REQUIREMENT 3.3: TEACHER LANGUAGE AND LITERACY ASSESSMENT (TLLA)

Final Report

Prepared for
United States Agency for International Development (USAID)

Prepared by
Dexis Consulting Group and RTI International

Research for Effective Education Programming – Africa (REEP–A)
Contract No. AID-OAA-I-15-00019
Task Order No. AID-OAA-TO-16-00024

June 2020

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

EGRA	Early Grade Reading Assessment
IDM	Item Difficulty Mean
L1, L2, Lx	First Language, Second Language, Xth Language
LOI	Language of Instruction
LOI1	Initial Language of Instruction
LOI2	Subsequent Language of Instruction
P	Primary Grade Level
REEP–A	Research for Effective Education Programming – Africa
RTI	RTI International (registered trademark and trade name of Research Triangle Institute)
SD	Standard Deviation
TAM	Teacher Ability Mean
TLLA	Teacher Language and Literacy Assessment
USAID	United States Agency for International Development

EXECUTIVE SUMMARY

The Research for Effective Education Programming – Africa (REEP–A) Task Order, awarded in September 2016, is a five-year project within the United States Agency for International Development (USAID) Africa Bureau. The primary objective of REEP–A is to generate and effectively disseminate Africa regional and country-specific education data, analysis, and research to inform the prioritization of needs and education investment decisions.

One research focus under REEP–A is to explore how teachers' language proficiency and literacy in the language of instruction (LOI) influence students' learning outcomes. It is hypothesized that the teachers' level of language proficiency and literacy in the LOI can either facilitate student learning, if high; or impede learning, if low. However, limited data are available on how teacher language and literacy skill levels precisely relate to student outcomes.

Exploring this relationship requires having a valid and reliable tool to measure teachers' language and literacy skills. USAID therefore commissioned the development of the Teacher Language and Literacy Assessment (TLLA) to assess teachers' language proficiency and literacy in the required LOI. The TLLA, adaptable to any language, consists of subtasks assessing speaking, listening, reading, and writing, as well as vocabulary and grammar, in the language(s) used for teaching and learning at the primary school level in a given context. It is envisioned that policymakers, researchers, and other education stakeholders can use the TLLA to collect data on teachers' linguistic assets and gaps in the languages that their role requires them to use. These data could be useful for identifying factors contributing to student learning outcomes, informing teacher training and professional development needs, designing teacher deployment policies, and evaluating the impact of interventions aimed at improving teachers' or students' language and literacy skills.

The aim of this report is to present the new tool and disseminate the initial findings around its technical adequacy. The international community has directed considerable effort to assessing and understanding the impact of language on students' literacy and language skills, and the TLLA is a complementary tool that shows promise for understanding teachers' language assets and needs.

COMPONENTS OF THE TLLA TOOL

The TLLA is a suite of tasks and subtasks designed to be administered to primary school teachers in multilingual contexts in sub-Saharan Africa. It is intended to delineate the teachers' linguistic and literacy assets as well as any gaps in the language or languages used in their role as teachers. The TLLA suite contains 14 tasks or subtasks, including an interview and 13 assessment exercises (see text box). The full TLLA can be administered in approximately one hour per language. Many of the tasks and subtasks are oral and must be administered individually. The written subtasks, however, may be administered either individually or in a group setting. Individual subtasks of the TLLA may be included or omitted depending on the specific research objectives of the assessment.

- Language background interview task
- Speaking task
- Vocabulary task
- Listening task (sentence repetition)
- Oral reading task:
  - Letter-sound identification subtask
  - Nonsense word reading subtask
  - Oral reading fluency subtask
  - Oral reading comprehension subtask
- Grammar task:
  - Structure and written expression subtask
  - Error identification subtask
- Silent reading comprehension task
- Writing task:
  - Correcting student writing subtask
  - Responding to a writing prompt subtask
  - Spelling subtask

RESEARCH PROCESS

The research process consisted of developing an initial TLLA instrument and an administration protocol in English, adapting the instrument into a second language, completing two rounds of cognitive interviewing and subsequently refining the instruments, field-testing and piloting the instruments, and then analyzing data collected using the sample versions of the tool in both languages.

After the initial development of the tool in English, Uganda was chosen as the location for the subsequent research stages. Uganda offered a linguistically complex environment and the possibility of assessing teachers in both English and a local language of instruction, Luganda. The implementer of this research activity was also able to leverage contacts and relationships from another project in-country to facilitate the research process.

Thirteen Ugandan teachers participated in the cognitive interviews, 36 in the field-test stage (18 in each language), and 298 in the pilot (149 in English and 147 in Luganda). The focus of this activity was to ensure that the tool performed as designed, thus this activity did not produce representative data for the locations where the assessment was field-tested or piloted.

FINDINGS FROM FIELD-TEST AND PILOT STAGES

The researchers examined a set of nine research questions at the field-test stage and an additional eight questions at the pilot stage. During the field-test stage, the research questions focused on the teachers' reactions to the purpose and format of the assessment; the effectiveness of the order, format, and instructions of the tasks and subtasks; and the administration time. At the pilot stage, the research questions focused on the internal consistency of the items in each task and subtask, possible redundancy among tasks, the distribution of and variability in scores, and the alignment between teachers' self-reported abilities and their performance on related tasks or subtasks. The results of all the research questions are presented in full in this report. Key findings include the following:

1. **Positive reactions from teachers.** Contrary to the researchers' expectations, the teachers' reactions to being assessed were overwhelmingly positive. Teachers did not appear wary or reluctant to participate; on the contrary, teachers repeatedly expressed their gratitude for the opportunity to take the assessment, noting that it was beneficial to know their own strengths and weaknesses so that they could improve.

2. **Assessment for literacy skills.** Related to the above finding, teachers were aware of the links between the assessment components and the language and literacy skills needed to perform well in their role as teachers in primary schools. Teachers did not object to their skills being tested explicitly, and it was not necessary to frame the tasks as marking a fictitious students' responses in order to make the assessment palatable to the participants.

3. **Item consistency.** Throughout this process, the item performance of the tool was examined. Ten of the 13 assessment components (i.e., tasks or subtasks) in English, and six of the 12

components in Luganda, had high internal consistency among items, as measured by a Cronbach's alpha of 0.70 or greater.

4. **Item redundancy.** Efficiency of testing items, namely getting the most information in the least amount of time, is a consideration in tool development. To ensure a reasonable test length, three pairs of subtasks—in writing, comprehension, and grammar—were examined for possible redundancy (i.e., producing similar information). The results revealed several relevant findings. First, correcting student writing was not a direct substitute for the writing prompt. Second, the oral reading and silent reading subtasks yielded different information in this pilot sample. Third, selecting just one of the two grammar subtasks is sufficient.

5. **Variations in scores by task/subtask and language.** The distribution of scores was calculated for each task and subtask based on data collected in the pilot. In general, teachers scored higher on the Luganda assessment than teachers on the English assessment. This was not surprising because Luganda served the role of a first language (L1) for most teachers, with English as the second language (L2). None of the tasks or subtasks presented a floor effect, except for the Luganda correcting student writing subtask (with a 35% mean score). Several tasks and subtasks presented an apparent ceiling effect. This high performance indicates that for both English and Luganda, most teachers in this sample were able to complete the oral and written tasks and subtasks expected of their primary grade students with high accuracy and comprehension.

6. **Interview results vs. assessment results.** The research team examined the degree of alignment between the teachers' self-reported speaking and reading abilities on the teacher language interview and their performance on related tasks or subtasks. Overall, most tasks and subtasks aligned fairly well with teachers' self-reported abilities even when the self-reported abilities were low. This finding suggests that teachers are generally aware of their abilities and did not inflate their rating. Therefore, depending on the research purpose, an interview could be used instead of an otherwise time-consuming assessment to gauge teachers' language skills.

7. **Administration time.** As noted above, the administration time for the full suite of tasks and subtasks was around one hour. To ensure efficiency during the administration process, the research team recommends that future users of the TLLA prioritize and select the tasks that are most useful and feasible according to their research purpose and the logistical constraints of their context. Specific guidelines for this selection process are offered in the full report.

RECOMMENDATIONS

This first pilot of the TLLA marks a promising start to the development of a valid and reliable tool that measures teachers' language abilities. Subsequent pilots of the TLLA could help to further refine each task and subtask and build a strong item bank of tested items. The research team encourages researchers and stakeholders to experiment with adapting the tasks and subtasks in the TLLA suite and administer them in other contexts and languages. Sharing the results of those experiences with local stakeholders and the larger international education community will improve the tool's effectiveness in assessing teacher language and literacy proficiency for a variety of purposes and contexts. The research team anticipates that the process described in this report can serve as an example of how to thoughtfully develop a tool for similar contexts and will spur other researchers and implementers to further experiment with and refine the proposed assessment.

INTRODUCTION

The Research for Effective Education Programming – Africa (REEP–A) Task Order, awarded in September 2016, is a five-year project within the United States Agency for International Development (USAID) Africa Bureau. The primary objective of REEP–A is to generate and effectively disseminate Africa regional and country-specific education data, analysis, and research to ensure the availability of evidence-based interventions that inform the prioritization of needs and education investment decisions.

One research focus under REEP–A is to explore how teachers' language proficiency and literacy in the language of instruction (LOI) influence students' learning outcomes. In Africa's densely multilingual societies, policies around the language(s) used in education are critical to education quality and equity. Numerous countries in sub-Saharan Africa have recently shifted toward adopting the language of the local community or the language spoken at home (referred to as first language, home language, mother tongue, or L1), as the language for initial literacy acquisition and, in some cases, as the initial language of instruction (LOI1) across all subjects in the early grades.[1] These policies typically include a transition to a second or additional language (L2 or Lx[2]), such as a regional, national, or international language, as the subsequent language of instruction (LOI2), usually starting in upper primary, although the timing of the transition varies. A wide array of political, financial, attitudinal, and logistical factors impact decisions made around a country's language of instruction policy. As a result, designing and implementing an evidence-based, effective language policy in a complex sociolinguistic context can present many challenges.

Much of the research on language-in-education policies has focused on students—how they learn best, what they need to learn, and how they fare under different policies. Less research has tackled the issue of language from the perspective of the teachers, including their own proficiencies in the different language options and how language proficiency and literacy relate to the quality of instruction offered in each language. It is hypothesized that the teachers' level of language proficiency and literacy in the LOI can facilitate student learning, if high; or impede it, if low. However, limited data are available on how precisely teacher language and literacy skills relate to student reading outcomes.

Exploring this relationship requires a valid and reliable tool to measure teachers' language and literacy skills. USAID therefore commissioned the development of the Teacher Language and Literacy Assessment (TLLA) to evaluate teachers' language proficiency and literacy in the required LOI. The TLLA consists of subtasks assessing speaking, listening, reading, and writing, as well as vocabulary and grammar, in the language(s) used for teaching and learning at the primary school level in a given context. Under this activity, two sample versions were developed for piloting in Uganda, one in English and one in Luganda, but the TLLA itself is adaptable to any language.

It is envisioned that policymakers, researchers, and other education stakeholders could use the TLLA to collect data on teachers' linguistic assets and gaps in the languages that their role requires them to use. These data would be useful for identifying factors that contribute to student learning outcomes, informing teacher training and professional development needs, designing teacher deployment policies, and evaluating the impact of interventions aimed at improving teachers' or students' language and literacy skills. The oral reading component of the TLLA includes subtasks identical to those within the Early Grade Reading Assessment (EGRA) so that the two assessments could be administered

---

[1] Albaugh, Ericka. A. 2014. *State-Building and Multilingual Education in Africa.* New York, NY: Cambridge University Press.

[2] Lx, meaning second, third, or any ordinal number of a language beyond the L1, is a useful designation in highly multilingual societies where learners have complex language proficiency profiles.

simultaneously to teachers and students in the same context, allowing for a direct comparison of scores if the research agenda so requires.

This report describes the process of developing, testing, and refining a valid and reliable tool that can fulfill the purposes outlined above. The research team anticipates that this body of work can serve as an example of how to thoughtfully develop a tool for similar contexts and will spur other researchers and implementers to further experiment with and refine the proposed instrument.

COMPONENTS OF THE TLLA

The TLLA consists of a suite of assessment tasks and subtasks designed to be administered to primary school teachers in multilingual contexts in sub-Saharan Africa. The TLLA is intended to delineate teachers' linguistic and literacy assets and gaps in the language or languages that they are required by policy or practice to use in their role as teachers. For that reason, the tool is designed to be adaptable into any language.

The TLLA suite contains 14 tasks and subtasks, including an interview and 13 assessment exercises. The assessment focuses on the following skills: speaking, listening, oral reading (including the four subtasks of letter identification, nonword reading, oral reading fluency, and reading comprehension), silent reading, writing (including three subtasks of correcting pupil writing, responding to a writing prompt, and spelling), vocabulary, and grammar (including the two subtasks of structure and written expression and error identification).

The full TLLA can be administered in approximately one hour. Many of the tasks are oral and must be administered individually. The written tasks, however, may be administered individually or in a group setting. Individual components of the TLLA may be included or omitted depending on the specific research objectives of the assessment. The English version that was piloted in Stage 6 (detailed below), can be found in **_Annex A_.** The interview and assessment exercises are described in detail below, by task and subtask.

TLLA TASKS AND SUBTASKS

- Language background interview task
- Speaking task
- Vocabulary task
- Listening task (sentence repetition)
- Oral reading task:
  - Letter-sound identification subtask
  - Nonsense word reading subtask
  - Oral reading fluency subtask
  - Oral reading comprehension subtask
- Grammar task:
  - Structure and written expression subtask
  - Error identification subtask
- Silent reading comprehension task
- Writing task:
  - Correcting student writing subtask
  - Responding to a writing prompt subtask
  - Spelling subtask

LANGUAGE BACKGROUND INTERVIEW TASK

The language background interview was developed under an independent activity designed to explore the relationship, if any, between teachers' self-reported language background and abilities and their students' learning outcomes.

This untimed interview elicits information about teachers' background in the language of assessment—for example, which language the teachers grew up using most frequently in their home environment; which language(s) are used at home today; whether they were taught to read in the language of assessment, and if so in what context; and whether teachers have received training in how to teach reading in that language. This background helps to provide context for interpreting the teachers' performance on tasks and subtasks in the given language.

In addition, for each skill—pronunciation, speaking, reading, and writing—teachers are presented with four statements and asked to select the one that best describes their ability in the language that the TLLA is assessing. The statements describe skill levels ranging from minimal to professionally competent. Asking teachers to judge and report their perception of their own abilities in the given language, and comparing those perceptions to their actual performance on related tasks, helps to elucidate how aware teachers are of their own strengths and weaknesses and how accurately they are able to report them. If teachers' self-awareness and accuracy in reporting are high, using just the interview instead of the assessment may be a more efficient way to elicit these data.

SPEAKING TASK

The speaking task is intended to assess teachers' expressive oral language skills in casual usage in both vocabulary and syntax. For this task, a teacher is presented with an illustration and asked to describe it orally "in as much detail as possible." The illustration is a black-and-white drawing depicting a familiar scene containing at least 25 items or actions that could be mentioned. The assessor records all the items that the teacher mentions, ticking the responses from a comprehensive list. This task is timed, giving the teacher two minutes to describe the picture. The task is scored as the number of items or actions mentioned. At the end, the assessor also records the teacher's dominant response pattern as either incomplete sentences, grammatically incorrect sentences, grammatically correct sentences, or grammatically correct sentences elaborating beyond the illustration.

VOCABULARY TASK

The vocabulary task is intended to assess teachers' breadth and depth of knowledge of academic vocabulary that they may need to use or explain in their role as teachers. The teacher is presented with 10 "Tier 2" vocabulary words[3] in both written and oral form (i.e., shown in the stimuli and read aloud by the assessor) and asked to orally provide a definition and an example for each word. The assessor evaluates the teacher's understanding of the word's meaning, "as evidenced by the explanation," as either Very Good (thorough and accurate), Fair (basic and/or partially inaccurate), Poor (inaccurate), or No Response (skipped). This task is untimed.

LISTENING TASK (SENTENCE REPETITION)

The listening task is intended to assess teachers' receptive oral language skills in casual usage. The assessor reads aloud 10 sentences, one at a time. The teacher is asked to repeat each sentence, word for word, immediately after listening to it. Because the number of speech sounds in a sentence exceeds the capacity of short-term memory, sentences require syntactic processing to parse, retain, and repeat. The sentences begin with simple syntactic structures for the given language and gradually increase in length and complexity. The sentence topics themselves require no specialized knowledge. The assessor marks which words, if any, the teacher does not repeat correctly. This task is untimed and is scored as the percentage of words repeated correctly.

---

[3] A Tier 2 vocabulary word is the type of word that is used across domains and is more descriptive than a word classified as Tier 1. Tier 1 words are used in everyday language (e.g., "sad"), while Tier 2 words are more sophisticated and generally learned through texts and in adult-child interactions (e.g., "astonished"). Academic settings require knowledge of Tier 2 words. Teachers should not only know the words' meanings themselves, but also know them well enough to be able to teach them to students.

ORAL READING TASK: LETTER-SOUND IDENTIFICATION SUBTASK

The TLLA contains four subtasks categorized under oral reading. These subtasks are identical to four tasks on the EGRA for students. These subtasks were included to permit a direct comparison between teacher and student results if the TLLA were to be administered simultaneously with an EGRA.

The first oral reading subtask is intended to assess teachers' knowledge of the letter-sound correspondences in the target language's orthography. A teacher is presented with a list of 100 letters and asked to pronounce the default sound that each letter represents. The letter-sound identification subtask is timed to one minute and is scored as number of correct letter sounds read per minute (rate) as well as the percentage correct of attempted (accuracy).

ORAL READING TASK: NONSENSE WORD READING SUBTASK

The second oral reading subtask is intended to assess teachers' ability to apply their knowledge of the letter-sound correspondences in the target language's orthography to decode unfamiliar words. A teacher is presented with 50 nonsense (invented) words that conform to the orthography and phonology of the target language and asked to read them aloud. The nonsense word reading subtask is timed to one minute and is scored as the number of correct nonsense words read per minute (rate) as well as the percentage correct of attempted (accuracy).

ORAL READING TASK: ORAL READING FLUENCY SUBTASK

The third oral reading subtask is intended to assess teachers' oral reading fluency in the target language. A teacher is asked to read aloud a short narrative of about 60 words. The oral reading fluency subtask is timed to one minute and is scored as the number of correct words read aloud per minute (rate and accuracy). The assessor also rates the teacher's use of vocal expressiveness, or intonation (a prosodic feature), on a scale of one to three.

ORAL READING TASK: ORAL READING COMPREHENSION SUBTASK

The fourth oral reading subtask is intended to gauge teachers' comprehension of the passage that they just read aloud. The teacher is asked five questions based on the passage—four explicit and one inferential. First, the teacher responds by memory, without looking back at the passage. If unable to respond correctly to any question, the teacher can consult the text again and make a second attempt. The oral reading comprehension subtask is untimed, and is scored as the percentage of questions answered correctly, with and without lookbacks, out of five total questions.

GRAMMAR TASK: STRUCTURE AND WRITTEN EXPRESSION SUBTASK

The TLLA measures teachers' implicit knowledge of grammar through two subtasks. In this usage, "grammar" is defined as the language's underlying structure and implicit rules for putting words together into meaningful sentences. Two subtasks were constructed that apply different modalities for assessing this knowledge.

In the first grammar subtask—structure and written expression—the teacher is presented with five written sentences, each with a missing part and four options for completing it. Only one of the four options conforms to the language's implicit syntactic and morphological structures. The assessor also reads the sentences and options aloud to avoid conflating reading ability with grammar ability. The task is untimed and is scored as the percentage of correct responses out of five.

GRAMMAR TASK: ERROR IDENTIFICATION SUBTASK

In the second grammar subtask, the teacher is presented with five written sentences. In each sentence, four parts are underlined. The teacher is asked to identify which underlined part, if any, contains an error. The errors, when present, violate the implicit syntactic or morphological structures of the target language. One sentence contains no errors, while the other four each contain one error. In the error identification subtask, the assessor does not read the sentences aloud; instead, the teacher reads them silently. The task is untimed and is scored as the percentage of correct responses out of five questions.

