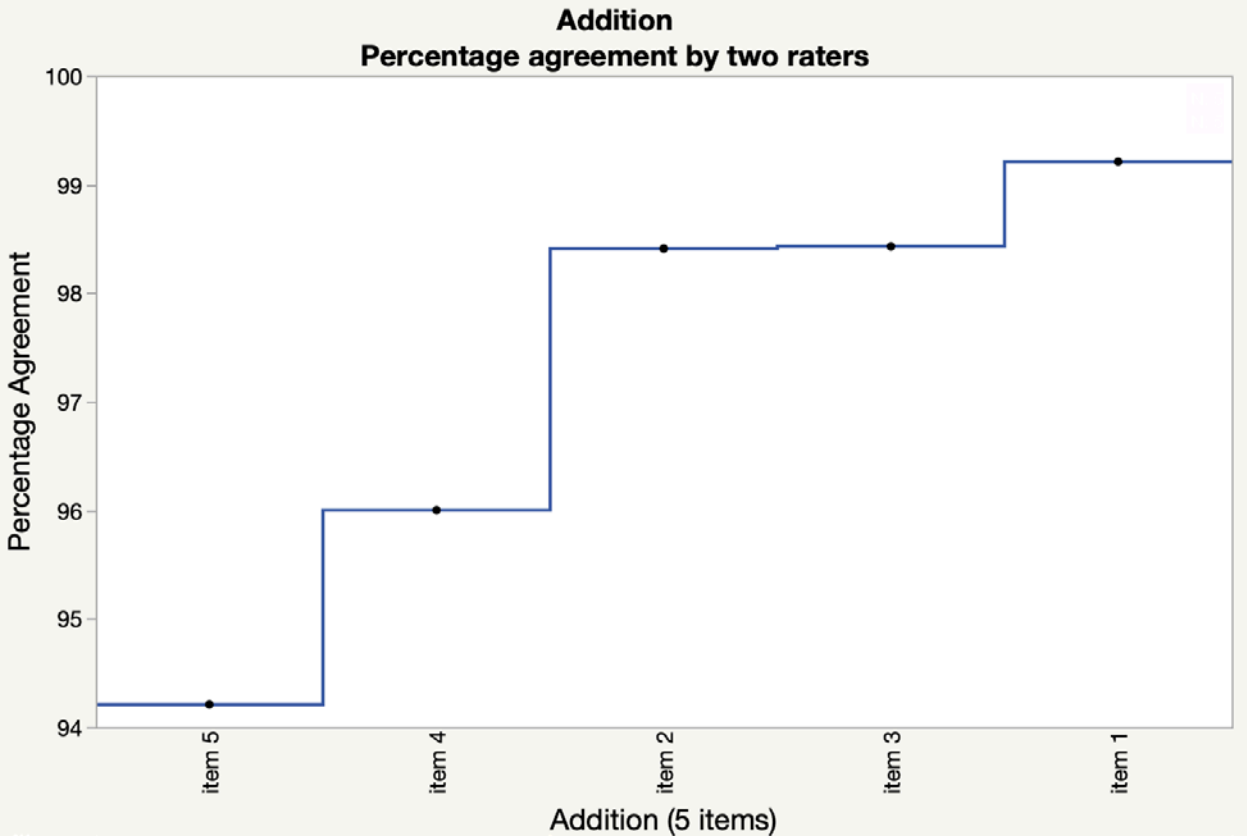


Figure 5. Item-level rater agreement for the Addition subtest

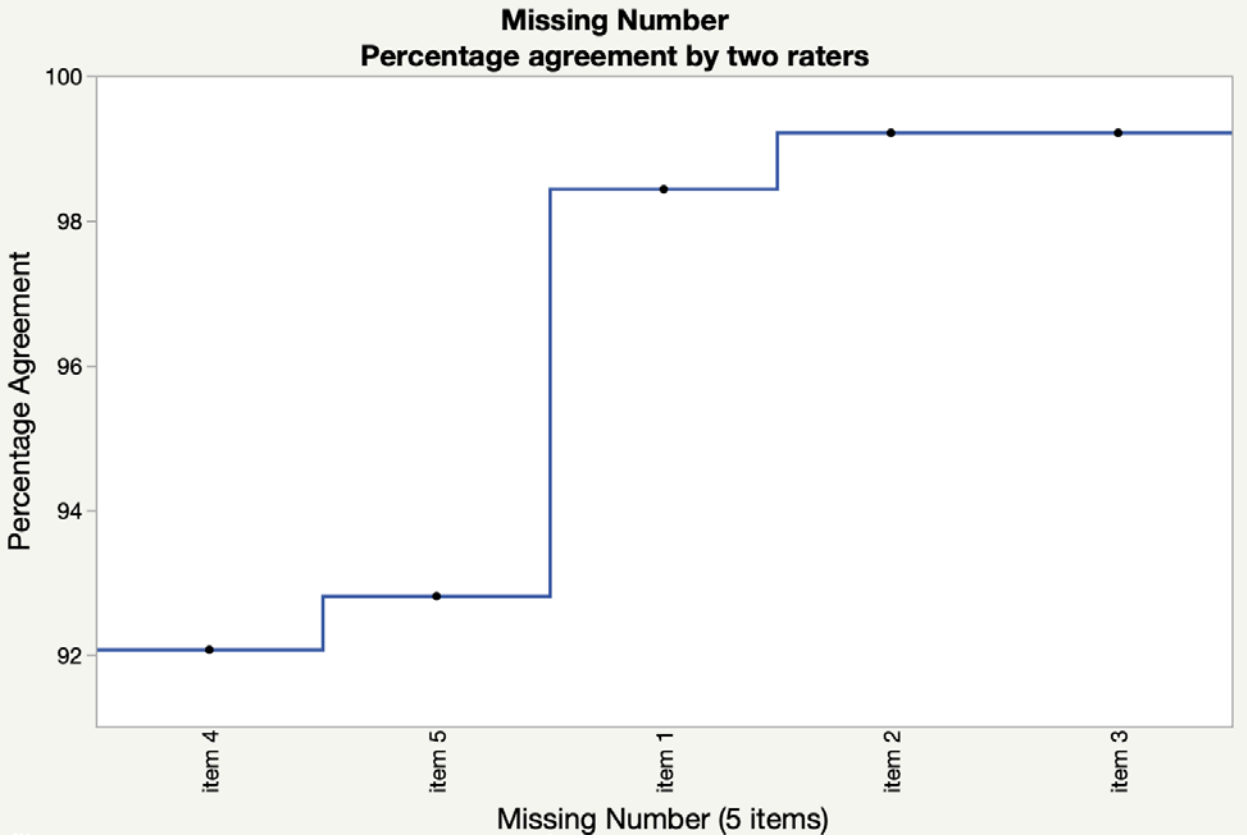


Figure 6. Item-level rater agreement for the Missing Number subtest

As figures 5 and 6 show, both the Missing Number and Addition subtests have high rates of agreement overall, but the final two of the individual items have a relatively lower level of agreement compared to the first three.

One potential explanation is that there may be some differences in enforcement of the rule related to ending the subtest. FLS protocol instructs that the subtest be aborted after two consecutive non-responses instances. While this process did occur with some regularity, it often was not the result of the discrepancy among the raters. For example, in the Missing Number subtest, only two of the 25 differences between the raters were due to one assessor marking an item as skipped due to a stopping rule (one on Item #3 and the other on Item #5). This is not the reason for the drop in the agreement in the last few items of these MICS subtests. It is not clear what is causing the larger disagreement at the end of some MICS subtests.

Concurrent validity results

Reading results

FLM analysis can go beyond collecting the overall averages by focusing on determining the proportion of children meeting particular benchmarks in both Reading and Numeracy. The benchmarks, outlined for MICS and EGRA, are:

- **Non-/Beginning reader:** Reading fluency greater than or equal to 0 words per minute and reading comprehension equal to 0 per cent
- **Emergent reader**: Reading fluency greater than 0 words per minute and reading comprehension greater than 0 per cent but less than 80 per cent
- **Reader**: Reading fluency greater than 0 words per minute and reading comprehension of at least 80 per cent

Table 8 shows the placement of the children on this scale for both the MICS and EGRA Reading subtasks.

Table 8. Placement of the children on MICS (reading) and EGRA measures

EGRA Reading Grouping	EGRA Reading Grouping based on the FLS instrument			
	Non-/Beginning	Emergent	Reader	Total
Non-/ Beginning	12	2	1	15
Emergent	4	31	64	99
Reader	0	6	72	78
Total	16	39	137	192

While the majority of children scored in the same performance category in both FLS and EGRA for this subtask (i.e., Non-/Beginning Reader on both tests), a significant number of children – nearly a third of those tested – scored in the Reader level on FLS but only in the Emergent Reader group in EGRA. These differences were largely due to children running out of time on the EGRA passage and thus not being presented with enough questions to have the chance to be placed in the Reader level.

The table below reflects the percentage of children in each EGRA Reading level that were able to successfully complete the FLS Reading indicator. Since the FLS indicators are binary, children in the lower EGRA groups should likely have a lower pass rate on the FLS indicator, and these pass rates should increase as the children exhibit greater reading skill and are placed in higher EGRA Reading groups.

Table 9. Percentages of children in each of the EGRA and MICS Reading performance levels

EGRA Reading Grouping	MICS Indicators				
	Accuracy	Literal comprehension	Inferential comprehension	Basic reading skills	Total count
Non-/ Beginning	12.5%	0.0%	0.0%	0.0%	16
Emergent	53.8%	33.3%	5.1%	0%	39
Reader	92.0%	94.2%	84.7%	74.5%	137
Total	77.6%	74.0%	61.5%	53.1%	192

Very few of the children deemed Non-reader/Beginning-reader (by RTI classification) were able to complete any of the four FLS indicators satisfactorily (2 out of the 16, or 12.5 per cent). However, 92 per cent of the children in the Reader-level classification were able to read with 90 per cent or greater accuracy, 94.2 per cent of these children were able to answer all three of the literal reading comprehension questions correctly, and 61.5 per cent of these children were able to answer the two inferential reading comprehension questions correctly. Overall, 77.6 per cent of the children were able to correctly read 90 per cent or more of the words in the Oral Reading Accuracy task, and 53.1 per cent completed all three skills successfully.

