



Contents lists available at ScienceDirect

International Journal of Educational Development

journal homepage: www.elsevier.com/locate/ijedudev



The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations

Margaret M. Dubeck*, Amber Gove**

RTI International, United States

ARTICLE INFO

Keywords:
International education
Development
Literacy
Educational policy

ABSTRACT

The rise and widespread adoption of the early grade reading assessment (EGRA) has produced an ample supply of critics and converts. This paper seeks to clarify the purpose of EGRA and its limitations. EGRA was created to inform education systems and programmes and alone, is not an intervention. Designed to measure some of the foundational literacy skills that readers need for beginning reading, EGRA is a collection of subtasks, each with a specific purpose. This paper includes a description and rationale for each subtask, as well as the conceptual framework that underpins the assessment. Key results from multiple surveys provide informative, grounded examples of how the assessment results are being used to inform both classroom practice and system-level policy. We conclude with a brief discussion on the potential uses of EGRA and similar oral assessments of early learning for informing the monitoring of the post-2015 education indicators.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The early grade reading assessment (EGRA) was created to provide a reliable and valid measure of skills that contribute to reading acquisition. Research on the development of reading skills has advanced the understanding of the reading process (August and Shanahan, 2006; Hoover and Gough, 1990; NICHD, 2000) and served as the underpinning of EGRA's conceptual framework. This research has focused on identifying basic literacy skills that are reliably measured to understand the reading process in typically achieving and struggling readers. Furthermore, these research efforts have aided the ability to identify and to respond to children who may have reading difficulties in the early school years. The motivation for EGRA's creation was to gather timely access to information to inform learning improvement in low-income countries.

Developed in 2006, EGRA has since been adapted for use in more than 65 countries and in over 100 languages. Existing versions are available in an open-source environment (www.eddataglobal.org) with guidance for adaptation based on the particular characteristics of a given language and country. Despite the differences in instrument development between contexts and countries, EGRA's widespread use has provided users with a shared language to describe results and monitor system-level changes,

knowing that direct comparisons between contexts should be avoided (primarily due to linguistic differences). EGRA can be used as a system-level progress monitoring tool or for programme evaluation purposes. EGRA is not an intervention or a curriculum. Nor is it a perfect assessment. No assessment is.

In writing this paper we are clarifying the intention of EGRA. Primarily, we are concerned that some critics may be assigning it purposes which were not intended. For example, motivation is understood to contribute to an individual's interest and persistence for reading (Guthrie, 2004). EGRA does not measure motivation. We are also concerned about the selection of subtasks for some intervention designs for which EGRA is not sensitive. An intervention that aims to improve social and emotional development, a domain of school readiness, will likely not show gains through most EGRA subtasks. This paper includes a description, rationale and targeted developmental level for each subtask, as well as the conceptual framework that underpins the assessment. Key results from multiple surveys provide grounded examples in how the assessment results are being used to inform classroom practice and policy. We conclude with a brief discussion on the potential uses of EGRA and similar oral assessments of early learning for informing the development of the post-2015 education indicators.

1.1. Purposes of EGRA

More broadly than the specific subtasks that measure discreet skills, EGRA, as a whole, offers several purposes. One, it can serve as

* Corresponding author.

** Corresponding author. Tel.: +1 2027282079.

E-mail addresses: dubeck@rti.org (M.M. Dubeck), agove@rti.org (A. Gove).

a baseline of early reading acquisition. We have seen this with large donors (UNESCO, 2014), NGOs (Concern Worldwide, 2014) and academics (Halliday et al., 2012). Two, it can guide the content that is included in an instructional programme. The *Early Grade Reading: Igniting Education for All* report (Gove and Cvelich, 2011) outlined the EGRA results from multiple continents, summarizing zero scores (i.e., the share of students unable to read a single word in a grade-level passage) for multiple languages and countries. These types of results served as the foundation for programme content funded by large-scale donors. Three, EGRA can evaluate programmes. We have seen this in pilots to inform country-wide scale-up (Piper et al., 2014) and academic research from various disciplines (Halliday et al., 2014). Because of its informed theoretical framework and consistent procedures, EGRA provides valid and reliable information for each of the purposes described above. Furthermore, EGRA provides a common language to discuss children's literacy abilities; for example, when presenting the results on the passage reading subtask, we have a means to understand the extent to which the children can handle grade-level text.

1.2. Limits of EGRA

Like all assessments, EGRA has limited functions. As stated in EGRA's inception documents (RTI International, 2009; RTI International and International Rescue Committee, 2011) it is not intended to be a high-stakes accountability measure to determine whether a child should advance to the next grade. Nor should it be used to evaluate individual teachers. Rather, the subtasks included in EGRA can be adapted for teachers to inform their instruction. As a formative assessment, teachers can either use EGRA in its entirety or select subtasks to monitor classroom progress, determine trends in performance, and adapt instruction to meet children's instructional needs.

There is some confusion about EGRA. It is not an instructional programme, although it can inform the content of programmes and interventions (some of which have used EGRA in their names). It is also not a translation of an existing international language assessment. Instead, its theoretical framework serves to support its adaptation to other languages. This framework was informed by existing reading assessments such as DIBELS (Dynamic Measurement Group, 2008), CTOPP (Wagner et al., 1999) and other assessments that measure literacy skills (Parker et al., 1992). See the EGRA toolkit (RTI International, 2009) for specifics about its initial development.

Furthermore, EGRA does not measure literacy behaviours, background knowledge or attitudes about reading. In the context of questionnaires typically administered alongside EGRA, assessors collect information about the child's literacy environment (i.e., language and family support). We recognize that the home environment and prior experiences contribute to children's reading acquisition and maintenance (Guthrie, 2004; Neuman, 2004). Nevertheless, understanding home and prior experiences is not a primary purpose, as capturing their effects reliably and efficiently is typically not possible through child interviews. Home visits are expensive and relying on families to self-report provides limited information (i.e., social desirability bias). Certainly, existing EGRA subtasks combined with a thoughtfully designed tool to measure these other constructs would be powerful.

