

Measurement

# Pitfalls of Measuring Comprehension with EGRA

Time Limits, Item Difficulty, and Inflated  
Fluency–Comprehension Relationships

**Authors:**

Cally Ardington (AFLEARN, University of Cape Town)  
Alicia Mendendez (University of Chicago)

## Contents

Abstract .....	1
1. Introduction .....	2
2. Oral Reading Fluency Timing and Its Impact on Comprehension .....	3
3. Misinterpretation of Fluency–Comprehension Relationships .....	7
4. Comprehension question difficulty .....	9
5. Discussion .....	11
6. Conclusion and Implications .....	12
References .....	14

## Abstract

This paper examines fundamental validity threats in the measurement of reading comprehension within the Early Grade Reading Assessment (EGRA). Although EGRA comprehension outcomes are widely used to identify proficient readers, analyse fluency–comprehension relationships, establish fluency benchmarks, and inform instructional guidance, the technical assumptions underpinning these measures remain largely unexamined. Drawing on published evidence and new data from early grade reading evaluations in South Africa and Nepal, we show that the standard one-minute time limit used in oral reading fluency (ORF) tasks introduces substantial bias into comprehension scores. While extending reading time does not meaningfully alter fluency measures, it significantly increases the number of comprehension questions learners can attempt and shifts the distribution of comprehension performance—particularly for slower readers who may understand the text but cannot read fast enough to reach key items. We demonstrate that both common scoring approaches—percent correct of attempted items and percent correct of total items—embed untenable assumptions that respectively inflate and underestimate comprehension. The one-minute time limit also induces mechanical correlations between ORF and comprehension, artificially strengthening the apparent fluency–comprehension relationship. Finally, analysis of PIRLS and EGRA items reveals wide variation in item difficulty, undermining the reliability of aggregate comprehension scores and the benchmarks derived from them. Together, these findings call for caution in interpreting existing EGRA-based evidence, highlight the need to re-examine benchmark estimates and cross-language comparisons, and underscore the importance of extending reading time and analysing item-level difficulty to improve the validity of comprehension measurement.

# 1. Introduction

National governments, donors, education researchers and policymakers have increasingly shifted focus from expanding access to education to improving the quality of education in developing countries. This has seen an increase in efforts to monitor system wide progress and evaluate specific interventions aimed at improving learning outcomes. Within these efforts, the Early Grade Reading Assessment (EGRA) has been used to assess early literacy outcomes in over 70 countries with results informing both classroom practice and education policy (Dowd and Bartlett 2019). EGRA data are increasingly being used to expand the empirical evidence base on reading development beyond studies on European languages in high income countries (Nagarajan et al. 2018, Campuzzano et al. 2018, Ardington et al. 2019, Kotze et al. 2019, Menendez et al. 2020, among many others).

The ultimate purpose of learning to read is to understand the information in a text and construct meaning from it. Reading comprehension is generally considered a complex phenomenon comprising many interacting components and processes. The relationship between these processes and their relative contributions to comprehension differ at the various stages of development. It is also likely that the relationship between these components and comprehension will be language specific, determined by the orthographic and linguistic features of the language. While there is broad agreement on the above, there is a robust and ongoing debate on the precise nature of these relationships (see for example Dowd and Bartlett 2019; Abadzi and Centanni 2020 and Dowd et al. 2020) and an active research agenda on understanding how these relationships differ by language and transfer between languages (Hoover and Gough 1990; Torppa et al. 2016; Kim 2017; Wang et al. 2019).

This paper does not engage directly with these debates but rather focusses on technical aspects of the existing EGRA measures of comprehension that inform much of this discussion. EGRA comprehension data have been widely used to identify learners who read with comprehension (Dowd et al 2019; Abadzi 2011); establish fluency benchmarks (Abadzi 2011; Jukes et al 2018a), examine the correlation between fluency and comprehension, investigate cross language transfer and structural relations between reading subskills; and to guide instructional practices (e.g. Dubeck and Gove 2015; Stern, Dubeck and Dick 2018).

Despite this wide use, insufficient attention has been paid to technical measurement issues and this paper seeks to contribute to knowledge on reading fluency and comprehension by highlighting the mis(use) and (mis)interpretation of EGRA data in the literature. Earlier papers in this spirit have considered the impact of variation in timing on assessments (Piper and Zuilkowski 2016; Zuilkowski et al. 2019; Dowd and Bartlett 2019; Betts et al. 2020), compared standard EGRA comprehension subtasks to alternatives (Zuilkowski et al. 2019) and raised selection bias issues around the exclusion of slow but comprehending readers (Bartlett, Dowd and Jonason, 2015; Dowd and Bartlett 2019). We reconsider the evidence from these earlier papers and draw on our own data from early grade reading external impact evaluations to revisit these issues.

This paper explicitly articulates the assumptions underlying each of the approaches to measuring oral reading comprehension and provides empirical evidence that suggests these assumptions are unlikely to hold. We show that, depending on which comprehension outcome measure is used, correlations

between fluency and comprehension are either attenuated or artificially inflated with the extent of the bias dependent on the learners' position in the reading fluency distribution.

The paper then considers the implications of these measurement issues for analyses that compare comprehension outcomes within and between samples and those that examine the association between fluency and comprehension. We highlight the potential misuse and misinterpretation of EGRA oral reading comprehension outcomes through the use of our own data and illustrative examples from the literature.

Our findings call in to question some technical recommendations from earlier studies. Beyond technical issues, we add to the knowledge base around reading with comprehension, both cautioning the interpretation of previous findings and highlighting potential pitfalls of using EGRA oral reading comprehension outcomes to inform applied research and policy.