SILENT READING COMPREHENSION TASK

The silent reading comprehension task is intended to assess teachers' reading comprehension in more depth than is possible through the oral reading comprehension task described above. The teacher reads a longer passage silently and answers 10 multiple choice comprehension questions based on the passage. The passage is an informational text around 150 to 200 words long, with complexity similar to that found in textbooks for primary grades (P) 4–6. The questions cover lower-order comprehension (e.g., items addressed directly in the text) as well as higher-order comprehension (e.g., main idea and inference). This task is timed to five minutes and scored as the number or percentage of correct responses out of 10 questions, as well as the percentage correct of attempted questions.

WRITING TASK: CORRECTING STUDENT WRITING SUBTASK

Lastly, the TLLA includes three writing subtasks. The first, the correcting student writing subtask, is intended to assess teachers' mastery of writing conventions in the target language. The teacher is presented with 10 sentences written by a fictitious student, each containing one error in either spelling, punctuation, capitalization, or grammar. The teacher is asked to both identify and correct the error. This task is timed to two minutes and is scored as the number or percentage of correct responses out of 10 total questions.

WRITING TASK: RESPONDING TO A WRITING PROMPT SUBTASK

In the second writing subtask, the teacher is asked to write a short composition in response to a writing prompt. The responding to a writing prompt subtask is timed to five minutes and is scored by a rubric on eight features: ideas, support/evidence, organization, introduction and conclusion, signal words, sentence structure and punctuation, word choice, and language usage.

WRITING TASK: SPELLING SUBTASK

The third writing subtask, added to the English version of the TLLA instruments during the field-test stage, is intended to assess teachers' mastery of letter-sound relationships in English. The assessor dictates 15 words, and the teacher writes the words. The words progress in complexity according to the known stages of spelling development in English. The spelling subtask is untimed and is scored on the percentage of correct responses, either at every possible letter junction, or as correct/incorrect for the whole word.

RESEARCH PROCESS

The research process consisted of the development of an initial TLLA instrument and administration protocol in English, followed by the adaptation of the instrument into a second language, Luganda; two rounds of cognitive interviewing and subsequent refinement of the instruments; field-testing; piloting;

and data analysis of the sample versions of the tool in both languages. This process unfolded across eight stages, as summarized in *Table 1* and described in detail in the subsections below.

STAGE	DESCRIPTION
TABLE 1. STAGES OF THE TLLA RESEARCH ACTIVITY	
1	Initial tool development in English
2	Round 1 of cognitive interviewing and subsequent refinement of English version
3	Adaptation of the tool into a second language
4	Round 2 of cognitive interviewing and subsequent refinement of both language versions
5	Field-testing with 36 teachers (18 per language) and subsequent refinement of both language versions
6	Pilot testing with 296 teachers (149 in English and 147 in Luganda)
7	Analysis of the results of the pilot data
8	Report writing and dissemination of results

Uganda was selected as the location for field-testing and piloting for several reasons. First, Uganda offered a linguistically complex environment and an opportunity to assess teachers in both English and one local language of instruction, Luganda. Luganda is widely spoken in Uganda as both a first and second language, and the country's language policy specifies Luganda as one of the languages of instruction in primary grades (P) 1–3. Nonetheless, not all teachers who are required to teach in Luganda speak it as a first language at home. While English is the language of teaching and learning beginning in P4, it is not a first or home language for most teachers. The variability in teachers' comfort levels in both languages was expected to result in corresponding variability in performance on the assessment results. These differences were extremely useful for gaining insights into the effectiveness of the assessment tool for a range of skill levels. Second, the implementer of this research activity had existing contacts in Uganda that were able to support the researchers in identifying a high-quality data collection firm. Third, the policy environment in Uganda was favorable to assessing teacher language proficiency, and the implementer was able to leverage existing positive Ministry relationships to facilitate the process.

STAGE 1: TLLA TOOL DEVELOPMENT

In Stage 1, a research team composed of education and linguistics experts developed the initial assessment tool based on best practices in language and literacy assessment, adapted for adult teacher participants in the sub-Saharan context. The TLLA tasks and subtasks are designed to cover essential components of language and literacy proficiency relevant for adult teacher professionals—that is, the items mirror the ways in which teachers are required to use language, reading, and writing in their teaching duties. The team was cognizant of the potential adverse effects that being assessed in the workplace might have on adults. To mitigate these concerns, the team aimed to frame the tasks in a way that would set teachers at ease, respect their dignity, and allow them to demonstrate their range of skills without undue emphasis on any deficiencies. The team also considered logistical factors, such as time constraints and participant fatigue, in designing the tool. The team created a suite of task descriptions, protocols, and sample items in English. As the TLLA is intended to be adaptable into any LOI in sub-Saharan African countries, the team also developed brief guidelines for adaptation.

STAGES 2 THROUGH 4: COGNITIVE INTERVIEWING AND ADAPTATION

A member of the U.S.-based research team led the work associated with research stages 2–4 in Uganda with the support of a Luganda language expert and a Ugandan logistics and data collection firm. Together they completed two rounds of cognitive interviewing, the adaptation of the instrument into Luganda, and refinement of the instrument based on the cognitive interviewing.

To ensure alignment with national and international research standards, the research plan for this study was reviewed and approved by the AIDS Support Organization Research Ethics Committee, a Ugandan institutional review board. The research team notified the districts where data collection was planned, presented the formal ethical approval, and shared the written permission obtained from the Ministry of Education.

For cognitive interviewing, the team utilized a convenience sample of three schools within the Wakiso District. Wakiso District was selected because Luganda is used as the LOI in this district, and it is close in proximity to Kampala, where the assessor training took place. The team conducted cognitive interviews with four teachers at one school using the English version of the assessment to verify that respondents understood the items as intended. The cognitive interview included probing questions such as "What are these instructions telling you to do?", "What made you say [that answer]?", "Now that you've done the exercise, is there any other information that we could add to the instructions to make them clearer?", and "How do you feel about this exercise?" Thirteen teachers participated in the cognitive interviewing *(Table 2).* All of the participating teachers were female, as no male teachers were available to participate in the cognitive interviews at the selected schools.

Utilizing the feedback gathered from the English cognitive interviews, the team revised some of the instructions and items in the English version to improve clarity and then adapted the instrument into Luganda based on the revised English version. Next, the team conducted another round of cognitive interviews in two additional schools, completing four additional English cognitive interviews and five Luganda cognitive interviews. The probes during the second round of cognitive interviews remained similar to those previously described to help ensure that the adjustments had adequately addressed any confusion.

TABLE 2. NUMBERS OF TEACHERS PARTICIPATING IN COGNITIVE INTERVIEWING			
LANGUAGE	MEN	WOMEN	TOTAL
English	0	8	8
Luganda	0	5	5
Total	0	13	13

The modifications made as a result of cognitive interviewing are described in the following section, alongside the additional modifications that resulted from the field-test stage.

STAGE 5: FIELD TEST

Immediately following the cognitive interview stage, the team proceeded to field-test both the English and Luganda versions of the tool. The goal for the field-test stage was to understand the

appropriateness of the task protocols and any variability in results, with a focus on nine initial research questions on the instrument itself. The information gleaned from this process informed updates to the TLLA in preparation for a larger pilot data collection.

For field testing, a convenience sample was used to select 12 schools in the Wakiso District. Similar to the cognitive interviews, this sampling strategy was utilized due to Luganda being the LOI in schools in this district and logistical convenience.

Three assessors were trained on both the English and Luganda instruments. All assessors were fluent in Luganda and had prior experience administering the EGRA. The three assessors administered the English and Luganda instruments to four teachers per day, two per language, for a total of 36 teachers *(Table 3).*

TABLE 3. NUMBERS OF TEACHERS PARTICIPATING IN FIELD-TESTING			
LANGUAGE	MEN	WOMEN	TOTAL
English	6	12	18
Luganda	0	18	18
Total	6	30	36

The nine research questions, and associated results, utilized for the field-test stage are detailed below.

STAGE 5 FIELD-TEST RESEARCH QUESTIONS

1. WHAT IS THE AMOUNT OF TIME TO ADMINISTER EACH TASK?

At the field-test stage, the average amount of time required to administer each task ranged from 1.8 minutes (the oral reading comprehension task in Luganda) to 8.4 minutes (the silent reading comprehension task in Luganda), with a total administration time of 60.1 minutes in English and 60.8 minutes in Luganda. *Table 4* presents the minimum, maximum, and average administration time, as well as the standard deviation (SD) per task and overall.

TASK	ASSESSMENT LANGUAGE	ADMINISTRATION TIME			
		MIN	MAX	AVERAGE	SD
Language background interview	English	4	10	6.2	1.63
	Luganda	4	10	7.0	1.46
Speaking	English	2	5	3.3	1.03
	Luganda	2	9	4.6	1.58
Vocabulary	English[1]	(6)	(15)	(9.7)	3.06
	Luganda	1	10	5.1	1.83

TABLE 4. AMOUNT OF TIME, IN MINUTES, IN WHICH EACH TASK WAS ADMINISTERED DURING FIELDTESTING

TABLE 4. AMOUNT OF TIME, IN MINUTES, IN WHICH EACH TASK WAS ADMINISTERED DURING FIELDTESTING

Listening (sentence repetition)	English	1	7	4.3	1.46
	Luganda	1	9	4.3	1.71
Oral reading: Letter-sound identification	English	1	5	3.1	1.08
	Luganda	2	9	3.8	1.59
Oral reading: Nonsense word reading	English	1	5	2.6	0.98
	Luganda	1	4	2.8	0.86
Oral reading: Fluency	English	1	4	2.0	0.84
	Luganda	1	4	2.3	0.77
Oral reading: Comprehension	English	1	4	2.1	1.05
	Luganda	1	3	1.8	0.71
Grammar: Structure and written expression	English	2	5	3.4	0.78
	Luganda	2	4	2.8	0.88
Grammar: Error identification	English	2	8	4.6	1.82
	Luganda	3	9	4.9	1.53
Silent reading comprehension	English	4	12	7.8	1.83
	Luganda	5	12	8.4	1.69
Writing: Correcting student writing	English	2	7	4.9	1.39
	Luganda	2	18	6.1	3.32
Writing: Responding to a prompt	English	4	9	6.8	1.34
	Luganda	5	10	6.9	1.63
Spelling[2]	English	1	5	3	0.48
Total (excluding teacher consent form)	English	32	79	60.1	10.51
	Luganda	50	76	60.8	7.49

[1] For the English vocabulary task, the assessors were asked to write down the teachers' answers. This was done to allow further insight into the items. The English task also contained 14 items instead of 10 so that the items could be further culled for the pilot based on the results. The Luganda vocabulary task contained only 10 items, and the assessors did not write down the answer but only scored it on the spot, which is the intended protocol for this task. Therefore, the timing of the Luganda vocabulary task is more indicative of the timing of this task when implemented with the intended protocol.

[2] The spelling subtask was administered to only four teachers in English during the field-test stage as an alternate writing subtask. Spelling was added to the pilot to include a quick assessment of writing.

2. WHAT IS AN EFFECTIVE ORDER IN WHICH TO ADMINISTER EACH TASK?

The tasks were administered in the order originally proposed, beginning with the oral tasks (speaking, vocabulary, listening/sentence repetition, oral reading, and grammar) and ending with the silent tasks (silent reading and writing). No issues were encountered with this order. On one occasion, in conjunction with testing the group administration format for the silent tasks, the silent tasks were administered first because that suited the participants' availability in that particular school. No issues

were perceived with that order either. Other than grouping the silent tasks together, since the teachers record their answers in a teacher booklet for the silent reading and writing tasks, the ordering of the tasks appears to be flexible. As noted under the next question, the silent tasks can also be administered in a group format, which offers the advantage of saving time.

3. WHAT IS AN EFFECTIVE FORMAT FOR ADMINISTERING THE WRITTEN TASKS?

The assessors tested administering the silent reading and written tasks to individually and as a group. The group option offers the advantage of reduced administration time. The group administration itself did not present any problems, but the logistics of scheduling multiple teachers outside of the classroom at the same time was challenging. Teacher availability aside, the researchers deemed that the silent reading and written tasks could be administered either individually or to a group.

4. HOW DO TEACHERS RESPOND TO BEING TOLD THE PURPOSE OF THE ASSESSMENT?

From the start of this process, the research team was concerned that assessing teachers in their workplace might evoke negative reactions. Contrary to the team's expectations, teachers were overwhelmingly receptive to the assessment and eager to participate. The teachers generally indicated that they did not fear negative consequences in taking the assessment because the assessor had explicitly stated that the results would be anonymous and have no bearing on their job. Instead, the teachers expressed gratitude for the opportunity to take the assessment and noted that it was beneficial to gain more awareness of the strengths and weaknesses in their language skills. Some teachers did express concerns that if they did too well they would not receive further opportunities for training. Many of the teachers indicated afterward that they had learned something new through the assessment process, and many asked to know their own results. One teacher said that their supervisors regularly observe their classroom but never share feedback, making it difficult for the teachers to know how to improve. Sharing the teachers' scores with them was not part of the protocol during the field-test stage; however, as teachers repeatedly indicated their desire for this information, future implementers should seriously consider whether and how to do so in a respectful way.

5. HOW DO TEACHERS RECEIVE EVALUATING FICTITIOUS STUDENT WRITING AND ANSWERS?

As mentioned above, the tasks were linked to the various ways that teachers are required to use language, reading, and writing in the classroom. For example, teachers need speaking and listening skills in their daily interactions with students; they must know letter sounds in order to teach reading through a phonics-based approach; must be able to read a text with comprehension themselves in order to guide students to do the same; and need to be able to model good writing and recognize and correct student errors. The teachers were very aware of the relevance of the skills being tested to their official duties. During the cognitive interview, teachers repeatedly mentioned the need to master these skills in order to help their students do the same. In other words, teachers themselves made explicit links between the tasks in the assessment and their role as teachers.

To link the teacher's duties with the skills being assessed, some of the tasks were framed as an evaluation of fictitious student answers. For example, one of the writing tasks asks teachers to correct a fictitious student's writing mistakes. In addition, in the field-test version of the silent reading comprehension task, the teachers were asked to mark a fictitious student's answers to a reading comprehension exercise and correct them where wrong. The teachers responded positively to this framing despite being aware that their own skills were being tested. During the cognitive interview, one teacher said, "You are asking us to mark the student's answers, but what you are really doing is testing our own understanding of the passage." Based on the teachers' awareness of and positive reaction to

being assessed in these skills, the researchers concluded that it was not necessary to frame a particular task as an evaluation of student answers to make the task more palatable to teachers. In fact, in most cases, it is likely more expedient to test the teachers' skills directly. In the case of the silent reading comprehension task, the instructions associated with marking the student's answers take a long time to explain, whereas a straightforward comprehension task would be familiar and potentially more efficient. Based on their experience in the field-test stage, the researchers decided to drop the "fictitious student" frame for the silent reading comprehension task, as noted under "Modifications" below.

6. DID THE ITEMS FOR THE SPEAKING SUBTASK EXTEND BEYOND THE PROVIDED ONES?

When the illustration from the initial tool development was used for the first round of cognitive interviewing, the assessors felt that the illustration was not conducive to eliciting enough language to last for the two minutes allotted to the task. Therefore, two new illustrations were selected for use in the field-test stage, one for English and one for Luganda. For both illustrations, the team administered the task to each other to generate an initial checklist of items that participants could potentially mention. Based on the field-test results, the list of items was further refined to include all the additional items that teachers mentioned when describing the pictures.

7. ARE THE TERMS USED IN THE INSTRUCTIONS UNDERSTOOD BY TEACHERS?

All of the terms used in the instructions were easily understood by the participants.

8. IS THERE OBVIOUS REDUNDANCY IN ANY OF THE SUBTASKS?

Based on general observations and hypotheses, the researchers suspected some redundancy between the two grammar tasks (structure and written expression and error identification), as both test similar knowledge (i.e., syntax) in different formats. The small sample size did not permit a psychometric analysis during the field-test stage, but psychometric analysis was later applied to check for redundancy with the larger sample in the pilot. Those results are addressed under Stage 7 below.

9. WHAT IS AN EFFICIENT WAY TO SCORE THE TEACHER WRITING SAMPLE THAT STILL PROVIDES USEFUL INFORMATION?

The field-test assessor training highlighted some issues with the scoring rubric used for the teacher writing sample. Though the terminology used in the rubric was familiar to most of the assessors, one assessor had an economics background and was not familiar with some of the education terms. This required additional training to ensure accurate scoring. Another issue was that some of the categories in the original version of the scoring rubric combined two or more criteria, which confounded scoring. For example, the original rubric included all of the following criteria together under one category, "Ideas:" *a focused and clearly identified main idea, stating the personal opinion in a fresh and original way, and supporting the opinion with three or more detailed reasons and examples.* The rubric was significantly revised for the field-test data collection to separate out the different criteria and then later adjusted for the pilot, as addressed in more detail under "Modifications" below.

MODIFICATIONS TO THE ASSESSMENT AFTER FIELD-TESTING

The cognitive interview and field-test stages resulted in multiple refinements to the tool that was initially developed during Stage 1. The following is a summary of the key refinements made to each task, where applicable.

LANGUAGE BACKGROUND INTERVIEW TASK

For most of the items, the cognitive interviews revealed that teachers interpreted the questions as intended, and only minor modifications were made to the wording. For the questions asking teachers to choose the statement that best described their skill level in speaking, reading, and writing, the results of the cognitive interviews suggested that many participants based their answer on their perception of the students' ability to understand them, rather than the teacher's own strengths or limitations. These questions were reworded several times to better guide teachers to rate their own skills instead of their students' skills.

SPEAKING TASK

During the first round of cognitive interviewing, the team had trouble getting the teachers to speak for longer than 30 seconds about the picture. Two adjustments were made. First, the instructions were updated to be more specific, namely:

> *Original*: "Please look at this picture and tell me what is happening in it."

> *Revised*: "For our first activity, please look at this picture. Please describe everything that you see in the picture in as much detail as possible. That is, describe the people, the objects, what they look like, and what they are doing." … If the teacher exhausts their initial description before the timer runs out, say: "What do you imagine life is like in this family?"

Second, the picture used for the English task was replaced with a picture with more people and actions for participants to describe. A similar picture was chosen for the Luganda task.

VOCABULARY TASK

Originally, the instructions for this task asked teachers to "provide an explanation in child-friendly language." However, during the cognitive interviewing, the researchers found that "child-friendly language" acted as a constraint on the participants' ability to demonstrate the full depth of their knowledge of the word. For example, in defining "curious," one participant said, "It's like inquisitive—but wait, a child probably wouldn't know inquisitive." After deliberation, the researchers felt that the purpose of the task, and every task, was not to evaluate participants' *teaching* ability (i.e., in this case, their ability to render a term understandable to the child), but rather to evaluate their own vocabulary knowledge. The decision was therefore made to change the instructions to "give a definition and an example" with no mention of "child-friendly language."

The cognitive interviewing process also revealed that for words with straightforward, one-word synonyms, some participants offered the synonym with no further explanation. This left assessors unsure of how well the participant knew the meaning of either the original word or the synonym. As a result, the researchers recommended updating the item development guidelines to include selecting words with no direct synonym, compelling participants to explain the meaning. They also added "and [give] an example" to the vocabulary instructions to further prompt the participants to expand on their explanation and give the assessors a better opportunity to judge the depth of their understanding. Assessors were trained to remind the participants to provide both a definition and an example if teachers initially only offered one of the two.

For English, after the first round of cognitive interviews, words with direct synonyms were removed and replaced with additional Tier 2 words found in teachers' guides from Ugandan primary schools. Both lists were culled from 15 words down to 10 words, based on the field-test results.

