



Developing Cross-Language Metrics for Reading Fluency Measurement: Some Issues and Options

Helen Abadzi¹

Global Partnership for Education

GPE Working Paper Series on Learning No. 6

July 10, 2012

¹ Helen Abadzi is a Greek psychologist, who has worked since 1987 as an education specialist and senior evaluation officer in the World Bank and the Global Partnership for Education. She explores cognitive neuroscience applications that may improve the education of the poor. Her publications helped raise early-grade reading fluency to a high-level international priority. Email: habadzi@worldbank.org, habadzi@gmail.com

Executive Summary

Since 2005, over 70 oral reading fluency tests have been given in many languages and scripts, either as part of the Early Grade Reading Assessment (EGRA) or as individual one-minute tests. Particularly in multilingual countries, reading speed and comprehension measures have been taken in multiple languages (e.g., Kenya) and also in multiple scripts (e.g., Ethiopia, India). As a result, various reading rate and comprehension measures appear side by side in various documents. But how comparable are the numbers? And why should they be compared?

Comparing words per minute across languages seems to be a somewhat controversial topic. Some educational researchers are of the belief that languages differ fundamentally, so it is impossible to compare the reading rates of students in Khmer vs. English (e.g. RTI 2010). Some others argue against using reading rates for assessment because of possible pressures and falsification (e.g. Wagner 2011).

This philosophical distrust of reading rates has turned attention to less precise measures that use time only implicitly, such as Uwezo in East Africa. It has also limited comparisons to non-readers, i.e. those who read zero words correct per minute. This measure has had much impact, but has also raised questions on whether a student with 1 correct word per minute is a reader. (To define a minimum lower estimate, the standard error of the estimate could instead be used.)

But reading fluency continues to have monitoring potential. At the very least, multilingual societies must plan how to improve literacy rates of all students. To design interventions that reach all students, technical specialists need relatively simple explanations of what these fluency and comprehension scores mean. They also need to get a sense of magnitude and an insight of how well students may understand text at various stages of fluency.

Research provides considerable support for comparing languages and even scripts at various points of proficiency. The support comes from the universal characteristics of human cognition. The development of language has a significant genetic component, which tends to create common grammatical structures. Then languages must conform to information processing limitations, notably to working memory capacity. On the basis of such features, it may be possible to develop common standards for performance improvement compare findings cross-linguistically. Languages are most comparable when large chunks are used rather than single words. And comparability would typically be rough; and it would not be wise, for example to rank countries along reading speed in the way that they are ranked in TIMSS or PISA. But it is possible to approximate and develop similar comparable interventions.

Research pertaining to working memory functions involving early-grade type materials has suggested that students would understand text well if they read 45-60 words correct per minute. Several EGRA studies by the Research Triangle Institute (e.g. RTI 2011) show that roughly 50-60 words per minute are needed to answer 4 out of 5 questions about a simple text. One-minute tests are simple and quick to administer, so this 50-60 wpm figure could be used as a rough benchmark of minimum acceptable fluency. However, units called words are written in different configurations in different languages. Multiple words may be clumped together, and the challenge arises how to count them in comparable ways.

To arrive at some comparisons, several methods may be tried. These include: (a) Counting actual words in connected texts or in lists, using some conventions if needed; (b) using computational solutions to arrive at coefficients of certain languages vis a vis others, such as 1 Swahili word being equivalent roughly to 1.3 English words; (c) using in multiple languages lists of words of a defined length, e.g. 4 letters; (d) measuring phonemes or syllables per minute, possibly dividing by average word length; (e) rapid serial visual presentation, potentially also measuring perception at the letter feature level. Simulation studies would be needed to compare these various methods for various languages and scripts and to provide guidance on which one(s) to use.