## 2. Oral Reading Fluency Timing and Its Impact on Comprehension

Evidence from the EGRA oral reading fluency subtask suggests that varying the assessment time makes little difference to the measure of fluency. Piper and Zuilkowski (2016) measured oral reading fluency in Kenya using both a time-delimited and a passage-delimited assessment. They observed a strong correlation between one-minute and extended time fluency of 0.9 for English and 0.89 for Kiswahili. Mean scores were similar (0.4 word difference in English and 0.23 word difference in Kiswahili)<sup>1</sup> as were the standard deviations suggesting that the one-minute assessment adequately captured oral reading fluency. Ardington et al. (2019) and Menendez (2020) allowed children up to three minutes to complete a passage and recorded oral reading fluency at both the one-minute mark and overall. They find high correlations between one- and three-minute measures of oral reading fluency for Nepali children and isiXhosa and isiZulu children in South Africa. It appears that extending the time to allow learners to complete the passage, has little effect on the measure of oral reading fluency.

However, changing the time limit for reading the passage has a substantial impact on the validity of oral reading comprehension measures. Children are only asked comprehension questions related to the parts of the passage that they attempted to read within the time limit, which is usually one minute (RTI 2016). Extending the time limit to three minutes allows children the opportunity to attempt more comprehension questions (USAID 2017; Betts et al. 2020). For example, in the South African study, the percentage of grade 3 learners completing the passage rises from two to 57 percent for isiXhosa and from seven to 66 percent for isiZulu when an additional two minutes are allowed (Ardington et al. 2019). In Nepal, the percentage of grade 3 children completing the passage increased from four to 41 percent.

Oral reading comprehension scores are typically reported in one of two ways:

1. Percentage of attempted questions answered correctly

---

<sup>1</sup> Mean correct words per minute with standard deviation in parentheses: English timed 13.91 (21); English untimed 14.32 (22.84); Kiswahili timed 10.05 (14.37); Kiswahili untimed 9.82 (14.99).

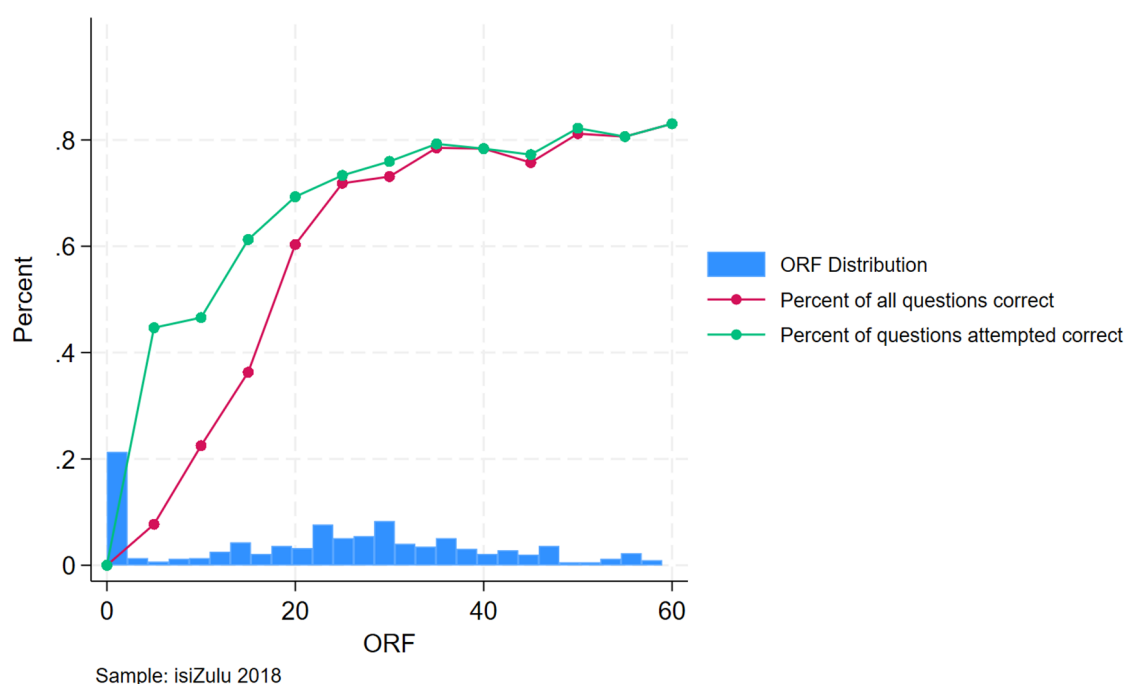
2. Percentage of total questions answered correctly, regardless of whether the learner reached that part of the passage

### *Inflation of comprehension via percent correct of attempted items*

The percent correct of attempted items is a measure that attempts to compensate for the fact that many children do not get to attempt all the questions. Using only attempted questions can artificially inflate comprehension scores, particularly for slower readers. This measure of reading comprehension implicitly assumes that learners would achieve the same average score for the questions for which they did not read far enough to attempt as they achieved on their answered questions. In other words, the difficulty level of the comprehension questions is assumed to be constant. However, this assumption is likely untenable as most EGRAs include several literal questions followed by one or more inferential questions.

It is not just the overall inflation of comprehension scores that is of concern but that the upward bias would be greatest for the learners attempting the least questions – those with the lowest fluency. Figure 1 compares the relationship between oral reading fluency (ORF) and reading comprehension scores calculated as a percentage of total questions and as a percentage of questions attempted for a 2018 sample of isiZulu learners to illustrate this point. In this South African study, learners in 2018 were only allowed to read for one minute while learners in 2019 were allowed up to three minutes to complete the passage. The inflation of scores at lower levels of fluency is very clear. In general, measuring reading comprehension as the percent correct of attempted questions is likely to attenuate the relationship between fluency and comprehension.