LISTENING TASK (SENTENCE REPETITION)

Some minor modifications were made to the English sentences after cognitive interviewing to eliminate some "Americanisms" by using terms and structures that were more familiar to non-native English speakers. The field-test results suggested that this task had a ceiling effect in both languages. A ceiling effect means that many teachers scored extremely high on this task, indicating the task was too easy for them. Given the extremely low error rates on the easier items in this task, the number of items was reduced from the original 15 down to the 10 most difficult items.

ORAL READING TASK: LETTER-SOUND IDENTIFICATION SUBTASK

Only minor edits were made to the instructions. For example, the original version of the instructions asked teachers to assume that they were presenting the letter sounds to their students for the first time. This resulted in the teachers reading the letter sounds very slowly with long pauses. In order to allow teachers to demonstrate their true fluency in letter sound identification, the instructions were revised to say: "Let's say you have taught each of these letter sounds to your pupils. Please tell me the most common sound that each letter makes in English."

GRAMMAR TASK: STRUCTURE AND WRITTEN EXPRESSION SUBTASK

Minor modifications were made after the cognitive interviewing to reduce unnecessary repetition in the instructions.

GRAMMAR TASK: ERROR IDENTIFICATION SUBTASK

Minor modifications were made to the instructions after the cognitive interviewing to clarify that each sentence contained at most only one error.

SILENT READING COMPREHENSION TASK

The original English passage for this task was adapted from a Ugandan student book. During cognitive interviewing, some of the teachers indicated that they were familiar with the passage from having taught it in class. The researchers decided to change the passage to one of similar length but different content for the field-test data collection to prevent bias due to previous exposure. Minor modifications were also made to the instructions to adapt the task for group administration.

As mentioned earlier, the field-test version of this task was framed as a marking exercise, whereby the teacher would read the passage and then mark the responses from a fictitious student as correct or incorrect. The researchers found that teachers understood that their own comprehension, rather than a students', was being tested, and explaining the instructions for the task was unnecessarily time-consuming. Thus, this task was reframed as a straightforward multiple choice comprehension activity for the pilot version, in which the teacher simply read the passage and selected the correct answer.

WRITING TASK: CORRECTING STUDENT WRITING SUBTASK

Minor modifications were made to the instructions to adapt the task for group administration.

WRITING TASK: RESPONDING TO A WRITING PROMPT SUBTASK

The writing prompt used for the cognitive interviewing in English asked participants to write a letter to the head teacher. The participants ended up spending a sizable portion of the allotted time writing out

conventions specific to formal letters, such as the sender's address, the recipient's address, and the date. In some cases, the teachers even drew the margins. This attention to formatting left less time to develop the content of the letter. To allow teachers more time to focus on content, the researchers changed the prompt to an opinion piece in which the teachers would present arguments for why music was more important than sports, or vice versa. For the pilot version, the decision was made to use the same prompt for English and Luganda.

As mentioned above, the scoring rubric also presented challenges for the assessors, some of whom were not professional educators and lacked experience in evaluating writing samples. Additionally, the assessor training allocated insufficient time to developing a strong consensus around scoring. Some of the wording on the rubric was revised for the pilot version, and more time was allocated to training the assessors on scoring using writing samples collected during the field-test stage.

WRITING TASK: SPELLING SUBTASK

The spelling subtask was added during the field-test stage but administered to only four teachers in English. No modifications were made from the field-test version to the pilot version.

OTHER TASKS

No issues were encountered in the administration of the nonsense word reading, oral reading fluency, and oral reading comprehension tasks, and as a result no modifications were made.

STAGE 6: PILOT

Following the field-test stage and resulting revisions of the instrument, another member of the U.S.-based research team traveled to Uganda to lead the assessor training for the pilot data collection, in collaboration with the Ugandan data collection firm. The training helped to ensure consistency during data collection, and thus statistical confidence in the results. Three of the seven assessors had been involved with the field-testing, six were experienced EGRA assessors, and all assessors had used the Tangerine™ data collection platform on tablets to collect survey data.

The assessors were told on the first day of training that the consistency of their scoring with that of other assessors would determine who would qualify to collect data. Several methods were used to measure consistency across the assessors' scoring. For instance, the assessors practiced scoring teacher writing samples that had been collected during the field-test stage. Participants also scored a full set of mock teacher results and examined their scoring with facilitators to discuss consistency and disagreements. Finally, the assessors received training on protocols and behaviors that are critical to maintain data quality, and their mastery of these protocols was checked with a written quiz.

The assessor agreement results of the training are presented in *Table 5*. The four subtasks adapted from the EGRA (letter sounds, nonsense words, oral reading fluency, and reading comprehension) were familiar to the assessors, and they achieved 99% agreement on their first submission, allowing the majority of the training to focus on the other tasks and subtasks. All of these consistency measures were within acceptable ranges for direct survey data. The vocabulary task and the writing prompt subtask required the most amount of time to reach consistency, as the scoring of both was partially subjective. Even though both had scoring criteria, it took time and debate to reach agreement. Agreement on the assessor protocol quiz was 98%.

TABLE 5. ASSESSOR AGREEMENT ON SCORING OF THE TLLA TASKS AND SUBTASKS

TASK OR SUBTASK	AGREEMENT
Language background interview	99%
Speaking	97%
Vocabulary	89%
Listening	95%
Oral reading: EGRA-adapted subtasks (letter sounds, nonsense words, oral reading fluency, and reading comprehension)	99%
Multiple-choice tasks and subtasks (grammar: written error identification, structure and written expression; silent reading comprehension; writing: correcting student writing)	100%
Writing: Responding to a prompt subtask	91%
Writing: Spelling subtask	100%
Assessor protocol quiz	98%

Three districts, Mukono, Mpigi, and Luwero, were selected to participate in the pilot because Luganda is used as the LOI in these districts and their relatively close proximity to Kampala. The proximity to Kampala allowed the assessor training to happen immediately before the data collection and also allowed the research manager to visit some of the schools to monitor the data collection.

As the purpose of this activity was not to provide representative data, the team used a convenience sample to select the 60 schools from among the three districts for the pilot data collection. First, each district was further divided into rural and urban sub-counties, Then, one urban sub-county and four rural sub-counties were selected from each district. From there, two parishes from each of the 15 sub-counties was selected, and two schools from each of these parishes was selected. This sampling method provided a total of 60 schools, 20 from each district, with a mix of rural and urban schools. The two schools per parish allowed assessors to travel in pairs throughout the data collection period.

The assessors conducted five interviews per day, with each lasting approximately one hour. Schools that had more P1–P4 teachers were prioritized in order to facilitate reaching 300 teachers (approximately 150 each in both English and Luganda) in a five day period. At each school, assessors prioritized administering the Luganda assessment to P1–P3 teachers since Luganda is the main LOI in those grades. Of the Luganda sample, 87% taught in P1–P3 and 13% taught in P4. For the English version of the tool, assessors prioritized P4 teachers because that is the grade in which the LOI transitions to English. Of the English sample, 77% taught P4 English or another mix of subjects, and 23% taught P1–P3. The numbers of teachers who participated in the pilot are presented in **Table 6.**

During the data collection, the research team sent assessors daily guidance tips for ensuring data quality on WhatsApp. These communications also encouraged collaboration during the data collection, as they provided a forum for asking questions and exchanging ideas. The data collection logistics firm also maintained daily communication with the assessors and addressed any issues immediately.

TABLE 6. NUMBERS OF TEACHERS PARTICIPATING IN PILOT TESTING			
LANGUAGE	MEN	WOMEN	TOTAL
English	44	105	149
Luganda	13	134	147
Total	57	239	296

STAGE 7: ANALYSIS

Utilizing the data collected during the pilot, the analysis focused on eight additional research questions, presented below. The analysis included calculating the distribution of scores for each task and subtask, as well as item analysis and Rasch analysis. This section presents the results of the analysis as well as information on the reliability and validity of the tasks and subtasks.

STAGE 7: PILOT TEST RESEARCH QUESTIONS

1. ARE THE ITEMS WITHIN EACH TASK AND SUBTASK CONTRIBUTING TO THE OVERALL SCORE OF THAT TASK/SUBTASK?

To examine reliability, the analysts calculated the commonly used Cronbach's alpha[4] to measure the internal consistency of each task and subtask. Cronbach's alpha is reported as a single number ranging from zero to one that represents how closely certain items measure the same construct (or characteristic). The higher the alpha, the more confidence in the internal consistency and the greater the likelihood that the individual items are all measuring the same underlying construct. In educational surveys, an alpha of 0.70 or higher is generally considered acceptable. Including more survey items in a task and restricting the item to measuring only one skill contributes to a higher alpha.

In the English version, 10 of the 13 assessment tasks or subtasks had an acceptable Cronbach's alpha of 0.70 or higher *(Table 7).* The exceptions were oral reading comprehension and the two grammar subtasks. In the Luganda version, six of the 12 assessment tasks (the Luganda version did not include a spelling task) had a Cronbach's alpha of 0.70 or above; the same three subtasks as in the English version, plus three additional subtasks—nonsense word reading, oral reading fluency, and correcting student writing—had an alpha lower than 0.70.

Each of the three tasks or subtasks with low alphas in both language versions included only five items. As mentioned above, a low number of items can contribute to a lower alpha, which is the most likely explanation those cases. Increasing the number of items could lead to higher internal consistency. However, lengthening the assessment might result in teacher fatigue, diminishing the reliability of the TLLA overall. In addition, the English oral reading comprehension subtask, which mirrored the same subtask on the student EGRA, contained one inferential question, as is standard practice on the EGRA, to measure this aspect of higher-level comprehension. On the TLLA, this inferential question caused the

---

[4] Streiner, David L. 2003. "Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency." *Journal of Personality Assessment* 80, no. 1, 99–103. https://doi.org/10.1207/s15327752jpa8001_18

Cronbach's alpha value to decrease by 0.23. This effect is not surprising, as answering inferential questions requires different skills than answering explicit questions.

	ENGLISH			LUGANDA		
TABLE 7. CRONBACH'S ALPHA ANALYSIS, BY TASK OR SUBTASK						
TASK OR SUBTASK	*n* OF ITEMS	ITEM ALPHA RANGE	CRONBACH'S ALPHA	*n* OF ITEMS	ITEM ALPHA RANGE	CRONBACH'S ALPHA
Speaking	117	[0.70, 0.71]	0.71	112	[0.83, 0.84]	0.84
Vocabulary	10	[0.73, 0.77]	0.77	10	[0.76, 0.80]	0.80
Listening	133	[0.97, 0.97]	0.97	80	[0.99, 0.99]	0.99
Oral reading: Letter-sound identification	100	[0.96, 0.96]	0.96	100	[0.96, 0.96]	0.96
Oral reading: Nonsense word reading	50	[0.78, 0.80]	0.80	50	[0.61, 0.64]	0.64
Oral reading: Oral reading fluency	59	[0.87, 0.89]	0.88	57	[0.53, 0.59]	0.57
Oral reading: Oral reading comprehension, with lookbacks	5	[0.38, 0.73]	0.50	5	[0.08, 0.55]	0.44
Grammar: Structure and written expression	5	[0.36, 0.47]	0.48	5	[0.01, 0.17]	0.11
Grammar: Error identification	5	[0.33, 0.57]	0.47	5	[0.22, 0.36]	0.35
Silent reading comprehension	10	[0.65, 0.71]	0.70	10	[0.67, 0.76]	0.74
Writing: Correcting student writing	10	[0.70, 0.75]	0.75	10	[0.59, 0.65]	0.65
Writing: Responding to a writing prompt	8	[0.78, 0.82]	0.82	8	[0.81, 0.84]	0.84
Writing: Spelling (whole word scoring)	15	[0.96, 0.96]	0.96	—	—	—

Additionally, all subtasks were subjected to a Rasch[5] analysis to understand the technical adequacy of the TLLA. Rasch analysis provides another method for understanding the constructs of a task or subtask by estimating the probability of a teacher giving a correct answer on a specific item. In this instance, the analysis employed a scale (it can be thought of as a ruler) with less difficult items placed on the bottom of the scale and more difficult items on the top. The analysis also placed teacher scores on the same

---

[5] Boone, William J. 2016. "Rasch Analysis for Instrument Development: Why, When, and How?" CBE—Life Sciences Education 15, no. 4, rm4. https://doi.org/10.1187/cbe.16-04-0148

scale, with lower scores on the bottom and higher scores on the top. For illustration, *Figure 1* shows a sample Rasch scale output.

FIGURE 1. SAMPLE RASCH SCALE OUTPUT

```
MEASURE                                          |                          MEASURE
  <more> -------------- TEACHER SCORES -+- ITEMS  --------------------- <rare>
    3               ############### S+S Q30                                3
                    ################  |  Q28 Q27
                 ####################### |  Q20
    2               ############ +  Q25 Q23                                 2
                    #############  |  Q26
                    #############  |  Q24
    1       ############################ +  Q16 Q19                         1
                ################### M|  Q21
                    #############  |  Q18 Q22
    0           ################### +M Q29                                 0
                   ################  |  Q15
              ######################## |  Q12 Q17
   -1          #################### +  Q6 Q10                             -1
                    #############  |  Q13
              #################### |  Q3 Q8
   -2            ############# +  Q14                                     -2
                    #########  |  Q11
                      ####### S|  Q2 Q7 Q5
   -3                 #######  +  Q9                                      -3
                      ########  |  Q7
                       ######  |  Q4
   -4                 ######## +S Q1                                      -4
  <less> -------------- TEACHER SCORES -+- ITEMS  --------------------<frequent>
Notes: Each '#' in the TEACHER SCORE column represents 1 teacher; M indicates the mean (teacher
ability mean on the left of the middle line; item difficulty mean on the right.)
```
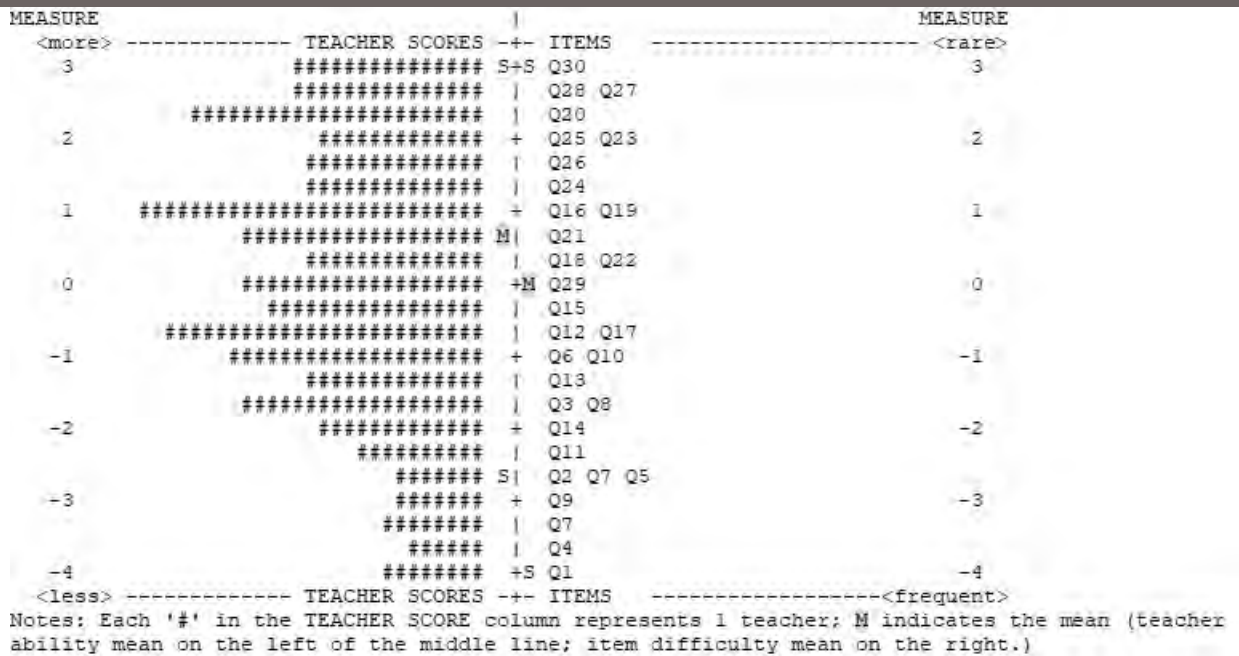
Placing the items and the teacher scores on the same scale helps to estimate the probability that a given respondent will answer an item correctly. This probability value is then used to examine which individual items did not perform as expected—in other words, where teachers of lower ability were able to give a correct response to a supposedly difficult item, or teachers of higher ability were unable to answer an easier item correctly. The following examples from the Rasch analysis on these instruments illustrate the kinds of insights that it can provide:

- In the English error identification subtask, item number 5 was relatively difficult compared to the other items, yet several low-ability teachers were able to respond correctly.

- In the English silent reading comprehension task, several teachers across the ability range were unable to provide correct responses to the first two items.

- In the Luganda vocabulary task, item number 7 was relatively difficult compared to the other items, yet several low-ability teachers received maximum credit for it.

- In the Luganda structure and written expression grammar subtask, several high-ability teachers were unable to provide the correct response to the first item.

To improve the tool, the next step would be to adjust or remove the problematic items and pilot again with the modifications. Although a second pilot was beyond the scope of this study, these findings underscore the importance of piloting each adaptation to ensure that all items function as intended.

2. IS THERE OBVIOUS REDUNDANCY IN ANY OF THE SUBTASKS?

Given the long administration time of the tool (about one hour) and the theoretical overlaps between some of the tasks or subtasks, the researchers wanted to examine the results of the pilot in terms of

redundancy. If two tasks or subtasks were found to yield redundant information, one could potentially be eliminated to ensure shorter administration time and greater efficiency. The research questions below focused on potential redundancy between three pairs of subtasks: the two writing subtasks (correcting student writing and responding to a writing prompt), the two comprehension tasks (oral reading comprehension subtask and silent reading comprehension task), and the two grammar subtasks (structure and written expression, and error identification). Each of these relationships is explored further below.

2A. WHAT IS THE RELATIONSHIP BETWEEN CORRECTING THE WRITING SAMPLE AND PRODUCING A WRITING SAMPLE?

The correcting student writing subtask presented the teachers with 10 sentences, each containing one error in spelling, punctuation, capitalization, or grammar. The teachers were asked to identify and correct the error. In the second writing subtask, teachers were given a prompt and asked to write a short composition in response. This subtask was scored against a rubric of eight writing subskills.[6] Although the writing prompt subtask generated data about a greater number of writing subskills, it was more challenging and time-consuming to score reliably than the correcting student writing subtask.

Due to the different formatting and scoring of each task and subtask, cross-tabulation was used to examine the relationship between the correcting student writing percentage score and the writing prompt subskills of sentence structure and punctuation, and language usage (see *Table 8*). For sentence structure and punctuation, the teachers who scored in the low and mid-range categories exhibited no apparent difference in their scores on the correcting student writing subtask in either language group. Those scoring in the highest category, however, had a mean score on the correcting student writing subtask that was 10% or more above the others. A relationship was slightly more evident between the correcting student writing score and the language usage score, in both languages. However, the very small number of teachers scoring in the lower categories (five in English and two in Luganda) limited the ability to draw any solid conclusions from the results. It is therefore not clear how well the correcting student writing subtask can serve as a substitute for the writing prompt's assessment of these subskills, or vice versa. Nonetheless, for situations in which logistical constraints prohibit the administration of a writing prompt, substituting the correcting student writing subtask may be a feasible way of collecting some data on writing skills.

---

[6] As indicated under Stage 1, the writing rubric categories were ideas, support/evidence, organization, introduction and conclusion, signal words, sentence structure and punctuation, word choice, and language usage.