This strong association between the EGRA and the FLS indicators lends validity to the use of the MICS instrument and its indicators. Generally speaking, most of the students who were labelled as Non-readers were not able to meet any of the MICS indicators successfully, and most of the children labelled as Readers were able to successfully meet the FLS standards.

By disaggregating the two subtests (Connected Text passage and the Reading Comprehension questions), it is possible to look at correlations between the scores on the two tests – visualized in figures 7 and 8, respectively. These graphs, while informative, indicate the ceiling effect of FLS tests. A large proportion of the children were able to score 80 per cent or more on the FLS, while a much smaller proportion scored 80 per cent or more in EGRA.

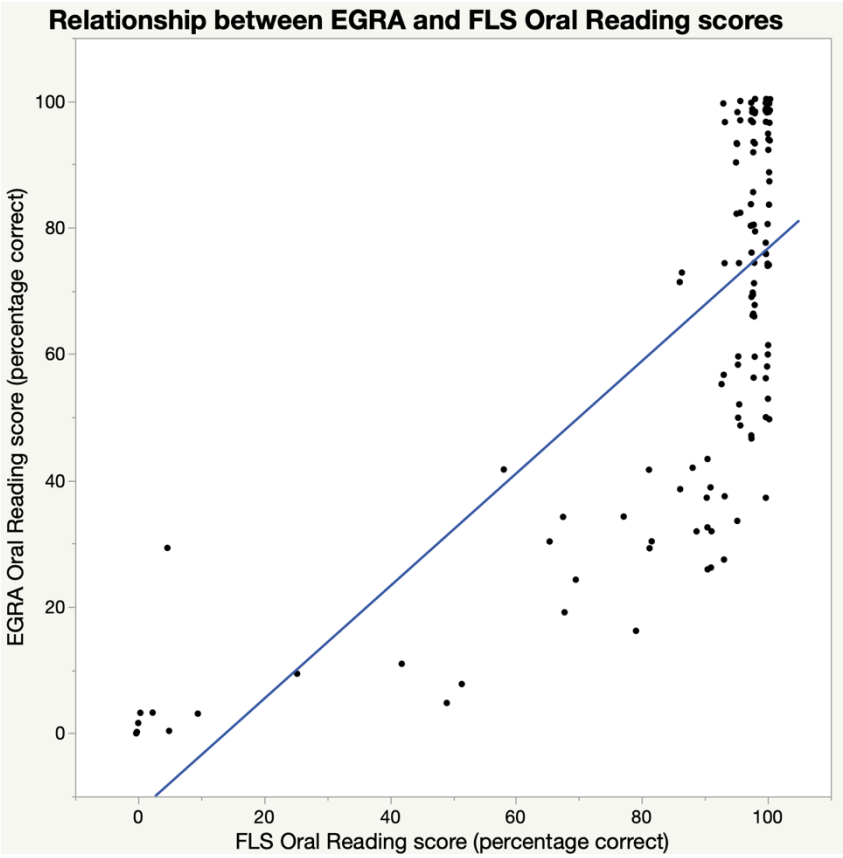


Figure 7. Relationship between EGRA and FLS Oral Reading scores (correlation=0.72 [CI: 0.63, 0.80])

It is more difficult to see the individual points in the Reading Comprehension subtasks because of the finite number of outcomes (between zero and five questions correct on each test), but by jittering the data (i.e., adding random values to avoid dots plotted exactly on top of each other), it is possible to see that there are greater and fewer points at each possible outcome.

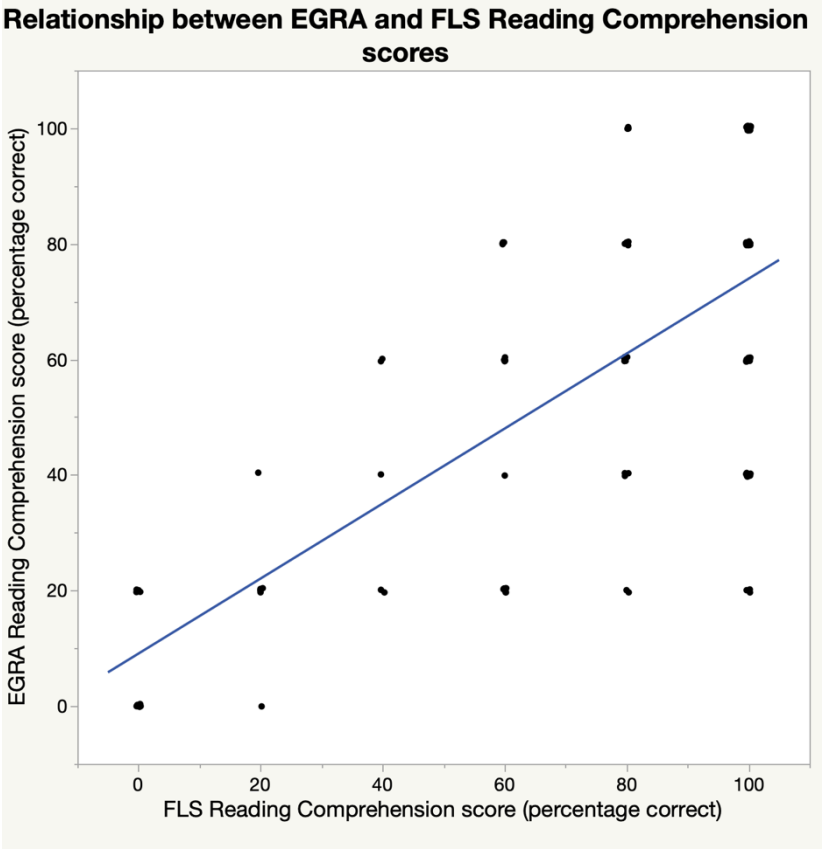


Figure 8. Relationship between EGRA and FLS Reading Comprehension scores (correlation=0.68 [CI: 0.58, 0.77])

Numeracy results

The modified skill groups for EGMA are:

- Non-/Early mathematician: Either missing number or addition (both level 1 and level 2 combined) below 30 per cent
- Emergent mathematician: Missing number and addition both above 30 per cent
- Mathematician: Missing number and addition both above 80 per cent

Table 10. Percentages of children in each of the EGRA and FLS numeracy performance levels

EGMA Numeracy Grouping	EGMA Numeracy Grouping based on the FLS instrument			
	Non-/Beginning	Emergent	Mathematician	Total
Non-/Beginning	4	22	3	29
Emergent	0	46	61	107
Mathematician	0	1	53	54
Total	4	69	117	190

Similar to the EGRA versus FLS group comparison, there are large numbers of children who did not achieve the same level of expertise on both numeracy tests (EGMA and FLS Numeracy). As Table 11 shows, of the 22 children classified as Emergent mathematician (FLS) and Non-/Beginning mathematician (EGMA), all but one performed at a level higher on the MICS addition questions compared to the EGMA. (Note that the levels are the same as in the above skill groups but are separated so that each skill –

Addition and Missing Number – can be independently evaluated. The lowest level was comprised of children answering fewer than 30 per cent of the items correctly, the upper answering more than 80 per cent correctly and the middle level scoring in between.) No such performance difference across the two measures was found in the Missing Number subtest.

Table 11. Children who achieved Emergent on MICS and Non-/Beginning on EGMA

MICS Addition	EGMA Addition	MICS Missing Number	EGMA Missing Number	Frequency
Emergent	Non-/Beginning	Emergent	Non-/Beginning	2
Emergent	Non-/Beginning	Emergent	Emergent	5
Emergent	Non-/Beginning	Mathematician	Non-/Beginning	2
Emergent	Non-/Beginning	Mathematician	Emergent	1
Mathematician	Non-/Beginning	Emergent	Emergent	11
Mathematician	Emergent	Emergent	Non-/Beginning	1
			Total	22

However, there were multiple factors for a set of 61 children who achieved Mathematician level on MICS and Emergent on EGMA, as seen in Table 12. For nearly half of this set, both of their EGMA scores were in the middle performance level while their MICS scores were in the higher performance level. The remaining half were evenly split between the children who were successful at one of these two EGMA subtests but not the other.

Table 12. Children who achieved Mathematician on MICS and Emergent on EGMA

MICS Addition	EGMA Addition	MICS Missing Number	EGMA Missing Number	Frequency
Mathematician	Emergent	Mathematician	Emergent	28
Mathematician	Emergent	Mathematician	Mathematician	15
Mathematician	Mathematician	Mathematician	Emergent	18
			Total	61

Table 13 reflects the percentage of children in each of the EGMA Numeracy performance levels who were able to successfully complete the FLS Numeracy indicators. Since the FLS indicators are binary, children in the lower EGMA groups should likely have a lower pass rate on the FLS indicator, and these pass rates should increase as the children exhibit greater mathematics skill and are placed at a higher EGMA Numeracy performance levels.

Table 13. Cross-tabulation of EGMA and FLS Numeracy performance levels

EGMA Reading Grouping	FLS Indicators: Number Identification	Number Discrimination	Addition	Missing Number	Basic numbers skills	Total count
Non-/ Beginning	25.0%	50.0%	25.0%	0.0%	0.0%	4
Emergent	40.6%	60.9%	53.6%	2.9%	0.0%	69
Mathematician	82.1%	88.9%	85.5%	72.6%	56.4%	117
Total	65.8%	77.9%	72.6%	45.8%	34.7%	190

Most children tested highly on the Number Identification task across both assessments, as shown in Figure 9. The EGMA test presents significantly more of the three-digit numbers at the end of the assessment, which may account for some of the variability in the last two columns (i.e., 80 per cent and 100 per cent on the FLS test).