*Figure 1. Percent correct of attempted items inflates comprehension scores for slower readers*



In summary, the assumption that question difficulty is constant is unlikely to be met for most EGRA assessments and researchers should refrain from using the percentage correct of attempted items as their measure of reading comprehension.

#### *Underestimation of comprehension via percent correct of total questions*

Conversely, calculating comprehension as a percentage of all questions may underestimate comprehension. Analyses where comprehension scores are calculated as a percent correct of the total number of questions make the assumption that learners who did not read far enough to attempt a question would not have been able to correctly answer that question. This is a very strong assumption, and the evidence suggests that this is often not the case.

For studies where learners were allowed to continue reading the passage for an additional two minutes, we can work out how many comprehension questions they would have been able to attempt after reading for only one minute and create a one-minute comprehension score<sup>2</sup>. This score can then be compared against the comprehension score after reading for up to three minutes. Figure 2 shows the relationship between the one- and three- minute comprehension scores for isiXhosa and isiZulu in 2019<sup>3</sup>. It is clear from the figures that a substantial proportion of learners continue to improve their comprehension scores when they are asked additional questions. The assumption that learners who had not read far enough in one minute to attempt a question would not have answered it correctly is clearly not supported by these data. Average comprehension scores increase by 9, 20 and 23 percentage points in Nepali, isiXhosa and isiZulu respectively when learners are allowed more time to read the passage.

This finding contrasts with that of Piper and Zuilkowski (2016) who conclude that allowing learners additional time (up to 3 minutes) to read a passage does not result in significantly better comprehension scores, compared to the standard one minute. Based on their results they recommend continued use of the one-minute limit for efficiency considerations. However, their analysis has several limitations. First, they use two different passages and acknowledge that the passages and corresponding comprehension questions may differ in difficulty<sup>4</sup>. Second, the majority of learners completed the three-minute passage within one minute (68 percent for the English passage and 58 percent for the Kiswahili passage). A comparison of comprehension scores across the two passages for these learners tells us nothing about whether allowing learners more time to read a passage results in improved performance on comprehension. Indeed, when the comparison is restricted to those learners who took longer than one minute to complete the three-minute passage, there are significant differences in the comprehension scores (English: 4.1 percentage points, 0.2 SD,  $p < 0.001$ , Kiswahili: 4.3 percentage points, 0.19 SD,  $p < 0.001$ )

---

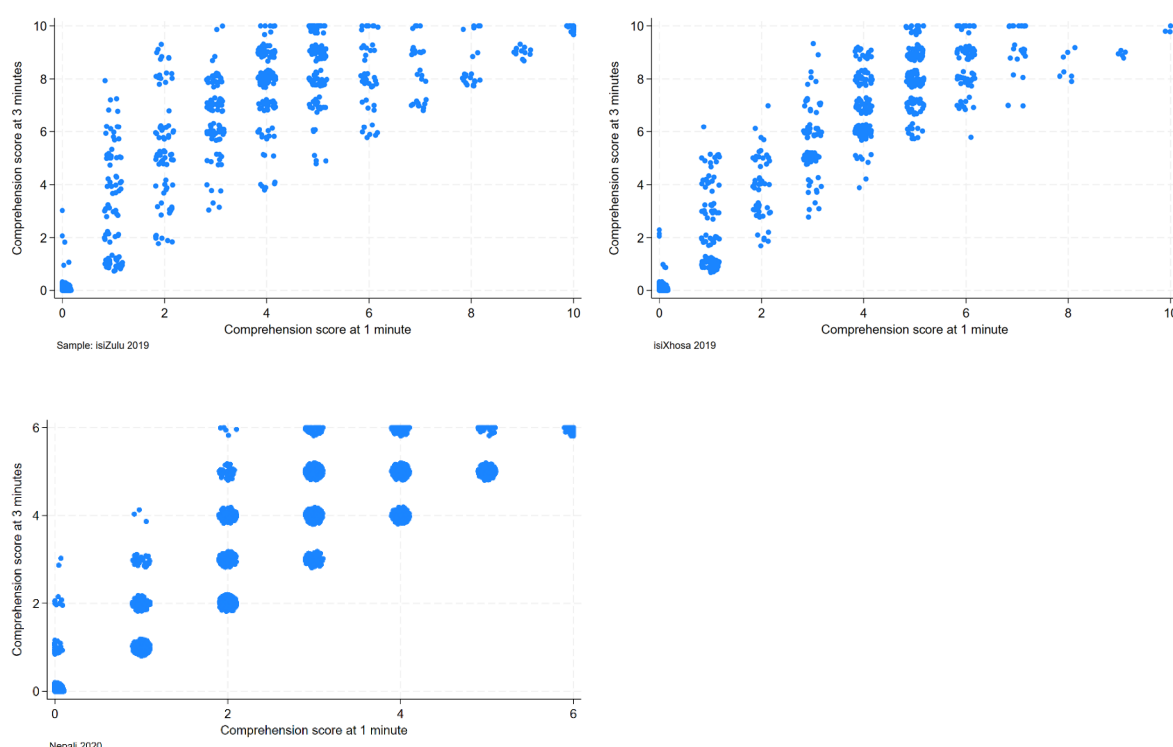
<sup>2</sup> This constructed one-minute comprehension score may be biased upwards if reading the whole passage aids comprehension by providing a more complete contextual frame for the story. On the other hand, it could be biased downwards as there are greater demands on working memory due to i) the time between reading the section of the passage covered in the first minute and answering the related questions and ii) the additional details of the story read in the extended time period. These biases are likely to be small and not relevant for the comparisons in this review.