TABLE 8. CROSS-TABULATIONS BETWEEN WRITING PROMPT SUBSKILLS AND CORRECTING STUDENT WRITING

WRITING PROMPT SUBSKILL	ENGLISH		LUGANDA	
	N	MEAN % SCORE FOR CORRECTING STUDENT WRITING	*N*	MEAN % SCORE FOR CORRECTING STUDENT WRITING
Sentence structure and punctuation				
The writer uses sentences of similar structures and lengths with little variety and little or no punctuation.	13	60	23	28
The writer uses some variety in sentence structure and length with some punctuation used appropriately.	62	62	55	30
The writer uses a wide variety of sentence structures and lengths with appropriate punctuation.	71	73	67	40
Language usage				
Four or more significant errors in language use make major portions of the text difficult to comprehend.	5	34	2	5
The writer has good control of language usage, with 2–3 significant errors in language use that make portions of the text difficult to comprehend.	20	59	25	28
The writer demonstrates strong control and correct usage of language, with 1–2 minor errors that do not impede comprehensibility.	121	70	118	37

2B. WHAT IS THE RELATIONSHIP BETWEEN CORRECTING COMPREHENSION QUESTIONS AND ANSWERING COMPREHENSION QUESTIONS?[7]

Two subtasks of the TLLA were intended to assess reading comprehension. In the first, the oral reading comprehension subtask, teachers were asked five questions orally about the short narrative passage they had read aloud for the oral reading fluency subtask. In the second, the silent reading comprehension task, teachers read a longer passage to themselves and answered 10 written multiple-choice questions about the passage. Though both subtasks assess the same skill, the oral reading subtask must be administered individually while the silent reading subtask can be administered individually or in a group setting. As such, the research team wanted to compare the two subtasks to ensure efficiency in future iterations of the tools

Teachers performed similarly on each subtask across the languages, and in both languages, performance was considerably stronger on the oral subtask than on the silent subtask. On the English oral reading comprehension subtask, teachers scored a mean of 91% correct with a standard deviation of 15 percentage points (i.e., less than one question). On the silent reading comprehension subtask, they scored a mean of 56%, with a standard deviation of 24 percentage points (i.e., approximately two

---

[7] For clarification, this research question refers to the silent reading comprehension task and the oral reading comprehension subtask. At the time the original research plan was written, the silent reading comprehension task involved correcting a hypothetical student's answers to comprehension questions. As indicated earlier, during the field test, this structure was seen to be unnecessarily cumbersome and was changed to a straightforward multiple-choice format.

questions). In Luganda, teachers scored a mean of 89% correct, with a standard deviation of 15 percentage points (i.e., less than one question) on the oral subtask, and a mean of 57%, with a standard deviation of 22 percentage points (i.e., approximately two questions) on the silent task.

Pearson correlations were calculated between the two comprehension tasks in each language. Pearson's is a widely used test statistic that measures the strength of the linear relationship between two variables and presents results on a −1 to +1 scale. A positive correlation indicates that as one score increases, the other score increases too, with +1 indicating a perfect correlation. In contrast, a negative correlation, or inverse relationship, means that as one score decreases, the other one increases. In Luganda, the relationship was moderate (0.57), as would be expected of two tasks measuring the same skill. However, the correlation was negligible (0.16) between the two comprehension tasks in English. The weak relationship between the tasks in English may indicate that, despite the intention, the English subtasks were measuring different skills. It seems likely that the oral reading version was more a measure of fluency than comprehension, and the silent reading subtask was more of a measure of vocabulary than comprehension. Assessors shared anecdotes to support this finding, noting that teachers said they were more worried about "sounding nice" when reading the passage aloud in the oral task and did not pay adequate attention to the meaning, even though the directions specified that questions would follow.

In both languages, the high mean scores and the Rasch model analysis of the fit between item difficulty and participant ability both pointed to a ceiling effect on the oral reading comprehension subtask. In other words, the fact that so many teachers received a perfect score on the task indicates that the task was too easy for them. Additionally, the ceiling effect made it impossible to distinguish high performers from low performers, so the oral subtask did not yield helpful information regarding teachers' reading comprehension skills. The silent reading comprehension task included twice as many items, increasing its potential sensitivity (i.e., the ability to distinguish high performers from low performers) compared to the oral subtask. For these two similar populations, the silent reading comprehension task generated more useful information.

2C. WHAT IS THE RELATIONSHIP BETWEEN CORRECTING GRAMMAR IN WRITING AND ORALLY?

The two grammar subtasks were intended to assess the same construct, namely the teachers' implicit knowledge of grammar. The structure and written expression subtask was presented as a multiple-choice, fill-in-the-blank exercise. The error identification subtask was also multiple choice, but the participants chose their response from among five underlined parts of the sentence. Both subtasks were presented in written form and were untimed. The assessor read aloud the sentences and the answer options in the structure and written expression subtask, but not in the error identification subtask.

In English, the teachers scored very similar means on the two tasks, namely 64% for structure and written expression and 62% for error identification. The standard deviations on the tasks were identical, at 27 percentage points (equating to a one question difference). In Luganda, the teachers performed better and with less variability on the structure and written expression subtask, with a mean of 89% and a standard deviation of 12 percentage points (which equates to less than a one question difference) than on the error identification subtask, with a mean of 68% and a standard deviation of 24 percentage points (equating to a one question difference).

Rasch analysis showed that in both languages, both grammar subtasks met the Rasch measurement assumptions for unidimensionality, meaning that the subtask was measuring a single construct, grammar ability; as well as for local independence, meaning that all the items were uncorrelated after accounting for the single dimension that they shared. Meeting these assumptions of unidimensionality and local

independence increases the confidence that these subtasks are valid measures of grammar and not another skill. However, neither subtask was sensitive enough to distinguish high performers from low performers, likely due to the low number of items in each. In both languages, and especially in Luganda, the item spread was better in the error identification subtask. The item spread was not perfectly redundant—that is, the items in the two subtasks were not at identical levels of difficulty. This means that rather than eliminating one subtask, combining them would give the best coverage and would increase sensitivity. For greater efficiency in administration, instrument developers could eliminate one of the subtasks and double the number of items in the remaining subtask, increasing the subtask's sensitivity while having to present only one set of instructions and examples.

Given its written format, the error identification subtask offers the advantage that it can be administered in a group. However, the written format has the potential to conflate reading ability with grammar ability. Specifically, poor readers might answer some questions incorrectly due to their inadequate reading skills rather than grammar deficiencies. However, this concern did not affect the performance of the English teachers in this sample, likely due to their adequate reading ability. The Luganda teachers performed similarly well on the silent reading comprehension task and the oral reading comprehension subtask. The primary cause of the Luganda teachers' lower performance on the error identification subtask is likely due to the adaptation of the task into Luganda at a more appropriate level of difficulty. These results highlight the challenges of adapting a grammar task across languages. Further experimentation in other contexts and with other languages may be helpful in determining which format, if either, is easier to adapt.

3. WHAT IS THE DISTRIBUTION (INCLUDING MEAN, MEDIAN, MODE) OF SCORES FOR EACH TASK AND SUBTASK?

The distribution of scores was calculated for each task and subtask and is presented in *Table 9* through *Table 11*. Generally, scores were higher on the Luganda assessment than on the English, although it is important to note that different individuals took the Luganda assessment than the English. The most striking difference in performance between the two language groups was in the speaking task. In English, teachers mentioned on average 19 items or actions (out of a possible 117), versus 94 in Luganda (out of a possible 112).

None of the tasks or subtasks presented a floor effect (i.e., with most of the teachers scoring very low), with the possible exception of the Luganda correcting student writing subtask, where the mean score was only 35% correct. In contrast, several tasks and subtasks may have presented a ceiling effect, with most of the teachers scoring very high. Teachers' mean score was at or above 85% for the following activities:

- English and Luganda nonsense word reading accuracy

- English and Luganda oral reading fluency accuracy

- English and Luganda oral reading comprehension with lookbacks

- English spelling (letter junction score)

- Luganda grammar: structure and written expression

This high performance indicates that for both English and Luganda, the vast majority of teachers in this sample were able to orally read a short narrative passage with high accuracy and comprehension. In the other tasks or subtasks, teacher performance showed greater variability.

TABLE 9. RANGE, MEAN, SD, MEDIAN, AND MODE, BY TASK OR SUBTASK, ENGLISH ASSESSMENT

TASK OR SUBTASK	METRIC	RANGE	MEAN	SD	MEDIAN	MODE
Speaking	*n* of items / actions mentioned	[0, 35]	19	6	20	21
Vocabulary	Score[1] out of 10	[1.3, 9.5]	5.3	2.1	5.3	6.5
Listening	% of words repeated correctly out of 133	[0, 100]	69	26	77	0
Oral reading: Letter-sound identification	*n* of correct letter sounds read per minute (rate)	[0, 86]	32	20	30	0
	% correct of attempted (accuracy)	[0, 100]	71	29	82	0
Oral reading: Nonsense word reading	*n* of correct nonsense words read per minute (rate)	[0, 95]	37	18	34	0, 15, 23, 39, 41
	% correct of attempted (accuracy)	[0, 100]	86	17	90	100
Oral reading: Oral reading fluency	*n* of correct words read per minute (rate)	[0, 183]	108	25	109	96.6, 112.2
	% correct of attempted (accuracy)	[0, 100]	98	12	100	100
Oral reading: Oral reading comprehension	% correct **without lookbacks,** out of 5	[0, 100]	53	30	60	60
Oral reading: Oral reading comprehension	% correct **with lookbacks,** out of 5	[0, 100]	91	15	100	100
Grammar: Structure and written expression	% correct out of 5	[0, 100]	64	27	60	60
Grammar: Error identification	% correct out of 5	[0, 100]	62	27	60	80
Silent reading comprehension	% correct out of 10	[10, 100]	56	24	50	40
	% correct of attempted	[10, 100]	58	24	56	40
Writing: Correcting student writing	% correct out of 10	[0, 100]	67	23	70	70
Writing: Spelling	% correct of 130 letter junctions	[7, 100]	91	10	94	96
	% correct of 15 whole words	[0, 100]	77	18	80	80

[1] Vocabulary score per item was calculated as "Very Good" = 1 point, "Fair" = 0.5 points, "Poor" = 0.25 points, "Skipped" = 0.

TABLE 10. RANGE, MEAN, SD, MEDIAN, AND MODE, BY TASK OR SUBTASK, LUGANDA ASSESSMENT

TASK OR SUBTASK	METRIC	RANGE	MEAN	SD	MEDIAN	MODE
Speaking	*n* of items / actions mentioned	[63, 108]	94	8	95	97
Vocabulary	score[1] out of 10	[2.8, 10]	7.9	1.9	8.5	10.0
Listening	% of words repeated correctly out of 80	[0, 100]	76	29	86	98
Oral reading: Letter-sound identification	*n* of correct letter sounds read per minute (rate)	[0, 85]	35	16	34	35
	% correct of attempted (accuracy)	[0, 100]	79	22	87	100
Oral reading: Nonsense word reading	*n* of correct nonsense words read per minute (rate)	[16, 80]	39	12	38	35
	% correct of attempted (accuracy)	[70, 100]	95	6	96	100
Oral reading: Oral reading fluency	*n* of correct words read per minute (rate)	[19, 104]	62	15	61	57
	% correct of attempted (accuracy)	[70, 100]	99	3	100	100
Oral reading: Oral reading comprehension	% correct **without lookbacks,** out of 5	[20, 100]	73	21	80	80
Oral reading: Oral reading comprehension	% correct **with lookbacks,** out of 5	[0, 100]	89	15	100	100
Grammar: Structure and written expression	% correct out of 5	[60, 100]	89	12	80	100
Grammar: Error identification	% correct out of 5	[0, 100]	68	24	60	60
Silent reading comprehension	% correct out of 10	[10, 100]	57	22	60	40
	% correct of attempted	[10, 100]	60	22	60	40
Writing: Correcting student writing	% correct out of 10	[0, 90]	35	17	30	30

[1] Vocabulary score per item was calculated as "Very Good" = 1 point, "Fair" = 0.5 points, "Poor" = 0.25 points, "Skipped" = 0.

TABLE 11. WRITING PROMPT SCORES, ENGLISH AND LUGANDA WRITING SAMPLES

WRITING SCORE COMPONENT	% OF ENGLISH TEACHERS (N = 149)	% OF LUGANDA TEACHERS (N = 146)
Ideas		
The writer does not give an opinion.	1	3
The writer's opinion is unclear.	11	5
The writer clearly states an opinion.	87	90
Sentence structure and punctuation		
The writer uses sentences of similar structures and lengths with little variety and little or no punctuation.	9	16
The writer uses some variety in sentence structure and length, with some punctuation used appropriately.	42	37
The writer uses a wide variety of sentence structures and lengths, with appropriate punctuation.	49	46
Word choice ("rich" vocabulary = descriptive, complex, vivid, precise words)		
The writer uses only simple vocabulary.	22	46
The writer uses 1–2 rich vocabulary words to support the opinion given.	43	35
The writer uses 3 or more rich vocabulary words to support the opinion given.	35	18
Support/evidence		
The writer does not give any reasons or examples to support the opinion presented.	7	3
The writer gives 1–2 reasons with examples to support the opinion presented.	30	29
The writer gives 3 or more reasons with specific examples to support the opinion presented.	63	67
Organization		
The ideas are disconnected.	3	3
The ideas are not always connected and logically arranged.	21	16
The ideas are connected and logically arranged.	76	80
Introduction and conclusion		
The writer does not introduce the topic or provide a concluding statement.	10	5
The writer's introduction and conclusion are unclear and weak but somewhat related to the opinion presented.	41	46

TABLE 11. WRITING PROMPT SCORES, ENGLISH AND LUGANDA WRITING SAMPLES		
The writer has a clearly stated introduction and conclusion related to the opinion given.	49	49
Signal words (e.g., I believe, in my opinion, I think, for example, in addition, etc.)		
The writer uses no signal words/phrases or uses them inappropriately.	30	37
The writer uses 1–2 signal words/phrases appropriately.	50	49
The writer uses 3 or more signal words/phrases appropriately.	21	12
Language usage		
Four or more significant errors in language use make major portions of the text difficult to comprehend.	3	1
The writer shows good control of language usage, with 2–3 significant errors in language use that make portions of the text difficult to comprehend.	14	17
The writer demonstrates strong control and correct usage of language, with 1–2 minor errors that do not impede comprehensibility.	83	81

Rasch analysis was also used to examine how well suited the items were to measuring the abilities of the sample that participated in the assessments *(Table 12).* A teacher ability mean (TAM) score that is higher than the item difficulty mean (IDM) indicates that a task is relatively easy for the population sampled. The opposite is also true; when the TAM score is lower than the IDM, a task is relatively difficult for the teachers. A TAM at 2 standard deviations above the IDM may be considered a ceiling effect. In the Rasch analysis, the TAM was close to 2 standard deviations above the IDM for the following tasks and subtasks:

- English and Luganda nonsense word reading
- English and Luganda oral reading fluency
- Luganda oral reading comprehension
- English oral reading comprehension
- Luganda listening

TABLE 12. RASCH MODEL ALIGNMENT BETWEEN TEACHER ABILITY MEAN (TAM) AND ITEM DIFFICULTY MEAN (IDM)

| TASK OR SUBTASK | VALUE OF TAM IN RELATION TO IDM, IN TERMS OF SD (RELATIVE DIFFICULTY OF TASK FOR THE POPULATION SAMPLED) | |
	ENGLISH	LUGANDA
Speaking	>1 SD below (difficult)	>1 SD above (easy)
Vocabulary	Perfectly aligned	>1 SD above (easy)
Listening	>1 SD above (easy)	Almost 2 SD above (very easy)
Oral reading: Letter-sound identification	>1 SD above (easy)	>1 SD above (easy)
Oral reading: Nonsense word reading	>2 SD above (too easy)	>1 SD above (easy)
Oral reading: Oral reading fluency	>2 SD above (too easy)	>2 SD above (too easy)
Oral reading: Oral reading comprehension	Almost 2 SD above (very easy)	>2 SD above (too easy)
Grammar: Structure and written expression	<1 SD above (a little easy)	<1 SD above (a little easy)
Grammar: Error identification	<1 SD above (a little easy)	<1 SD above (a little easy)
Silent reading comprehension	<1 SD above (a little easy)	<1 SD above (a little easy)
Writing: Correcting student writing	<1 SD above (a little easy)	<1 SD below (a little difficult)
Writing: Answering a writing prompt	1 SD above (easy)	Almost 1 SD above (easy)

In the Rasch analysis of the English assessment, the subscale was sensitive enough to distinguish between high and low performers only for the letter-sound and spelling tasks. In Luganda, the subscale was sensitive enough in the speaking task, the listening task, and the writing prompt subtask, but not the others. In some cases, especially the nonsense word and oral reading fluency subtasks, sensitivity was likely hindered by the large number of high performers. In other cases where there was an acceptable spread of item difficulty and range of representation, sensitivity may have been hindered by the low number of items, including for the vocabulary task, the two grammar subtasks, and the correcting student writing subtask.

4. WHAT ARE TEACHERS' SELF-REPORTED LANGUAGE BACKGROUNDS?

The results of the teacher language background interview are presented in *Table 13* through *Table 15*. Of the teachers taking the Luganda assessment, 89% reported Luganda as their most frequently spoken home language as a child, and 92% as an adult. For these teachers, Luganda functioned as their first language. Of the teachers taking the English assessment, none reported English as their most frequent home language as a child, and only 7% reported English as their most frequently spoken home language as an adult. For the overwhelming majority of these teachers, English functioned as their second language. The teachers assessed in Luganda tended to be somewhat more confident about their speaking and pronunciation skills in Luganda than in English. However, the English teachers tended to rate their ability to write in English higher than the Luganda teachers rated their ability to write in Luganda.

TABLE 13. TEACHERS' SELF-REPORTED LANGUAGE BACKGROUNDS

LANGUAGE	% OF ENGLISH TEACHERS (N = 149)	% OF LUGANDA TEACHERS (N = 147)
Language used most frequently at home with family as a child:		
Acholi	2	—
Ateso	4	—
Kiswahili	1	—
Lubarati	1	—
Luganda	64	89
Lugwere	5	1
Lukhonzo	—	1
Lumasaba	2	1
Lusoga	9	6
Runyankore-Rukiga	5	1
Runyoro-Rutoro	3	1
Other	5	-
Language used most frequently at home with family nowadays:		
Acholi	1	—
Ateso	1	—
English	7	5
Leblango	1	—
Lubarati	1	—
Luganda	81	92
Lugwere	1	—
Lukhonzo	—	1
Lumasaba	1	—
Lusoga	2	1
Runyankore-Rukiga	3	1
Runyoro-Rutoro	1	—
Other	1	—

TABLE 13. TEACHERS' SELF-REPORTED LANGUAGE BACKGROUNDS

Additional language(s) used frequently at home with family nowadays:

English	70	64
Kiswahili	1	—
Lubarati	1	—
Luganda	5	3
Lugwere	—	1
Lukhonzo	1	—
Lusoga	1	1
Runyankore-Rukiga	1	1
Runyoro-Rutoro	1	1
Other	2	1
None	17	27

TABLE 14. TEACHERS' SELF-REPORTED LANGUAGE SKILL LEVELS

QUESTION: "CHOOSE THE ONE SENTENCE THAT BEST DESCRIBES YOUR …"	LANGUAGE	PERCENTAGE OF TEACHERS, BY ANSWER OPTION			
		"I CANNOT"	"WITH SOME EFFORT"	"AT EASE WITH BASICS"	"AT EASE WITH BASICS AND PROFESSIONALLY"
Speaking ability in	English (*n* = 149)	—	7	42	52
	Luganda (*n* = 147)	—	6	24	69
Pronunciation in	English (*n* = 149)	2	9	46	44
	Luganda (*n* = 147)	1	8	27	65
Reading ability in	English (*n* = 149)	1	13	35	51
	Luganda (*n* = 147)	28[1]	9	22	41
Writing ability in	English (*n* = 149)	1	13	28	59
	Luganda (*n* = 147)	2	24	18	56

[1] A translation error in the first response option for reading ability was discovered in the Luganda instrument after the fact. This option read "I can read Luganda with understanding" instead of the intended "I cannot read Luganda with understanding." The results for this question were therefore skewed.

