AN ADDITION TO THE TOOLBOX FOR MEASURING LITERACY SKILLS OF THE YOUNGEST STUDENTS

The Group Administered Literacy Assessment (GALA)

Why create a new instrument?

For the past decade, donors and development organizations have applied significant resources to create and administer early grade reading and mathematics assessments. These assessments, designed initially to raise awareness about education quality in the early grades, have proven critical in determining proficiency levels of student reading and math ability in developing countries, as well as measuring program impact across a wide range of interventions. Nearly all of the assessments have been individually and orally administered. Although oral administration allows researchers to measure student fluency rates, however, rigorous training is required to ensure reliable test administration.

As work began on developing locally administered school monitoring systems in which a single school inspector would be responsible for assessing a number of children in a short amount of time,[1] it became clear that a different kind of assessment was needed. This new assessment would have to be administered quickly and reliably with little administrator

> The GALA was designed to maximize the benefits of a group-administered instrument, while also taking advantage of lessons learned from the individually administered Early Grade Reading Assessment (EGRA). It limits administration time; can be administered with minimal training; has a simplified scoring system; is based on a previously validated tool; assesses a full range of early grade literacy skills; does not rely on passage reading; and can be adapted for new contexts and languages.

training. In response to this need, a group of researchers at RTI International designed and piloted Group Administered Literacy Assessments (GALAs) for use with students in grades 1 through 3, in a range of countries and languages.

Group-administered assessments are not a new phenomenon. As a matter of fact, many of the most well-known educational assessments across the globe are administered in group settings.[2] All of these assessments have been designed to

reduce time and costs by testing large groups of students in approximately the same time it takes an assessor to test a single student one-on-one. This ultimately means being able to collect data on large numbers of students in a fraction of the time for a fraction of the cost (often with less assessor training), thus providing significant benefits with regard to scale and sustainability.

The international assessments, however, typically have had one major shortcoming for measuring early grade reading skills: a reliance on basic passage reading ability as a starting point. This assumption can lead to significant floor effects for primary school students in the world's lowest-performing countries, as well as a lack of information on lower-order skills that are the building blocks for reading with comprehension.

This was the landscape that led to the GALA. It was designed to maximize the benefits of a group-administered instrument, while also taking advantage of the lessons learned from 10 years of experience with the individually administered Early Grade Reading Assessment (EGRA), which had been validated and used in over 70 countries and more than 120 languages. Accordingly, the GALA

- limits administration time to approximately 45 minutes for a full classroom assessment;
- can be administered with minimal training;
- has a simplified scoring system;
- is based on a previously validated tool;
- assesses a full range of early grade literacy skills;
- does not rely on passage reading; and
- can be adapted for new contexts and languages.

Design of the GALA

The first GALA, in 2013, was trialed by RTI in Ghana as part of a Lot Quality Assurance Sampling pilot under the Education Data for Decision Making (EdData II) project (Mulcahy-Dunn et al., 2013). We created the GALA by adapting EGRA subtasks for use in a group-administered setting. We then applied the lessons learned from that initial pilot in a revised instrument and more rigorous pilot in Egypt in 2014 (RTI International, 2014).

Results of the Egypt pilot pointed to four main challenges in designing and implementing GALAs: (1) limiting student guessing; (2) limiting student cheating (i.e., copying from neighboring students); (3) measuring reading fluency; and (4) creating appropriate and relevant distractors for multiple-

choice response sets. These lessons were important for the redesign of the instrument, and led to its current structure. Subsequent applications in Ghana, Tanzania, and Ethiopia all contributed to continued refinements of GALA.

The GALA instrument has two main components:

1. **Student sheets.** A booklet containing response options for each subtask, to be marked directly by students. Each multiple-choice question includes five response options, in order to reduce the likelihood of guessing correctly. Multiple forms of the student sheets are created and distributed within classrooms in order to reduce the likelihood of student copying or cheating to increase scores.

2. **Assessor tool.** A booklet of instructions, examples, and prompts to be used by the assessors for proper administration of the assessment. The assessor tool is designed as a complete protocol, and test administrators are trained to follow the tool explicitly, for ease of administration and standardization across administrations. Accordingly, the assessor tool is designed such that an untrained assessor could pick up the instrument, study it, and administer it from start to finish simply by reading the instructions and prompts throughout.

Additionally, the tool includes examples for each subtask, which the assessor is expected to write on the board for the entire classroom to see. Their purpose is to ensure that all students are familiar with each subtask before they begin.

While it is possible for any silent-reading or assessor-read task to be included in a group administered reading assessment, 12 subtasks have been created and piloted for use with the GALA.[3] Eight of these subtasks were adapted from EGRA subtasks, while the remaining three were newly developed tasks that we created specifically for the GALA.

These subtasks were designed to measure as wide a range of early grade learning skills as possible, although due to time constraints and concerns about testing fatigue, we do not recommend that all subtasks be used within a single GALA. The intention is that full test administration for a classroom of approximately 20–30 students should take no more than about 40–45 minutes. Accordingly, GALAs should be limited to a maximum of six subtasks. Furthermore, a limit of four subtasks is recommended whenever possible, focusing specifically on the prominent curricular skills for the given grade level. Table 1 provides an overview of each subtask, with a description of the original EGRA task (where applicable) and the newly developed GALA subtasks.