<sup>3</sup> We have added some random noise to the data in order to be able to visualize the sample size at each point.

<sup>4</sup> They do not discuss differences in how far a learner needed to read to reach each comprehension question. For example, the third question required the learner to read 38 words in the Kiswahili untimed passage versus 29 words in the Kiswahili timed passage.

across the two passages. In a more recent paper, Zuilkowski et al. (2019) again recommend the continued use of the standard one-minute oral reading passage to measure comprehension as they find that allowing more time does not provide more information on reading comprehension. However, this conclusion is based on comparing reading comprehension for a short passage (66 words for English and 60 words for Kiswahili) with a time limit of one-minute against comprehension for a longer passage (228 words for English and 176 words for Kiswahili) with a time-limit of three minutes. This comparison speaks to issues such as whether speed, accuracy and working memory falter as time passes or whether exposure to more contextual information through a longer reading period aids comprehension. But is not instructive for considering whether to allow more than the standard one-minute time in order to include slower readers who may comprehend.

Figure 2. Relationship between one- and three- minute comprehension scores.



In the Philippines, Betts et al. (2020) compare comprehension scores from a passage with a one-minute reading limit against scores from a second slightly longer passage that learners were given three minutes to complete. For the one-minute reading limit, the standard EGRA protocol was followed and the passage was removed before the questions were asked, whereas the second passage remained in front of the learner as they responded to the comprehension questions. This makes it impossible to separate out the effects of being allowed more time from the opportunity to refer back to the text<sup>5</sup>. In addition, a clean comparison requires the two passages and corresponding questions to be at exactly the same level.

Dowd and Bartlett (2019) argue strongly against limiting learners' oral reading to one minute as this precludes the possibility that slower readers will be able to demonstrate comprehension. They claim to

<sup>5</sup> Allowing the learners to look back "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 (Dubeck and Gove 2015, p.319)".

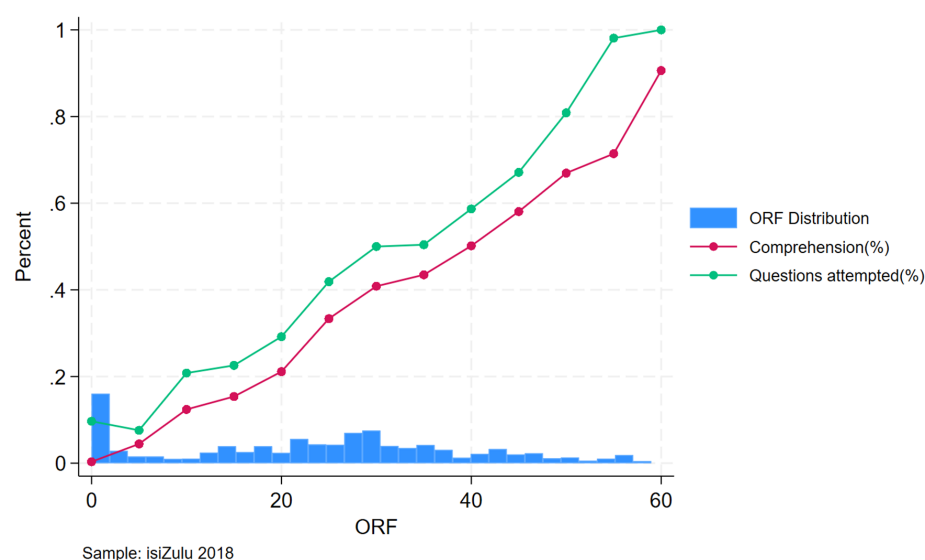
have “showed that, with more time to read a passage students demonstrated greater comprehension<sup>6</sup>”. However, of the 11 countries for which they have data there is only a comparison of a time limit of one minute against a limit of three minutes for Rwanda and unfortunately the implementation makes it impossible to separate out the effects of being allowed more time from being allowed a second chance to read the text and the opportunity to refer back to the text<sup>7</sup>.

A reconsideration of the evidence in Piper and Zuilkowski (2016), together with the evidence presented in Figure 2 suggest that limiting passage reading time to one-minute may bias estimates of the percentage of learners reading with comprehension downwards.

### 3. Misinterpretation of Fluency–Comprehension Relationships

However, the concern is not merely that we may exclude slow readers who comprehend, but limiting reading time distorts the relationship between fluency and comprehension. The relationship between fluency and comprehension is artificially inflated as learners are only asked questions relating to the parts of the passage they have read. To illustrate this point, Figure 3 shows the average number of questions attempted and the average comprehension score at each level of ORF for Grade 3 isiZulu learners in 2018<sup>8</sup>. The mechanistic relationship between fluency, number of questions attempted and comprehension scores is clearly evident with the one-minute time limit.

*Figure 3. Fluency, number of questions attempted, and mean comprehension score with one-minute reading time*



<sup>6</sup> This claim is made in response (Bartlett et al. 2020, 313) to a commentary (Abadzi and Centanni 2020) on the original paper.