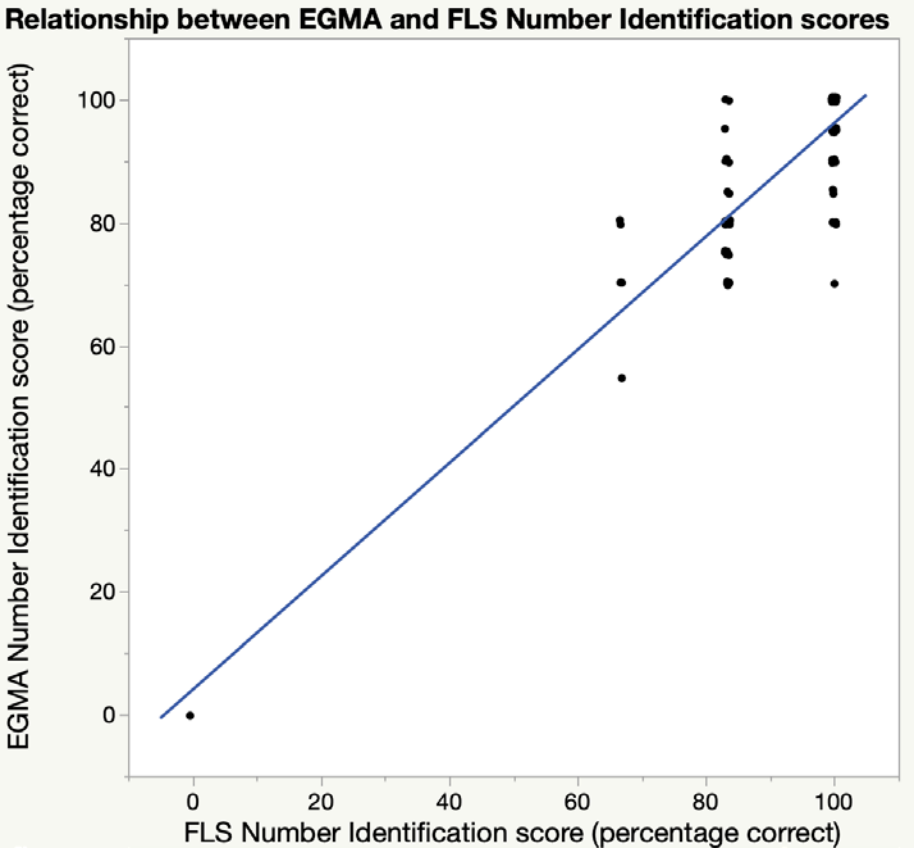


Figure 9. Relationship between EGMA and MICS FLS Number Identification scores (correlation=0.87 [CI: 0.82, 0.91])

Similar to findings in the previous subtests, the children's results tended to have a ceiling effect on the MICS FLS data in the Quantitative Comparisons subtests, as shown in Figure 10.

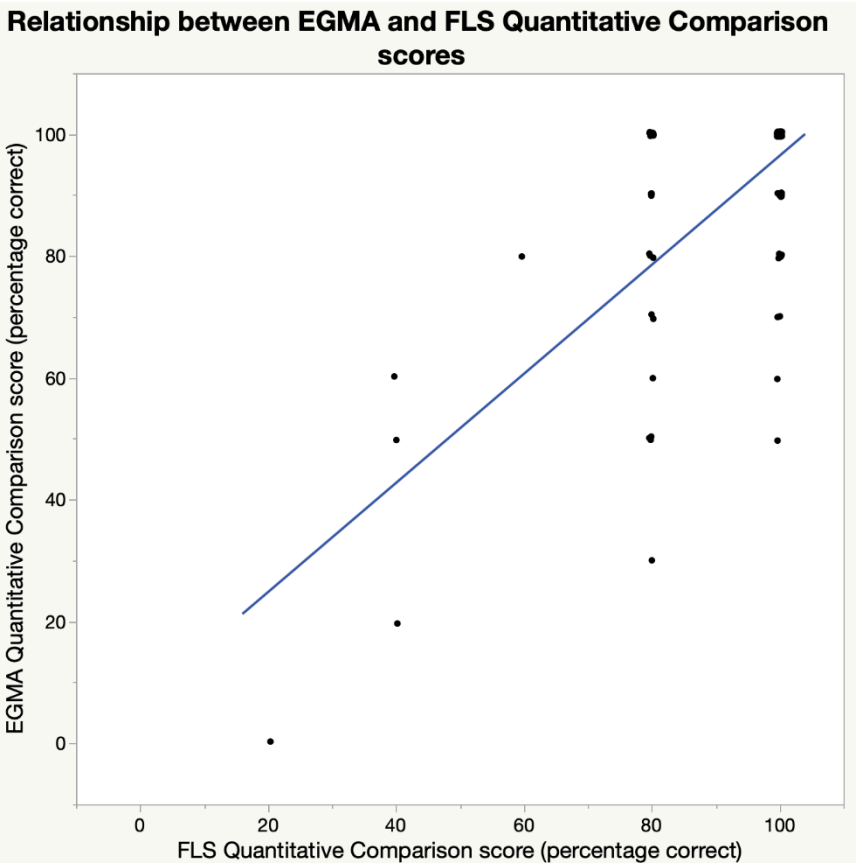


Figure 10. Relationship between EGMA and MICS FLS Quantitative Comparison scores (correlation=0.69 [CI: 0.59, 0.77])

In the Pattern Recognition subtest (i.e., Missing Number recognition) the EGMA scores were significantly lower than the MICS scores for most of the children, as shown in Figure 11. Several children were only able to answer two or fewer (of 10) EGMA questions correctly, while answering three or more questions in the FLS correctly.

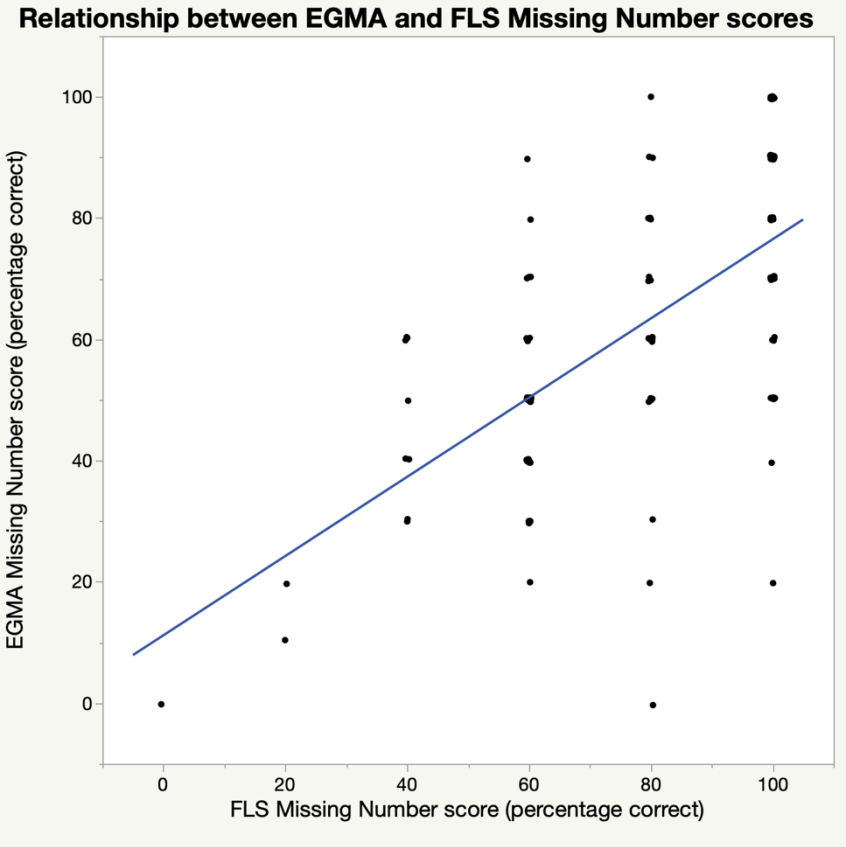


Figure 11. Relationship between EGMA and MICS FLS Missing Number scores (correlation=0.63 [CI: 0.51, 0.73])

The MICS test uses only five addition questions, whereas the EGMA uses 25 questions broken into Level 1 and Level 2 difficulty categories. The 20 Level 1 questions are related to the sums of 2 one-digit numbers, many of which have answers that are also in single digits. The five Level 2 questions are more complex, involving the addition of the two-digit numbers and carrying values to the next column. These skills are important indicators of the higher performance level in mathematical operations and thinking skills.

The FLS test uses only five questions. The first four are similar to the EGMA Addition Level 1, and the last one is similar to EGMA Addition Level 2. Because of this asymmetry in similarity across the two tests, it is difficult to compare the FLS and the EGMA.

One possible way for a fair comparison of EGMA and MICS scores might be by combining the EGMA Addition questions into one group and looking at the percentage scores. Notice that 80 per cent of the questions in the EGMA and MICS FLS Numeracy tests are of the Level 1 type and the remaining 20 per cent are of Level 2 type (i.e., higher in complexity).