TABLE 15. TEACHERS' SELF-REPORTED TRAINING

RESPONSE OPTION	% OF ENGLISH TEACHERS (n = 149)	% OF LUGANDA TEACHERS (n = 147)
Was taught how to read in [English/Luganda]	99	86
Received training to teach reading	77	96
Received training to teach reading specifically in English	60	33
Received training to teach reading specifically in Luganda	24	80

5. DO TEACHERS' SELF-REPORTED LANGUAGE AND LITERACY SKILLS ALIGN WITH THEIR MEASURED SKILLS? (I.E., DO TEACHERS ACHIEVE THEIR BEST RESULTS IN THE TASKS AND SUBTASKS THAT ALIGN WITH THE SKILLS FOR THEIR SELF-REPORTED STRENGTH?)

Asking teachers to self-report their language skills offers several advantages over administering an assessment of those skills, including the time needed for administration and the avoidance of any discomfort associated with being assessed. However, self-reported data are always subject to validity and reliability challenges. Subjects may not be able to accurately discern or communicate their own abilities. For instance, the Dunning–Kruger effect, a specific cognitive bias, suggests that people at lower ability levels have trouble recognizing their own shortfalls and tend to overestimate their ability.[8] Further, people may feel subconscious pressure to give a socially desirable response, regardless of their own assessment of their skill levels, in an interview context. The TLLA pilot offered the opportunity to examine how well the teachers' self-reported skills lined up with their performance on tasks intended to assess those skills.

For the language background interview, teachers were asked to select from four options that best described their speaking, pronunciation, reading, and writing abilities. The statements (also shown in *Table 14* above) reflected the following levels: no ability, able to do the basics "with some effort," at ease with casual (but not professional) usage, and at ease with both casual and professional usage. Discounting the data from the Luganda reading ability question, which were skewed due to a translation error, at most only 3% of the 296 participants selected the "no ability" statement for any given skill, leaving the remaining three levels of distinction (i.e., low, mid, and high). Even then, for many tasks, the number of teachers self-selecting the lowest skill level was much lower than the numbers selecting the mid and high skill levels.

The degree of alignment between teachers' self-reported skills and measured skills varied by task and language. In some tasks, the performance lined up fairly consistently with all three self-reported skill levels. For example, as shown in *Figure 2,* teachers who self-reported their English speaking ability at the lowest skill level (blue line) performed fairly consistently lower on the English listening task than those who self-reported their speaking ability at the mid skill level (red line); this second group, in turn, performed lower than those who self-reported their speaking ability at the highest level (gray line).

---

[8] Kruger, Justin, and David Dunning. 1999. "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments." *Journal of Personality and Social Psychology* 77, no. 6, 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

For some tasks, the performance of those self-reporting low skill levels lagged behind the other two groups, but the performance of the groups with the mid and high self-reported levels was intermixed and often indistinguishable. One such example is the English speaking task, shown in *Figure 3,* where the red and gray lines are intertwined.

FIGURE 3. CUMULATIVE DISTRIBUTION OF ENGLISH SPEAKING SCORE BY SELF-REPORTED SPEAKING ABILITY
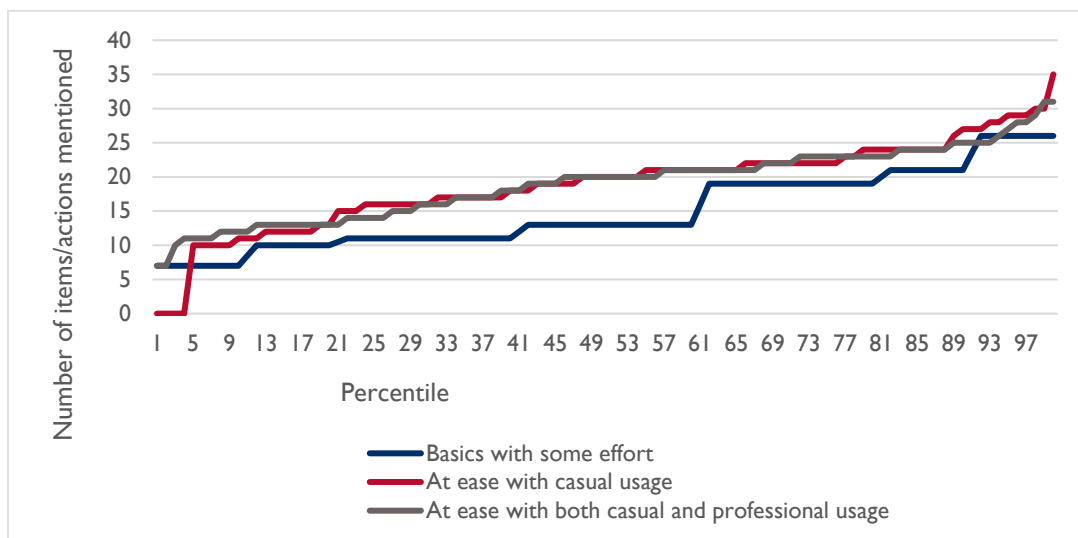
In tasks with a ceiling effect, except at the very lowest percentiles, teachers' performance was universally high and therefore indistinguishable among the self-reported skill levels. This can be seen in the example of the Luganda speaking task, shown in *Figure 4.*
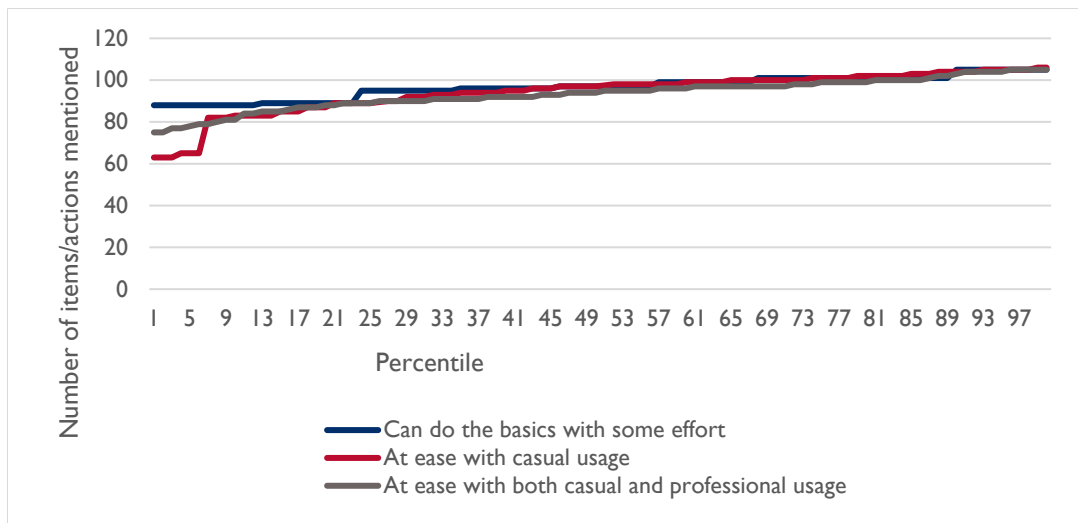
In only one task, English vocabulary, performance was noticeably misaligned with the teacher's self-reported level. In that task, the teachers self-reporting the lowest speaking ability actually performed as well as or better than the higher two groups, except in the 90th percentile and above.

Table 16 summarizes the degree of alignment observed between the self-reported speaking and reading abilities and the performance on related tasks or subtasks. Overall, most tasks and subtasks aligned fairly well with self-reported skill levels, particularly in distinguishing low performers from the mid and high performers, though less well in distinguishing between mid and high performers. The tasks or subtasks with a ceiling effect provided no useful information toward this analysis. These results suggest that the teachers of lower abilities in the sample were indeed generally aware of their lower abilities. Second, they were able to truthfully report their lower ability despite any social desirability pressures to inflate their response.

TABLE 16. DEGREE OF ALIGNMENT BETWEEN SELF-REPORTED SPEAKING AND READING ABILITIES AND ACTUAL PERFORMANCE ON RELATED TASKS OR SUBTASKS

ALIGNMENT BETWEEN SELF-REPORTED SKILL LEVEL AND PERFORMANCE	SELF-REPORTED SPEAKING ABILITY IN ENGLISH/LUGANDA	SELF-REPORTED READING ABILITY (ENGLISH ONLY[1])
Aligned at all three self-reported skill levels	English listening task Luganda vocabulary task	English silent reading comprehension task English grammar error identification subtask
Aligned at the lowest self-reported skill level, but mid and high levels' performance was intermixed	English speaking task Luganda vocabulary task	English nonsense word reading subtask English oral reading fluency subtask
Ceiling effect (all self-reported levels' performances were high and indistinguishable)	Luganda speaking task Luganda listening task (above the 34th percentile)	English oral reading comprehension subtask
Unclear	—	English letter-sound subtask[2]
Misaligned	English vocabulary task	—

[1] Due to a translation error in the interview, the Luganda data related to self-reported reading ability were skewed and could not be analyzed.

[2] The high self-reported skill-level group's performance was consistently the highest, and the low group's was consistently the lowest. However, the mid group's performance sometimes overlapped with the low group's and sometimes with the high group's.

6. WHAT ARE THE SYNTAX RESPONSE PATTERNS FOR DESCRIBING A PICTURE?

For the speaking task, teachers were asked to describe a picture in as much detail as possible. In addition to marking the number of different items or actions mentioned, the assessors also scored the teachers on the dominant response pattern they used when speaking from among the following options: incomplete sentences (i.e., listing items), grammatically incorrect sentences, grammatically correct sentences, and grammatically correct sentences that extended beyond the illustration (see *Table 17*). This metric was designed to complement the quantitative data gathered during the speaking task with a qualitative characterization of the teachers' spontaneous syntactic patterns. The results in Table 17 align with the teacher language background interview in terms of the home language data that show Luganda as the L1 of most Luganda teachers and English as an L2 of most of the English teachers (*Table 13*). The results also align with the teachers' relative self-reported comfort levels in speaking each language (*Table 14*).

For this task, the Luganda teachers mentioned substantially more items/actions in Luganda than the English teachers did in English (*Table 9* and *Table 10*) were more likely to use the most sophisticated response pattern.

TABLE 17. DOMINANT RESPONSE PATTERNS FROM THE SPEAKING TASK

DOMINANT RESPONSE PATTERN	% OF ENGLISH TEACHERS (*n* = 149)	% OF LUGANDA TEACHERS (*n* = 147)
The teacher just listed the items and actions in mostly incomplete sentences, possibly with pauses between items.	15	14
The teacher used sentences but with some grammatical errors.	9	5
The teacher used sentences that were grammatically correct.	60	53
The teacher used sentences that were grammatically correct and extended the illustration.	16	27

7. DOES A PRIMARY GRADE 4 PIECE OF WRITING SHOW VARIABILITY IN TEACHERS' ABILITY TO IDENTIFY AND CORRECT ERRORS?

Scores on this subtask exhibited wide variability, ranging from 0% to 100% in English and from 0% to 90% in Luganda, with a mean of 67% correct in English and 35% in Luganda *(Table 18)*. The Rasch analysis showed that for the English version, the TAM was higher than the IDM, indicating that this task was relatively easy for the teachers who took it. The opposite was true for the Luganda version; the TAM was slightly lower the IDM, meaning that it was relatively difficult for the teachers who took it. These results suggest that the teachers who took the English test were more skilled at correcting student writing errors in English than the teachers who took the Luganda test. Although the same skills were targeted in both versions (i.e., spelling, capitalization, punctuation, and grammar), the task may not have been adapted into Luganda at a difficulty level comparable to the English version. Further experimentation with adapting the task into new languages may clarify these results. In both versions of the assessment, the low number of items prevented the subscale from being sensitive enough to distinguish between high and low performers. Increasing the number of items would increase the sensitivity of the task, though this would also increase the administration time.

TABLE 18. SCORES ON THE CORRECTING STUDENT WRITING SUBTASK, BY LANGUAGE

ASSESSMENT LANGUAGE	RANGE	MEAN (%)	SD	MEDIAN (%)	MODE (%)
English	[0, 100]	67	23	70	70
Luganda	[0, 90]	35	17	30	30

8. WHAT IS THE RELATIONSHIP BETWEEN ORAL READING RATES AND READING COMPREHENSION?

Fluency is known to have a strong influence on comprehension, especially in the early stages of reading development. The mind must be able to decode and recognize words automatically enough to be able to divert conscious attention from the decoding process itself to comprehension.

The Pearson correlation was calculated between the oral reading fluency rate and the oral reading comprehension score in order to examine the relationship between these two skills in the population sampled. The Pearson correlation between these two subtasks was weak in English (0.39) and strong in Luganda (0.70). The relatively high oral reading fluency rates that the English teachers demonstrated in

English, with a mean reading rate of 108 correct words per minute (with 98% accuracy) may account for this difference. At that rate, reading comprehension skills, rather than slow decoding, were hindering understanding. With the caveat that it is not appropriate to compare oral reading fluency rates directly across languages due to substantial differences in average word lengths and structures, the Luganda teachers read at a mean rate of 62 correct words per minute (and 99% accuracy). At that mean rate, the strong correlation between the oral reading fluency rate and oral reading comprehension in Luganda suggests that decoding speed still played a role in those teachers' ability to understand. If teachers decode fast enough, little else would block their comprehension. This finding aligns with the data from the language background interview and suggests that Luganda was a first language for most of these Luganda teachers.

RESULTING RECOMMENDATIONS, BY TASK

The pilot TLLA includes 14 tasks and subtasks to measure reading, writing, listening, and speaking skills of teachers in their required language of instruction. As it stands, the assessment requires about one hour of a teacher's time to complete. The pilot purposely included more subtasks than is likely feasible to administer in a typical data collection. Some of the subtasks are administered individually and others can be administered in a group setting with multiple teachers. This section of the report makes recommendations about each task to guide researchers and stakeholders in deciding which tasks they might include in future studies, depending on their research purpose and available resources.

LANGUAGE BACKGROUND INTERVIEW TASK

The language background interview provided valuable information that complemented and aligned with teacher performance data on the assessment tasks. Including this interview with each assessment would significantly strengthen the results and analysis.

In this study, the results suggested that teachers were at least moderately able to recognize and report their own strengths and weaknesses. Replicating this study elsewhere would make it feasible to learn whether different contexts produce similar results. If so, depending on the research purpose, an interview might be sufficient, and could stand in for an otherwise time-consuming and cost-prohibitive assessment of teachers' language skills. Self-reported data will never be as thorough, granular, or quantifiable as assessment data, and theoretical bias cannot be completely eliminated. Nonetheless, if the research purpose aims to ascertain how teachers feel about their own language skill levels in the various languages of instruction, the language interview alone has the potential to identify places where there is a large mismatch between teachers' comfort levels and the requirements of their positions.

SPEAKING TASK

In this study, the speaking task resulted in much greater variability *between* the language groups than it did *within* them. Unfortunately, due to logistical considerations introduced at the pilot stage, each language version was administered to a different population. Therefore, the analysts could not draw any strong conclusions about the relative speaking abilities of the teachers in English and Luganda. In addition, because the original intent was to administer the task in both languages to each participant, two different picture stimuli were chosen. The drawback was that this decision limited the ability to ascertain the influence of the stimulus itself on the quantity of items mentioned. An area for future research would be to administer the task with one stimulus in both languages in order to better gauge relative speaking ability.

The research team found the speaking task to be worthwhile. The two scoring methods—i.e., counting the number of items or actions mentioned, and characterizing the dominant response pattern—complemented each other. The teachers were not intimidated by the task, and it functioned well as the opening task to set them at ease. Further, the task was not difficult to administer. The main challenge was in culling the list of possible items/actions that respondents could mention to a number that would be manageable for the assessor, while still retaining enough response options to allow the task to distinguish between high and low performers. When developing such a list, instrument designers will need to consult with speakers with advanced vocabulary first to create a comprehensive checklist. They must then pilot and refine the list, ultimately producing an optimal mixture of frequently and rarely mentioned items.

VOCABULARY TASK

This task performed well in both languages in this study. However, the scoring rubric was subjective, and it was time-consuming to train the assessors to score it reliably. This task would be particularly useful for research purposes directly related to vocabulary knowledge and growth, such as specialized vocabulary in a given domain. It may not be absolutely essential for research purposes targeting language ability in general. The speaking and silent reading comprehension tasks also have the potential to offer indirect insights about a teacher's vocabulary. For future consideration, another simple method for understanding vocabulary knowledge would be presenting a list of both real words and nonsense words that appear real to serve as distractors, and asking participants if they know the words.[9] The research team opted not to use this format because of validity issues (i.e., some people dismiss the format as a measure of language knowledge). However, the challenge of reliably scoring vocabulary via an expressive task makes this real/nonsense methodology worth future consideration.

LISTENING TASK (SENTENCE REPETITION)

For this study, the listening task showed a ceiling effect in Luganda, indicating that it may be more appropriate for assessing teachers in a language that is not their L1. Even in English, it showed strong alignment with the teachers' self-reported speaking ability. Thus, for general research purposes, the language background interview may suffice. Unless listening comprehension is specifically targeted in the research purpose, it may be advisable to omit this task from the assessment. If it is retained, a next step for any tailored instrument should be experimenting and piloting with a wider array of sentence structures and lengths in order to ensure that the item difficulty spread is a good fit for the target audience.

ORAL READING TASK: LETTER-SOUND IDENTIFICATION SUBTASK

The four oral reading subtasks, beginning with this one, mirrored the same subtasks within the student EGRA. They would, therefore, be useful if the TLLA were to be given in conjunction with an EGRA for the purpose of exploring the relationship between the teachers' mastery of these skills and students' outcomes.

Otherwise, the letter-sound subtask would be useful whenever the research purpose is to ascertain how well lower primary teachers know their letter sounds or if they are expected to teach phonics. Some teachers expressed their dislike for this subtask if they "had not been trained by [the project]."

---

[9] Anderson, Richard C., and Peter Freebody. 1982. *Reading Comprehension and the Assessment and Acquisition of Word Knowledge.* Technical Report No. 249. Urbana: University of Illinois, Center for the Study of Reading. https://www.ideals.illinois.edu/bitstream/handle/2142/17540/ctrstreadtechrepv01982i00249_opt.pdf?sequence=1&origin=publication_detail

Nevertheless, they also acknowledged that they need to know the letter sounds if they are expected to teach them.

ORAL READING TASK: NONSENSE WORD READING SUBTASK

Similar to the letter-sound identification subtask, the nonsense word reading subtask is useful for assessing teachers' knowledge of phonics in contexts where they are expected to teach phonics.

ORAL READING TASK: ORAL READING FLUENCY SUBTASK

Oral reading fluency has a strong theoretical link to comprehension, which itself is quite challenging to measure. This subtask, however, was easy and efficient to administer. It is important to note that oral reading fluency is highly variable even within an individual; that is, it can fluctuate considerably based on the difficulty level and genre of the text, the familiarity and appeal of the topic to the reader, and the reader's attention. Another factor to consider when deciding whether to utilize this subtask is teachers' willingness to be assessed with a subtask they know is also used with their students (the sampled teachers did not object to it in this instance). In this study, a simple P2 narrative, as employed in the EGRA, resulted in ceiling effects in both languages. Based on these results, the same passage needs to be used with both teachers and students if the TLLA is to be administered in conjunction with an EGRA to permit direct comparison. But if the research purpose is not tied to an EGRA, a more challenging passage could render more useful information—assuming that the target population has reading abilities similar to those of the teachers who participated in the pilot sample. Piloting several passages at different difficulty levels could help to pinpoint the best fit for the target population.

ORAL READING TASK: ORAL READING COMPREHENSION SUBTASK AND SILENT READING COMPREHENSION TASK

The oral reading comprehension subtask in this study had a ceiling effect in both languages, indicating that it was too easy for the population sampled. The silent reading comprehension task was more appropriately leveled for this population, had a greater number of items, and showed greater variability in results. In populations where teachers have oral reading abilities similar to those in this sample, the oral reading comprehension subtask may not be necessary, and the silent reading comprehension task may be more informative. However, if one purpose of an assessment was to analyze teacher scores against student scores in oral reading comprehension, the teachers would need to complete the same task as the students. One distinct efficiency advantage of the silent reading comprehension task is the potential to administer it to a group of teachers simultaneously, perhaps at a teacher training.

