Review of Early Grade Reading Assessment (EGRA)

Tore Bernt Sørensen, Graduate School of Education, University of Bristol
t.b.sorensen@bristol.ac.uk

Introduction

This brief review of Early Grade Reading Assessment (EGRA) is based on desk research and a review of relevant literature. The purpose of the review is to introduce EGRA and outline the main concerns associated with this assessment format. For those interested in more details, depth and context, references are provided.[1]

The review proceeds as follows:
1. The institutional origins of the Early Grade Reading Assessment (EGRA) are outlined.
2. A brief description of the EGRA assessment.
3. Main concerns related to EGRA.
4. Concluding remarks

As pointed out by Wagner (2011, p.130), assessments are here to stay, and they will increasingly be used globally and locally for a variety of political and practical purposes. The central question is not whether a particular type of assessment is better or worse. It depends on their purpose and assessments are only as good as the uses that are made of them. Therefore, effective use does not only refer to the technical parameters or statistical methodologies. Wagner suggests that the design, planning and management of assessments today must "*put a greater priority on near term, stakeholder diverse, culturally sensitive, and high-in-local-impact assessments*" (Wagner, 2011, p.145).

According to Bartlett *et al.* (2015, p.312), EGRA currently exert "*matchless influence*" on assessment policies. EGRA appears to have cultivated an important niche in the contemporary learning assessment industry. In contrast to other international assessments like the OECD Programme for International Student Assessment (PISA), EGRA focuses on learning in the early grades of education. A related assessment type is the Early Grade Mathematics Assessment (EGMA). By mid-2014, more than 20 countries had conducted EGMAs (UNESCO, 2015, p.192). However, this note limits itself to focus on EGRA.

---

[1] Two accounts stand out in particular: Wagner (2011) and Bartlett *et al.* (2015). Gove and Cvelich (2011), Gove and Wetterberg (2011), Dubeck and Gove (2015) are key accounts by individuals directly involved in developing and implementing EGRA.
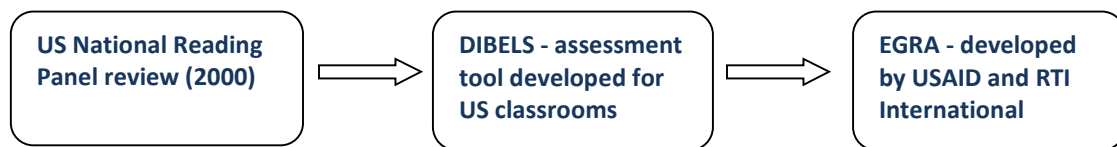
Within the context of the Education 2030 agenda and the Sustainable Development Goals, the origins, impact and possibilities of EGRA are particularly important to consider. The emphasis on quality education in for instance the Incheon Declaration means that EGRA-hybrid type "smaller, quicker, cheaper" assessment programmes might increasingly appear as a viable tool to many decision-makers seeking to reinforce the learning of basic skills.

Recent developments in the field of education and development are likely to further the use of EGRA, and the crude quality measures of standards-based performance it relies on, especially in low income countries. The major agencies of the Learning Metrics Task Force (2013; see also Learning Metrics Task Force, no date, on the *International Platform to Support Learning Assessment Systems*), the World Bank, Center for Universal Education at Brookings, and the Global Partnership for Education are thus all directly engaged with promoting EGRA type assessments.

Combined with the current drive towards Results-Based Financing (RBF) in education development (see Center for Universal Education at Brookings, 2011; Global Partnership for Education, 2014; World Bank, 2015), the SDG and Education 2030 agendas in particular are likely to prove instrumental for the further adoption or imposition of EGRA, with financing becoming conditional upon results as measured on the basis of EGRA type assessment.