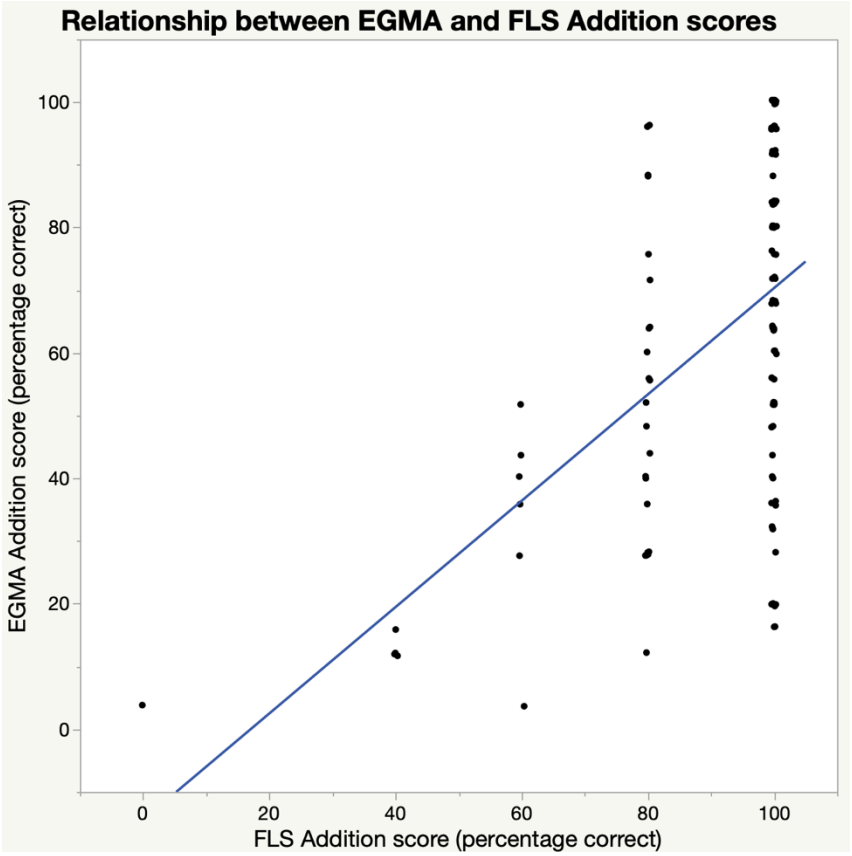


Figure 12. Relationship between EGMA and MICS FLS Addition scores (correlation=0.52 [CI: 0.38, 0.63])

Figure 12 depicts the strong ceiling effect in FLS, with over 90 per cent of children scoring 80 per cent or more on the test, compared to only 44 per cent doing so in EGMA. There is also a relatively large distribution of EGMA Addition scores for the children who answered all of the FLS questions correctly, ranging from 20 per cent to 100 per cent. Variability of this sort shows that FLS test items may need to be more difficult to draw a proper comparison to EGMA.

Table 14. Summary of results from reliability and validity analyses

	Number of items	Reliability		Concurrent validity: correlation (Pearson's r)
		Inter-rater reliability (PNA)	Internal consistency (Cronbach's alpha)	
Reading				
Oral Reading Accuracy of connected text	43	0.76	0.91	0.72
Reading Comprehension	5	0.88	0.83	0.68
Numeracy				
Number Identification	6	0.94	0.54	0.87
Quantitative Comparison	5	0.94	0.61	0.69
Addition	5	0.90	0.58	0.52
Missing Number	5	0.90	0.62	0.63

5

Conclusions

This paper presents results of a concurrent validity study of the new FLS module as compared to EGRA/EGMA, as well as an inter-rater reliability study. In addition, internal consistency results for the new FLS module are presented. Findings from this study provide evidence on the FLS instrument's validity and reliability and shed light on areas for further improvement when the instrument is administered.

Correlations with external criteria – measures of similar constructs such as EGRA or EGMA – are high and provide support regarding the concurrent validity of the instrument. Inter-rater reliability estimates also range from very good to excellent. It will be instructive to evaluate the instrument's internal consistency (i.e., Cronbach's alpha) – which was not the central focus of this study – using a larger sample size before making definitive conclusions.

Internal consistency, as measured by Cronbach's alpha, shows some variation across the two assessment domains, namely reading and numeracy. For the reading domain, Cronbach's alpha is high for both Reading Accuracy (.91) and Comprehension (.83) subdomains, and hence even higher for the composite reading score, which combines Accuracy and Comprehension (.92) subdomains. For the numeracy domain, Cronbach's alpha was acceptable for the composite (.76) but lower than desired for some of the specific subtests, such as Number Identification (.54). This may have been due to scoring practices that perhaps were too stringent, particularly for the three-digit numbers. In summary, internal consistency is overall acceptable, but some of the tasks may require further attention during training of interviewers to avoid overly stringent scoring practices (particularly in the Number Identification subtest).

Inter-rater reliability, on average, is high. The lowest estimated value was for the Reading Accuracy (.76) component, followed by Reading Comprehension (.88). On the numeracy domain, the kappa values for all tasks were .90 or higher. Therefore, the priority for improvement during the training should be on the Reading Accuracy. Specifically, interviewers, supervisors and trainers should reach an agreement about which pronunciations are deemed acceptable for words in the story. This should help improve inter-rater reliability for this task.

Concurrent validity with EGRA and EGMA was evaluated in several ways, including by using correlation coefficients, which ranged from .52 to .87, all of which were statistically significant. Considering that the FLS module does not take exactly the same approach to the measurement of these competencies as EGRA and EGMA, we should not expect these correlation coefficients to be substantively higher. Specifically, as we noted above, EGRA and EGMA are mostly timed and focus on fluency, whereas FLS is

not timed and does not focus on fluency. Therefore, a moderate amount of variation across the results of the two instruments is to be expected, which we confirmed.

In summary, the new module is a reliable measure with strong support for its validity. As long as the FLS module is properly administered, one should be able to obtain the intended outcome. Based on the results of this study, attention should be paid to training interviewers in scoring practices, and particularly in the (1) Number Identification (especially for three-digit numbers), (2) Missing Number and (3) Reading Accuracy components.

6

References

1. See <mics-surveys-prod.s3.amazonaws.com/MICS6/West%20and%20Central%20Africa/Sierra%20Leone/2017/Survey%20findings/Sierra%20Leone%202017%20MICS%20Survey%20Findings%20Report_English.pdf>, accessed 24 April 2019.
2. McHugh, Mary L., 'Interrater Reliability: The kappa statistic', *Biochemia Medica*, vol. 22, no. 3, 2012, pp. 276-82.
3. Hattori, Hiroyuki, Manuel Cardoso and Blandine Ledoux, 'Collecting Data on Foundational Learning Skills and Parental Involvement in Education', MICS Methodological Papers, no. 5, Data and Analytics Section, Division of Data, Research and Policy, UNICEF New York, 2017.
4. Rigney, Daniel, *The Matthew Effect: How advantage begets further advantage*, Columbia University Press, 2010.
5. Wilson, Mark and Perman Gochyyev, 'Psychometrics', ch. 1 in *Handbook of Quantitative Methods for Educational Research*, edited by Timothy Teo, Sense, Rotterdam, 2013, pp. 3-30.
6. Cohen, Jacob, 'A Coefficient of Agreement for Nominal Scales', *Educational and Psychological Measurement*, vol. 20, no. 1, 1960, pp. 37-46.
7. Feinstein, Alvan R. and Domenic V. Cicchetti, 'High Agreement but Low Kappa: I. The problems of two paradoxes', *Journal of Clinical Epidemiology*, vol. 43, no. 6, 1990, pp. 543-549.
8. Cicchetti, Domenic V. and Alvan R. Feinstein, 'High Agreement but Low Kappa: II. Resolving the paradoxes', *Journal of Clinical Epidemiology*, vol. 43, no. 6, 1990, pp. 551-558.
9. Brennan, Robert L. and Dale J. Prediger, 'Coefficient Kappa: Some uses, misuses, and alternatives', *Educational and Psychological Measurement*, vol. 41, no. 3, pp. 687-699.
10. Gwet, Kilem Li, *Handbook of Inter-rater Reliability*, STATAXIS Publishing Company, Gaithersburg, 2002.
11. Ebel, Robert L., *Measuring Educational Achievement*, Prentice-Hall, Englewood Cliffs, 1965.
12. See <www.rti.org/sites/default/files/resources/rti-publication-file-ee6f12c7-a962-4285-b4a1-4e139eb12545.pdf>, <shared.rti.org/content/early-grade-mathematics-assessment-egma-toolkit> and <https://ierc-publicfiles.s3.amazonaws.com/public/resources/Jordan%20RAMP%20LQAS%20Summary%20Report%20FINAL.pdf>, all accessed 2 May 2019.
13. Ebel, 1965.

7

Appendices

Appendix 1: Instruments used

MICS

CHILD LEARNING QUESTIONNAIRE

Kenya Concurrent Validity Study 2016

Schoolchildren – Assessor Version – Literacy – MF

CHILD LEARNING INFORMATION PANEL	**CL**
This questionnaire is to be administered to the selected child age 7-14 years in school, as identified by the field test team.	
CL1. Village / locality _____	**CL2**. Interviewer's name and number: Name..... ___ ___
CL3. School name: Name_____	**CL4**. School code: ..... ___ ___ ___
CL5. Child's name: Name_____	**CL6**. Child number: .....___ 0 ___
CL7. Role ☐ Team Assessor ☐ Solo Assessor ☐ Rater	**CL8**. Day / Month / Year of interview: ___ ___ /___ ___ / 2 0 1 ___

CL11. Record the time the interview with the child started.	Hour and minutes ..... __ __ : __ __

It is important to establish a playful and relaxed atmosphere with the child, using some simple initial conversation among topics of interest to the child (icebreakers). Ideally the child should think of the assessment as a game to be enjoyed. Be friendly to put the child at ease. Start by greeting the child and saying

JINA LANGU NI (*na jina lako ni nani*). **SASA NINGEPENDA KUKUAMBIA KUHUSUS MIMI.** [*Mimi ninatoka _____ /Idai na miaka ya watoto/Mambo wanapenda kufanya na kadhalika.]*

JE, UNAWEZA KUNIAMBIA MAMBO MACHACHE KUHUSU WEWE

If child is reluctant continue with an icebreaker such as

JE, UNAPENDA KUFANYA NINI WAKATI WAKO WA KUPUMZIKA?
JE, JINA LAKO LINA MAANA GANI?
JE, WAKATI GANI KATIKA SIKU YAKO UNAPENDA SANA?