2. Development of reading acquisition

Studies have shown that literacy skills in alphabetic languages develop along predictable patterns (Adams, 1990; Goikoetxea, 2005; Hachén, 2002). Some people acquire these skills independently, others require intense interventions, yet the majority of people learn with instruction (Vellutino et al., 1996). Sufficient

research (August and Shanahan, 2006) defines effective reading instruction as that which is provided at a reader's developmental level (not too difficult nor too easy); is informed and adjusted based on informal classroom-based assessments; and includes frequent opportunities to read and write a variety of text types both new to the child and self-selected.

Moreover, developmental literacy research has identified specific early language and literacy skills associated with successful literacy acquisition (NICHD, 2000; Snow et al., 1998). The literacy skills that can be easily measured that contribute to reading achievement fall under three major domains: (a) phonological awareness, (b) print knowledge, and (c) orthographic knowledge.

These three domains of knowledge are influenced by various contexts. A structured context, such as classroom instruction or regular tutelage, is an environment in which these skills are taught in a deliberate way. Informal contexts, such as market transactions (Olateju, 2010) or artistic endeavours (Marsick and Watkins, 2001) are also known to be motivators for literacy acquisition. It is in these informal contexts that the trajectories and the patterns of acquisition may diverge from the research previously cited.

Nevertheless, within these three domains are early skills known to be consistently strong predictors to later reading achievement. Many of these skills are measured in EGRA.

Phonological awareness is a collection of skills that contributes to early reading. At the most basic level, it is defined as sensitivity to language at the phonological level. Many studies (Badian, 2001; Denton et al., 2000; McBride-Chang and Kail, 2002; Muter et al., 2004; Wang et al., 2006) have supported its role in predicting early reading achievement in both opaque and transparent languages.

Researchers (Adams, 1990; Stahl and Murray, 1994) have described the development of phonological skills and the difficulty of the various tasks. Adams (1990) has established levels of complexity for phonological awareness tasks from easier to more difficult: (a) knowledge of rhymes, (b) oddity tasks (i.e., a sensitivity to similarities and differences between words), (c) blending and syllable splitting, (d) segmentation (i.e., identifying the individual phonemes that comprise a word), and (e) manipulation (i.e., deletion or reordering). Phonological awareness tasks vary in difficulty due to the size of the linguistic unit that can be manipulated. The focus unit of manipulation would vary depending on the language.

In the same way that researchers have identified a typical trajectory for phonological awareness skills, other researchers (Lomax and McGee, 1987) have examined the development of *print knowledge*, the domain that describes an understanding about the orthographic system and written language. Through investigations, print knowledge is understood to advance in a hierarchical, yet recursive way. In other words, each print knowledge component is a prerequisite of another component, but skills are not necessarily mastered before new learning commences.

Within print knowledge are multiple skills. Print concepts include a variety of understandings about print, including book orientation (e.g., the cover; where to start reading), directionality (e.g., left to right; top to bottom), and a purpose for reading (e.g., to inform; to entertain).

An understanding of the distinctive features and names of individual alphabet letters also appears under print knowledge. Besides letter recognition, alphabet knowledge encompasses knowledge of letter names and their corresponding sounds. Letter knowledge has been consistently shown to be a strong predictor of early reading (Adams, 1990; Ehri and Wilce, 1985; Piper and Korda, 2011; RTI International, 2013; Wagner et al., 1994; Yesil-Dagli, 2011).

Research has demonstrated that alphabet knowledge is also a robust predictor for non-native language literacy acquisition

(Chiappe et al., 2002; McBride-Chang and Suk-Han Ho, 2005). More specifically, alphabetic knowledge is a stronger predictor for non-native speakers than for native speakers and predicts reading achievement one year (Chiappe et al., 2002) and two years later (McBride-Chang and Suk-Han Ho, 2005; Manis et al., 2004; Marsick and Watkins, 2001).

Orthographic knowledge, an understanding about words in their written form, is a third domain relevant to early reading acquisition. It includes the knowledge that certain sequences of letters compose words that represent spoken sounds. Applying this knowledge helps to identify familiar and decode unfamiliar words in isolation and in connected text. An explanation of this knowledge will be used to describe literacy acquisition models further.

Literacy acquisition models suggest that learners come to recognize progressively more complex aspects of written language, specifically through orthography (Ehri, 1995; Frith, 1986). Students advance through a series of phases, or stages, that represent their understanding about printed words (Chall, 1983; Henderson and Templeton, 1986). The concept of orthographic knowledge as a developmental phenomenon was introduced in the early 1970s (Chomsky, 1970; Read, 1971). Later, Ehri (1995) proposed five phases to capture this phenomenon: (a) prealphabetic, (b) partial alphabetic, (c) alphabetic, (d) consolidated-alphabetic, and (e) automatic. Although developed to explain children's English reading acquisition, the phases describe a student's evolving knowledge about symbols to represent sounds in a word, which is a common organizing principle across languages.

In the earliest phase, *prealphabetic*, students do not yet understand the *alphabetic principle*. They have yet to learn that speech can be turned into print or that letters represent speech sounds in the language. They can memorize text, associate meaning with pictures and environmental print, and identify words by their unique shapes. Students at this emergent stage notice the phonological features of spoken words (e.g., word length) and learn to navigate a book—its direction, purpose of pictures, and the like.

In the second phase, *partial alphabetic*, students now understand the alphabetic principle, meaning they use some of the word's letters (i.e., symbols) to cue corresponding sounds, and this, in conjunction with memory for that word's initial unit of sound, allows them to "read" the word. Students in this phase can give the names and sounds of some letters, recognize a few words, and match spoken to written words. They learn to combine picture cues with initial sounds or syllables to read new words and are developing a sight-word vocabulary (i.e., words read automatically).