GRAMMAR TASK: STRUCTURE AND WRITTEN EXPRESSION SUBTASK AND ERROR IDENTIFICATION SUBTASK

The grammar subtasks assessed the same underlying construct in two different ways. In this study, the English subtasks performed well and almost identically to each other. In Luganda, the error identification subtask appeared to be a slightly better fit for the population sampled. Constructing a useful grammar task in an additional language is challenging, and piloting multiple items, multiple times will likely be necessary to achieve a good fit.

For the pilot, the assessors administered two grammar subtasks with different formats and five items each. In the future, a better approach would entail choosing only one format and increasing the number of items from five to 10 to increase the task's sensitivity. This modification would shorten administration time because there would be only one set of instructions and examples. In this study, the error

identification subtask took slightly longer to administer one-on-one than the structure and written expression subtask; however, because of its written format, it could be administered to a group. The choice between the two formats might come down to preference for an oral administration versus a written administration. The error identification subtask, on the other hand, should not be administered orally because it is difficult for assessors to read errors aloud naturally, without subtly cueing their presence (e.g., by pausing or stumbling over the error). If prohibitively low reading ability is suspected in a target population, the structure and written expression subtask could be administered both in writing and orally (as it was done here) to help to mitigate the conflation of reading and grammar ability.

WRITING TASK: CORRECTING STUDENT WRITING SUBTASK

Although not as informative as the writing prompt subtask, the correcting student writing subtask was an efficient way to measure some writing subskills, particularly those related to writing conventions and mechanics such as spelling, punctuation, capitalization, and grammar. If these subskills were important to a given study's research purpose but logistical constraints did not allow for a writing prompt subtask, this subtask would be another option.

WRITING TASK: RESPONDING TO A WRITING PROMPT SUBTASK

The writing prompt subtask provided the richest information about teachers' writing skills. However, it was difficult to achieve scoring agreement among the assessors and time-consuming to mark, so it may not be very practical for future users to administer. Some of the skills addressed in the writing prompt scoring rubric are partially assessed under other tasks. For example, the skills targeted by the word-choice criterion are addressed by the vocabulary task, and the language usage criteria are addressed by the grammar subtasks. If a study's purpose was directly tied to composition writing, such as an impact evaluation of an intervention aiming to improve teacher writing skills, this task might be worthwhile to administer. In that case, there is much room for piloting and refining the scoring rubric. For most general large-scale assessments, the logistical challenges associated with administering this subtask probably would be prohibitive.

WRITING TASK: SPELLING SUBTASK

Spelling skills are closely linked to reading, writing, and comprehension skills. In some respects, spelling is to writing as reading fluency is to reading comprehension. When writers have trouble with spelling, their attention is diverted from the higher-level thought processes needed to organize and produce good writing. The spelling subtask would be particularly appropriate in a context where writing skills were suspected to be weak and logistical constraints would prevent the administration of a composition task. The spelling subtask offers many advantages, such as easy development, easy administration (including in groups), high validity, and high internal consistency. Piloting of multiple items should be done to ensure the best fit for the target audience.

GENERAL CONSIDERATIONS FOR FUTURE USE

Striking the right balance between gathering enough data without taking too much of the teachers' valuable time was one of the primary concerns throughout the process of developing this tool. Researchers and stakeholders who intend to use the TLLA may need to prioritize a subset of tasks and subtasks from the full suite, depending on the logistical constraints of their data collection context and the specific data needs of their research agenda. The research team offers the following general recommendations to future users of the TLLA to guide that selection.

Individually administered assessments incur higher assessor costs and administration time. Group administration is one way to assess a large number of teachers at once with only one assessor, including during a teacher training or other gathering. The following three tasks may be efficiently administered and sufficient for general research purposes: silent reading comprehension, grammar (both subtasks), and spelling. If the teachers' reading abilities are adequate, the language background interview could also be administered in a guided format with the assessors reading the questions and the teachers following along on paper and selecting their response to each item. These four tasks would take approximately 25 minutes to administer in a group in written format.

Conversely, individual oral administration may be more feasible in certain contexts, such as in a school setting in which teachers cannot leave students unattended and thus are available only one at a time. In that case, the following tasks and subtasks would serve general research purposes: the language background interview, speaking, oral reading fluency subtask (at a grade level higher than P2), silent reading comprehension, grammar (both subtasks), and spelling. These five tasks and subtasks would take about 27 minutes to administer in an individual, oral format.

RECOMMENDED NEXT STEPS

This report has outlined the process undertaken to develop and test the two language versions of the TLLA. The tool refinement benefited from insights gained through two rounds of cognitive interviewing, a field test, and one pilot in Uganda. In keeping with best practices for assessment development, and in order to confirm that the results are replicable in other populations, the TLLA should be rigorously adapted and piloted in various contexts. With each iteration, further refinements can be made to improve the reliability of the TLLA. Of particular interest is whether the relationship between the self-reported data in the language background interview and the participants' actual performance on related tasks is replicable.

Future pilots would permit further refinement of the tasks and subtasks, as shown in *Table 19.*

TABLE 19. AGENDA FOR ADDITIONAL PILOT, BY TASK	
TASK OR SUBTASK	AGENDA FOR AN ADDITIONAL PILOT
Speaking	Refine the list of items in the original task; pilot additional stimuli; administer the task to the same participants in two languages. Although there likely would be some inflated scores the second time the assessor administered the task (i.e., priming effects), it would be worth understanding the differences within a teacher expected to teach in two languages.
Listening	Refine the original list of items and pilot additional items of varying lengths and complexities.
Vocabulary	Pilot additional items to create a larger item bank (in English). Use alongside other tool with known/unknown format.
Oral reading fluency and comprehension	Pilot additional texts and comprehension questions written at a higher reading level.
Grammar	Pilot additional items to create a larger item bank for each format.
Silent reading comprehension	Refine the problem items in the original task and re-pilot; pilot additional texts and comprehension questions.

TABLE 19. AGENDA FOR ADDITIONAL PILOT, BY TASK

Writing: Correcting student writing	Pilot additional items.
Writing: Responding to a prompt	Pilot additional prompts. Test scoring the writing prompt compositions in terms of number of words, average sentence length, and average word length, and compare those quantitative metrics with the results of the writing rubric to see if the quantitative method of scoring offers a more efficient way to get useful data.
Spelling	Pilot additional items.

One eventual goal for the TLLA is to explore how teachers' language and literacy proficiency influence students' learning outcomes. This first pilot of the TLLA is a promising start toward the development of a valid and reliable tool to measure teachers' language abilities. Subsequent pilots of the TLLA could help to identify the types of research questions this tool can address. For example, future research streams might compare students' EGRA performance with teachers' performance on the TLLA to more fully explore the link between teachers' language proficiency and students' reading acquisition.

The TLLA has the potential to be and informative for understanding the language assets and needs of teachers, just as the EGRA has been helpful to understand student reading abilities. [10] As such, the research team recommends that education researchers and stakeholders experiment with adapting the tasks and subtasks in the TLLA suite and administering them in other contexts and languages.

---

[10] Dubeck, M. M., & Gove, A. (2015). The Early Grade Reading Assessment (EGRA): Its theoretical foundation, purpose, and limitations. International Journal of Education Development, 40, 315–322. http://dx.doi.org/10.1016/j.ijedudev.2014.11.004

GLOSSARY

Adaptation. The process of transforming an assessment from the source language/culture into a target language/culture. It is not a direct translation from the original language to another.

Agreement. The extent to which assessors made the same ratings about how to score a participant's response to an assessment item.

Ceiling effect. Occurs when there is an artificial upper limit on the possible values for a variable and a large concentration of participants score at or near this limit. This is the opposite of the *floor effect* (see below).

Cognitive interviewing. A process to improve the quality of survey items and testing protocols. A participant is asked to share their thinking on why they responded in a certain way and how they interpreted the instructions.

Construct. The subject or the skills that is intended to be measured with an assessment item or an assessment subtask.

Cronbach's alpha. A measure of the internal consistency of a test containing items that are not scored dichotomously, based on the extent to which test-takers who answer a given test time one way respond to other items in a similar way.

Floor effect. Occurs when there is an artificial lower limit on the possible values for a variable and a large concentration of participants score at or near this limit. This is the opposite of the *ceiling effect* (see above).

Internal consistency. A reliability measurement to determine if the items in a subtask are related to determine how well they measure the same concept or construct.

Fluency. Being able to read words quickly, accurately, and with expression (i.e., prosody).

Lookback. A procedure used in a reading test to see if the reader has the ability to return to the text they just read to skim for the answer.

Orthography. The written representation of the sounds of a language; spelling.

Phonics. An instructional approach that teaches the relationship between letters and the sounds they represent. It supports decoding (reading) and encoding (spelling) skills.

Population. The theoretical group of subjects (individuals or units) to whom a study's results can be generalized. The *sample* (see below) and the population share similar characteristics, and the sample is a part of the population of interest.

Prosody. Reading with expression, proper intonation, and phrasing.

Rasch analysis. A statistical technique to examine the effectiveness of each item within an assessment subtask. It is based on a probabilistic model where the likelihood of a participant's correct response is a function of the participant's skill or difficulty of the item.

Redundancy. When developing an assessment, items or subtasks are examined to understand if they are duplicating measurement in another part of the assessment. Reducing redundancy can shorten the administration time or inform the administration protocol.

Reliability. Refers to the consistency or stability of the data.

Rubric. A set of criteria to rate performance levels. Among other skills, a rubric can be used to rate expressive language or responding to a writing prompt.

Sample. The group of subjects (individuals or units), from a population, selected to be in a study.

Syntax. The rules governing the ordering of words in a sentence.

Tier 2. A categorization of vocabulary words that are used across domains (e.g., commotion). As a comparison Tier 1 words are everyday words (e.g., dog, happy) and Tier 3 words are technically specific words (e.g., stethoscope).

Validity. The extent an instrument measures what it is intended and designed to measure.

Writing prompt. In educational assessment, a direction to the participant to write about a particular topic in a particular format. Using one helps to increase the consistency in scoring across participants.

REFERENCES

Albaugh, E. A. (2014). *State-Building and Multilingual Education in Africa*. New York, NY: Cambridge University Press.

Anderson, R. C., & Freebody, P. (1982). Reading Comprehension and the Assessment and Acquisition of Word Knowledge (Vol. 249). Urbana: University of Illinois, Center for the Study of Reading.

Boone, W. J. (2016). Rasch analysis for instrument development: why, when, and how? *CBE Life Sciences Education, 15*(4), 4. http://dx.doi.org/10.1187/cbe.16-04-0148

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

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134. http://dx.doi.org/10.1037/0022-3514.77.6.1121

Streiner, D. L. (2003, Feb). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99-103. http://dx.doi.org/10.1207/S15327752JPA8001_18

Teacher Language and Literacy Assessment (TLLA)

English

Assessor Protocol Booklet

Pilot Test – February 2020

Establish Rapport, Obtain Participant Consent, and Assign Code

- Establish rapport by greeting the teacher warmly, exchanging pleasantries, and thanking them for their time.
- Give the teacher a copy of the Teacher Consent form. Read it aloud to the teacher.
- If the teacher consents to participate and signs the form, assign the teacher a temporary identification code (i.e. your initials and the number of teachers you have interviewed so far, e.g. KH02) and write it in the box below.
- Write the same code on the front of a copy of the Teacher Booklet for Silent Reading and Writing Tasks. Keep the booklet in your possession until it is time for those tasks.
- If the Silent Reading and Writing Tasks will be administered in a group, write the same code on a separate piece of paper and give it to the teacher to keep until the time of the group-administered tasks.

Teacher Code:	
Assessor Name:	
Date:	
School Name:	
Is the teacher female?	□ Yes □ No
Which grade(s) does the teacher teach?	□ P1 □ P2 □ P3 □ P4 □ P5 □ P6 □ Other

Set the Teacher Booklet for Oral Tasks in front of the Teacher.

Language Background Interview

[Return to Language Background Interview Task section of report]

Start Time: _____

Instructions to the Assessor:

> Show the teacher the Language Background questions on Pages 1-2 in the Teacher Booklet for Oral Tasks. Read the instructions below to the teacher.
>
>
> **Let's get started. First, I am going to ask you some questions about your language background. You can follow along here** [point to the questions] **as I read the questions and answers aloud. Some of these questions are based on your own judgment, and there is no right or wrong answer.**

Read the questions and all the response options aloud to the teacher, one question at a time. Do not read aloud any parts in brackets. Write or tick the teacher's response. For each multiple-choice question, tick only one response, unless otherwise indicated. If the teacher refuses to answer any question, tick No Response.

1a. When you were a young child, which language did you use the most frequently at home with your family? [Select only one]

[□ Acholi □ <u>Ateso</u> □ English □ Kiswahili □ Leblango □ Luganda

□ Lugbarati □ Lugwere □ Lukhonzo □ Lumasaba □ Lusoga □ Nkarimojong

□ Runyankore-Rukiga □ Runyoro-Rutoro □ Other_____ □ No Response]

1b. Nowadays, which language do you use the most frequently at home with your family? [Select only one]

[□ Acholi □ <u>Ateso</u> □ English □ Kiswahili □ Leblango □ Luganda

□ Lugbarati □ Lugwere □ Lukhonzo □ Lumasaba □ Lusoga □ Nkarimojong

□ Runyankore-Rukiga □ Runyoro-Rutoro □ Other_____ □ No Response]

1c. In addition to the language given in 1b, do you use any other languages frequently at home? □ Yes □ No [If no, skip Question 1d] [□ No Response]

1d. If yes, which language or languages? [Multiple responses allowed]

[□ Acholi □ <u>Ateso</u> □ English □ Kiswahili □ Leblango □ Luganda

□ Lugbarati □ Lugwere □ Lukhonzo □ Lumasaba □ Lusoga □ Nkarimojong

□ Runyankore-Rukiga □ Runyoro-Rutoro □ Other_____ □ No Response]

2. **Choose the one sentence that best describes your <u>speaking ability</u> in English.**

   □ *A) I do not speak English.* [If A is selected, skip Question 3]

   □ *B) With some effort, I can use English in conversation.*

   □ *C) I can easily use English in conversation, but I sometimes have challenges knowing all the words I need to know when teaching in English.*

   □ *D) I can easily use English both for conversation and for teaching.*

   [□ No Response]

3. *Choose the one sentence that best describes your <u>pronunciation</u> in English.*

□ *A) My pronunciation in English often makes it difficult for other adults to understand me.*

□ *B) I sometimes struggle to pronounce some specific letter sounds and words in conversation in English.*

□ *C) Other adults generally understand my pronunciation in conversation in English, but I sometimes have challenges pronouncing some letter sounds and words when teaching in English.*

□ *D) Other adults understand my pronunciation in English, and I can easily pronounce letter sounds and words when teaching in English.*

[□ No Response]

4. *Choose the one sentence that best describes your <u>reading ability</u> in English.*

□ *A) I cannot read with understanding in English.*

□ *B) With some effort, I can read and understand basic English.*

□ *C) I can easily read and understand basic English, but I sometimes have challenges reading unfamiliar words or understanding parts of the teacher's guide, pupil textbooks, or other materials that we read in English in class.*

□ *D) I can easily read and understand both basic English and all of the materials that we read in English in class.*

[□ No Response]

5. *Choose the one sentence that best describes your <u>writing ability</u> in English.*

□ *A) I cannot write in English.*

□ *B) With some effort, I can write basic English.*

□ *C) I can easily write basic English, but I sometimes have challenges modeling good writing in English in class.*

□ *D) I can both easily write basic English and model good writing in English in class.*

[□ No Response]

6a. *Were you ever taught <u>how to read</u> in English?* □ Yes   □ No [□ No Response]

[If no, skip Question 6b]

6b. *If so, in what context were you taught <u>how to read</u> in English? (You can choose   more than one.)*

　　□ *At primary school*

　　□ *At secondary school*

　　□ *At a teacher training college or university*

　　□ *In a non-formal setting (e.g. adult literacy class, at church, at mosque, by a family   member or friend, etc.)*

　　□ *At an in-service training workshop*

　　□ *Other, please specify:*

　[□ N/A or □ No Response]

7a. Have you received any training to teach reading in any language?

　　□ Yes　　□ No　[If no, skip Question 7b] [□ No Response]

7b. Was the training you received for teaching reading designed to teach reading in a   specific language? □ Yes　　□ No　[If no, skip Question 7c]

　　[□ N/A or □ No Response]

7c. *If yes, in which language or languages?*

　　[□ N/A or □ No Response]

| When finished, say: ***Thank you! Let's go on to the next section.*** |

End Time:

A. Speaking Task

[Return to Speaking Task section of report]

Start Time: _____

Instructions to the Assessor:

Show the teacher the picture for the Speaking task on Page 3 in the Teacher Booklet for Oral Tasks and say: ***For our first activity, please look at this picture. Please describe everything that you see in the picture in as much detail as possible. That is, describe the people, the objects, what they look like, and what they are doing.***

☺ Set the timer for 2 minutes.

Are you ready? Please begin.

Start the timer when the teacher starts talking. Follow along on your score sheet and tick all the items that the teacher mentions from the list. Stay quiet, except if the teacher hesitates for 10 seconds, say: *You can tell me about anything you see in the picture.*

If the teacher exhausts their initial description before the timer runs out, say: *What do you imagine life is like in this family?*

Early stop rule: If the teacher does not say anything at all for 20 seconds, even after the second prompt, say: *Thank you, that's all.* Discontinue this task, tick the box at the bottom of the task, and continue to the next task.

If the timer runs out before the teacher finishes speaking, say: *Thank you, that's all.* If the teacher is almost finished, you may let them finish; you do not have to interrupt them. Either way, tick the final item mentioned when the timer ran out; do not count any items mentioned after the end of the timer.

At the end of the timed portion, tick the **one** option under the Utterances Description section of your score sheet that best describes the overall response pattern used by the teacher.

Task Stimulus:

Assessor Score Sheet:

February 2020 pilot: This checklist was printed for administration and responses were entered in Tangerine after the teacher was dismissed.

Items Mentioned (Tick all that are said. For this phase only, write in any additional nouns, verbs, adjectives or adverbs that the teacher mentions that are not on this list.)

☐	☐	☐
☐ activities	☐ gentleman	☐ rack
☐ baby	☐ girl	☐ responsible
☐ ball	☐ grass	☐ road
☐ barefoot	☐ ground	☐ rubbish
☐ basket	☐ hair	☐ sandals
☐ bird	☐ happy	☐ sanitation
☐ boil	☐ helping	☐ sauce
☐ boy	☐ hem	☐ saucepan
☐ broom	☐ hen	☐ seat/seated
☐ brother	☐ hoe	☐ seed
☐ busy	☐ hold	☐ shoes
☐ carry	☐ home	☐ shorts
☐ catering skills	☐ house	☐ sister
☐ chair	☐ husband	☐ sit
☐ charcoal	☐ hygienic	☐ slipper (s)
☐ chicken	☐ knife	☐ small
☐ child(ren)	☐ lady	☐ smart
☐ clean	☐ little girl	☐ smile
☐ clear	☐ man	☐ spoon
☐ cock	☐ matoke	☐ stand
☐ communal work	☐ middle	☐ standing
☐ compound	☐ mother	☐ stool
☐ container	☐ move	☐ stove
☐ cook	☐ nuclear	☐ strand
☐ cooperative	☐ objects	☐ sweep
☐ crowing	☐ organized	☐ table
☐ cups	☐ outside	☐ taught
☐ cut	☐ parents	☐ trousers
☐ dig	☐ peel	☐ utensils
☐ dishes	☐ people	☐ veranda
☐ domestic work	☐ picking	☐ village
☐ dress	☐ place	☐ walk
☐ environment	☐ plaited	☐ wash
☐ exercise	☐ plate stand	☐ water
☐ family	☐ play	☐ weed
☐ farmer	☐ pot	☐ woman
☐ father	☐ prepare	☐ wrapper
☐ folded		☐ young
☐ food		☐ other
☐ foreground		

If the teacher finished with time remaining, record the seconds remaining: |____|____|____|

Utterances Description Tick the dominant response pattern.	Example of each pattern.
☐ The teacher just listed the items and actions in mostly incomplete sentences, possibly with pauses between items.	*a girl, boy, stove, father, hen…* *sweeping, washing, cooking …*
☐ The teacher used sentences but with some grammatical errors.	*The girl wash utensils. Baby carry ball.*
☐ The teacher used sentences that were grammatically correct.	*The girl is nice. She is washing some utensils. A baby is carrying a ball.*
☐ The teacher used sentences that were grammatically correct and extended the illustration.	*The girl is washing utensils and smiling. Maybe she is thinking about something pleasant or singing a song while she works.*
☐ Task discontinued because the teacher had no response. ☐ Task skipped because teacher refused.	
When finished, say: **Thank you! Let's go on to the next section.**	

End Time: _____

B. Vocabulary Task

[Return to Vocabulary Task section of report]

Start Time: _____

Instructions to the Assessor:

Show the teacher the Vocabulary task on Page 4 in the Teacher Booklet for Oral Tasks as you read the instructions below.