When the child is comfortable continue with the verbal consent

Begin the verbal consent by saying to the child

WACHA NIKWAMBIE KWA NINI NIMEKUJA HAPA LEO. MIMI NINATOKA *WOMEN EDUCATIONAL RESEARCHERS OF KENYA*. NA MIMI NI MMOJA WA WATU AMBAO WANAZUNGUMZA NA WATOTO KUHUSU MASOMO YAO NA KUWAULIZA KUFANYA MAZOEZI YA KUSOMA NA YA HESABU. IKIWA UNGEPENDA KUTUSAIDIA KUFANYA MAZOEZI HAYA, NITAKUULIZA MASWALI MACHACHE NA KUKUPA MAZOEZI KIDOGO AMBAYO UTAFANYA. NITAKUELEZA KUHUSU KILA ZOEZI WAKATI TUNATAKA KULIFANYA NA UNAWEZA KUNIULIZA SWALI WAKATI WOWOTE. HUHITAJI KUFANYA ZOEZI LOLOTE AMBALO HUTAKI KUFANYA. MARA TUNAPOANZA, IKIWA HUTAKI KUJIBU SWALI AU HUTAKI KUENDELEA NI SAWA. JE, UKO TAYARI KUANZA?

☐ *Yes, consent is given ⇨ Continue.*

☐ *No, consent is not given ⇨ Circle '04' in CL12. Discuss this result with your supervisor.*

CL12. Result of interview with selected child (7-14 years).	Completed ..... 01 Child refused ..... 04 Partly completed ..... 05 Incapacitated ..... 06 Other (*specify*) _____ 96

CHILD INFORMATION MODULE — CI

CI1. MARK THE GENDER OF THE CHILD.	Male .....1 Female 2	
CI2. JE, ULIZALIWA LINI? *If response is Don't Know record 98 for Month & Year*	Month .....__ __ Year.....2 0 __ __	
CI3. JE, UNA MIAKA MINGAPI? *Record in completed years* *Record 98 for Don't Know*	Age.....__ __	
CI4. JE UNAENDA SHULE GANI?	Pre-school .....1 Primary.....2 Secondary .....3 Vocational (post primary) .....4 Alternative education program 5	
CI5. JE, UKO DARASA GANI?	Grade ……____ Don't know …… 98	

FOUNDATIONAL LEARNING SKILLS — FL

READING SKILLS

FL1. NINGEPENDA TUZUNGUMZE KUHUSU KUSOMA. [A] JE, HUWA UNASOMA VITABU NYUMBANI?	Yes No	

[B] JE, KUNA MTU AMBAYE HUKUSOMEA VITABU UKIWA NYUMBANI?	Reading books at home.....1 2 Read to at home 1 2	
FL2. JE, NI LUGHA GANI WEWE HUZUNGUMZA UKIWA NYUMBANI MARA NYINGI? NI LUGHA IPI (*chagua kutoka kwa orodha ya lugha hapa*)?	Kiswahili.....1 Kikuyu .....2 Dholuo .....3 English……4 Other.....7 Don't know 8	
FL3. JE, MARA NYINGI UKIWA SHULENI, WALIMU WAKO WANATUMIA LUGHA GANI WAKIFUNDISHA DARASANI? NI LUGHA IPI? (*chagua moja kutoka kwa orodha ya lugha hapa*)	Kiswahili.....1 Kikuyu .....2 Dholuo .....3 English……4 Other.....7 Don't know 8	

KISWAHILI READING & COMPREHENSION

Give the child the Reading & Number *Book, open to page 1.*

Is the child aged 9 years or younger? Check the answers to CI3.

- ☐ *9 years or younger (CI3) ⇨ Continue*
- ☐ *Other ⇨ Turn the page to the reading passage. Go to FL5*

Open the page showing the reading practice item and say:

SASA TUTASOMA KIDOGO. NINGEPENDA WEWE USOME SENTENSI HIZI KWA SAUTI. (*ukimwonyesha mtoto setensi*)**. UKIMALIZA KUSOMA NITAKUULIZA SWALI.**

Tina ni kuku. Tina ana miaka 6.

If the child does not start, you can read the first word, monitor the child's reaction, and continue to read only if needed.

If the child starts reading confidently, let the child continue.

If the child seems to prefer reading along with you, the two of you can read together.

Once the reading is done (by the child, the interviewer, or both),say:

TINA ANA MIAKA MINGAPI?

If the child does not answer after 5 seconds,, say:

TINA ANA MIAKA SITA

If the child does or does not answer, say:

ASANTE. SASA NINGEPENDA USOME SEHEMU HII WEWE MWENYEWE

Turn the page to reveal the reading passage and go to FL5.

FL5. **HAPA KUNA HADITHI FUPI NINGEPENDA USOME HADITHI HII KWA SAUTI. JITAHIDI KUSOMA VIZURI KABISA. UTAANZA KUSOMA KUANZIA HAPA.** (*onyesha mtoto neno la kwanza kwenye mstari wa kwanza*) **NA UTASOMA MSTARI BAADA YA MSTARI** (*onyesha mtoto mstari wa kusoma kwa kuelekeza na kidole*). **WAKATI UTAMALIZA KUSOMA NITAKUULIZA MASWAL KUHUSU MAMBO AMBAYO UMESOMA. NIKISEMA 'ANZA' SOMA HADITHI VIZURI KABISA. KAMA UTAONA NENO AMBALO HUJUI KUSOMA NENDA KWA NENO LINALOFUATA. WEKA KIDOLE CHAKO KWA NENO LA KWANZA? HAYA UKO TAYARI? ANZA .**

Mark any incorrect or missed words by a '/' through the number underneath the word.

STOP RULE 1: No correct words in the first line. Say **ASANTE. HIYO NI SAWA. TUENDELEE.**

STOP RULE 2: Not near end after 3 minutes Say **ASANTE. HIYO NI SAWA. TUENDELEE.**

Mark the final word with a bracket ']' after its number.

Maria	ana	miaka	saba.	Siku	moja,
1	2	3	4	5	6
Nyanyake	alimtuma	sokoni	kununua	karoti.	Alimpa
7	8	9	10	11	12
pesa.	Maria	aliziweka	mkobani	mwake.	Mkoba
13	14	15	16	17	18
Huo	ulikuwa	na	shimo	kubwa.	Alipokuwa
19	20	21	22	23	24
njiani,	Maria	alizipoteza	pesa.	Juma	aliziona
25	26	27	28	29	30
Pesa	hizo	na	akampa	Maria.	Alifurahi
31	32	33	34	35	36
sana.	Maria	alimshukuru	Juma	na	akaenda
37	38	39	40	41	42
sokoni.					
43					

Total number of words attempted..... __ __

Total number of words incorrect or missed .....__ __

FL6. *How well did (name) read the story?*

The child read at least one word correct.....1	
The child did not read any word correctly.....2	2⇨FL12
The child did not try to read the story 3	3⇨FL12 4⇨FL12

	Refusal 4	
Say **SASA NITAKUULIZA MASWALI KIDOGO KUHUSU HADITHI UMESOMA**		
Make sure the child can still see the passage. Say FL7. **MARIA ANA MIAKA MINGAPI?** How old is Maria? [Maria/She is 7 years old/ seven] *If the child does not provide a response after 5 seconds, repeat the question. If no response after a further 10 seconds, mark 'No response'. Say* THANK YOU. THAT IS OK. WE WILL MOVE ON.	Correct [Maria ana miaka saba / saba] ..... 1 Incorrect ..... 2 No response / Says 'I don't know' 3	
FL8. **NANI ALIYEMTUMA MARIA SOKONI?** Who sent Maria to the market? [Maria was sent to the market by her grandmother] *If the child does not provide a response after 5 seconds, repeat the question. If no response after a further 10 seconds, mark 'No response'. Say* THANK YOU. THAT IS OK. WE WILL MOVE ON.	Correct [nyanyake alimtuma Maria sokoni / nyanyake] ..... 1 Incorrect ..... 2 No response / Says 'I don't know' 3	
FL9. **MARIA ALITUMWA KUNUNUA NINI?** What was Maria asked to buy? [She was sent to buy carrots / carrots] *If the child does not provide a response after 5 seconds, repeat the question. If no response after a further 10 seconds, mark 'No response'. Say* THANK YOU. THAT IS OK. WE WILL MOVE ON.	Correct [Alitumwa kununua karoti / karoti] .... 1 Incorrect ..... 2 No response / Says 'I don't know' 3	
FL10. **KWA NINI MARIA ALIPOTEZA PESA?** Why did Maria lose the money? [Because it fell through the hole in the bag / The bag had a hole] *If the child does not provide a response after 5 seconds, repeat the question. If no response after a further 10 seconds, mark 'No response'. Say* THANK YOU. THAT IS OK. WE WILL MOVE ON.	Correct [Kwa sababu mkoba wake ulikuwa na shimo kubwa.] ..... 1 Incorrect ..... 2 No response / Says 'I don't know' 3	
FL11. **KWA NINI MARIA ALIMSHUKURU JUMA?**	Correct [Kwa sababu alimrudishia pesa zake.] 1	

Why did Maria thank Juma? [Because he gave her back the money] *If the child does not provide a response after 5 seconds, repeat the question. If no response after a further 10 seconds, mark 'No response'. Say* THANK YOU. THAT IS OK. WE WILL MOVE ON.	Incorrect ..... 2 No response / Says 'I don't know' 3	

Task 3 : Syllable Sounds	🕮 **Page 3**	🕐 ***60 seconds***
	✋ If you marked as incorrect all of the answers on the first line, say "Thank you!" and discontinue the exercise.	➲ If the child hesitates for 3 seconds, point to the next letter and say "Please go on." Mark the skipped letter as incorrect.