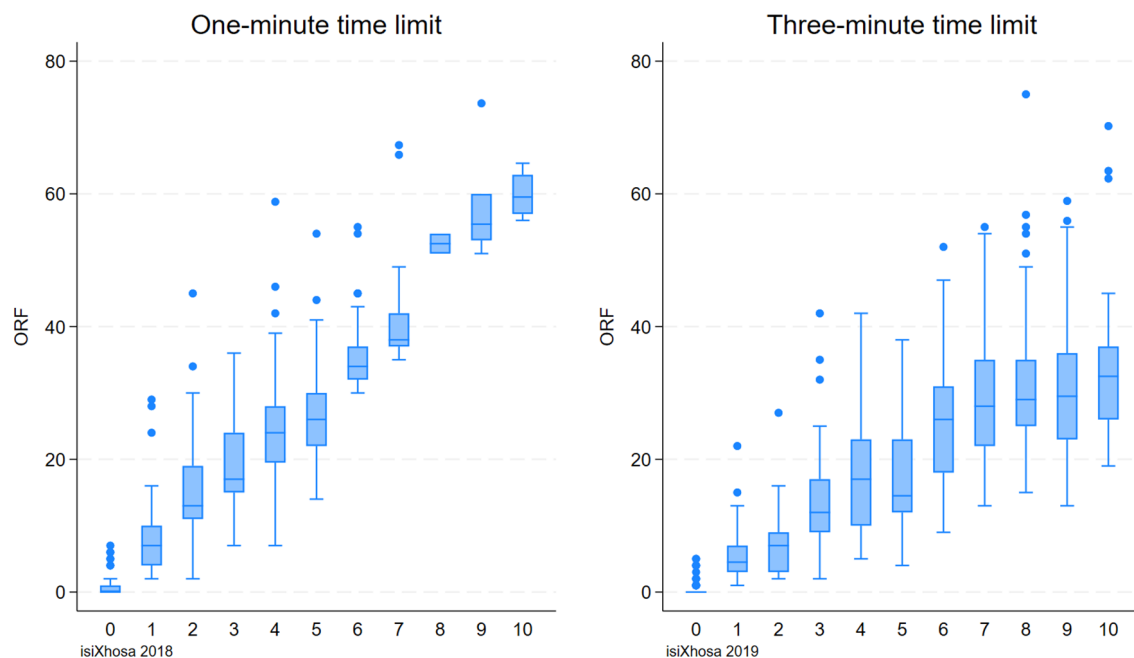
<sup>7</sup> Comprehension was tested in two rounds with the first round following the standard one-minute limit for passage reading and removing the passage before asking the comprehension questions. In the second round, learners who had not finished the passage were given an additional two minutes to complete the passage and were then asked the comprehension questions with the text in front of them.

<sup>8</sup> Figures 3 and 4 show similar patterns across all three languages.



Figure 4 further illustrates this point by showing the distribution of ORF by comprehension score (out of a maximum of 10) for Grade 3 isiXhosa learners in 2018 (left panel) and 2019 (right panel). The left and right panels suggest a very different relationship between fluency and comprehension. The left panel shows a very strong linear relationship between fluency and comprehension. Learners are only asked comprehension questions related to the parts of the passage that they read within the time limit. The majority of learners do not complete the passage within one-minute resulting in a mechanical relationship between fluency and comprehension induced by the number of questions attempted. When learners are allowed more time to read (right panel), the relationship between ORF and comprehension flattens out at lower levels of ORF. There is now a much larger sample, including slower readers, reaching each comprehension question.

*Figure 4. Relationship between fluency and comprehension scores with one-minute and three-minute reading times*



In support of their conclusion that allowing more time does not lead to significantly better performance on comprehension, Piper and Zuilkowski (2016) compare the correlation between fluency and comprehension across the one-minute and three-minute passages. This is problematic as the correlation between fluency and comprehension will be artificially inflated for the standard one minute but not for the three minutes where the vast majority of students were able to complete the passage. As expected, correlations were higher for the standard one minute than for the three minutes (English: 0.77 versus 0.69, Kiswahili: 0.88 versus 0.80).

## 4. Comprehension question difficulty

A further limitation of EGRA comprehension lies in the difficulty of the items themselves. Although EGRA typically distinguishes between literal and inferential questions, empirical evidence shows substantial overlap in their difficulty. This complexity is usefully illustrated by the PIRLS comprehension taxonomy and empirical anchoring of each question to the various benchmark levels. A priori PIRLS classifies questions according to four broad-based comprehension processes that range from literal (Focus on and Retrieve Explicitly Stated Information) to inferential (Make Straightforward Inferences; Interpret and Integrate Ideas and Information; and Evaluate and Critique Content and Textual Elements). Empirical assessment data is then used to classify the difficulty level of each item according to the lowest benchmark at which a specified minimum percent of learners answered correctly<sup>9</sup>. Table 1 shows for each comprehension process, the classification of 2016 PIRLS Literacy questions by benchmark level. Although the majority of literal questions are anchored at the low international benchmark, some literal questions present a greater challenge than questions tapping into higher order comprehension processes. Similarly, some inferential questions are accessible to learners performing at the low international benchmark.

*Table 1. Percentage of questions classified at each benchmark level by comprehension process category – PIRLS Literacy 2016*

Benchmark level	Comprehension process				
	Focus on and Retrieve Explicitly Stated Information	Make Straightforward Inferences	Interpret and Integrate Ideas and Information	Evaluate and Critique Content and Textual Elements	All
Low	80%	51%	23%	9%	56%
Intermediate	14%	33%	38%	55%	26%
High	5%	16%	21%	36%	13%
Advanced	1%	0%	18%	0%	4%
All	100%	100%	100%	100%	100%