The above difficulties have pointed to comprehension as an attractive benchmark. However this concept presents different challenges. Many educators use the term to mean not literal comprehension that the mind does automatically, but prediction and inference, which are higher order functions and require some instruction. Test items may be written for literal or for advanced comprehension, and these are not comparable. Also working memory has serial position effects, so the first and last parts of a text tend to be remembered better than others. Thus a very brief text may have better comprehension scores than a slightly longer text. An alternative to consider is retelling the text, if a reliable scoring system can be established.

Overall, reading rate as words per minute seems to be a valid and reliable indicator of achievement, with 45-60 words being a range that is usable as a benchmark. In the low ranges, e.g. 100 wpm or less, fluency predicts comprehension reliably if the vocabulary is known. The challenge arriving at reasonably comparable rates suggesting that students in various languages process roughly the same amount of text per unit of time.

Due to many educators' uncertainty regarding memory functions, stakeholders are often asked to determine how many words per minute they would like their students to read. However, this is tantamount to negotiating with the limits of working memory. If text is read too slowly, for example 25 words per minute, little of it may be understood. It is more viable to ask stakeholders to plan for the percentage of students who ought to attain 45-60 wpm over a period of years.

Memory, Measurement, and Cross-Cultural-Comparability

Understanding fluency and developing benchmarks requires some knowledge of memory research and its implications. This domain of knowledge is rarely taught in faculties of education or economics, so this section explains the essential concepts and issues in simple language. It covers the following topics:

- How memory encodes and retains a message
- How fluency develops and the importance of practice
- Why comprehension results from fluency
- Knowledge acquisition from reading (reading to read)
- Comparability of memory and linguistic reading mechanisms across languages
- Common oral reading fluency tests
- Psychometric and sampling issues of oral reading fluency tests
- Norms and benchmarks pertinent to oral reading fluency tests in various countries

Why is reading fluency a prerequisite for comprehension?

The faster one reads a message, the more time one has to put it into the working memory, connect it with prior knowledge, and thus understand it. People who read beyond a minimum speed can understand the text because the mind is able to hold a message long enough in working memory, contemplate it and bring out of long-term memory relevant knowledge to interpret it. (See more on working memory below.)

Many studies have demonstrated a close relationship between speed and comprehension (review in Fuchs et al. 2001).² The relationship is moderated by accuracy; working memory gets cluttered with the failed attempts, so a 5 percent error rate is associated with comprehension test scores of only 75 percent (Barr et al. 2002: 253).

Thus students learn from books only when they can read them fluently. As grades advance, texts become more complex and if students keep reading, speed should rise in tandem. By the time they reach grade 6 students ought to be reading about 150 words per minute on average.³ If they only read 90, they cannot read volumes of material, as current modern jobs demand. For example, they may be unable to read computer screens fast enough or make sense of the material. Thus, early-grade reading strongly affects the efficiency of an education system. Much repetition and dropout can be avoided if students read at the required speed.

² Many studies have found that speed was a prerequisite for comprehension (Laberge and Samuels 1974, Daneman and Carpenter 1980, Breznitz 1997a and 1997b, Snow et al. 1998; Pikulski and Chard 2005). Struggling English readers who practiced recognition to the point of automaticity answered more comprehension questions than students who merely were instructed on word meanings (Tan and Nicholson 1997). Oral reading fluency measures correlate not only with simple recall questions but also with the comprehension of more complex texts (correlation 0.91; Fuchs, Fuchs and Maxwell 1988). Similarly the Equipo de Orientación de Marbella (Andalusia, Spain) found a 0.45 correlation between grades and reading fluency in language, 0.56 with overall grades. Furthermore, the correlation between fluency and the comprehension measures of a Chilean standardized examination was 0.4 up to grade 4, and 0.6 starting grade 5. (Torres 2007).

³ Norms are from the US, Hasbrouck and Tindal 2006. In grades 1-2 of the US, speed increases about a word every week (Dowrick et al. 2006).

The speed-comprehension relationship is particularly strong in the earlier primary grades and is most obvious in languages with consistent spelling rules. In such languages, students generate pronunciations of novel orthographic strings independently once they learn the rules. Thus, they “self-teach” basic comprehension (Share 1999, 2010). By contrast, in English (the subject of the vast majority of studies) students may see words that they cannot pronounce, so comprehension may require considerable vocabulary development.