Table 1. GALA and EGRA subtask comparisons

Subtask	EGRA subtask definition	GALA subtask definition	No. of GALA items
Letter Names	Children are asked to read letter names aloud	Assessors read letter names and ask students to tick or mark the letter they hear from the multiple-choice options on their student sheets Example: "Tick the letter 'r.'"	10
Letter Sounds	Children are asked to read letter sounds aloud	Assessors read a letter sound and corresponding word that begins with that sound, then ask students to select the correct letter from the multiple-choice options on their student sheets Example: "Tick the letter that makes the sound /t/, as in 'turtle.'"	10
Familiar Word Reading	Children are asked to read familiar words aloud	Assessors read familiar words and ask students to select the correct word from the multiple-choice options on their student sheets Example: "Tick the word 'hot.'"	10
Nonword Reading	Children are asked to read invented words aloud	Assessors read invented words and ask students to select the correct nonword from the multiple-choice options on their student sheets Example: "Tick the word 'faz.'"	10
Word-Picture Match	Not applicable	Students are provided with a word and four corresponding pictures; students select the picture that displays the prompt word	5
Sentence-Picture Match	Not applicable	Students are provided with six short sentences and six corresponding pictures; students have to draw a line to match each phrase with the correct picture	~6
Reading Comprehension	Children are given one minute to read a story aloud. They are then asked comprehension questions related to the story	Students are given two minutes to read a story silently; students then read comprehension questions silently and select the correct answers from multiple-choice options on their student sheets	5
Maze	Children are asked to read a story silently and to select the correct choice for missing words in each sentence	Students are asked to read a story silently and to select the correct choice for missing words in each sentence	7
Word Dictation	Not applicable	Assessors read individual words and children are asked to write what they hear Example: "Write the word 'home.'"	10
Sentence Dictation	Assessors read a short sentence and children are asked to write what they hear	Assessors read three short sentences and children are asked to write what they hear Example: "Write the sentence: 'Learning at school is fun.'"	10–15

There are two major differences between the EGRA and GALA subtasks. First is the reduced number of items for most GALA subtasks; second is the role of the assessor. That is, for the EGRA, students speak most of their responses for the assessor to mark, whether digitally or on paper. For the GALA, students mark their own worksheets based on assessor prompts. The assessor then collects all the completed worksheets and scores them for the whole classroom.

Since all subtasks aside from dictation are in multiple-choice format on the GALA, minimal training is required for assessment scoring, which simplifies and differentiates it from the EGRA. Additionally, pilot tests that calculated concurrent validity (see next section) showed that students tended to perform slightly better on the GALA than the EGRA, providing initial evidence that the GALA may be useful for measuring skills of the lowest-performing students, while also delivering more information on students who produced zero scores on the EGRA.

Test reliability and validity

After the preliminary version in Ghana and the subsequent tool developed for Egypt, we adapted the GALA for use in seven additional countries: Ghana (7 languages), Ethiopia (2 languages), Tanzania, Jordan, Nepal, Zambia, and Lebanon. As of early 2017, pilot results were available for Ethiopia, Tanzania, and Ghana (see summary in Table 2).[4] For each country, we conducted the same set of analyses: (1) item-level analyses using item response theory (IRT; including item and test characteristic curves); (2) internal consistency (Cronbach's alpha); (3) factor analysis; and (4) concurrent validity with the EGRA.

In the majority of instances, these analyses pointed to strong evidence that the GALA was a reliable test (Cronbach's alpha) that assessed a unidimensional construct of early grade reading ability (factor analysis), was appropriate for the tested grade level (item-level analyses), and was correlated with the EGRA (concurrent validity).

Table 2 shows that the internal consistency of the GALA was strong (above the traditional 0.70 cutoff for Cronbach's alpha) in all three countries, and for all but one language (Ga in Ghana).

Table 2. GALA reliability and validity estimates across countries and languages

Country	Language	Cronbach's alpha	Factor ratio	GALA-EGRA concurrent validity
Ethiopia	Amharic	0.87	16.5	0.82
	Tigrinya	0.85	6.8	0.73
Tanzania	Kiswahili	0.84	8.3	0.50
Ghana	English	0.88	26.5	0.85
	Akuapem Twi	0.78	4.4	
	Asante Twi	0.74	4.6	
	Dangme	0.75	5.8	
	Ewe	0.88	12.4	
	Ga	0.69	15.0	
	Mfantse	0.76	8.8	

The factor ratio column displays the ratio of the first factor eigenvalue to that of the second, which provides an estimate of whether or not each assessment appears to be measuring a single construct. The standard rule of thumb for unidimensionality is that the first factor eigenvalue should be at least three times the second. All versions of the GALA met this criterion.

The final column shows the concurrent validity estimate for total scores on the GALA and EGRA. Three of the four estimates provided strong evidence of concurrent validity (i.e., strong correlations across scores on the two assessments, showing that they appeared to be measuring the same constructs). The Kiswahili version of the GALA in Tanzania, however, showed a smaller correlations. Much of this weaker relationship is explained by the ceiling effects on the pilot sample, which severely limited the variation in scores.