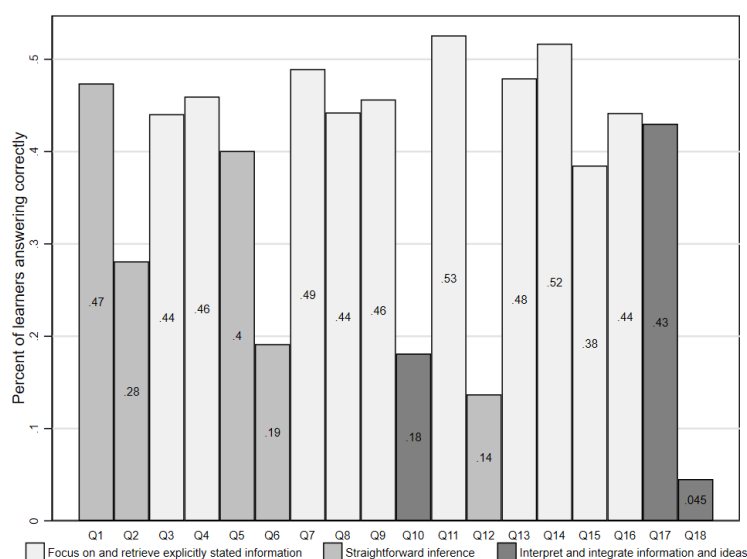
*Note: authors calculation from Appendix 13B and Appendix 13C in Martin et al. (2017)*

Figure 5 presents the percentage of learners correctly answering each question about a narrative passage for the most populous South African language, isiZulu<sup>10</sup>. Not only is there large variation in difficulty within each comprehension process, there appears to be substantial overlap in the range of difficulty between the processes. For example, in Grade 3 samples from Nepal, isiXhosa, and isiZulu, performance on individual literal questions ranged from 38 to 53% correct, while inferential questions ranged from 5 to 47%.

<sup>9</sup> For example, a multiple-choice question is classified as low international benchmark if at least 65 percent of learners with scores between 390 and 410 answer the question correctly.

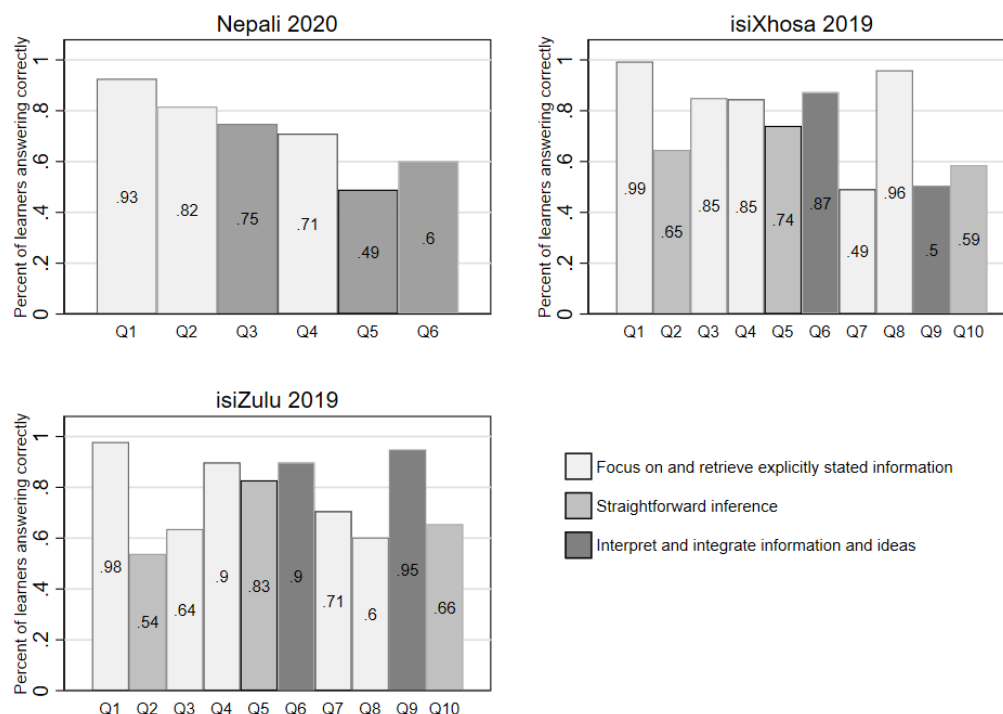
<sup>10</sup> Patterns are similar for all 11 official South African languages. That said there are differences in item difficulty across languages that further highlight the challenge of establishing comparability across questions.

Figure 5. Percent of isiZulu learners answering each question on *The Pearl* passage correctly - PIRLS Literacy 2016



Applying the PIRLS comprehension taxonomy to our EGRA comprehension questions, we observe a similar pattern with considerable variation within each comprehension process and some overlap between processes. Figure 2 shows the percentage of learners correctly answering each comprehension question for our Grade 3 Nepali, isiXhosa and isiZulu samples<sup>11</sup>.

Figure 6. Percent correct by comprehension question



<sup>11</sup> In order to keep the sample constant across questions, the samples are restricted to learners attempting all the questions.

Although some papers using EGRA data suggest that learners find the inferential questions more challenging, there is very little interrogation of question difficulty or comparability and it is exceedingly rare to see the results for individual questions. The focus on questions, if any, tends to be on their collective reliability or uni-dimensionality. The concerns with unexamined question difficulty are exacerbated in analyses that focus on a specific comprehension level (e.g. 80 percent correct) to identify proficient readers or derive fluency benchmarks.

## 5. Discussion

In contrast to previous evidence, we show that the one-minute limit used in many oral reading fluency assessments, has a substantial impact on the measurement of reading comprehension. Extended reading time (e.g., up to three minutes) does not meaningfully alter fluency scores but greatly improves the validity of reading comprehension scores—by allowing slower readers to demonstrate understanding and reducing mechanical inflation of fluency-comprehension correlations.

These findings highlight three major pitfalls of EGRA comprehension measurement: (1) bias introduced by the one-minute time limit, (2) artificial inflation of fluency-comprehension correlations, and (3) variability in item difficulty that undermines score validity. Together, these issues compromise the reliability of EGRA as a tool for assessing comprehension in the early grades. Crucially, these are not merely technical concerns: they shape how learners are classified, how programme effects are interpreted, how reading benchmarks are set, and how previous work on the interrelationships between accuracy, fluency, and comprehension should be understood.