The relationship between speed and comprehension weakens in more advanced grades and higher proficiency levels, because vocabulary and background knowledge become more important than speed. This can be the case for students in the US who read more than about 85 words per minute.⁴ This reading rate is achieved on average in the US by the end of grade 2 (Hasbruck and Tindal 2006). In lower-income countries, however, this rate may not be achieved even in upper-primary.

Worldwide, fluency is expected even when this goal is not explicitly stated. Curricula prescribe that students must learn to read in grades 1-2 and perform various tasks related to reading content.⁵ Research in European languages and middle-income environments suggests that by the end of grade 1, students should be able to crack the code and read, albeit haltingly (Seymour et al. 2003). Curricula typically expect students to be reading significant amounts of material by the end of grade 2, so by that stage at the latest, students should be reading fluently.

Fluency is necessary for comprehension but it is not sufficient (Chen and Vellutino 1997). Students must know the language and the subject matter. So comprehension depends on vocabulary depth (Ouellette 2006), morphological and syntactical awareness (Karatzas 2005), and also on prior knowledge about the situations described in the text. Working memory capacity is also an important variable; this may be more limited in lower-income children due to the impact of health problems (Swanson and Alloway 2010; Simeon and Grantham-McGregor 1989, Beaulieu et al. 2005). Inordinately limited working memory capacity may account for the low scores occasionally seen among readers with satisfactory fluency rates (Georgiou et al. 2008, 2009).⁶

⁴ For students reading more than 85 words per minute, comprehension is best predicted not from a single reading rate measure but from a combination of three-minute tests, reading rate, error rate, and prosody ratings (Valencia et al. 2010). Since low-income students do not attain this reading rate for years, a single one-minute test may provide sufficient information quickly and cheaply.

⁵ For example, the Kenyan syllabus states that: By Standard 2 children are expected to “read simple sentences/passages related to greetings and polite language” (Objective 1.2.d) as well as a about colors (2.2.f), numbers (4.2.e), time (5.2.e), position and direction (6.2.e), home and home activities (7.2.e), shopping (8.2.c), their bodies (9.2.e), health and hygiene (10.2.c), travel (11.2.f), clothes (12.2.c), food (13.2.d), wild animals (14.2.c), weather (15.2.c), the farm (16.2.c.) and home equipment (17.2.c. and d.). In many of these cases, the child is also expected to “answer comprehension questions.” (RTI 2007, in Malindi). Thus fluency goals can clearly be inferred.

⁶ Measures of memory form a good index for differentiating good readers from poor readers (Abu-Rabia & Siegel, 2002; Geva & Siegel, 2000; Just & Carpenter, 1992). Perhaps differences in working memory capacity are caused by the ability to hold and manipulate phonemes in memory (Crain & Shankweiler, 1988).

Learning to read vs. “reading to Learn”. Beginning readers may not acquire much information from a text because they struggle to retain the content in working memory. In order to interchange information between the long-term and working memory effortlessly, readers must decode rapidly and automatically. The percentage of children in various countries able to make this “reading to learn” transition is important (Willms 2008). This transition is a particular concern in English, due to spelling complexity, and it must be achieved by the end of grade 3.⁷ If children cannot read English with ease and understand what they are reading by the time they enter fourth grade, they are less able to take advantage of the learning opportunities that lie ahead. For languages with consistent spelling systems, students may acquire information from print within a few months. But since lower-income students worldwide tend to learn reading later, the English-based guidelines may be useful for them.⁸

Why Does Speed Result in Comprehension? The Role of Working Memory

To comprehend a message, people must be able to hold it in their mind long enough to make sense out of it. They must connect its contents to previous knowledge, and store the outcome in their memory for long-term use. Human memory has many complexities, but it can be visualized as a very fat bottle with a very narrow neck. The body of the bottle – long-term memory – is essentially bottomless. It contains an infinite number of memories, intricately connected into cognitive networks. The neck of the bottle, short-term memory or working memory, has very limited capacity. Very roughly, it lasts only about 12 seconds and may hold about 7 verbal items (more likely 4 - the capacity varies depending on whether they are words, letters, short phrases; see more details below). Within that timeframe, text must be read and knowledge must be brought out of long-term memory to create comprehension.