General Instructions:

The answer is "correct" if the student gives the name of the letter.

✍ (/) Cross out each item for which the student has given an incorrect answer.

(O) Circle the item if the student self-corrects.

( ]) After the last item read.

Karatasi hii ina silabi mbali mbali. Tafadhali zitamke silabi zote unazozijua.
Kwa mfano, silabi hii *[kisha mwonyeshe silabi]* **ni "Ja"**

Hebu tufanye mazoezi: Nitamkie silabi hii *[mwonyeshe silabi Ni]:*
Iwapo jawabu la mwanafunzi ni sahihi, sema : **Vyema, silabi hii ni "Ni"**
Iwapo jawabu la mwanafunzi sio sahihi, sema: **Silabi hii ni "Ni"**

Sasa, hebu jaribu silabi nyingine: nitamkie silabi hii *[mwonyeshe silabi Ku]:*
Iwapo jawabu la mwanafunzi ni sahihi, sema: **Vyema, silabi hii ni "Ku."**
Iwapo jawabu la mwanafunzi sio sahihi, sema: **Silabi hii ni "Ku."**

Je, umeelewa unavyopaswa kufanya?

Nikisema "Anza", tafadhali zitamke silabi hizi haraka iwezekenavyo lakini kwa makini. Nitamkie silabi, kuanzia hapa kisha kuendelea hivi. *[Elekeza kidole chako katika silabi ya kwanza katika mstari wa juu baada ya mfano kisha uendelee hadi mwisho wa mstari huo].* **Nitanyamaza nikusikilize. Uko tayari? Anza.**

Mifano: ja ni ku

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	
pa	nya	mbi	sa	ti	ho	ha	ii	wa	ja	*(10)*
ka	yu	da	ba	la	me	ye	cho	mu	te	*(20)*
pe	ndi	de	o	za	je	bi	mwa	fu	msi	*(30)*
no	ua	pi	fi	se	bwa	cha	li	sha	bu	*(40)*
so	ji	mwe	he	ko	di	ra	vu	ru	do	*(50)*
nda	nga	re	hu	we	nyu	to	na	mi	mbe	*(60)*
su	nzi	ku	ne	ri	tu	ma	au	fa	mba	*(70)*

aa	wi	hi	du	zo	po	si	yo	shi	ga	*(80)*
nu	mo	nye	ni	nde	le	ju	zu	gu	vi	*(90)*
be	mto	che	ke	zi	ya	ki	go	ngu	ngi	*(100)*

Muda uliosalia katika saa ya kupima kasi kufikia mwisho wa kusoma (idadi ya SEKUNDE) : ☐

Tia alama katika kisanduku hiki iwapo shughuli ya kusoma ilisitishwa kwa sababu mwanafunzi hakupata jawabu sahihi katika mstari wa kwanza. ☐

Task 4a : Oral Reading Passage	🕮 Page 4	🕐 *60 seconds*
	🖐 If you marked as incorrect all of the answers on the first line, say "Thank you!" and discontinue the exercise.	➲ If the child hesitates for 3 seconds, point to the next word and say "Please go on." Mark the skipped word as incorrect.

✍ (/) Cross out each item for which the student has given an incorrect answer.
(O) Circle the item if the student self-corrects.
( ]) After the last item read.

When the student has finished reading, ***remove the text of their possession***.

Task 4b : Reading Comprehension	🕮 ×	🕐 ×
	🖐 Ask questions until the last line that the child has completed.	➲ If the child remains silent after 10 SECONDS repeat the question and give the child another 5 seconds, then mark No Response.

The correct answers can be provided in English or any other language. Ask question
until the last line that the child has completed.

✍ (✓) Correct, Incorrect or No Response

[Put the text in front of the student. Point to the first line.] **Hii hapa ni hadithi fupi. Ningependa uisome kwa sauti, haraka iwezekanavyo lakini kwa makini. Ukimaliza kuisoma, nitakuuliza maswali kuhusu yale uliyosoma. Je, umeelewa jinsi unavyopaswa kufanya? Nikisema "Anza," isome hadithi vizuri kadri ya uwezo wako. Nitanyamaza nikusilikilize. Uko tayari? Anza.**		**Sasa nitakuuliza maswali machache kuhusu hadithi ambayo umesoma. Jaribu kujil maswali vyema kabisa.** When the student has finished reading, ***remove the text of their possession***.			
		QUESTIONS	**Student Responses**		
			Correct	**Incorrect**	**No Respo**
Kamau anapenda kucheza mpira wa miguu.	6	**Kamau anapenda kucheza mchezo gani?** [mpira wa miguu]			
Yeye ni mchezaji hodari. Wiki iliyopita, timu yao ilicheza na timu jirani.	18	**Timu ya Kamau ilicheza na timu gani?** [Ilicheza na timu jirani, timu ya jirani]			
Kamau alifunga mabao matatu. Alishangiliwa na kupewa zawadi. Juzi alipokuwa akifanya mazoezi, aliumia	31	**Ni nini kilifanyika kwa Kamau alipokuwa akifanya mazoezi?** [Aliumia, aliumia mguu, aliumizwa mguu]			
mguu. Alipelekwa hospitalini. Daktari alimwambia apumzike	37	**Daktari alimwambia Kamau afanye nini?** [Apumzike/apumzike kwa siku mbili]			
kwa muda wa siku mbili. Kamau alihuzunika sana. Alifikiri angekosa kushiriki katika mechi ya Jumamosi. Ijumaa aliporudi hospitalini, daktari alimwambia mguu umepona.	59	**Kwa nini Kamau alifurahi kabisa?** [kwa sababu aliambiwa amepona / sasa angecheza mpira tena]			
Kamau alifurahi kabisa.					

Check this box if the exercise was discontinued because the child had no correct answers in the first line.	

Time remaining on stopwatch at completion (number of SECONDS):	

Task 5 : Listening Comprehension	🕮 ×	🕐 ×
	✋ ×	➲ If the child remains silent after 10 SECONDS repeat the question and give the child another 5 seconds, then mark No Response.

General Instructions:

You will read aloud a story **once**, then ask students some comprehension questions.

The correct answers can be provided in any language.

✍ (✓) Correct, incorrect or no response

[Tell the student :]

Hii hapa ni hadithi fupi. Nitaisoma kwa sauti. Nitaisoma mara moja tu. Halafu nitakuuliza maswali. Tafadhali sikiliza kwa makini kisha ujaribu kujibu maswali. Je, umeelewa jinsi unavyopaswa kufanya? Uko tayari? Naanza.

Recho anapenda kusukwa nywele. Nywele zake ni ndefu na za kupendeza. Siku moja, rafiki yake akaja kuwatembelea. Recho hakujua kuwa ana chawa kichwani. Siku chache baadaye, Recho akaanza kujikuna kichwani. Mama akasema ana chawa. Recho akahuzunika. Lakini mama akaleta dawa.

QUESTIONS	Correct answers (DO NOT READ TO THE STUDENT)	STUDENT RESPONSES		
		Correct	Incorrect	No Response
Recho anapenda kufanya nini?	*[kusuka nywele]*			
Je, nani alimtembelea Recho?	*[Rafiki yake]*			
Je, nini kilisababisha Recho kujikuna kichwani?	*[Nywele iliingiwa na chawa]*			
Recho alihisi vipi baada ya kupatikana na chawa?	*[Alihuzunika, aliaibika] [Alisikia kujikuna]*			
Unafikiri mama alitoa dawa wapi?	*[Kwa daktari, dukani, hospitali, kwa jirani, kwa nyumba. Na mangineo]*			

CL13. Record the time when the literacy activities were completed.	Hour and minutes..... __ __ : __ __

To end the interview:

Thank the child for their participation and check through the entire questionnaire to ensure that no information is missing.

Remember to ask the child not to tell anyone about the specific details of the reading and number activities the child was asked to complete.

OBSERVATIONS

MICS

CHILD LEARNING QUESTIONNAIRE

Kenya Concurrent Validity Study 2016

School children – Assessor Version – Numeracy – MF

CHILD LEARNING INFORMATION PANEL	**CL**
This questionnaire is to be administered to the selected child age 7-14 years in school, as identified by the field test team.	
CL1. Village / locality _____	**CL2**. Interviewer's name and number: Name..... ___ ___
CL3. School name: Name_____	**CL4**. School code: ..... ___ ___ ___
CL5. Child's name: Name_____	**CL6**. Child number: .....___ 0 ___
CL7. Role ☐ Team Assessor ☐ Solo Assessor ☐ Rater	**CL8**. Day / Month / Year of interview: ___ ___ /___ ___ / 2 0 1 ___

CL11. Record the time the interview with the child started.	Hour and minutes ..... __ __ : __ __

It is important to establish a playful and relaxed atmosphere with the child, using some simple initial conversation among topics of interest to the child (icebreakers).Ideally the child should think of the assessment as a game to be enjoyed. Be friendly to put the child at ease. Start by greeting the child and saying

JINA LANGU NI (*na jina lako ni nani*). **SASA NINGEPENDA KUKUAMBIA KUHUSUS MIMI.** [*Mimi ninatoka _____ /Idai na miaka ya watoto/Mambo wanapenda kufanya na kadhalika.]*

JE, UNAWEZA KUNIAMBIA MAMBO MACHACHE KUHUSU WEWE

If child is reluctant continue with an icebreaker such as

JE, UNAPENDA KUFANYA NINI WAKATI WAKO WA KUPUMZIKA?

JE, JINA LAKO LINA MAANA GANI?

JE, WAKATI GANI KATIKA SIKU YAKO UNAPENDA SANA?