The next phase, *alphabetic*, is marked by an ability to learn new words using several strategies. Students attend to more parts of the word and use their knowledge of grapheme–phoneme correspondences (i.e., symbol–sound) to decode unfamiliar words, and to develop an extensive sight-word vocabulary. Students in this phase can read some text independently and are learning to read fluently.

The fourth phase, *consolidated-alphabetic*, considers the importance of using larger spelling patterns within a word to read unfamiliar words. Students in this phase continue to learn new words through reading and writing, and attend mostly to comprehension instead of decoding.

The last phase, *automatic*, is marked by proficient word reading. Students recognize nearly all the words they see in print and when they encounter a new word, they can identify it independently. Furthermore, they recognize more words in print than they use in typical conversations. The focus of attention is almost entirely on comprehension.

EGRA includes a battery of subtasks that are aligned with these five phases. It measures phonological, print and orthographic

knowledge as well other skills understood to contribute to reading with understanding, such as receptive language and several types of comprehension (e.g., explicit, inferential, informed by syntax). These subtasks are described in the next section.

3. Description of EGRA

EGRA is a research-based collection of individual subtasks that measure some of the foundational skills needed for reading acquisition in alphabetic languages. The purpose (i.e., research question) and the context for which EGRA is used contribute to its parameters. Its theoretical framework allows for it to be adapted to other languages relatively quickly to have it available soon after the need is identified.

The context in which EGRA is used determines some of its parameters. It has been used primarily with children in the early primary grades so the subtasks seek to gather a maximum amount of information in the least amount of time. Depending on the number of subtasks, the entire battery could be administered in 10–20 min with the assessor actively aiming to maintain the child's attention. For each subtask that could be added to provide arguably useful information, administration time is lengthened, requiring more attention from the child and potentially reducing the validity of the results. Therefore, subtasks should be selected to give the most useful information and should be limited to avoid overloading the child.

EGRA administration procedures are child-centred to increase the child's comfort and increase the validity of the results. Individually administering the assessment allows the assessor some flexibility to adapt to a child's response. The assessor begins by establishing rapport and describing what will happen during the assessment to help the child realize that this will be a safe and supportive interaction; at this time the assessor also obtains the child's agreement to participate. Many of the subtasks have discontinuation rules (i.e., early stop) to limit the discomfort children may feel if they are unable to perform. If a child initially responds incorrectly for an item but then changes the response before proceeding to the next item, the response is considered a self-correction and scored as correct.

Consistent procedures for administering each subtask increase the confidence in the results. To increase comparability, the child is prompted at pre-established intervals (i.e., 3 seconds) to attempt the next item. This ensures that children are all exposed to an equal number of items to determine their score. Furthermore, the oral comments that are used are intended to encourage the children that their effort is noted (e.g., I can tell you are doing your best). Comments that provide evaluative statements about correctness (e.g., You got them all correct!) or that instruct are unacceptable.

EGRA is adaptable to languages and to the grade level of interest. The items are informed by creating a grade-level corpus of words for that language generated from existing text. Items are then selected based on their frequency in the corpus which, again, increases confidence in the validity of the subtask. For example, the non-word reading subtask measures the ability to apply the knowledge of letter–sound relationships to decode unfamiliar words. The orthographic structure of the non-words is determined by the orthographic structure of actual words in the corpus. A structure that appears frequently and is grade-appropriate is followed to create the non-words.

For nearly all of the subtasks, the items within it are considered to be of equal difficulty and, therefore, measuring the same construct (i.e., grade 2 familiar words). This implies that the items that appear in the first row of text (e.g., individual letters or words) have the same level of difficulty as the items in the last row of text for that subtask. Ultimately, when the early stop rule is employed because the child had zero correct in the first row of items, it is

done because the child would have performed similarly on the remaining items.

3.1. Description of EGRA subtasks

All of the EGRA subtasks are administered individually in an interaction between a trained assessor and an individual (typically a child in primary school). Items incorporate age-appropriate words taken from the corpus of grade-level text. Word and letter frequencies are calculated to inform the development of the instruments. Selected words have a similar orthographic structure, represent the most common features of the language and align with expectations for student reading in that grade. Table 1 gives an overview of each subtask. Note that not all subtasks are administered in every country or context; instead EGRA is a flexible template from which assessment designers can select the most appropriate subtasks. Those noted with an asterisk below are considered to be “core” to the assessment and it is recommended that these subtasks be included in every assessment.

3.2. Purpose and limitations of subtasks

As with any instrument, EGRA has a particular purpose. Each of the subtasks yields certain information and has limitations. Some of these are explained next.

Most of the subtasks are instructionally transparent, which can inform the next steps of instruction. For example, the results on the familiar word reading subtask help to shape the amount of practice a child needs with grade-level text. Specifically, a child who identifies a limited number of words (i.e., less than 10) needs practice interacting with grade-level text. A child who can automatically read about 50 words from the same word list is ready to continue to more advanced text.

The orientation to print subtask is attractive because it is not intimidating and it is useful for measuring a child’s familiarity with print. However, it is often not included because of time considerations and the ceiling effects that often occur.

Presently the phonemic awareness subtasks are presented orally. This means that the child listens to the stimuli word or

Table 1
Description of EGRA subtasks.