Explaining new vocabulary words to pupils is an important part of our role as teachers. In this activity, I will read each word to you, and you will provide a definition and an example for that word.

[Point to the first example word and say:] ***For example, this word is exhausted. A definition of exhausted may be "feeling so tired you can hardly move." An example may be "I was exhausted after working in the garden all day."***

You can provide the explanation in any language, as you might do in the classroom. Some of the words are less common than others. If you are not familiar with a word and want to skip it, that is fine.   Are you ready? Let's begin.

Read each word aloud, record the teacher's explanation, and rate the explanation on your score sheet. If the teacher provides only a definition, ask them to add an example. If the teacher provides only an example, ask them to add a definition. Then continue to the next word.

Stay quiet, except if the teacher hesitates for 10 seconds. If so, say: *How would you explain _____?* If the teacher hesitates for another 5 seconds, move on to the next word.

Early stop rule: If the teacher does not provide any response at all for the first three items, say: *Thank you, that's all.* Discontinue this task, tick the box at the bottom of the task, and continue to the next task.

Word	Sample Definition (DO NOT READ ALOUD)	Evaluate understanding of the word's meaning, as evidenced by the explanation
1. **effect**	*a change; to cause something* *Example: The effect of global warming is changing weather patterns.*	☐ Very good (thorough and accurate) ☐ Fair (basic and/or partially inaccurate) ☐ Poor (inaccurate) ☐ No response
2. **expose**	*to make something visible; to show the truth* *Example: The mini dress exposes too much skin to wear to church.*	☐ Very good (thorough and accurate) ☐ Fair (basic and/or partially inaccurate) ☐ Poor (inaccurate) ☐ No response
3. **regret**	*to feel sad or disappointed; a feeling* *Example: I regret that I played football yesterday. I was not ready for the test.*	☐ Very good (thorough and accurate) ☐ Fair (basic and/or partially inaccurate) ☐ Poor (inaccurate) ☐ No response
4. **advantage**	*to be a better position* *Example: An advantage of living in the city are better schools.*	☐ Very good (thorough and accurate) ☐ Fair (basic and/or partially inaccurate) ☐ Poor (inaccurate) ☐ No response
5. **curious**	*to want to know how something works* *Example: The girl stayed up late to finish the mystery book because she was curious how it would end.*	☐ Very good (thorough and accurate) ☐ Fair (basic and/or partially inaccurate) ☐ Poor (inaccurate) ☐ No response
6. **adapt**	*to change to fit the situation* *Example: The nursery teacher will adapt the long story so her young learners don't have to sit still too long.*	☐ Very good (thorough and accurate) ☐ Fair (basic and/or partially inaccurate) ☐ Poor (inaccurate) ☐ No response
7. **innovator**	*a person who creates or tries new solutions to solve problems. Example The person who created mobile money was an innovator who gave us new ways to manage money.*	☐ Very good (thorough and accurate) ☐ Fair (basic and/or partially inaccurate) ☐ Poor (inaccurate) ☐ No response
8. **precaution**	*to take steps to avoid problems later Example: As a precaution, the mother took all the small things off the table so the baby would not put them in his mouth.*	☐ Very good (thorough and accurate) ☐ Fair (basic and/or partially inaccurate) ☐ Poor (inaccurate) ☐ No response
9. **erode**	*to gradually wear away; it can be natural like soil or human relationships.* *Example: The girl was not dependable to her friends. She broke promises and was late. Her mother warned her that this could erode those friendships and she will be left without friends.*	☐ Very good (thorough and accurate) ☐ Fair (basic and/or partially inaccurate) ☐ Poor (inaccurate) ☐ No response

Word	Sample Definition (DO NOT READ ALOUD)	Evaluate understanding of the word's meaning, as evidenced by the explanation
10. **inspiration**	*A source of hope or motivation* *Example: The professional runner Stephen Kiprotich who covers long distances is an inspiration for me to do shorter runs to stay healthy.*	☐ Very good (thorough and accurate) ☐ Fair (basic and/or partially inaccurate) ☐ Poor (inaccurate) ☐ No response

☐ Task discontinued (items 1-3 were incorrect)   End Time: _____

C. Listening Task

Start Time: _____

Instructions to the Assessor:

> There is no stimulus for the teacher for this task. Set the Teacher Booklet aside so as to not be a distraction. Read the instructions below to the teacher.
>
> ***For this activity, I will read some sentences aloud in English. Please listen very closely because I will read each sentence only once. After I read each one, please repeat the same exact sentence back to me word for word. The sentences will gradually get longer as we go. If you cannot remember the whole sentence exactly, that's fine, just repeat as much of it as you remember.***
>
> ***Let's start with an example. If I say, "We sing songs in the classroom", then you will just repeat back to me: "We sing songs in the classroom."***
>
> ***Now you try an example. "Children like to go to school."***
>
> > [If the teacher repeats some or all of it, say:] ***That's right.***
> >
> > [If the teacher does not repeat or does not understand what to do, say:] ***OK, all you need to do is repeat the sentence back to me word for word as much as you remember. Let's try again: "Children like to go to school." …***
>
> ***Remember, I can only say each sentence once. Are you ready? Let's begin.***
>
> Read each sentence aloud clearly and naturally. Do not read too fast or too slowly. Read each sentence only one time. Listen as the teacher repeats the sentence and follow along with your pen. Clearly mark any incorrect, transposed, or omitted words with a slash (/). Count self-corrections as correct. If you already marked the self-corrected word as incorrect, circle it (ø) and continue. If the teacher cannot repeat any words in a given sentence, draw a line through the whole sentence.
>
> Stay quiet, except if the teacher hesitates for 3 seconds. If so, say: ***Please repeat as much as you remember.***
>
> *Early stop rule:* If the teacher provides no response for any 3 sentences in a row, say: ***Thank you, that's all.*** Discontinue this task, tick the box at the bottom of the task, mark with a bracket (]) the end of the final sentence that you read, and continue to the next task.

Assessor Stimulus and Score Sheet:

1. The girls bought fruits at the market.

2. They arrived in town just as the sun was setting.

3. It was the head teacher who said we should do it this way.

4. The girls won the game because they had trained harder than the others.

5. When John woke up this morning, he discovered that his bicycle had been stolen in the night.

6. Every morning Mother rises early in order to heat water for our bath.

7. She was a capable president, so it's understandable why she was elected three times.

8. As soon as my brother finished his work, he left to visit his friends.

9. The reason they slept on a mat on the floor was because there was no room in the bed.

10. The police have a checkpoint that slows traffic between here and the city.

☐ Task discontinued because the teacher had no response for three sentences in a row.

☐ Task skipped because teacher refused.

When finished, say: **Thank you! Let's go on to the next section.**

End Time: _____

D. Oral Reading Tasks

D1. Letter-Sound Identification Subtask

[Return to Letter-Sound Identification Subtask section of report]

Start Time: _____

Instructions to the Assessor:

Show the teacher the Letter Sounds task on Page 5 in the Teacher Booklet for Oral Tasks as you read the instructions below.

Learning the letters of the alphabet is one of the first steps that our pupils take in learning to read. When we teach a new letter, we teach the sound that it makes.

Here is a page of letters of the English alphabet. Let's say you have taught each of these letter sounds to your pupils. Please tell me the most common sound that each letter makes in English. Not the name of the letter, but the sound that it makes.

Let's start with some examples. [Point to the letter *u*.] *In English, the sound of this letter is /u/ ("uh"). If I were teaching my pupils this letter, I would tell them that it makes the sound /u/ in English.*

Try this one: [Point to the letter *f*.] *What is the sound of this letter in English?*

   [If the teacher says /f/, say:] *That's right.*

   [If the teacher does not say /f/, say:] *OK, in English the sound of this letter is /f/.*

Let's try one more: [Point to the letter *C*.] *What is the sound of this letter in English?*

   [If the teacher says /k/, say:] *That's right.*

   [If the teacher says /s/, say:] *That is one sound of this letter, but can you tell me the most common sound? …*

   [If the teacher does not say /k/ or /s/, say:] *OK, in English the most common sound of this letter is /k/.*

When I say "Begin," start here [point to first letter] *and go across the page* [point]. *Then continue with the next line and so on* [point]. *Point to each letter and tell me the sound of that letter in English. Read as quickly as you can. If you come to a letter that you do not know how to pronounce, just skip it and go on to the next letter.*

☺ Set the timer for 1 minute.

Are you ready? Please begin.

Start the timer when the teacher reads the first letter. Follow along with your pen and clearly mark any incorrect letter sounds with a slash (/). Count self-corrections as correct. If you already marked the self-corrected letter as incorrect, circle it (ø) and continue.

Stay quiet, except if the teacher hesitates for 3 seconds. Point to the next letter and say: *Please go on.* Mark the skipped letter as incorrect.

If the teacher provides the letter name rather than the sound, or if the teacher adds other explanation as if she were teaching, (e.g., "Pupils, this is sound /m/"), say: **Please just say the sound of the letter in English.**

Early stop rule: If the teacher does not provide a single correct response for the first 10 items, say: **Thank you, that's all.** Discontinue this task, tick the box at the bottom of the task, and continue to the next task.

If the timer runs out before the last item is read, say: **Thank you, that's all.** If the teacher is almost finished, you may let them finish; you do not have to interrupt them. Either way, mark with a bracket (]) the final letter read when the timer ran out; do not count any letters that they read after the end of the timer.

If the teacher reaches the last item before the time runs out, stop the timer as soon as the teacher reads the last letter. Note the number of seconds remaining and record it at the bottom of the task.

Assessor Score Sheet:

h	i	A	L	h	S	X	A	L	c	(10)
N	r	c	d	i	T	r	y	s	P	(20)
D	T	s	N	R	O	J	e	H	i	(30)
a	e	L	u	V	g	E	U	t	Z	(40)
e	t	o	E	I	t	S	n	w	e	(50)
W	F	A	n	o	E	G	T	N	R	(60)
M	h	T	b	E	i	n	H	m	T	(70)
O	e	L	D	Y	d	a	f	E	a	(80)
i	U	p	i	N	t	O	Q	h	R	(90)
e	o	C	A	O	e	S	a	K	S	(100)

Time remaining on stopwatch at completion (number of SECONDS): |___|___|
☐ Task discontinued because the teacher had no correct answers in the first line.

☐ Task skipped because teacher refused.

When finished, say: **Thank you! Let's go on to the next section.**

End Time: _____

D2. Nonsense Word Reading Subtask

[Return to Nonsense Word Reading Subtask section of report]

<div align="right">Start Time: _____</div>

Instructions to the Assessor:

Show the teacher the sheet of Nonsense words on Page 6 in the Teacher Booklet for Oral Tasks as you read the instructions below.

One important skill we teach our pupils is how to read new words by sounding them out.

For this activity, let's say that you are teaching your pupils to read some new words they have never seen before, using their knowledge of the letter sounds. Here we will use some nonsense words that are spelled like real words in English. Please tell me the correct way that you would expect the pupils to read each word in English according to the way that word is spelled.

Let's start with some examples. [Point to the word *ud*.] ***This is not a real word in English, but if it were, we would expect a pupil to read this word as /ud/.***

Try this one: [Point to the word *bif*.] ***How would you expect a pupil to read this word?***

    [If the teacher says /bif/, say:] ***That's right.***

    [If the teacher does not say /bif/, say:] ***OK, in English we would read this word as /bif/.***

Let's try one more: [Point to the word *mep*.] ***How would you expect a pupil to read this word?***

    [If the teacher says /mep/, say:] ***That's right.***

    [If the teacher does not say /mep/, say:] ***OK, in English we would read this word as /mep/.***

When I say "Begin," start here [point to first letter] ***and go across the page*** [point]. ***Then continue with the next line and so on*** [point]. ***Point to each word and tell me the correct way that you would expect the pupils to read that word in English. Read as quickly as you can. If you come to a word that you do not know how to pronounce, just skip it and go on to the next word.***

🕐 Set the timer for 1 minute.

Are you ready? You may begin.

Start the timer when the teacher reads the first word. Follow along with your pen and clearly mark any incorrect words with a slash (/). Count self-corrections as correct. If you already marked the self-corrected word as incorrect, circle it (ø) and continue.

Stay quiet, except if the teacher hesitates for 3 seconds. Point to the next word and say: ***Please go on.*** Mark the skipped word as incorrect.

Early stop rule: If the teacher does not provide a single correct response for the first five items, say: ***Thank you, that's all.*** Discontinue this task, tick the box at the bottom of the task, and continue to the next task.

If the timer runs out before the last item is read, say: ***Thank you, that's all.*** If the teacher is in the final five items, you may let them finish; you do not have to interrupt them. Either way, mark with a

---

bracket (]) the final word read when the timer ran out; do not count any words that they read after the end of the timer.

If the teacher reaches the last item before the timer runs out, stop the timer as soon as the teacher reads the last word. Note the number of seconds remaining and record it at the bottom of the task.

Assessor Score Sheet:

lus	paf	sim	zon	maz	(5)
ver	lut	ral	Fid	gax	(10)
rop	teb	fut	Et	sal	(15)
sen	tib	lef	huz	leb	(20)
bif	wix	fim	Riz	ret	(25)
yag	hig	tat	tup	ved	(30)
nad	gof	zib	fol	reg	(35)
dit	san	nep	jod	mib	(40)
sig	peb	dag	nom	nup	(45)
vom	yod	kad	tob	kib	(50)

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

☐ Task discontinued because the teacher had no correct answers in the first line.

☐ Task skipped because teacher refused.

When finished, say: ***Thank you! Let's go on to the next section.***

End Time: _____

D3. Oral Reading Fluency Subtask

Start Time: _____

Instructions to the Assessor:

Show the teacher the oral reading fluency (ORF) passage on Page 7 in the Teacher Booklet for Oral Tasks as you read the instructions below.

We often read stories aloud to our pupils. Reading aloud to them lets us model for them what fluent and expressive reading sounds like.

For this activity, let's say you are going to read the following story to your pupils. First, you will take a moment to skim the story. Then you will read the story aloud to me as if you were reading it to your pupils to model fluent and expressive reading. If there are any words you don't know, just skip them and go on to the next word.   Pay attention to what you are reading, because when you finish, we will do a comprehension activity about the story.

🕐 Set the time for 1 minute.

Are you ready? You may begin.

Start the timer when the teacher reads the first word aloud. Follow along with your pen and clearly mark any incorrect words with a slash (/). Count self-corrections as correct. If you already marked the self-corrected word as incorrect, circle it (ø) and continue.

Stay quiet, except if the teacher hesitates for 3 seconds. Point to the next word and say: ***Please go on.*** Mark the skipped word as incorrect.

Early stop rule: If the teacher does not read a single word correctly in the first line, say: ***Thank you, that's all.*** Discontinue this task, tick the box at the bottom of the task, skip the Oral Reading Comprehension task, and continue to the Silent Reading task.

If the timer runs out before the last item is read, say: ***Thank you, that's all.*** If the teacher is almost finished, you may let them finish; you do not have to interrupt them. Either way, mark with a bracket (]) the final word read when the timer ran out; do not count any words that they read after the end of the timer.

If the teacher reaches the last item before the screen flashes red, stop the timer as soon as the teacher reads the last word. Note the number of seconds remaining and record it at the bottom of the task.

Mark the rating that best characterizes how expressive the teacher was in reading the passage using intonation (i.e. expressing meaning through strategic variation in vocal pitch and volume).

Assessor Score Sheet:

My name is Pat. I live on a farm with my mother, father, and brother. The land gets very dry. Every year we watch the sky and look for the rain. One afternoon as I sat outside, I saw dark clouds. Then something hit my head, lightly at first and then harder. The rains had come at last.

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

☐ Task discontinued because the teacher had no correct answers in the first line.

☐ Task skipped because teacher refused.

Assessor Scoring Guide for Intonation (Tick one):

3	2	1
Consistently read with vocal expressiveness, conveying meaning and emotional content through appropriate intonation.	Used occasional expressiveness and/or used it sometimes inappropriately (i.e. intonation did not match meaning).	Used little to no vocal expressiveness. Mostly dull and monotonous. Intonation offered little support toward conveying meaning.
When finished, say: ***Thank you! Let's go on to the next section.***		

End Time: _____

D4. Oral Reading Comprehension Subtask

[Return to Oral Reading Comprehension Subtask section of report]

Start Time: _____

Instructions to the Assessor:

Remove the Oral Reading Fluency task from in front of the teacher as you read the instructions below.

As teachers, we ask our pupils a lot of questions to check their understanding of the lesson. Now I will ask you some questions that you might ask your pupils about the story you just read. Please tell me what answer you would expect them to give if they had understood the story. If you have forgotten the answer, I will give you back the paper at the end so you can find it in the story.   Are you ready? Let's begin.

Ask the provided questions in the table. Mark the provided box according to the teacher's answer.

Look Back: This activity is used if the teacher did not answer a question correctly. Give the passage again to the teacher and say: ***Now you can use the passage to help you remember the answer.*** Ask only the questions that were answered incorrectly the first time. If correct, tick the Correct with Lookback box.

Assessor Score Sheet:

Questions	1st Attempt (from Memory)			2nd Attempt (with Lookback)		
	Correct	**Incorrect**	**No Response**	**Correct**	**Incorrect**	**No Response**
1. **Where does Pat live?** *[on a farm]*						
2. **What gets very dry?** *[the land or the ground]*						
3. **Why do Pat and his family watch the sky?** *[hoping the rains come; waiting for the rain, looking or watching for rain]*						
4. **What did Pat see as he sat outside?** *[clouds, dark clouds]*						
5. **How did Pat feel when the rains came?** *[excited; thankful; happy; reasonable answer]*						
When finished, say: ***Thank you! Let's go on to the next section.***						

End Time: _____

E. Grammar Tasks

E1. Structure and Written Expression Subtask

[Return to Structure and Written Expression Subtask section of report]

Start Time: _____

Instructions to the Assessor:

Show the teacher the Structure and Written Expression task on Page 8 in the Teacher Booklet for Oral Tasks as you read the instructions below.

For this activity, here are some sentences in English. Each sentence is missing one part. Below each sentence are four ways to complete the sentence. I will read each sentence while you follow along. Then you tell me which answer best completes the sentence.

Let's start with an example. [Point to the example.] *This sentence says, "My father [dash] beans every year."* [Point to the response options.] *The four answer options are: "A. planting; B. has plants; C. plants; D. is planted." Which answer best completes the sentence?*

[If the teacher says C, say:] *That's right.*

[If the teacher does not say C, say:] *OK, consider this. If we chose C, the sentence would read "My father plants beans every year." Answer C completes the sentence better than the other answers.*

I will start reading here. [Point to the first sentence.] *You will follow along and choose the answer that best completes the sentence. You can just say A, B, C, or D. If you come to one that you do not know the answer to, just skip it and we will go on to the next sentence.*

Are you ready? Let's begin.

Read each sentence and all the answer options aloud. Clearly mark the answer that the teacher indicates for each item. If the teacher says one answer and then changes his or her mind, mark the new answer. If you already marked the first answer, cross out the old answer (ø) and mark the new answer.

Stay quiet, except if the teacher hesitates for more than 20 seconds on one item. Point to the next item and say: *Please go on.* Mark the skipped item as incorrect.