Educated people in high-income countries usually process text fast and within the capacity of their working memory, so its limitations are not obvious. But in low-income countries students may recognize text slowly, if at all, and they may have little knowledge to retrieve quickly. Thus, working memory capacity becomes an important limitation. If we read too slowly, by the end of a sentence we may forget the beginning.

Working memory capacity implies that a minimum reading speed is needed in order to make sense of text. For simple text and limited prior knowledge, we ought to read a simple sentence of seven known words within 12 seconds in order to make sense of it. Dividing 7 words into 12

⁷ This competency is referred to in the 2010 “early warning” report by the Annie B. Casey Foundation, which highlights links between early literacy (proficiency by the end of grade 3), high school graduation rates, and future economic success (www.aecf.org). “Reading to learn” corresponds to Jeanne Chall’s stage 3, which normally happens in grade 4 in English (1982), but students of simply spelled scripts may achieve this competency earlier because they have item-based rather than stage-based development (Share 1994, 1999). “Reading to Learn to read” could be assessed through questions that require comprehension and extraction of conclusions rather than merely recall of information items read (as in “learning to read”).

⁸ The International Evaluation Association uses this “Read to Learn” competency as a specification for the PIRLS test, which is given in grade 4 (<http://timss.bc.edu/>, Campbell et al. 2001).

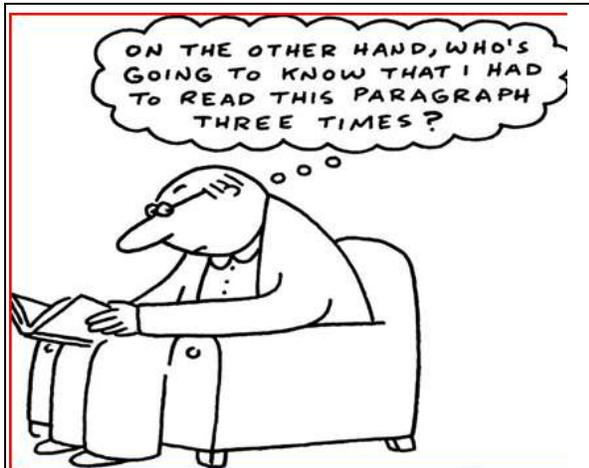
seconds and adding some milliseconds of processing time for messages to travel through the brain gives a rough estimate of a minimum benchmark: about one word 1.5 second or about 45 words per minute (Royer et al. 2004). More complex and longer text requires higher speeds for retention and understanding (Tindal et al. 2005); 60 words per minute is a more realistic minimum benchmark for simple connected text. As shown in subsequent sections, this theoretical relationship has been held up in measurements across various languages. (See for example, RTI 2010b).⁹

Initially students read letter by letter, using parts of the brain that specialize in this function. As they get more practice, their reaction time drops and they begin to recognize letters rapidly and in clusters. Then, a part in the brain becomes activated (the visual word form), that specializes in the identification of entire words (McCandliss et al. 2003, Shaywitz 2003, Dehaene and Cohen 2010, Dehaene et al. 2010). Activation of the visual word form is a prerequisite for fast and effortless reading. Performance before and after activation sounds different; the latter is more like natural speech. Therefore, common people can hear students read and reasonably judge if they are fluent or read letter by letter. In middle-class students the transition to fluency may take place within a few days in grade 1, but children receiving inadequate instruction and practice may be stuck in intermediate stages for years. To ensure that students progress to learning from texts (see subsequent section), their visual word form should be activated by the end of grade 2.