Name	Description	Phase (s) of development
<i>Orientation to print</i>	Measures knowledge of early print concepts such directionality. It is untimed and does not have a discontinuation rule.	Pre-alphabetic
<i>Letter name identification</i>	Measures knowledge of letter names. 100 letters are presented in random order in both upper and lower case. It is timed to 60s and is discontinued if none of the letters in the first line (i.e., 10 letters) is read correctly.	Partial alphabetic
<i>Letter-sound identification*</i>	Measures knowledge of letter-sound correspondences. 100 letters are presented in random order in both upper and lower case. It is timed to 60s and is discontinued if none of the sounds in the first line (i.e., 10 letters) is produced correctly.	Partial alphabetic
<i>Initial-sound identification</i>	Measures the ability to discriminate beginning sounds. Three words are presented and the aim is to identify the word that begins with a different sound from the other two. It is oral and has 10 sets of words. It is discontinued if no points are earned in the first five items.	Pre-alphabetic Partial alphabetic
<i>Segmentation (phoneme or syllables)</i>	Measures the ability to segment a word into individual phonemes or syllables. This subtask is oral and has 10 items. It is discontinued if no points are earned in the first five items.	Pre-alphabetic Partial alphabetic
<i>Syllable identification</i>	Measures the ability to read individual syllables. 50 syllables are presented. It is timed to 60s and is discontinued if none of the first five syllables is read correctly.	Partial alphabetic
<i>Familiar word reading</i>	Measures the ability to identify individual words from grade-level text. 50 words are presented. It is timed to 60s and is discontinued if none of the words in the first line (i.e., five words) is read correctly.	Partial alphabetic Alphabetic
<i>Non-word reading*</i>	Measures the ability to decode individual non-words following common orthographic structure from grade-level text. 50 non-words are presented. It is timed to 60s and is discontinued if none of the words in the first line (i.e., five words) is read correctly.	Partial alphabetic Alphabetic
<i>Oral reading fluency*</i>	Measures the ability to read a grade-level passage of approximately 60 words. It is scored for accuracy and rate. It is timed to 60s and is discontinued if none of the words in the first line (i.e., about 10 words) is read correctly.	Consolidated-alphabetic
<i>Reading comprehension (with or without lookbacks)*</i>	Measures the ability to answer questions about the grade-level passage. Question types include explicit and inferential, and lookbacks (i.e., referencing the passage for the answer) can be used if appropriate.	Consolidated-alphabetic Automatic
<i>Cloze</i>	Measures the ability to identify a word among several choices that would complete the sentence using the correct part of speech. It is untimed and does not have a discontinuation rule.	Consolidated-alphabetic Automatic
<i>Listening comprehension*</i>	Measures receptive language of an orally read passage with both explicit and inferential questions. It is untimed and does not have a discontinuation rule.	Used diagnostically across various phrases
<i>Vocabulary</i>	Measures receptive language skills of individual words and phrases related to body parts, common objects, and spatial relationships. It is untimed and does not have a discontinuation rule.	Used diagnostically across various phrases
<i>Dictation</i>	Measures the ability to spell and use grammar in a grade-level sentence. Words can be scored for partial representation.	Partial alphabetic Alphabetic Consolidated-alphabetic
<i>Interview</i>	Gathers information about the child that is related to literacy and language development (e.g., first language; access to print). It is self-reported by the child.	Any phase of interest

* Those sub-tasks noted with an asterisk are considered to be core to the assessment.

words and has to manipulate them in some way (e.g., identify the odd beginning sound in the words dog, map, desk). This also means that the child has to hold the words in short-term memory, which adds a level of complexity to interpreting the results. An alternative to reduce the memory load would be to provide and identify illustrations that represent the words to be manipulated, as has been done elsewhere (Invernizzi et al., 2003–2007). Yet, acquiring valid images and presenting them in a consistent manner when adaptation time is limited is a challenge, which is why this approach were ultimately rejected by the expert panel during EGRA’s development phase (RTI International, 2009).

The reading comprehension subtask has also been criticized as being a memory task. Typically, after the passage is orally read, it is removed from view and the corresponding comprehension questions are asked. This procedure measures recall and is used to reduce administration time. However, in a situation where the research question wants more diagnostic information, it is acceptable to return the written passage to the children to see if they can look back to find the answer. Doing so, however, changes the task from one that assesses recall and comprehension to one that assesses whether or not a child is able to negotiate text and skim to locate the answer.

Some of the other subtasks are also diagnostic. They complement each other to offer explanatory information as to why a child may have struggled with a particular subtask. For example, if a child reads the grade-level passage at a rate and accuracy level that suggests fluency, it could be expected that the child would be able to answer related comprehension questions. When that is not the case, the results on the listening comprehension or the vocabulary subtask can be analyzed. If the results on those two subtasks are also low, it suggests that the child needs to develop language skills to support comprehension. Conversely, if a child did well on the language measures but poorly on comprehension, this might suggest that the child is not monitoring the text for purposes of comprehension.

3.3. Timing of subtasks

The timing of the EGRA subtasks has been criticized. If done carelessly, timing can add pressure to the child, who may already be intimidated by the assessment. Also, reporting the timed results can give the impression that speed is of primary importance. Furthermore, timing is not seen in a natural reading context. Although the critiques of timing deserve consideration, its benefits

for an evaluation to inform classroom instruction outweigh the negatives.

There are multiple reasons for timing some EGRA subtasks. The ones that are timed assess skills that become more accurate and automatic (i.e., faster) as that skill develops. In other words, the rate at which a reader recognizes particular items increases as reading skills advance. Therefore, the same passage (or set of letters or words) can be used with readers at different reading levels to show individual differences. This has the benefit of avoiding ceiling effects that would occur if the items were scored only on accuracy. This also avoids the need to create multiple passages or word lists when resources and time for assessment development are limited. Additionally, the results based on time can inform intervention and instructional design (i.e., the child who can identify 15 grade-level words in a minute has different instructional needs than the child who can identify 50 in the same amount of time).

Another benefit of timing is consistency. With it, we know that the individual EGRAs are administered under the same conditions to achieve consistency so that information is comparable. Most importantly, this consistency provides a common language to discuss children’s results. Ultimately, if a child reads 10 words in a minute, the discussion should not focus on the time; the message is that the child could read only 10 words from grade-level text. Timing some EGRA subtasks yields valuable information about individual child differences.