Given this, fluency-comprehension relationships derived from one-minute EGRA data warrant careful re-examination, particularly where they have informed benchmark-setting, policy guidance, or cross-language comparisons.

A recent meta-analysis by Leachman, Wolters, and Kim (2025) shows that the magnitude of the fluency-comprehension relationship varies systematically by orthographic transparency, grade level, and stage of reading development. Extending this work to include more evidence from transparent orthographies in LMICs—as well as from assessments that permit additional reading time—would substantially deepen understanding of these dynamics.

Crawford et al. (2025) synthesise findings from more than 230 EGRA reports across 48 countries with a primary focus on fluency but their analysis includes correlations between oral reading fluency and comprehension, and between reading and listening comprehension. In most of the underlying studies, the one-minute limit applied to the fluency task, which would systematically depress observed comprehension scores and inflate the apparent strength of the fluency-comprehension association.

An area where EGRAs one-minute timing have been consequential is the work on establishing fluency benchmarks (RTI 2016, RTI 2017, Jukes et al. 2018b, Abadzi 2011). Benchmarking methods typically focus on identifying the fluency level associated with a specific comprehension score, most commonly 80 percent. In contexts, where very few learners complete reading the passage within one minute and therefore do not attempt all the comprehension questions. This results in benchmarks being set beyond

reach of the vast majority of learners, limiting the usefulness for tracking incremental improvements. For example, RTI (2017) find the percentage of learners reaching benchmarks to be only around five percent in the Middle Eastern and Africa.

Approaches to benchmarking fluency to a specified comprehension level are also very sensitive to the placement of questions in relation to how far the learner has to read in the text. For example, Jukes et al. (2018) calculate a logistic benchmark of 32.2 words per minute for Nepali. The instrument includes six comprehension questions, and a learner can only possibly achieve 80 percent comprehension if they read far enough in the passage to attempt the fifth comprehension question (45 words). If we remove the fifth question, then a learner has the potential to achieve 80 percent comprehension if they attempt the fourth question, which is asked of learners who read at least 36 words of the passage. Using the same methodology and recalculating the benchmark, we obtain an estimate of 24.8<sup>12</sup> words per minute. It is clear that the position of the question, in terms of the number of words that learners need to reach in order to have a chance of scoring 80 percent has an enormous impact on the benchmark. See Ardington et al. (2021) and Mohohlwane et al. (2022) for further potential problems with this approach to benchmarking fluency.

## 6. Conclusion and Implications

The widespread reliance on EGRA comprehension outcomes for monitoring foundational learning requires urgent re-examination. Evidence shows that extending the time limit beyond one minute significantly improves comprehension measurement, capturing the abilities of slower readers and reducing artefactual correlations with fluency. At the same time, substantial variation in item difficulty challenges the validity of using simple comprehension cut-offs. Future work should therefore: (i) revisit findings based on one-minute EGRA data, (ii) develop clearer guidelines for analysing and reporting comprehension item difficulty, and (iii) extend timing of the oral reading fluency to allow for the valid measurement of comprehension. Without such adjustments, comprehension outcomes derived from EGRA risk systematically misrepresenting learners' skills and misinforming education policy.

The widespread reliance on EGRA comprehension outcomes for policy and research requires urgent re-examination. Evidence from South Africa and Nepal shows that extending passage reading beyond one minute substantially improves the validity of comprehension measures, both by capturing the abilities of slower readers and by reducing artefactual correlations with fluency. At the same time, analyses of PIRLS and EGRA data reveal significant variation in comprehension item difficulty, undermining confidence in aggregate scores and benchmarks.

Future work should focus on three priorities:

- Revisiting findings based on one-minute EGRA comprehension outcomes, particularly those establishing fluency benchmarks or modelling fluency–comprehension relationships.
- Developing guidelines for handling comprehension item difficulty, including calibration and reporting of individual question performance.

---

<sup>12</sup> This difference is not due to there being only five questions. If we remove question four and leave in question five, then the benchmark estimate is 30.2.

- Ensuring that children are given additional time to complete reading the EGRA passage in order to validly measure comprehension.

In sum, one-minute EGRA comprehension outcomes risk systematic bias. By extending reading time and explicitly addressing item difficulty, assessments can provide a fairer, more accurate picture of children's comprehension skills — and better guide education systems seeking to ensure all children learn to read with understanding.