Assessor Stimulus and Score Sheet: Circle A, B, C, or D according to the answer indicated by the teacher, or tick "No response".

1. The government recognizes education _____ human right and strives to provide free primary education to all children. *A. basic* *B. as basic* *C. as a basic* *D. basically as*	□ No response
2. A tropical cyclone is expected to hit the region on Thursday evening, _____ many domestic flights have been cancelled. *A. so* *B. because* *C. provided* *D. due to*	□ No response
3. For the first time, the manufacturer has revealed that it _____ three million tons of plastic packaging in one year. *A. use* *B. was used* *C. used to* *D. used*	□ No response
4. Our school plans to review _____ official policies on pupil absenteeism. *A. its* *B. an* *C. ourselves* *D. one*	□ No response
5. With a quarter million people _____ in its many villages in northern Uganda, Bidibidi is the second largest refugee camp in the world. *A. life* *B. have lived* *C. living* *D. lived*	□ No response
When finished, say: *Thank you! Let's go on to the next section.*	

End Time: _____

E2. Error Identification Task

[Return to Error Identification Subtask section of report]

<div style="text-align: right;">Start Time: _____</div>

Instructions to the Assessor:

Show the teacher the Error Identification task on Page 9 in the Teacher Booklet for Oral Tasks as you read the instructions below.

An important part of our role as teachers is to identify and correct pupil errors. For this activity, here are some sentences in English. Let's say that your pupils wrote these sentences. Some of the sentences contain an error in one of the underlined parts, A, B, C, or D. Please read each sentence and tell me which part contains the error, if any. You do not have to correct the error, just find it, if there is one. Choose only one answer for each sentence.

Let's start with an example. [Point to the first example.] *This sentence says, "They was very interested in what they had heard on the news." There is an error in Part A; "was" is not correct in this sentence. So the answer here is A.*

Now you try an example. [Point to the second example.] *This sentence says, "She has been working as an accountant in this office for two year." Does any underlined part contain an error?*

   [If the teacher says D, say:] *That's right.*

   [If the teacher does not say D say:] *OK, consider this: "two year" is not correct, so the answer here is D.*

You will start here. [Point to the first sentence.] *You will read each sentence to yourself and find the underlined part that contains an error, if any. You can just say A, B, C, D, or E for No Error. If you come to one that you do not know the answer to, just skip it and continue to the next sentence.*

Are you ready? You may begin.

DO NOT READ THE SENTENCES ALOUD. Clearly mark the answer that the teacher indicates for each item. If the teacher says one answer and then changes his or her mind, mark the new answer. If you already marked the first answer, circle the old answer (ø) and mark the new answer. If the teacher gives two answers for one sentence, say: *Please choose only one answer for each sentence.*

Stay quiet, except if the teacher gets stuck on one item for more than 20 seconds. Point to the next item and say: *Please go on.* Mark the skipped item as incorrect.

Assessor Score Sheet: Circle A, B, C, D, or E according to the answer indicated by the teacher, or tick "No response."

1. Most areas <u>in Uganda</u> usually receive plenty of rain. <u>Some areas</u> 　　　　　　　A　　　　　　　　　　　　　　　　　　　　　B of the Southeast and Southwest average <u>more than</u> 150 millimeters 　　　　　　　　　　　　　　　　　　　　　C <u>per months</u> in the rainy season.　<u>No Error</u> 　　D　　　　　　　　　　　　　　　　E	□ No response
2. Female <u>school attendance</u> is <u>low than</u> that of males <u>at all</u> levels 　　　　　　　　A　　　　　　B　　　　　　　　　　C <u>of education</u>.　<u>No Error</u> 　　D　　　　　　E	□ No response
3. The <u>increase population</u> in the city <u>in recent years</u> <u>has</u> put a lot 　　　　A　　　　　　　　　　　　B　　　　　　C of stress on the limited water <u>resources</u>.　<u>No Error</u> 　　　　　　　　　　　　　　　D　　　　　　E	□ No response
4. In 2004, <u>a team of</u> government <u>scientists</u> at the Ministry of the 　　　　　A　　　　　　　　　　B Environment <u>find that</u> chemicals from the local factory 　　　　　　　C <u>had contaminated</u> the river.　<u>No Error</u> 　　　D　　　　　　　　　　　E	□ No response
5. The PTA <u>decided to</u> provide more pupil desks <u>because</u> 　　　　A　　　　　　　　　　　　　　　B enrollment was high and <u>there were</u> not <u>enough</u> seats.　<u>No Error</u> 　　　　　　　　　　　C　　　　　　D　　　　　　　E	□ No response
When finished, say: ***Thank you! Let's go on to the next section.***	

End Time: _____

- Remove the Teacher Booklet for Oral Tasks.
- Ensure that you have copied the Teacher Code from Page 2 of this assessor protocol onto the front of the Teacher Booklet for Silent Reading and Writing Tasks.
- Set the Teacher Booklet for Silent Reading and Writing Tasks in front of the teacher.

F. Silent Reading Comprehension Task

Start Time: _____

Instructions to the Assessor:

Have the teacher turn to the Silent Reading: Example task on Page 1 of the Teacher Booklet for Silent Reading and Writing Tasks. Read the following instructions to the teacher.

One of our roles as teachers is to build our pupils' reading comprehension. For this activity, here is a reading comprehension exercise that you might give to your pupils. You will read the text silently and then circle the correct answer to each question. Let's start with a short example. I will read the text to you.

[Point to the example.] ***"Mary was excited to go to her new school because it had good teachers, and it was close to her house."***

[Point to the instructions.] ***The instructions say, " Circle the correct answer."***

[Point to the question.] ***The question asks, "Why was Mary excited to go to a new school?"***

[Point to the answer options.] ***The answer options are: "A. It had good teachers. B. It had many teachers. C. It was far from her house." Which answer is correct?***

  [If the teacher(s) say(s) A, say:] ***That's right. So circle A.***

  [If the teacher says B or C, say:] ***Let's look at the text and read the response options again. … Answer A answers the question better than the other options. Circle answer A.***

When I say begin, you will do a longer text on your own. First you will read the text. Then you will answer 10 questions about the text. Circle the correct answer.

You will have 5 minutes to complete this exercise. Just do as much as you can in the 5 minutes. When time is up, I will say Stop. When I say stop, please stop writing and put your pen down even if you have not finished. If you come to the end of the 10 questions before time is up, you will see the word STOP. Stop and wait for my instructions. Do not go on to the next exercise until I tell you.

🕐 Set the timer for 5 minutes.

Please turn to the next page (Page 2). Are you ready? Let's begin.

Start the timer when you say "Let's begin." When the timer is up, say: ***Stop. Thank you. Please put your pen down, and let's go on to the next section.***

Text:

Malala Yusafzai was born in Pakistan in 1997. She was named after a woman warrior from Afghanistan. Malala's father was a schoolteacher who believed that education, especially for girls, was important. Malala loved going to school and wanted to become a doctor.

When Malala was 11, a group of fighters took over her town. This group had different ideas about how people should live. For example, they believed that people should not take photographs. Unfortunately for Malala, they also believed that girls should not go to school. So, they destroyed all the schools for girls. Malala felt that this was unfair. She spoke out against this by giving interviews on television. She said all children had the right go to school. Soon, the Pakistani army chased the fighters away. Schools for girls were reopened, and Malala started going to school. But the fighters were still angry, and one day they attacked Malala and almost killed her. This made Malala famous. When she became better, she continued to fight for children's right to education. She spoke about this all over the world. In 2014, at the age of 17, she became the youngest person to be awarded the Nobel Peace Prize.

F. Silent Reading Comprehension Questions (Scored after the teacher is dismissed)

1. What happened when Malala was 11 years old?
   a. A group of fighters visited her school.
   b. A group of fighters visited her town.
   c. A group of fighters captured her town.

2. What was the occupation of Malala's father?
   a. soldier
   b. schoolteacher
   c. farmer

3. Why did Malala have to stop going to school?
   a. There was a war in her town.
   b. Her father did not believe girls should go to school.
   c. The group of fighters in her town did not permit girls to go to school.

4. What did the group of fighters believe?
   a. They believed that only boys should go to school and that people should not take photographs.
   b. They believed that both boys and girls should go to school and that people should take photographs.
   c. They believed that people should not go to school and should not take photographs.

5. Why did Malala give interviews on television?
    a. She hated the fighters who destroyed schools.
    b. She wanted to speak out about the importance girls going to school.
    c. She wanted to speak out about the army chasing the fighters away.

6. Schools for girls were reopened in Malala's town because
    a. the army fought with the group of fighters and won.
    b. the army closed the schools for boys in the town.
    c. the army became friends with the fighters so there was peace in the town.

7. Why was Malala attacked?
    a. Because she dared to speak out against the Pakistani army.
    b. Because she dared to speak out about the right to for all children to go to school.
    c. Because she loved school.

8. Why did Malala become famous?
    a. She became famous because she wrote a book.
    b. She became famous because she was attacked for her beliefs.
    c. She became famous because she was a young woman.

9. Choose the most appropriate title for this passage.
    a. Malala, the Brave Warrior Woman
    b. Malala and the Nobel Peace Prize
    c. No School, No Photographs

10. What is the main idea of the passage?
    a. Malala loved school and wanted to become a fighter.
    b. Malala made sure schools were reopened for girls.
    c. Malala was brave and fought for the right of all children to go to school.

End Time: _____

G. Writing Task

G1. Correcting Student Writing Subtask

[Return to Correcting Student Writing Subtask section of report]

Start Time: _____

Instructions to the Assessor:

Have the teacher(s) turn to the example for Correcting Pupil Writing task on Page 4 in the Teacher Booklet for Silent Reading and Writing Tasks as you read the instructions below.

We help build our pupils' writing skills when we mark their papers. For this activity, you will correct a letter written by a primary 4 pupil. Read the letter carefully and correct the letter for mistakes in grammar, capitalization, punctuation, and spelling.

Underline the mistakes the pupil made in the pupil answer column [point] ***and then write the correct answer in the teacher feedback column*** [point]***. If there is a punctuation mark missing, draw a line where the punctuation mark should be and then write the correct punctuation mark in the teacher feedback column. You will correct one error per row.***

Let's look at the examples together. [Point to the first example.] ***This sentence says "Some children like playing football and netball." Where is the pupil error in the sentence?***

> [If the teacher says 'football,' say:] ***That's right. So underline the word fotball in the pupil answer column and write the correct spelling answer in the teacher feedback column.***

> [If the teacher does not say 'football,' say:] ***OK, let's look at the word fotbal. It should have two o's and two l's. So, the spelling of the word football is the error in the sentence. Now underline the word fotbal in the pupil answer column and write the correct spelling in the teacher feedback column.***

Let's look at the next example. [Point to the second example sentence.] ***This sentence says, "Can we play other schools". Where is the pupil error in the sentence?***

> [If the teacher says there is no question mark, say:] ***That's right. Now underline the space after the word schools in the pupil answer column and write the question mark in the teacher feedback column.***

> [If the teacher does not say there is no question mark, OR if the teacher says any other answer, say:] ***OK, let's look at the sentence. The pupil is asking a question. Look at the end of the sentence. It does not have a question mark. The missing question mark is the punctuation error in this sentence. Now, underline the space after the word "schools", in the pupil answer column and write the question mark in the teacher feedback column.***

You will have 2 minutes to complete the task. Do as much as you can in the 2 minutes. When time is up, I will say Stop. When I say stop, please stop writing and put your pen down even if you have not finished. If you come to the end of the 10 questions before time is up, you will see the word STOP. Stop and wait for my instructions. Do not go on to the next activity until I tell you.

☺ Set the timer for 2 minutes.

> ***Please turn to the next page (Page 5). You will start here**. [Point to the first sentence.] **Are you ready? Let's begin.***
>
> Start the timer when you say, "Let's begin." When the timer is up, say: ***Stop. Thank you. Please put your pen down, and let's go on to the next section.***

Correcting Pupil Writing (Scored after the teacher is dismissed)

Pupil Answer Underline the pupil error in each sentence.		Assessor Evaluation of Teacher Response
1.	Dear headmistres, I talk with my friends every morning.	□ Correct (headmistress) □ Underlined correctly but error in feedback □ Some other correction □ No response
2.	They tells me what they like about school.	□ Correct (tell) □ Underlined correctly but error in feedback □ Some other correction □ No response
3.	They have some ideas to make our school beter.	□ Correct (better) □ Underlined correctly but error in feedback □ Some other correction □ No response
4.	I will share one ideas with you.	□ Correct (idea) □ Underlined correctly but error in feedback □ Some other correction □ No response
5.	Much pupils like to read.	□ Correct (Most, Many, Some, Few, No, Etc.) □ Underlined correctly but error in feedback □ Some other correction □ No response
6.	They read befor school and during lunch.	□ Correct (before) □ Underlined correctly but error in feedback □ Some other correction □ No response

7.	They want more books for the libray.	□ Correct (library) □ Underlined correctly but error in feedback □ Some other correction □ No response
8.	Can the school get the pupils more books.	□ Correct (added question mark) □ Underlined correctly but error in feedback □ Some other correction □ No response
9.	we think this will improve our school.	□ Correct (We uppercase) □ Underlined correctly but error in feedback □ Some other correction □ No response
10.	Sincerely Mary Benson	□ Correct (Sincerely,) □ Underlined correctly but error in feedback □ Some other correction □ No response

End Time: _____

G2. Responding to a Writing Prompt Subtask

[Return to Responding to a Writing Prompt Subtask section of report]

Start Time: _____

Instructions to the Assessor:

Have the teacher turn to the Writing Prompt task on Page 6 in the Teacher Booklet for Silent Reading and Writing Tasks as you read the instructions below.

We also build our pupils' writing skills when we model how to write correctly. For this activity you will write a short composition based on these instructions. [Point to the writing prompt]. *Follow along as I read it.*

Which is more important for people, music or sports? Write a short composition in which you state your opinion and then argue for your position in response to the question. Be sure to give reasons with examples to support your position.

You will have 5 minutes to write your response here. [Point to the space below the writing prompt]. *Just do as much as you can in the 5 minutes. When time is up, I will say Stop. When I say stop, please put your pen down even if you have not finished.*

🕐 Set the timer for 5 minutes.

Are you ready? Let's begin.

Start the timer when you say, "Let's begin." When there are 60 seconds showing on the timer, announce that there is one more minute remaining. When the timer is up, say: *Stop. Please put your pen down.*

End Time: _____

Mark at the end of the assessment, after the teacher has been dismissed.

Responding to a Writing Prompt (Opinion)
Circle one box per row in columns A, B, or C that **best** describes the teacher's writing sample.

Criteria	A	B	C
Ideas	The writer clearly states an opinion.	The writer's opinion is unclear.	The writer does not give an opinion.
Support/Evidence	The writer gives 3 or more reasons, with specific examples to support the opinion presented.	The writer gives 1-2 reasons with examples to support the opinion presented.	The writer does not give any reasons, nor examples to support the opinion presented.
Organization	The ideas are connected and logically arranged.	The ideas are not always connected and logically arranged.	The ideas are disconnected.
Introduction and Conclusion	The writer has a clearly stated introduction and conclusion related to the opinion given.	The writer's introduction and conclusion are unclear, and weak but somewhat related to the opinion presented.	The writer does not introduce the topic nor provides a concluding statement.
Signal words I believe, in my opinion, I think, for example, in addition, etc.	The writer uses 3 or more signal words/phrases appropriately.	The writer uses 1-2 signal words/phrases appropriately.	The writer uses no signal words/phrases or uses them inappropriately.
Sentence structure punctuation	The writer uses a wide variety of sentence structures and lengths with appropriate punctuation.	The writer uses some variety in sentence structure and length with some punctuation used appropriately.	The writer uses sentences of similar structures and lengths, with little variety and little or no punctuation.
Word choice "rich" vocabulary = descriptive, complex, vivid, precise words	The writer uses 3 or more rich vocabulary words to support the opinion given	The writer uses 1-2 rich vocabulary words to support the opinion given	The writer uses only simple vocabulary.
Language Usage (sentences make sense)	Strong control and correct usage of language with 1-2 minor errors which do not impede comprehensibility.	Good control of language usage with 2-3 of significant errors in language use, which make portions of the text difficult to comprehend.	4 or more significant errors in language use make major portions of the text difficult to comprehend.

G3. Spelling Subtask

Instructions to the Assessor:

Have the teacher turn to the final page in the writing booklet.

Say, **As teachers we often use dictation to help our pupils learn to write words correctly. Now I will read some English words aloud and would like for you to write them down here, just as you might ask your pupils to do. I will say each word once. Then I will use the word in a sentence. Then I will say the word a second time and third time. Write the word - not the whole sentence, just the target word. If you do not know how to spell the whole word, that's fine, just spell it as best as you can. If you would rather skip the word, you may, and I will go on to the next word.**

Let's do an example. Listen as I say the word and use it in a sentence. Then write the word. Big. … The boy is big. … Big. … Big. [Check the teacher's answer and where she is writing….]

Start the timer after completing the example.

No.	Time	Word	Sentence
1	:00	**full**	**He was full after he ate a big meal. full full**
2	:20	**dock**	**The boats are at the dock. dock dock**
3	:40	**theft**	**Yesterday there was a theft in the market. theft theft**
4	1:00	**spices**	**We cook with spices to make a delicious meal. spices spices**
5	1:40	**scratch**	**He got a scratch when he walked in the bush. scratch scratch**
6	2:00	**mood**	**The mood was sad at the funeral. mood mood**
7	2:20	**ripen**	**Bananas are sweeter after they have had time to ripen. ripen ripen**
8	2:40	**laughter**	**When the comedian tells a joke, there is always laughter. laughter laughter**
9	3:00	**marched**	**The military marched in the parade. marched marched**
10	3:00	**sharing**	**I thanked him for sharing when he lent me his pencil. sharing sharing**
11	3:40	**fortunate**	**In Uganda, we are fortunate to have many kinds of food. fortunate fortunate**
12	4:00	**opposition**	**In the debate, the opposition was well-prepared. opposition opposition**
13	4:20	**laziest**	**The laziest boy sat all day and did nothing. laziest laziest**
14	4:40	**visible**	**Without headlights, the oncoming car was not visible at night. visible visible**
15	5:00	prosperity	**If you have prosperity, you have all that you need in life. prosperity prosperity**

- Take the Teacher Booklet for Silent Reading and Writing Tasks.
- Thank the teacher for their participation and dismiss them.
- After the teacher is dismissed score the writing and speaking tasks.

Spelling Subtask Scoring

Score after the teacher is dismissed. Tick the boxes that do not match the way the word was written. If the entire word was written correctly, tick yes. If there is a double letter that should not be there (two p's in ripen) select spot 3 or spot 4.

Word	Spot 1	Spot 2	Spot 3	Spot 4	Spot 5	Spot 6	Spot 7	Spot 8	Spot 9	Spot 10	Spot 11	Entire word correct
1. full	_f	fu	ul	ll	l_							Yes
2. dock	_d	do	oc	ck	k_							Yes
3. theft	_t	th	he	ef	ft	t_						Yes
4. spices	_s	sp	pi	ic	ce	es	s_					Yes
5. scratch	_s	sc	cr	ra	at	tc	ch	h_				Yes
6. mood	_m	mo	oo	od	d_							Yes
7. ripen	_r	ri	ip	pe	en	n_						Yes
8. laughter	_l	la	au	ug	gh	ht	te	er	r_			Yes
9. marched	_m	ma	ar	rc	ch	he	ed	d_				Yes
10. sharing	_s	sh	ha	ar	ri	in	ng	g_				Yes
11. fortunate	_f	fo	or	rt	tu	un	na	at	te	e_		Yes
12. opposition	_o	op	pp	po	os	si	it	ti	io	on	n_	Yes
13. laziest	_l	la	az	zi	ie	es	st	t_				Yes
14. visible	_v	vi	is	si	ib	bl	le	e_				Yes
15. prosperity	_P	pr	ro	os	sp	pe	er	ri	it	ty	y_	Yes