4. Applications and results

As mentioned in the Introduction, EGRA has been used for programme monitoring and evaluation as well as system-level reporting. For detailed findings please see published academic journals (Jiménez et al., 2010; Piper et al., 2014), country-level reports and tools (www.eddataglobal.org; www.earlygradereadingbarometer.org) and reports that consolidate information from various contexts (Gove and Cvelich, 2011). This present paper briefly shows results from two contexts, highlighting how those results were used to inform instruction.

The EGRA results from a national survey from an Asian country illustrate the role of a timed subtask to differentiate children’s abilities. Fig. 1 shows a relatively normal distribution of the oral reading fluency performance for a nationally representative sample of grade 2 students reading a grade 2 passage. This spread shows differences in automaticity for word recognition and provides information that instruction and reading materials should differ depending on individual child needs.

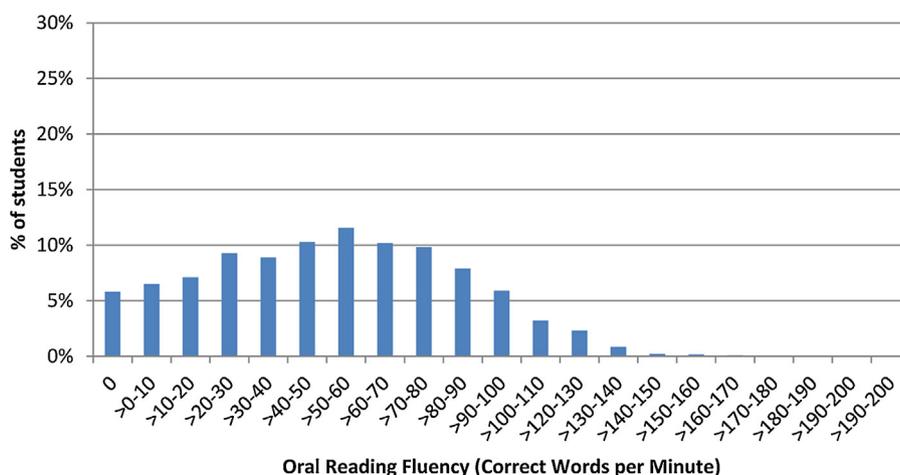


Fig. 1. Distribution of oral reading fluency, grade 2 students, Asian country.

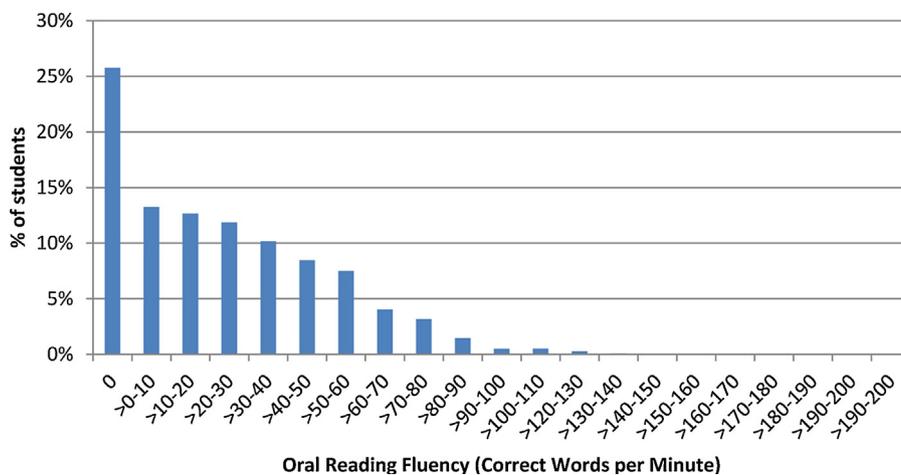


Fig. 2. Distribution of oral reading fluency, grade 2 students, sub-Saharan African country.

Based on the presented oral reading fluency distribution (Fig. 1) and the associated comprehension results, four reading profiles were created (Nonreaders, Reading with limited comprehension, Reading with comprehension and Reading fluently with comprehension). The lowest category of “Nonreaders” corresponds to the bottom 6% of grade 2 students unable to read any words at all, while the highest level, “Reading fluently with comprehension”, accounts for the 47% of students who read the grade 2 passage with at least 80% comprehension and with a fluency rate greater than 60 words per minute. For “Nonreaders” the implications for instruction are to address basic decoding skills to get the students on track for successful reading. For the highest skilled readers, teachers were advised to continue to challenge students with both student-selected and teacher-guided materials, while explicitly building their comprehension strategies. For all four profiles, teachers should provide strategies for understanding and relating to printed materials. At the system level, the student profiles can inform approaches to support teachers to improve and differentiate their classroom instruction.

In contrast, the distribution (Fig. 2) from a sub-Saharan African country of oral reading fluency indicates that significant attention is needed to address the high number of zero scores (i.e., no correct words read in the connected-text passage). With more than one in four grade 2 students unable to read a single word in a grade 2 passage, the recommendations for instruction are quite different from those of the country distribution represented in Fig. 1 (for ease of comparison, Figs. 1 and 2 distributions are presented on the same scale). Furthermore in this sample, because the subtask was timed, distinct reading profiles are seen—children reading 20 words from the passage have distinctly different instructional needs from those reading 60 words.

The diagnostic nature of EGRA allows users to explore the results for students in the lowest performing group to better understand which foundational skills present difficulties. In this case, oral reading fluency was highly correlated with the non-word reading task—those students who were able to decode non-words were also able to read connected text fluently and with understanding.