1. The institutional origins of EGRA

Bartlett *et al*. (2015) and Wagner (2011) trace the origins of EGRA concept of reading assessment to the US around 2000. Their accounts might be visualized in the following way:

US National Reading Panel review (2000)	→	DIBELS - assessment tool developed for US classrooms	→	EGRA - developed by USAID and RTI International

In 2000, the US National Reading Panel issued the influential report *Teaching Children to Read*. This report simplified complex debates about reading acquisition and language development and proposed 5 pillars of reading:

1. phonemic awareness, or the ability to identify the individual sounds in spoken words
2. phonics, or the correspondence of letters (graphemes) to sounds (phonemes)
3. fluency, which is the ability to read text accurately and quickly, with natural prosody
4. vocabulary
5. comprehension, which is the ability to understand and communicate meaning from what is read.

Subsequently, the Reading First guidelines, part of the US reform *No Child Left Behind* (NCLB), rested heavily on this five-pillar model. This was for instance reflected in the guidelines' focus on fluency.

Another major inspiration for the EGRA assessment was the *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS) assessment. DIBELS is a continuous assessment classroom tool developed for use in the US. The assessment reduces reading to discrete skills and then condenses those skills to isolated, quantitative measures. DIBELS consists of a set of short, timed tests meant to measure phonemic awareness, alphabetic principle, fluency, reading comprehension, and vocabulary.

DIBELS has been widely criticized for making claims not based in evidence; distorting the skills required to read and then testing only a fragment of those skills; emphasizing speed over accuracy; proving difficult to administer consistently; and benefiting financially from the inappropriate promotion of tests as part of the federal Reading First program in the wake of NCLB (see e.g. Goodman, 2006; Kamii and Manning, 2005; Samuels, 2007; Shelton *et al,* 2009).

In 2006, the United States Agency for International Development (USAID) commissioned a desk review of early grades reading that also adopted the five-pillar model of reading. The USAID review recommendations endorsed tasks developed for English monolinguals. Later in 2006, USAID, through its Education Data for Decision Making (EdData II) project, contracted with Research Triangle Institute (RTI) International to develop an instrument for assessing early grade reading. RTI International, with the input of an expert panel, took the DIBELS subtests, modified them slightly to different languages, and field tested them. The result was EGRA (Gove and Cvelich, 2011; RTI, 2009).

According to Bartlett *et al.* (2015, pp.309-310) and Bloch (2014), EGRA struck a chord in USAID and the World Bank and expanded rapidly from the mid-2000s. EGRA proved instrumental for RTI International and partner organizations as a policy lever to demonstrate disappointing learning levels in the early grades ("*Assessment: Illuminating the reading deficit*", as it is put in Gove and Cvelich, 2011, p.19) to government officials and politicians and the resulting need for further monitoring, early grade reading interventions, and wider reform initiatives in teacher training and continuing professional development.

It is remarkable that the very term Early Grade Reading Assessment, with its particular origins in the US, NCLB, USAID and RTI International, apparently has come to be adopted as a generic concept for this type of assessment programmes. The trenchancy and impact of EGRA is reflected in the fact that the *Education For All Global Monitoring Reports* (see UNESCO, 2014; UNESCO, 2015) refer to EGRA and use the assessment data in their analyses. According to UNESCO, more than 60 countries carried out one or more EGRAs between 2007 and mid-2014 (UNESCO, 2015, p.192). In 2014, USAID funded at least 25 three-to-five year projects to the tune of $700 million.

2. The EGRA assessment

The Early Grade Reading Assessment (EGRA) is an individually administered oral assessment of the most basic foundation skills for literacy acquisition in early grades. The assessment requires about 15 minutes per child. It has been designed as an inexpensive and simple diagnostic of individual student progress in reading. The assessment focuses on what it labels the ''three early stages of reading acquisition":

Table 4. Early Grade Reading Assessment Components

The reading skills measured by early grade reading assessments focus on three early stages of reading acquisition (see Table 1). Although the average rate at which children pass through these phases varies by country and language, the following provides rough guidance for when most children should acquire these skills:

Stage	Test Components*
Stage 0: Emergent Literacy Birth to grade 1	• Concepts about print • Phonemic awareness • Listening comprehension
Stage 1: Decoding Beginning grade 1	• Letter naming • Letter sounds • Syllable naming • Nonsense word reading • Familiar word reading
Stage 2: Confirmation and Fluency End of grade 1 to end of grade 3	• Paragraph reading (oral reading fluency) with comprehension • Dictation

* Not all components are tested in all languages.