## References

- Ardington, C., Wills, G., Pretorius, E., Mohohlwane, N. and Menendez, A. (2021) "Benchmarking oral reading fluency in the early grades in Nguni languages" *International Journal of Educational Development* 84
- Abadzi, H. 2011. "Reading Fluency Measurements in EFA FTI Partner Countries: Outcomes and Improvement Prospects." Global Partnership for Education, GPE Working Paper Series on Learning, No. 1.
- Abadzi, H. and T. Centanni. 2020. "Why Fast and Effortless Reading Is Indispensable for Comprehension." *Comparative Education Review* 64(2): 299-308.
- Bartlett, L., A. J. Dowd, and C. Jonason. 2015. "Problematizing Early-Grade Reading: Should the Post-2015 Agenda Treasure What Is Measured?" *International Journal of Educational Development* 40:308–14.
- Betts K., M. Punjabi, S. Pouezevara, and C. Cummiskey. 2020. "2019 Regional Early Grade Reading Assessment (EGRA): Bahasa Sug, Chavacano, Magindanawn, and Mëranaw Findings Report." USAID All Children Reading–Philippines.
- Campuzano, L., C. Fernandez, J. Lugo-Gil, S. Glazerman, N. Murray and I. Padilla. 2018. "Evaluation of Amazonia Lee Reading Intervention in Peru, Final Report." Mathematica Policy Research. USAID. [https://pdf.usaid.gov/pdf\\_docs/PA00TCQ1.pdf](https://pdf.usaid.gov/pdf_docs/PA00TCQ1.pdf)
- Crawford, M., Raheel, N., Korochkina, M., & Rastle, K. 2025. Inadequate foundational decoding skills constrain global literacy goals for pupils in low- and middle-income countries. *Nature Human Behaviour* 9, 74–83. <https://doi.org/10.1038/s41562-024-02028-x>
- Dowd, A. J., and L. Bartlett. 2019. "The Need for Speed: Interrogating the Dominance of Oral Reading Fluency in International Reading Efforts." *Comparative Education Review* 63 (2): 189–212.
- Dowd, A. J., L. Bartlett, R. Khamis-Dakwar and K. Froud. 2020 "Measure Fluency, Vocabulary, and Comprehension in International Reading Efforts." *Comparative Education Review* 64 (2): 309-315.
- Dubeck, M. and A. Gove. 2015. "The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations." *International Journal of Educational Development* 40: 315–322.
- Gough, P and W. Tunmer. 1986. "Decoding, reading, and reading disability." *Remedial and Special Education* 7: 6-10.
- Jukes, M., C. Cummiskey, M. Gargano and P. Dubeck. 2018a "Guidance Note: Setting Data-Driven Oral Reading Fluency Benchmarks." RTI International, Research Triangle Office. Accessed from [https://www.roomtoread.org/media/984470/room-to-read\\_fluency-benchmarking-guidance-note\\_published-may-2018.pdf](https://www.roomtoread.org/media/984470/room-to-read_fluency-benchmarking-guidance-note_published-may-2018.pdf)
- Jukes, M., C. Cummiskey, M. Gargano and P. Dubeck. 2018b. "Data-Driven Methods for Setting Reading Proficiency Benchmarks." RTI International, Research Triangle Office. Accessed from [https://www.roomtoread.org/media/984466/room-to-read\\_fluency-benchmarking-analysis\\_report\\_may-2018.pdf](https://www.roomtoread.org/media/984466/room-to-read_fluency-benchmarking-analysis_report_may-2018.pdf)
- Kim, Y.G. 2017. "Why the simple view of reading is not simplistic: Unpacking the simple view of reading using a direct and indirect effect model of reading (DIER)." *Scientific Studies of Reading* 21: 310–333.
- Kotze, J., B. Fleisch and S. Taylor. 2019. "Alternative forms of early grade instructional coaching: Emerging evidence from field experiments in South Africa." *International Journal of Educational Development*, 66(1), 203-213.

- Leachman, M., Wolters, A., & Kim, Y.-S. G. 2025. The relation between text reading and reading comprehension varies as a function of developmental phase, orthographic depth, and measurement characteristics: Evidence from a meta-analysis. *Journal of Educational Psychology*, 117(3), 508–528. <https://doi.org/10.1037/edu0000932>
- Martin, M. O., Mullis, I. V. S. and M. Hooper (Eds.). 2017. “Methods and Procedures in PIRLS 2016.” TIMSS & PIRLS International Study Center, Boston College. Accessed from <https://timssandpirls.bc.edu/publications/pirls/2016-methods.html>
- Mohohlwane, M., Wills, G. and Ardington, C. (2022) A review of recent efforts to benchmark early reading skills in South African languages. In Spaull, N. & Pretorius, E. (Eds) Early Grade Reading in South Africa. Cape Town: Oxford University Press.
- Nagarajan, G., P., Carneiro, A. Hur. 2018. “Impact evaluation of the early grade reading activity (EGRA) : final report.” Social Impact, Inc., USAID. Mission to Malawi - [https://pdf.usaid.gov/pdf\\_docs/PA00T3Q6.pdf](https://pdf.usaid.gov/pdf_docs/PA00T3Q6.pdf)
- Piper, B. and S. Zuilkowski. 2016. “The role of timing in assessing oral reading fluency and comprehension in Kenya.” *Language Testing* 33(1): 75–98.
- RTI International. 2016. “Early Grade Reading Assessment (EGRA) Toolkit, Second Edition.” Washington, DC: United States Agency for International Development.
- RTI International. 2017. “ACR-Asia EGRA Benchmarks and Standards Research Report.” Washington, DC: United States Agency for International Development.
- Stern, J., M. Dubeck and A. Dick. 2018. “Using Early Grade Reading Assessment (EGRA) data for targeted instructional support: Learning profiles and instructional needs in Indonesia.” *International Journal of Educational Development* 61 64–71.
- Torppa, M., GK. Georgiou, M-K. Lerkkanen, P. Niemi, A-M. Poikkeus, and JE. Nurmi. 2016. “Examining the simple view of reading in a transparent orthography: A longitudinal study from kindergarten to grade 3.” *Merrill Palmer Quarterly*, 62, 179–206. <http://digitalcommons.wayne.edu/mpq/vol62/iss2/4/>
- USAID. 2017. “Literacy, Language and Learning Initiative (L3). National Fluency and Mathematics Assessment of Rwandan Schools: Endline Report.” Education Development Center. USAID
- Wang, Z., J. Sabatini, T. O’Reilly, and J. Weeks. 2019. “Decoding and Reading Comprehension: A Test of the Decoding Threshold Hypothesis.” *Journal of Educational Psychology*, 111(3), 387-401.
- Zuilkowski S., B. Piper, D. Kwayumba, and M. Dubeck. 2019. “Examining options for reading comprehension assessment in international contexts.” *Journal of Research in Reading*, Volume 42, Issue 3-4, pp 583–599