To support students to acquire the missing reading skills, the project responded by modifying the classroom instruction. This included teachers’ guides with structured lesson plans and revised student books introduced through teacher training and then reinforced by pedagogical coaches who offered ongoing classroom support visits at least once per month. The results in impact, measured by EGRA, ranged between 0.3 and 0.5 SD for effect size (compared to a similarly randomly selected set of control classrooms). The results of the pilot programme then informed plans for a national scale-up.

In the example cases from Asia and sub-Saharan Africa, the EGRA results motivated discussion with counterparts at all levels, from the Ministry of Education to the classroom. Results informed changes in both high-level policy and classroom-based practice: from curriculum and textbook redesign to access to supplemental readers to developing to an intervention using mother-tongue instruction. Individual subtask results helped to identify which skill areas needed reinforcement, enabling instructional designers to include varied tasks for teachers and students to practice in the classroom.

5. Use as a global indicator

In the discussions of the post-2015 education goals, a number of tools, including EGRA, have been considered for measuring and reporting against a to-be-determined global indicator of early reading performance. While leaders on the global stage work to create clear, communications-driven goals, behind the scenes, the development of actionable and measurable indicators for monitoring performance against those goals is under way. Key global agencies, including the World Bank, the UNESCO Institute for Statistics and increasingly, the Brookings Institute’s Center for Universal Education-led Learning Metrics Task Force (LMTF) are leading the discussion for both what and how education and learning should be measured in the post-2015 framework. Central to the debate is the issue of “comparability” of results, with some in the international community insisting that comparability implies the development of equivalent instruments in all contexts. As defined in a recent LMTF report:

“Comparable information refers to the ability of the metrics to generate evidence that is *equivalent* across populations (age, gender, residence, ethnic, socioeconomic, linguistic background, country) and/or across time. Comparability can be desirable at both the national and international levels in order to enable tracking progress and identify equity gaps against a common metric. . . . When comparable information is required, technical mechanisms must be in place to ensure that a *common metric with equivalent properties* is used to measure across a variety of analysis units. This entails the identification of: a *common operationalization of the constructs*; *similar quality control mechanisms for all the phases of the data generation process*; a *careful review of the construct equivalence across contexts (language or settings)*; and a way to monitor measurement equivalence.” (LMTF, 2013, pp. 35–36; authors’ emphasis)

Using the above definition, EGRA is described as “not suitable for international comparison” (LMTF, 2013, p. 45). Based on the

authors' experiences to foster the use of assessment to inform policy and practice in more than a dozen countries, we are concerned that restricting the definition of "comparability" to metrics that are deemed psychometrically "equivalent" risks excluding measurement approaches such as EGRA or other country-led initiatives to inform future global indicators. As noted in this paper, EGRA and similar tools (including civil-society based assessments such as those applied by Uwezo and ASER) follow a common framework to generate comparable results but are not necessarily equivalent across countries. We are concerned that an insistence on equivalency may undermine the effort to increase the number of countries measuring early reading. To move from virtually no globally comparable metrics on learning for low-income countries to a system of "equivalent" metrics may well be a case of letting the perfect be the enemy of the good.

We understand that the call for equivalent measures may be rising from concerns with some EGRAs. Issues of validity can result if EGRA adaptations do not follow the procedures outlined in the guidance notes. For example, if appropriate experts (e.g., reading and language) are not involved or subtasks are selected that are inappropriate for the research question, the results may not yield useful information. Yet, when the adaptation process is thoughtful and procedures are sound, EGRA will offer valid information quickly and reliably.

As documented elsewhere (Gove and Wetterberg, 2011), countries have elected to conduct early grade reading assessments for myriad reasons, including accessibility, available guidance, adaptability and the ability to quickly and relatively easily mobilize the assessment. The rapid expansion of EGRA is partially explained by support from the United States Agency for International Development (USAID), which accounts for roughly half of the country applications to date. Other donors, including the Department for International Development (DFID) and the Global Partnership for Education (GPE), have also made use of EGRA in their programming (see the EGRA Tracker at www.eddataglobal.org under "Reading" for specific details).

Other countries, NGOs and academics have elected to use EGRA for various reasons, principal of which is the tool's accessibility and open-source approach. But we have also heard from country counterparts that EGRA is of interest because they know they will not appear on a globally published league-table of results. Furthermore, these counterparts identify that EGRA's adaptability to additional languages and contexts of interest does not require 3–5 years of development. For many countries, EGRA is a tool that has been useful in informing policy and improving classroom practice.

Additionally, requiring that metrics for global reporting against a common learning indicator be equivalent may be pushing the definition of comparability so far that it extends beyond common practice. Consider an example from the health sector, the indicator for MDG Goal 4, the under-five child mortality rate (with the proportion of newborn deaths). An excerpt from the World Health Organization (WHO) guidance document highlighting the use of the word "comparable" is provided below:

"Under-five mortality rates are computed from data collected in vital registration systems, local demographic surveillance systems, household surveys (full or summary birth histories) and censuses (summary birth histories). In most of the 74 countries, household surveys conducted as part of DHS and MICS are the main data source. The United Nations Inter-agency Group for Child Mortality Estimation (IGME) produces *comparable* estimates for 195 countries on an annual basis. The IGME uses all available national-level data after reviewing data quality, and produces country specific child mortality estimates with publicly available data inputs, adjustments and a

replicable statistical model (<http://www.childmortality.org>, accessed 3 November 2011). IHME has produced estimates using an alternative set of statistical assumptions. The two sets of estimates are *generally consistent* in terms of measures of overall global trends in mortality declines." (WHO, 2011, p. 14, *authors' emphasis*)

As noted above, WHO and the United Nations use the available data to calculate progress against what is surely one of the most important goals in the Millennium Development Framework. The document goes on to state that countries need to strengthen their health systems to better monitor under-five deaths. But nowhere does the WHO document insist on the degree of equivalence that seems to be emerging from the LMTF recommendations. While we respect the aspiration that education measurement be at least as rigorous as health sector monitoring, exceeding the standards set in health may be unnecessarily restrictive. Additionally, emulating the flexible methods in health may be an equally important goal. Can comparable instead mean similar in the approach to measurement, but not equivalent in the statistical sense?