The fluency range associated with the activation of the visual form could be used as a minimal goal for all to attain. However, this can only be measured through sophisticated brain imaging equipment, which is rarely used outside high-income countries.¹⁰ If about seven “average” words can be recognized in about 12 seconds, then letters have already become clustered into words. Thus, the memory and neuropsychological estimates may roughly seem to coincide.

⁹ Words per minute in this and in other educationally oriented measures typically refers to meaningful text or sometimes lists of words. The reader’s eyes move across the text, and longer words are averaged. (Word lists are read more slowly). Cognitive psychologists measuring speed in the laboratory use *rapid visual serial presentation* (RSVP) which keeps text stationary on the screen and controls for word length. Under such circumstances, it is possible to read three times faster than real-world circumstances (e.g. 300 wpm; e.g. Pelli et al. 2006).

¹⁰ Brain imaging methods, such as functional magnetic resonance imaging, are expensive, but event-related potentials could be easily used in low-income countries. The perceptual aspects of reading are also assessed by studying eye movements through computerized displays (e.g. Sereno and Rayner 2003).



If you read a text too slowly for its level of complexity, by the end of a unit you may forget the beginning. (By Charles Barsotti, The New Yorker 7/4/1988)

Relevant studies are few, but they seem to concur about the reading rate. For example, Peruvian third graders who read 45 words per minute or more score understand text better than those who read less fluently (Kudo and Bazan 2008; Figures 1 and 2). Among the low-income Spanish-speakers in the U.S., 30–60 words per minute in Spanish in Grades 1 and 2 signal disadvantage (de la Colina 2001). Also U.S. students reading below 40 words per minute at the end of grade 1 are considered at risk of failure (Davidson and Towner 2005; Riedel 2007). Adults reading at lower speeds have showed limited comprehension (Royer et al. 2004).

The working memory model taught for decades in psychology courses has become more nuanced over the years. The 7 items in working memory

were based on the recall of English-language digits (Miller 1956), but biophysical research suggests that a recall of 7 items has a basis in neural code used for the recollection of persons, objects or places (Migliore et al. 2008). The 12 seconds resulted from reading of trigrams (Peterson and Peterson 1959), and newer studies suggest that information load influences time.

Newer models of working memory emphasize the dynamic exchange between cognitive networks and the central executive function. Under such circumstances, experiments controlling for context cues, attention, and interference show a smaller buffer of about 4 digits and a variable timespan of 10-20 seconds (e.g. Cowan 2001, 2005; Alloway et al. 2006). Depending on prior knowledge, large amounts of information can be retrieved in milliseconds, obscuring the size of the verbal store. Thus working memory studies have produced variable results, depending to some extent on the paradigms chosen. And because the amounts of prior knowledge are variable, newer research has not focused much on the timeframe that information is available in working memory.

However, the older studies and models still seem appropriate. They used protocols similar to fluency tests of beginning students, who also have little prior academic knowledge to bring into working memory (e.g. Daneman and Carpenter 1980). Such students may hardly know the languages they use, thus the old trigram paradigms have some ecological validity.

Furthermore, reading a text serially word for word resonates with the older “short term” memory models. These may be the reasons why reading simple text at 50-60 words per minute helps answer correctly 4 out of 5 literal comprehension questions on average. Exceptions may be tonal languages that convey two bits of information in one syllable, and may thus more “efficient”. They often have shorter words and may be and understood faster by beginners (e.

Lao). On the opposite end, reading various languages in the unvoiced Arabic script may require higher speed in order to maintain sufficient material into working memory and decide on likely meanings.

The functions described above best fit the lower reading rates and levels of education, which are most pertinent for low-income populations. Thus, the number of words that can theoretically fit into working memory can be used to set minimal benchmarks at no extra cost. Also time, research has shown the increased commonalities among languages due to the underlying neurological mechanisms (Bolger et al. 2005, see below).