Source: **Adopted from Gove and Cvelich (2011, p.20)**

Different tasks can be adapted or excluded in different locations, so the test components mentioned are available but may not be employed. For example, the listening comprehension task is often not included. According to Bartlett et al. (2015), most of the EGRA reports on tests implemented internationally include measures across all three stages. However, the greatest variation in skills shows up in emergent (Stage 0) and decoding (Stage 1) skills, while confirmation and fluency (Stage 3) scores are regularly limited or non-existent. Accordingly, decoding, phonics and phonemic awareness have become stressed in EGRA assessment practices more than comprehension.

In contrast to programmes such as PISA or TIMSS, the dimension of competitive comparison between countries is not emphasised with EGRA. EGRA is currently not being presented as comparable in the traditional psychometric sense across different languages or tests. So, each EGRA is in principle locally tailored and developed (Gove 2012). Yet, with

Wagner (2011, p.43) argues that EGRA represent a hybrid type of assessment. Wagner puts EGRA into context and distinguishes between four main types of assessments: national, regional, international and hybrid. He points out that hybrid assessments in recent years have come to focus more directly on the needs of low-income countries assessment contexts. Hybrid assessments were initially conceptualized under the acronym "smaller, quicker, cheaper" (SQC) methods of literacy assessment. The basic idea is to see whether large-scale educational assessment methodologies could be transformed into hybrid methods that are just big enough, faster at capturing, analyzing and disseminating data, and cheaper in terms of personnel and cost outlays. As such, with their particular profile EGRA-type assessments might provide benefits not available with other assessment types in terms of scale, timeliness and cost-efficiency. These three parameters summarize the cost issues of assessments and should always be considered, especially in the context of resource-constrained low income countries (please, see appendix for benefits and limitations according to Wagner).

3. Concerns with EGRA

Wagner (2012) points out that there is a lot of consensus to be found among test-makers globally concerning the technical parameters of a 'good' assessment. However, the following issues all remain contested, particularly in low-income countries where the growth in assessments is most rapidly expanding and where the empirical base is least developed:

- who gets tested
- what gets tested
- when testing occurs
- how tests are conducted
- why is the test conducted

These issues are closely associated with each other. For example, if minority language speakers or refugees are included in the assessment population (the *who*), it is likely to have implications for the other issues, such as the type of content (the *what*) that should be included in the assessment. The *why* question is particularly contentious, since there are often multiple reasons why learning assessments are put into place.

In their discussion of the origins of EGRA and as the testing format has been implemented so far internationally, Bartlett *et al.* (2015, p.310) argues that EGRA warrants far greater debate than it has received to date. Yet, before turning to the main concerns, the potential strengths of EGRA and what is known about the impact of EGRA should be addressed.

In terms of strengths, Wagner (2011, pp.43, 101) and Gove (2012) point out that EGRA-type assessments offer the following possibilities:

- EGRA focuses on early grade learning and interventions unlike large scale educational assessments which in general do not target children until fourth grade and often much later. At these later point children may be far behind in reading development. This is clearly an important point in terms of social justice and equality (see also Barrett 2014).
- EGRA is flexible, adaptable and can be tailored to any language and orthography without the constraints of strict comparability between countries and systems
- population samples can be smaller as EGRA is currently designed to be a monitoring device, rather than a national representative high-stakes assessment
- the time taken between design and reporting can be considerably less than other major assessments, because of its smaller overall size
- currently, the cost per learner in EGRA appears similar to large scale educational assessments. However, costs are likely to drop for EGRA in the future as tools become more familiar and staff better trained.

Concerning the impact of EGRA assessments, Bartlett *et al.* (2015) review all eight publically available impact evaluations of EGRA-inspired programs. These programmes took place in eight African countries: Egypt, Gambia, Kenya, Liberia, Malawi, Mali, Mozambique, and South Africa.
It should be noted that the interventions assessed in these evaluations focused on training and equipping teachers to teach reading skills to students in grades 1, 2, and/or 3. The evidence is sparse as the eight evaluations varied in rigor. A particular limitation of the evaluation designs was that the promotion of reading skills acquisition in nearly all cases was framed as happening solely through schools. Moreover, the programmes varied in duration from two months to two years and in the intensity of the support provided in terms of number of teacher training sessions and amount of lesson materials provided to the schools.