Allowing countries to use *comparable* (in the WHO sense) forms of assessment to report against a to-be-determined learning indicator, while supporting their increased capacity and eventual participation in globally *equivalent* measurement approaches may be the best route to ensuring that more countries monitor and improve learning outcomes, particularly in the early grades.

6. Conclusion

The early grade reading assessment, through a collection of subtasks, assesses some of the skills that are needed for literacy acquisition. Depending on the research question, some of the subtasks are included to offer explanatory information on other subtasks. EGRA's theoretical framework was informed by research that describes the skills that can be measured reliably to predict initial and later reading acquisition. Primarily it examines the print, phonological, and orthographic knowledge readers need to become successful readers of alphabetic languages. These skills attend to the developmental nature of reading acquisition and therefore are valid measures.

The results from EGRA for more than 65 countries are useful for understanding current reading skills or tracking longitudinal growth. Paramount is its purpose to provide timely information to better understand abilities and to inform system-level policy and programme design that can be influenced by classroom instruction.

In the long term, increasing country capacity to use information to inform instruction is critical to improving learning for the estimated 250 million children around the world who are not acquiring basic skills (UNESCO, 2014). The global community should encourage all countries to develop and improve the capacity of their assessment systems and the use of assessments, particular those that are best suited to their needs (Wagner, 2011). EGRA and similar assessments should be considered as viable, comparable metrics for tracking against global indicators. Just as the quality of the access metrics and indicators have improved over time, so will the learning metrics and indicators (Chabbott, 2014). Ultimately, EGRA aims to inform so that literacy acquisition is sufficient to support additional learning and engaged citizens.

References

- Adams, M.J., 1990. *Beginning to Read: Thinking and Learning About Print*. MIT Press, Cambridge, MA.
- August, D., Shanahan, T. (Eds.), 2006. *Developing Literacy in Second-Language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Lawrence Erlbaum Associates, Mahwah, NJ.