When the child is comfortable continue with the verbal consent

Begin the verbal consent by saying to the child

WACHA NIKWAMBIE KWA NINI NIMEKUJA HAPA LEO. MIMI NINATOKA *WOMEN EDUCATIONAL RESEARCHERS OF KENYA*. NA MIMI NI MMOJA WA WATU AMBAO WANAZUNGUMZA NA WATOTO KUHUSU MASOMO YAO NA KUWAULIZA KUFANYA MAZOEZI YA KUSOMA NA YA HESABU. IKIWA UNGEPENDA KUTUSAIDIA KUFANYA MAZOEZI HAYA, NITAKUULIZA MASWALI MACHACHE NA KUKUPA MAZOEZI KIDOGO AMBAYO UTAFANYA. NITAKUELEZA KUHUSU KILA ZOEZI WAKATI TUNATAKA KULIFANYA NA UNAWEZA KUNIULIZA SWALI WAKATI WOWOTE. HUHITAJI KUFANYA ZOEZI LOLOTE AMBALO HUTAKI KUFANYA. MARA TUNAPOANZA, IKIWA HUTAKI KUJIBU SWALI AU HUTAKI KUENDELEA NI SAWA. JE, UKO TAYARI KUANZA?

☐ *Yes, consent is given ⇨ Continue.*

☐ *No, consent is not given ⇨ Circle '04' in CL12. Discuss this result with your supervisor.*

CL12. Result of interview with selected child (7-14 years).	Completed..... 01 Child refused..... 04 Partly completed ..... 05 Incapacitated..... 06 Other (*specify*) _____ 96

CHILD INFORMATION MODULE		**CI**
CI1. MARK THE GENDER OF THE CHILD.	Male .....1 Female 2	
CI2. JE, ULIZALIWA LINI? *If response is Don't Know record 98 for Month & Year*	Month .....__ __ Year.....2 0 __ __	
CI3. JE, UNA MIAKA MINGAPI? *Record in completed years* *Record 98 for Don't Know*	Age.....__ __	
CI4. JE UNAENDA SHULE GANI?	Pre-school .....1 Primary.....2 Secondary .....3 Vocational (post primary) .....4 Alternative education program.....5	
CI5. JE, UKO DARASA GANI?	Grade ……____ Don't know …… 98	

NUMBER SKILLS		
Turn the page in the Reading & Numbers Book so the child is looking at the list of numbers. Make sure the child is looking at this page Say *FL12.* **SASA HAPA KUNA NAMBARI. NINGEPENDA UONYESHE KWA KIDOLE CHAKO KILA NAMBARI NA UNIAMBIE NI NAMBARI GANI.** *Point to the first number and say* **ANZA HAPA.** *If a child stops on a number for more than 10 seconds, tell the child what the number is, mark the number as missed, point to the next number and say,* **JE, HII NI NAMBARI GANI?** *Mark any number the child misses or reads incorrectly by putting a '/' through the number.* *STOP RULE: 2 'No attempts' Say* **ASANTE. HIYO NI SAWA. TUENDELE.** *Go to next activity.*	9, 12, 30, 48, 74, 731 All numbers correct.....1 One error.....2 Two errors.....3 No attempt 4	

Turn the page so the child is looking at the first pair of numbers. Make sure the child is looking at this page. Say *FL13.* **ANGALIA NAMBARI HIZI. NIAMBIE NAMBARI GANI NI KUBWA KULIKO ZINGINE. NIAMBIE NAMBARI HIYO.** *Record the child's answer before turning the page in the book and repeating the question for the next pair of numbers.* *If the child does not provide a response after 5 seconds, repeat the question. If no response after a further 10 seconds, mark an 'X' for the answer on the appropriate row on the questionnaire.* *<u>STOP RULE:</u> 2 'No attempts'. Say* **ASANTE. HIYO NI SAWA. TUENDELEE.** *Nenda kwenye zoezi linalofuata.*	7 5 _____ (7) 11 24 _____ (24) 58 49 _____ (58) 65 67 _____ (67) 146 154 _____ (154) All answers correct..... 1 One error..... 2 Two errors ..... 3 No attempt 4	
Give the child a pencil and paper. Turn the page so the child is looking at the first addition. Make sure the child is looking at this page. Say *FL14.* **ANGALIA HESABU HII. JE, MAJIBU YAKE NI NINI** (*nambari moja ongeza nambari nyingine*)**? NIAMBIE JIBU. UNAWEZA KUTUMIA PENSELI NA KARATASI KUKUSAIDIA KUFANYA HESABU.** *Record the child's answer before turning the page in the book and repeating the question for the next sum.* *If the child does not provide a response after 5 seconds, repeat the question. If no response after a further 10 seconds, mark an 'X' for the answer on the appropriate row on the questionnaire.* *<u>STOP RULE:</u> 2 'No attempts' Say* **ASANTE, HIYO NI SAWA. TUENDELEE.** *Go to next activity.*	3 + 2 = _____ (5) 8 + 6 = _____ (14) 7 + 3 = _____ (10) 13 + 6 = _____ (19) 12 + 24 = _____ (36) All answers correct..... 1 One error..... 2 Two errors..... 3 No attempt 4	

Turn the page to the practice sheet for missing numbers. Say

HAPA KUNA NAMBARI FULANI. 1, 2, NA 4. NAMBARI GANI HUJA HAPA?

If the child answers <u>correctly</u> say:

HII NI SAHIHI, 3. HEBU TUFANYE HESABU NYINGINE.

If the child answers <u>incorrectly</u> say:

NAMBARI YA 3 HUJA HAPA. SEMA NAMBARI HIZI NA MIMI. [*point to each number*]

1, 2, 3, 4. 3 HUJA HAPA. TUFANYE HESABU NYINGINE.

Now turn the page to the next practice sheet. Say

HAPA KUNA NAMBARI ZAIDI. 5, 10, 15 NA ____. JE, NI NAMBARI GANI HUJA HAPA?

If the child answers <u>correctly</u> say:

SAHIHI, 20 NI SAHIHI. SASA NINGEPENDA UJARIBU ZOEZI HILI WEWE MWENYEWE.

If the child answers <u>incorrectly</u> say:

IKIWA NAMBARI YA 20 HUJA HAPA. SEMA NAMBARI HIZI NA MIMI. [*onyesha kila nambari kwa kutumia kidole*] **5, 10, 15, 20. 20 HUJA HAPA. SASA NINGEPENDA UJARIBU ZOEZI HILI WEWE MWENYEWE**

Now turn the page in the Reading & Numbers Book with the first missing number activity. Say

FL15. **HAPA KUNA NAMBARI ZAIDI. NIAMBIE NI NAMBARI IPI HUJA BAADA YA NAMBARI HIZI.** *Record the child's answer before turning the page in the book and repeating the question.* *If the child does not provide a response after 5 seconds, repeat the question. If no response after a further 10 seconds, mark an 'X' for the answer on the appropriate row on the questionnaire.* *STOP RULE: 2 'No attempts'. Say* **ASANTE. HIYO NI SAWA. TUMEMALIZA MAZOEZI HAYA.**	5 6 7 _____ (8) 14 15 _____ 17 (16) 20 _____ 40 50 (30) 2 4 6 _____ (8) 5 8 11 _____ (14) All answers correct..... 1 One error..... 2 Two errors..... 3 No attempt 4	

Task 5: Number Identification - EXERCISE | 🕮 **Sheet 23** | 🕐 60 seconds (Timed)

🗣 **Hapa pana nambari kadhaa. Nataka uonyeshe kila nambari kwa kidole na uniambie ni nambari gani. Nitakuhesabia wakati, nitakueleza wakati wa kuanza na wakumalizia.**

[Onyesha nambari ya kwanza kwa kidole] **Anza hapa.** *[Teleza mkono kutoka upande wa kushoto hadi wa kulia].* **Je, uko tayari?... Anza.**

✍ (/) Cross out each item for which the student has given an incorrect answer.
(O) Circle the item if the student self-corrects.
( ]) After the last item read.

2	9	0	17	23
31	55	49	11	20
95	73	46	87	64
121	403	300	711	919

✍ Record time left (seconds):

✋ (Stop)
- If the time runs out (60 seconds), Tablet will turn red.

➲ (Move on)
- If a child stops on a number for **5 SECONDS,** mark as wrong then prompt pupil to move on.

Task 6: Number Discrimination – PRACTICE | 🕮 **Sheet 24** | 🕐 × (Not Timed)

P1:

🗣 **Tazama nambari hizi. Niambie ni nambari gani kubwa?**

8 4

✓🗣 **Sahihi! 8 ndio kubwa. Tujaribu mfano mwingine.**

×🗣 **Nambari kubwa ni 8.** *[elekeza kidole kwa kwa 8].* **Hii ni 8.** *[elekeza kidole kwa 4].* **Hii ni 4. '8' ni kubwa kuliko '4'. Tujaribu mfano mwingine.**

P2:

🗣 ***Tazama nambari hizi. Niambie ni nambari gani kubwa?***

12 22

✓🗣 ***Ndivyo! 22 ni kubwa. Ebu tuendelee.***

×🗣 **Nambari kubwa ni 22.** *[Elekeza kidole kwa 22].* **Hii ni 12.**

[Elekeza kidole kwa 12]. **22 ni kubwa kuliko 12. Ebu tuendelee.**

Task 6: Number Discrimination - EXERCISE | 🕮 **Sheets 24 & 25** | 🕐 **× (Not Timed)**

🗣 **Tazama nambari hizi. Nionyeshe nambari gani kubwa kuliko nyingine.**

[Repeat for each item]

Circle 1 if correct, circle 0 if incorrect.