The results from the programmes are ambiguous. The evaluations show positive results for some reading skills. Longer terms of EGRA interventions may have shown larger impact. Yet, there is little evidence that these interventions enabled much reading comprehension. So, it is possible that alternative assessment programmes based on other models of reading and learning, tailored to local contexts and the target language(s) would result in greater reading growth (Bartlett *et al*, 2015; for a less critical account, see Gove and Cvelich, 2011, and Davidson and Hobbs, 2013).

The short overview below points out a number of contested issues associated with EGRA.

3a. The conception of reading and language development
The stage model of reading underlying EGRA has been widely criticized in the literature. RTI International staff themselves are aware that the model has some limitations (Gove and Cvelich, 2011, p.20):

> "*Although the average rate at which children pass through these phases varies by country and language, the following* [three stages 0, 1, and 2] *provides rough guidance for when most children should acquire these skills.*"

However, the idea that reading is acquired in 'stages,' is not universally embraced. Accordingly, the framing of reading acquisition as 'stages' is problematic. For example, many reading experts insist that comprehension and fluency must be taught simultaneously with decoding skills. The problem is further aggravated by the stress on the initial two stages (0 and 1) in the practical implementation of EGRA internationally (Bartlett *et al.* 2015; Bloch 2014).

Bias towards phonics
Effectively, EGRA has isolated 'core reading components' rather than emphasizing how they develop together. This is represented in the controversial task where children read 'nonsense words', which pits phonics against comprehension (Bartlett *et al.* 2015; Bloch 2014).

Framing of reading as isolated domain of language development
EGRA ignores the important and developmentally essential interactions between print exposure, reading and oral language development, as well as the links between reading and writing. Oral language development often cultivated at home and in the community, helps children to expand their vocabulary (Bartlett *et al.* 2015; Wagner, 2011; Bloch, 2014).

Timing of tasks undermine demonstration of comprehension
The assessment requires that a child read aloud for 1 min, and then answer questions based on that reading. The use of stopwatch is likely be misinterpreted by students, professionals, parents and decision-makers. While stopwatches assure uniformity and improved speed of testing time, it also signals that the speed of reading is a critical benchmark of good reading (Wagner 2011, p.96). The strict timing also means that slower readers are not allowed to demonstrate fully their comprehension due to the imperative to time the task. Accordingly, EGRA has missed the opportunity to capture global variation in how rates of reading relate to comprehension (Bartlett *et al.* 2015).

3b. EGRA in languages other than English and multilingual contexts: Anglo-centrism, linguistic and pedagogical imperialism
Wagner (2011, p.43) argue that hybrid type assessments like EGRA have particular strengths since they offer opportunities for adaptation to local contexts and languages, unlike large-scale educational assessments like PISA. Yet, according to Bartlett et al. (2015), Trudell and Schroeder (2007), and Bloch (2014) this potential for meaningful adaptation does not tend to be put into practice.

Different alphabetic languages require different reading strategies, and EGRA-based approaches may not be appropriate (Bartlett *et al*. 2015, p.311; Bloch, 2014). Bartlett *et al.* (2015, pp.310-311) ask how appropriate it is to apply the stage model of reading, based on Anglo-centric reviews of the empirical evidence regarding how monolingual English-speaking children learn to read, to the processes of learning to read in other languages and scripts.

Trudell and Schroeder (2007), Schroeder (2013) and Bloch (2014) argue for taking local contexts and languages seriously, and, in particular, for matching reading methodologies in Africa to the linguistic characteristics of the learners' languages. Thereby, the extent to which EGRA is applied as a lever for linguistic and pedagogical imperialism would be reduced. Reading pedagogies designed in other contexts for other languages would be less relied upon when designing policies.