While these reliability and validity estimates are promising, we recommend that some additional measures be explored to further strengthen the claim that the GALA is a valid and reliable tool. The following approaches are currently in the discussion phase for future pilots: test-retest reliability, predictive validity, and additional measures of concurrent validity with other pre-existing early grade reading measures. Finally, we are planning to reanalyze existing concurrent validity data in order to obtain more nuanced estimates, such as weighted composite scores in lieu of basic total scores.

Limitations of the instrument

One of the major limitations of the GALA is the lack of a reading fluency measure. The EGRA results most commonly reported come from the oral reading fluency subtask, but this measure cannot be replicated in a group-administered setting, for obvious reasons—It would not be possible to have all students read aloud at the same time. While several silent reading fluency measures do exist in the literature, only one has been piloted thus far with the GALA (Word Separation, in Tanzania) and it was found to be very unreliable. Continued work on this front is required.

It is also essential to understand what type of information the GALA can and cannot provide. While the main benefit of the GALA is its simplicity, the corresponding drawback is more limited information on any given subtask (due to the smaller number of items tested). Thus far, the GALA has typically been used to ascertain whether or not students are meeting proficiency standards across subtasks, not for generating point estimates regarding student performance. Therefore, the GALA is not intended as a replacement for the EGRA. Rather it should be seen as a useful complement for more routine monitoring of schools; or as a quick snapshot of student performance that can be administered by a school inspector or district officer and scored immediately afterward, without any software or statistical expertise.

Conclusion

Although the GALA was being used in 2016–2017 as part of larger routine monitoring instruments for district-level officials in Ghana, Jordan, and Tanzania (with expectations of the same in Nepal), the implications of this work expand beyond these contexts. We have now created GALAs in eight countries, improving on them each time based on lessons learned from each successive adaptation and administration. These newly created assessments always require work with in country staff, input from local language experts, and availability of grade-appropriate reading materials in the language of choice. That said, the basic building blocks of the GALA allow for somewhat easier adaptation than for the EGRA, particularly with our new templates and guidelines for GALA instrument designs (available upon request).

Ultimately, the reality is that funding agencies are increasingly calling for less costly alternatives to individually administered assessments, and governments are seeking to incorporate reading assessments at scale within their own educational systems. Group-administered reading assessments such as the GALA can provide both researchers and practitioners with an efficient and cost-effective means of monitoring early reading abilities of students across developing countries.

Endnotes

1　Local school monitoring refers in particular to Lot Quality Assurance Sampling (LQAS) and the Local Education Monitoring Approach (LEMA); see RTI International (2016).

2　Examples of well-known group assessments in the United States are the SAT, ACT, and National Assessment of Educational Progress (NAEP). Internationally, studies using group instruments include the Programme for International Student Assessment (PISA); Trends in International Math and Science Study (TIMSS); Progress in International Reading Literacy Study (PIRLS); Latin American Laboratory for Assessment of the Quality of Education (LLECE); surveys of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ); and the Programme d'analyse des systèmes éducatifs de la CONFEMEN [Conférence des ministres de l'Éducation des Etats et gouvernements de la Francophonie], or PASEC.

3　Two GALA subtasks (Listening Comprehension and Word Separation) have proven problematic and therefore are not recommended for use in their current state. Accordingly, they do not appear in Table 1.

4　Preliminary results from Jordan (along with initial analyses from Egypt) provided evidence of unreliability in Arabic GALA instruments.

References

Mulcahy-Dunn, A., Valadez, J. J., Cummiskey, C., & Hartwell, A. (2013). *Report on the pilot application of lot quality assurance sampling (LQAS) in Ghana to assess literacy and teaching in primary grade 3*. Prepared for USAID under the Education Data for Decision Making (EdData II) project, Task Order No. EHC-E-07-04-00004-00 (RTI Task 7). Research Triangle Park, NC: RTI. http://pdf.usaid.gov/pdf_docs/PA00K2DT.pdf

RTI International. (2014). *Egypt Grade 3 Early Grade Reading 2nd National Assessment*. Analysis report prepared for USAID/Egypt under the EdData II project, Task Order No. AID-263-BC-14-00002 (RTI Task 27). Research Triangle Park, NC: RTI. http://pdf.usaid.gov/pdf_docs/PA00K78G.pdf

RTI International. (2016). *Toolkit for the Local Education Management Approach (LEMA)*. Prepared for USAID under the EdData II project, Task Order No. AID-OAA-12-BC-00003 (RTI Task 20, Activity 5). Washington, DC: USAID. https://globalreadingnetwork.net/sites/default/files/eddata/LEMA%20Manual_23Nov2016_SubmittedtoUSAID_Final.pdf

More Information

Jonathan Stern
+1 919 541 6747
jstern@rti.org

Follow @RTI_EdWork
Follow @RTI_Intl_Dev
http://SharEd.rti.org

RTI International
3040 E. Cornwallis Road, PO Box 12194
Research Triangle Park, NC 27709-2194 USA

RT 10741 0217