- Badian, N., 2001. Phonological and orthographic processing: their roles in reading prediction. *Ann. Dyslexia* 51, 179–202.
- Chabbot, C., 2014. Institutionalizing Health and Education for All: Global Goals, Innovations, and Scaling Up. *International Perspectives on Education Reform Series*. Gita Steiner-Khamsi Series Ed, Teachers' College Press, New York.
- Chall, J.S., 1983. *Stages of Reading Development*. McGraw-Hill, New York.
- Chiappe, P., Siegel, L.S., Wade-Woolley, L., 2002. Linguistic diversity and the development of reading skills: a longitudinal study. *Scientific Studies of Reading* 6, 369–400.
- Chomsky, C., 1970. Reading, writing, and phonology. *Harvard Educational Review* 40, 287–309.
- Concern Worldwide, 2014. EGRA Instruments: Somalia. www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&id=630 (accessed 08.08.14).
- Denton, C.A., Hasbrouck, J.E., Weaver, L.R., Riccio, C.A., 2000. What do we know about phonological awareness in Spanish? *Read. Psychol.* 21, 335–352.
- Dynamic Measurement Group, 2008. DIBELS 6th Edition Technical Adequacy Information (Tech. Rep. No. 6) Eugene, OR. <http://dibels.org/pubs.html>.
- Ehri, L.C., Wilce, L.S., 1985. Movement into reading: is the first stage of printed word learning visual or phonetic? *Read. Res. Q.* 20, 163–179.
- Ehri, L.C., 1995. Phases of development in learning to read words by sight. *J. Res. Read.* 18, 116–125.
- Frith, U., 1986. A developmental framework for developmental dyslexia. *Ann. Dyslexia* 36, 67–81.
- Goikoetxea, E., 2005. Levels of phonological awareness in preliterate and literate Spanish-speaking children. *Read. Writ.* 18, 51–79.
- Gove, A., Cvelich, P., 2011. *Early Reading: Igniting Education for All. A Report by the Early Grade Learning Community of Practice*, rev. ed. RTI International, Research Triangle Park, NC.
- Gove, A., Wetterberg, A., 2011. The early grade reading assessment: an introduction. In: Gove, A., Wetterberg, A. (Eds.), *The Early Grade Reading Assessment: Applications and Interventions to Improve Basic Literacy*. RTI Press, Research Triangle Park, NC, pp. 1–37.
- Guthrie, J.T., 2004. Teaching for literacy engagement. *J. Lit. Res.* 36, 1–29.
- Hachén, R., 2002. Conocimiento lingüístico y reflexión metalingüística: El rol de las conceptualizaciones en torno a la estructura de la sílaba en el proceso de alfabetización. (Linguistic knowledge and metalinguistic reflections: the role of conceptualization with regard to syllable structures in teaching reading and writing). *Lect. vida* 23, 6–17.
- Halliday, K.E., Karanja, P., Turner, E.L., Okello, G., Njagi, K., Dubeck, M.M., Allen, E., Jukes, M.C.H., Brooker, S., 2012. *Plasmodium falciparum*, anaemia and cognitive and educational achievement in school children in an area of moderate malaria transmission: baseline results from a cluster randomised controlled trial on the coast of Kenya. *Trop. Med. Int. Health* 17, 532–549. <http://dx.doi.org/10.1111/j.1365-3156.2012.02971>.
- Halliday, K.E., Okello, G., Turner, E.L., Njagi, K., Mcharo, C., Kengo, J., Allen, E., Dubeck, M.M., Jukes, M.C.H., Brooker, S.J., 2014. Impact of intermittent screening and treatment for malaria among school children in Kenya: a cluster randomised trial. *PLoS Med.* 11 (1), e1001594.
- Henderson, E., Templeton, S., 1986. The development of spelling ability through alphabet, pattern, and meaning. *Elem. School J.* 86, 305–316.
- Hoover, W.A., Gough, P.B., 1990. The simple view of reading. *Read. Writ.* 2, 127–160.
- Invernizzi, M., Meier, J., Juel, C., 2003–2007. *Phonological Awareness Literacy Screening (PALS): Grades 1–3*. University of Virginia, Charlottesville, VA.
- Jiménez, J.E., Rodríguez, C., Crespo, P., González, D., Artiles, C., Alfonso, M., 2010. Implementation of Response to Intervention (RtI) Model in Spain: an example of a collaboration between Canarian universities and the department of education of the Canary Islands. *Psicothema* 22 (4), 935–942.
- LMTF (Learning Metrics Task Force), 2013. *Toward Universal Learning: A Global Framework for Measuring Learning*. Report No. 2 of the Learning Metrics Task Force. UNESCO Institute for Statistics and Center for Universal Education at the Brookings Institution, Montreal and Washington.
- Lomax, C., McGee, L., 1987. Young children's concepts about print and reading: toward a model of reading acquisition. *Read. Res. Q.* 22, 237–256.
- Manis, F.R., Lindsey, K.A., Bailey, C.E., 2004. Development of reading in grades K-2 in Spanish-speaking English-language learners. *Learn. Disabil. Res. Pract.* 19, 214–224.
- Marsick, V.J., Watkins, K.E., 2001. Informal and incidental learning. *New Dir. Adult Contin. Educ.* 89, 25–34.
- McBride-Chang, C., Kail, R.V., 2002. Cross-cultural similarities in the predictors of reading acquisition. *Child Dev.* 73, 1392–1407.
- McBride-Chang, C., Suk-Han Ho, C., 2005. Predictors of beginning reading in Chinese and English: a 2-year longitudinal study of Chinese kindergartners. *Sci. Stud. Read.* 9, 117–144.
- Muter, V., Hulme, C., Snowling, M.J., 2004. Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study. *Dev. Psychol.* 40, 665–681.
- National Institute of Child Health and Human Development (NICHD), 2000. *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769) U.S. Government Printing Office, Washington, DC.
- Neuman, S.B., 2004. The effect of print-rich classroom environments on early literacy growth. *Read. Teach.* 58 (1), 89–91.
- Olateju, M.A., 2010. Functional literacy empowerment for nomadic herders in Osun State, Nigeria. *Lang. Cult. Curric.* 23 (2), 109–121.
- Parker, R., Hasbrouck, J.E., Tindal, G., 1992. The maze as a classroom-based reading measure: construction methods, reliability, and validity. *J. Spec. Educ.* 26 (2), 195–218.
- Piper, B., Korda, M., 2011. EGRA Plus: Liberia. Program Evaluation Report. RTI International. <https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=541>.
- Piper, B., Zuilkowski, S.S., Mugenda, A., 2014. Improving reading outcomes in Kenya: first-year effects of the PRIMR Initiative. *Int. J. Educ. Dev.* 37, 11–21.
- Read, C., 1971. Pre-school children's knowledge of English phonology. *Harv. Educ. Rev.* 41, 1–34.
- RTI International, 2009. *Early Grade Reading Assessment toolkit*. Prepared for the World Bank, Office of Human Development, under Contract No. 7141961. RTI International, Research Triangle Park, NC. <https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&id=149>.
- RTI International, 2013. *Indonesia PRIORITAS Baseline Monitoring Report: An Assessment of Early Grade Reading*. <https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=511>.
- RTI International and International Rescue Committee, 2011. *Guidance Notes for Planning and Implementing EGRA*. <https://www.eddataglobal.org/reading/index.cfm?fuseaction=pubDetail&ID=318>.
- Snow, C.E., Burns, M.S., Griffin, P. (Eds.), 1998. *Preventing Reading Difficulties in Young Children*. National Academy Press, Washington, DC.
- Stahl, S.A., Murray, B.A., 1994. Defining phonological awareness and its relationship to early reading. *J. Educ. Psychol.* 86, 221–234.
- UNESCO, 2014. *Teaching and Learning—Achieving Quality for All. Education for All Global Monitoring Report*. <http://unesdoc.unesco.org/images/0022/002256/225660e.pdf> (accessed 07.08.14).
- Vellutino, F.R., Scanlon, D.L., Sipay, E.R., Small, S.G., Pratt, A., Chen, R., Denckla, M.B., 1996. Cognitive profiles of difficult-to-remediate and readily remediated poor readers: early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *J. Educ. Psychol.* 88, 601–638.
- Wagner, D.A., 2011. *Smaller, Quicker, Cheaper: Improving Learning Assessments for Developing Countries*. UNESCO International Institute for Educational Planning, Paris, France.
- Wagner, R.K., Torgesen, J.K., Rashotte, C.A., 1994. The development of reading-related phonological processing abilities: new evidence of bi-directional causality from a latent variable longitudinal study. *Dev. Psychol.* 30, 73–87.
- Wagner, R.K., Torgesen, J.K., Rashotte, C.A., 1999. *Comprehensive Test of Phonological Processing: CTOPP*. Pro-ed, Austin, TX.
- Wang, M., Park, Y., Lee, K.R., 2006. Korean-English biliteracy acquisition: cross-language phonological and orthographic transfer. *J. Educ. Psychol.* 98, 148–158.
- WHO (World Health Organization), 2011. *Monitoring maternal, newborn and child health: understanding key progress indicators*. http://www.who.int/healthmetrics/news/monitoring_maternal_newborn_child_health.pdf (accessed 06.08.14).
- Yesil-Dagli, U., 2011. Predicting ELL students' beginning first grade English oral reading fluency from initial kindergarten vocabulary, letter naming, and phonological awareness skills. *Early Child. Res. Q.* 26 (1), 15–29.