3c. Local capacity-building needs greater attention
Wagner (2011, pp.108-109) points out that high-level skills are required in the design, planning and management of assessments. These skills are in short supply globally and especially in low income countries. Globally, the variety of assessments available today is wide, and officials and politicians are confronted with difficult decisions with regard to choosing, developing, and adapting assessments to national and local contexts and languages. Ensuring that policymakers and key stakeholders have accurate information on the actual costs of assessments is a critical step in identifying appropriate tools to inform and influence initiatives aimed at improving educational outcomes.

4. Concluding remarks

In way of conclusion, it is clear that the various concerns mentioned above overlap and have mutual implications. Thereby, they illustrate Wagner's (2012) point concerning the mutual implications of the *who, what, when, how* and *why* of assessment programmes.

It should be stressed that the issues might have wider effects on policy and thinking about learning and educational practices among teachers, teacher educators and politicians. Due to the EGRA model of reading acquisition, reform initiatives and pedagogical responses might become biased towards decoding and phonics with insufficient attention to other components of reading. This has potential implications for national curricula, textbooks, teacher training, teaching practices and may overall lead to a failure to promote language- and literacy-rich environments inside and outside school (Bartlett *et al,* 2015; Wagner, 2011, pp.87-88). Focusing on South Africa, Bloch (2014) provides a damning critique of EGRA and the detrimental effects the assessment format has had on early literacy pedagogy. On the other side, monitoring learning in early grades help to draw attention to discrepancies in learning outcomes in the lower years of primary education which is clearly positive from a social justice perspective (Barrett, 2014; Gove, 2012; Wagner, 2011).

Acknowledging that EGRA-type assessments have particular strengths, this review has pointed out some of the main issues with EGRA. Both Bartlett *et al.* (2015) and Wagner (2011) provide helpful recommendations for how EGRA-type assessments may be developed and used in more constructive ways.

Summative external assessments are here to stay. This appears to be the case too with EGRA, at least for years to come. Their popularity should be discussed with reference to power structures in global educational governance, the relations between donors and recipients, international business interests in education, and post-colonialism. The main implementer of EGRA so far has been RTI International, based in North Carolina, US. Yet, these issues do not appear to have been treated in depth in the literature so far.

References

Barrett, A.M. (2014). Measuring literacy post-2015: some social justice issues. *In*: McIlwraith, H. (ed). *Language Rich Africa. Policy dialogue. The Cape Town Language and Development Conference: Looking beyond 2015.* London: British Council, 71-77.

Bartlett, L., A.J. Dowd & C. Jonason (2015). Problematizing early grade reading: Should the post-2015 agenda treasure what is measured? *International Journal of Educational Development,* 40, 308–314.

Bloch, C. (2014). Growing young readers and writers: underpinnings of the Nal'ibali National Reading-for-Enjoyment Campaign. *In*: McIlwraith, H. (ed). *Language Rich Africa. Policy dialogue. The Cape Town Language and Development Conference: Looking beyond 2015.* London: British Council, 50-56.

Center for Universal Education at Brookings. (2011). *A global compact on learning: taking action on education in developing countries.* Washington DC: Center for Universal Education at Brookings.

Davidson, M. & J. Hobbs (2013). Delivering reading intervention to the poorest children: The case of Liberia and EGRA-Plus, a primary grade reading assessment and intervention. *International Journal of Educational Development*, 33, 283–293.

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.

Global Partnership for Education (2014). *The New Funding Model: A Results-Based Approach for the Education Sector*. Available from [July 2015]: http://www.globalpartnership.org/fr/download/file/fid/44686%20

Goodman, K.S. (ed. 2006). *The truth about DIBELS: what it is, what it does*. Portsmouth, NH: Heinemann. Available from [July 2015]: http://www.heinemann.com/shared/onlineresources/E01050/wilde.pdf

Gove, A. (2012). Think global, act local: how early reading assessments can improve learning for all. In Wagner, D.A., M. Lockheed, I. Mullis, M.O. Martin, A. Kanjee, A. Gove & A.J. Dowd (2012). The debate on learning assessments in developing countries. *Compare: A Journal of Comparative and International Education*, (42)3, 509-545.