7	2	*7*	1	0	91	81	*91*	1	0
16	23	*23*	1	0	325	620	*620*	1	0
51	15	*51*	1	0	864	963	*963*	1	0
88	78	*88*	1	0	419	219	*419*	1	0

✋ (Stop)
- If the child makes 4 successive errors, the Tablet will turn red

➲ (Move on)
- If the child doesn't respond after **5**

32	42	*42*	1	0	681	981	*981*	1	0	

SECONDS, mark as wrong then prompt pupil to move on.

Task 7: Missing number – PRACTICE

📖 **Sheet 26** 🕐 × (Not Timed)

P1:

🗣 **Hapa pana nambari kadhaa.1, 2, pengo, 4. Ni nambari gani itaenda hapa?**

1	2	(3)	4

✓🗣 **Ndivyo, 3! Tujaribu mfano mwingine.**

×🗣 **Nambari 3 itawekwa hapa. Tuseme nambari hizi pamoja.**

[Elekeza kidole kwa kila nambari]. **1, 2, 3, 4. Nambari 3 itawekwa hapa. Tujaribu mfano mwingine.**

P2:

🗣 **Hapa pana nambari kadhaa: 5, 10, 15, pengo. Ni nambari gani itaenda hapa?**

5	10	15	(20)

✓🗣 **Ndivyo, 20! Tujaribu mifano zaidi.**

×🗣 **Nambari 20 itawekwa hapa. Tuseme nambari hizi pamoja**

*[elekeza kidole kwa kila nambari]***. 5, 10, 15, 20. 20 inawekwa hapa. Tujaribu mifano zaidi.**

Task 7: Missing number - EXERCISE

📖 **Sheets 26 & 27** 🕐 × (Not Timed)

🗣 **Hapa pana nambari zaidi.** *[elekeza kidole kwa sanduku] …* **Ni nambari gani itaenda hapa.** *[Repeat for each item]*

✋ (Stop)

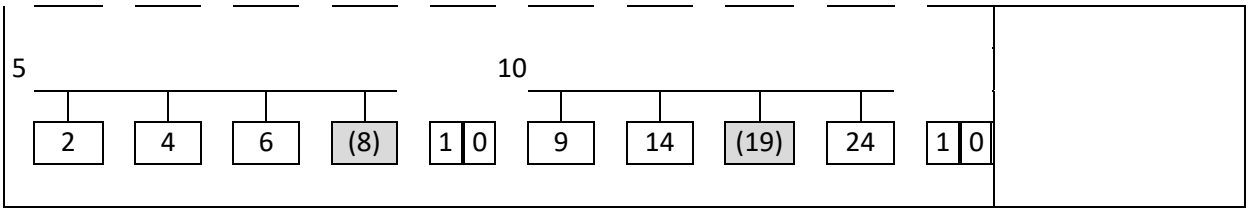
- If the child gets 4 successive errors

➲ (Move on)

- If the child doesn't respond after **5 SECONDS,** mark as wrong then prompt pupil to move on.

Circle 1 if correct, circle 0 if incorrect.

1	3	4	5	(6)	1	0	6	623	624	(625)	626	1	0
2	12	13	(14)	15	1	0	7	68	(66)	64	62	1	0
3	30	(40)	50	60	1	0	8	75	80	(85)	90	1	0
4	(200)	300	400	500	1	0	9	450	440	430	(420)	1	0



Task 8A: Addition: Level 1 - EXERCISE | 🕮 **Sheets 28 & 29** | 🕐 60 seconds (Timed)

🗣 **Hapa kuna mazoezi ya kuongeza**. *[Pitisha mkono kutoka juu hadi chini.]* **Nitakuhesabia wakati na nitakuambia wakati wa kuanza na wakati wa kumaliza. Sema jibu kwa kila swali. Kama hauna jibu, endelea na swali linalofuatia. Je, uko Tayari? . . .**

Anzia hapa *[elekeza kidole kwa swali la kwanza]*

✋ (Stop)

- If the time runs out (60 seconds), the Tablet will turn red.

✎ (/) Incorrect or no response
(O) Circle the item if the student self-corrects.
( ]) After last problem attempted

➲ (Move on)

- If a child stops on an item for **5 SECONDS,** mark as wrong then prompt pupil to move on.

1 + 3 = (4)	7 + 8 = (15)
2 + 3 = (5)	4 + 7 = (11)
6 + 2 = (8)	7 + 5 = (12)
4 + 5 = (9)	8 + 6 = (14)
3 + 3 = (6)	9 + 8 = (17)
8 + 1 = (9)	6 + 7 = (13)
7 + 3 = (10)	8 + 8 = (16)
3 + 6 = (9)	8 + 5 = (13)
2 + 7 = (9)	8 + 10 = (18)
9 + 1 = (10)	10 + 2 = (12)

✎ Record time left (seconds):

To solve the problems, indicate the method the child used (tick all that apply):

□ Solved the problems in his/her head
□ Fingers
□ Counters
□ Tick marks on paper with a pencil
□ Other (✍ describe) _____

Task 8B: Addition: Level 2 – EXERCISE | **Sheet 30** | × (Not Timed)

❖ Paper and pencil.

Hapa kuna mazoezi mengine ya kuongeza. Ukipenda, waweza kutumia penseli na karatasi. Lakini sio lazima.

Anzia hapa *[elekeza kidole kwa swali la kwanza].*

Circle: 1 = Correct.
0 = Incorrect or no response.

12 + 7 = (19)	1	0
17 + 8 = (25)	1	0
18+ 11 = (29)	1	0
22+ 37 = (59)	1	0
38 + 26 = (64)	1	0

(Stop)

- If the child did not answer any Level 1 question correctly.
- If the child makes 4 consecutive errors.

➲ (Move on)

- If a child uses an inefficient strategy (e.g., tick marks), ask the child "Do you know another way to solve the problem?"
- If a child continues to use an inefficient strategy or stops on an item **for 5 SECONDS.**

To solve the problems, indicate the method the child used [(✓) tick all that apply]:

□ Solved the problems in his/her head
□ Calculation method
□ Fingers
□ Counters
□ Tick marks on paper with a pencil
□ Other (describe) _____

To end the interview:

OBSERVATIONS

Thank the child for their participation and check through the entire questionnaire to ensure that no information is missing.

Remember to ask the child not to tell anyone about the specific details of the reading and number activities the child was asked to complete.

- -

CL13. Record the time when the literacy activities were completed.	Hour and minutes................................ __ __ : __ __

Appendix 2: A note on inter-rater reliability measures

Cohen's Kappa is built upon the idea of having each assessor mark an item as one of two possible outcomes. MICS's connected text passage uses a system where correct answers are left blank and only incorrect answers are marked on the paper or data entry interface. This allows a rater who is not actively paying attention to appear to perform well if the child also does relatively well on the test items. Similarly, an overly aggressive rater will appear to do well if a child is not successfully able to respond to items in the test. Lastly, Cohen's kappa does not take into account disagreement between the raters on whether an item was tested. For example, if one rater correctly applies an early stop rule when the other does not, one rater will have a missing value for the remaining test items, whereas the other rater will continue to mark those items as correct or incorrect. This third possible outcome makes Cohen's kappa an imperfect metric by which to judge the reliability of the raters on some subtests.

The table below shows the possible configurations of outcomes for two raters marking a child as correct, incorrect or with missing/no response. Cohen's kappa measures only the responses for correct and incorrect by both raters, thus cells W, X, Y and Z.

		Rater #1		
		Correct	Incorrect	Missing/No Response
Rater #2	Correct	W	Y	A
	Incorrect	X	Z	B
	Missing/No Response	C	D	F

PNA discards cell W as that is the default if neither rater is paying attention and places a higher value on cell Z, where both assessors are paying attention and agreed that the child was wrong. It also takes into account other disagreements, such as in cells A, B, C and D, where one of the raters is marking items either as correct or incorrect, while the other rater has decided that the item should be left blank. This most often occurs in MICS when a child has triggered one of the auto-stop rules. These situations, while rarer in MICS than other types of early primary testing, are important to note as they could significantly contribute to large rater-effects among child test scores. In addition to avoiding the problems of Cohen's kappa, PNA also has a distribution with finite limits that make it somewhat easier to understand intuitively.

The equation used to calculate the PNA for each rater pair is:

$$PNA = \frac{2Z}{2Z + X + Y + A + B + C + D}$$

Cell F is not used in the above equation because if both raters agree that the child did not answer the question (perhaps because of the auto-stop rule), this item should neither count for nor against the reliability metric.

For tests where there is a 'default' answer, as in the Oral Reading Accuracy task, PNA would be the best metric to apply to measure agreement between raters. When there is no default answer, Cohen's kappa would be the best metric. However, both metrics have drawbacks. For both PNA and Cohen's kappa, the formula returns 0/0 (undefined) as the calculation when both assessors agree that the child answered all items correctly. This high level of agreement should be treated as a success, so these cases will be recoded as '1' or perfect agreement between the raters.

		Rater #1	
		Correct	Incorrect
Rater #2	Correct	5	0
	Incorrect	0	0

$$PNA = \frac{0}{0} \rightarrow 1$$

$$Kappa = \frac{0}{0} \rightarrow 1$$

Similarly, there is a second situation where Cohen's kappa returns an undefined value for the calculation: when both raters agree that the child did not answer any items correctly. In this situation, the PNA formula produces a '1' – perfect inter-rater reliability – but Cohen's kappa again returns ''0'. This situation will also be treated as a success, so these cases will also be recorded as '1' or perfect agreement between the raters.

		Rater #1	
		Correct	Incorrect
Rater #2	Correct	0	0
	Incorrect	0	5

$$PNA = \frac{10}{10} = 1$$

$$Kappa = \frac{0}{0} \rightarrow 1$$

For information on the report, please contact:

United Nations Children's Fund
Data and Analytics Section
Division of Data, Research and Policy
3 United Nations Plaza, New York, NY 10017, USA
Tel: +1 (212)-326-7000
Email: mics@unicef.org