Gove, A. & P. Cvelich (2011). *Early Reading: Igniting Education for All. A report by the Early Grade Learning Community of Practice* (Revised Edition). Research Triangle Institute.

Gove, A. & A. Wetterberg (2011). *The Early Grade Reading Assessment: Applications and Interventions to Improve Basic Literary*. Research Triangle Institute,

Kamii, C. & M. Manning (2005). Dynamic Indicators of Basic Early Literacy Skills (DIBELS): A Tool for Evaluating Student Learning? *Journal of Research in Childhood Education*, 20(2), 75-90.

Learning Metrics Task Force (2013). *Towards universal learning: recommendations from the Learning Metrics Task Force*. Montreal and Washington DC: UNESCO Institute of Statistics and Center for Universal Education at Brookings.

Learning Metrics Task Force (no date). *Assessment for Learning (A4L): An International Platform to Support Learning Assessment Systems*. Available from [July 2015]: http://www.uis.unesco.org/Education/Documents/assessment-for-learning.pdf

Samuels, S.J., 2007. Is speed of barking at text what we mean by fluency? Read. Res. Q. 42, 563–566.

Schroeder, L. (2013). Teaching and assessing independent reading skills in multi-lingual African countries: not as simple as ABC. *In*: C. Benson, K. Kosonen (Eds.) *Language Issues in*

Comparative Education: Inclusive Teaching and Learning in Non-dominant Languages and Cultures: Rotterdam: Sense, 245–264.

Shelton, N.R., B. Altwerger & N. Jordan (2009). Does DIBELS Put Reading First? *Literacy Research and Instruction*, 48(2), 137-148.

Trudell, B. (2013). Early grade literacy in African schools: Lessons learned. *In*: McIlwraith, H. (ed.) Multilingual Education in Africa: Lessons from the Juba Language-in-Education Conference. British Council.

Trudell, B. & L. Schroeder (2007). Reading methodologies for African languages: avoiding linguistic and pedagogical imperialism. *Language Culture Curriculum*, 20(3), 165–180.

UNESCO (2014). *Teaching and Learning: Achieving quality for all. EFA Global Monitoring Report 2013/4*. UNESCO.

UNESCO (2015). *Education For All 2000-2015: Achievements and Challenges*. UNESCO.

Wagner, D.A. (2011). *Smaller, Quicker, Cheaper: Improving Learning Assessments for Developing Countries*. Paris: UNESCO.

Wagner, D.A. (2012). What should be learned from learning assessments? *In*: Wagner, D.A., M. Lockheed, I. Mullis, M.O. Martin, A. Kanjee, A. Gove & A.J. Dowd (2012). The debate on learning assessments in developing countries. *Compare: A Journal of Comparative and International Education*, (42)3, 509-545.

World Bank (2015). *The Rise of Results-Based Financing in Education.* Available from [July 2015]: http://www.worldbank.org/content/dam/Worldbank/Brief/Education/RBF_ResultsBasedFinancing_v9_web.pdf

APPENDIX

Benefits and limitations of EGRA type assessments

TABLE 9.1. Summary of benefits and limitations of various assessments		
Type of Assessment	Benefits	Limitations
SQC/EGRA	• Localized design and content of test items, including in mother tongue • Sample-based assessment • Data may be collected by teachers • Ability to 'target' particular populations (e.g. by language, ethnicity, citizenship, out of school youth) • Value placed on core cognitive skills as needed for building reading proficiency • Assessment can begin at young age (first grade), making early detection possible. • Potential to affect instruction at individual level due to individualized testing • Can support directed professional development • Individualized approach can be used at other ages (such as with adult literacy education) • Ability to have policy impact not only at national level, but also at provincial, school and instructor level • Costs in time to completion, as well as cost/pupil assessed is likely to be lower than other assessments. • Can be undertaken by NGOs in collaboration with government – fast startup and turnaround	• Global credibility is still modest as of 2011, though growing • Local capacity building needs greater attention • Focus mainly limited to only first 3 years of schooling • Limited concurrent validity with curricular content • If undertaken principally by NGO, may be ignored by Ministry of Education. • Secondary analyses unlikely

Source: **Adopted from Wagner, 2011, p.132**