To determine parameters precisely, studies are needed that would encompass multiple languages and writing systems. However, studies require years and considerable amounts, that donors are hesitant to spend. Policy and implementation decisions must be made quickly, so reasonably developed hypotheses from current research become crucial. Therefore all available studies must be used to build a causal chain.

To what extent can reading speed and comprehension be compared across languages?

When proficient readers read, cognitive networks are activated in milliseconds, as if the readers had heard speech. Background knowledge and context rush into working memory (Devlin 2010). Languages and scripts conform to the cognitive resources available to humans, such as working memory capacity and the transmission speed of the nervous system. Linguistic variables vary, but do so within relatively narrow limits and may involve redundancies and compensation (Miestamo et al. 2008).¹¹ In all languages and scripts, an identical site (the visual word form) gets recruited for reading (Bolger et al. 2005). One reason is that that human beings are born with knowledge of certain syntactical rules and intuitively identify word order and which parts of speech constitute “words” (e.g. Culbertson and Legendre 2010). Linguistic and memory similarities across cultures may account for the similarities in curricular prescriptions worldwide for reading fluency in the first one or two grades (Bonnet et al., 2001).

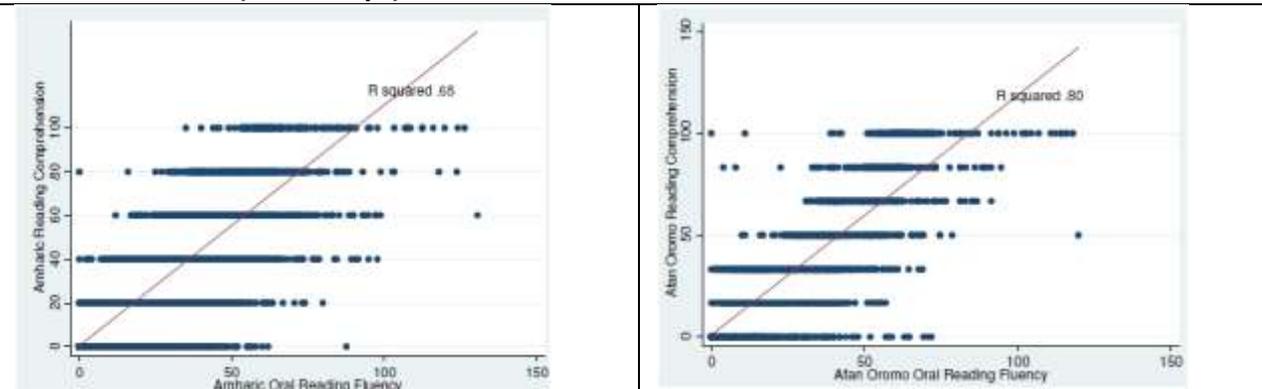
Several cross-linguistic comparison studies have conducted. These show that comparisons are possible within specific parameters. One important parameter is spelling complexity. For example, a 16-country comparative study of mainly European languages showed that grade 1 reading rates depend on orthographic transparency (Seymour et al. 2003). The more complex scripts or spelling systems take longer to master, and students keep learning symbols for years. For example students learning fundamental English reading require about 2.5 times more years of literacy than German (Seymour et al. 2003). Accuracy among 7-year old German children in grade 1 was 92 percent compared to 69 percent for English-speaking children (Wimmer et al. 1999).

¹¹ A study of word lists showed that the Chinese who have short words could remember about nine and the Welsh who have long words could remember about five (Ellis 1980). However, sentences even out the variability because some expressions are longer than others.

International tests have shown similarly close correlations between reading fluency rates across languages and comprehension. The faster the children read across each Ethiopian language, the higher were their comprehension scores. Moreover, in order to reach 80% to 100% comprehension scores, oral reading fluency levels need to be somewhere between 50 and 80 wpm (RTI 2010). Regressions showed that the range of oral reading fluency needed for 80% to 100% comprehension was between 53 wpm to 59 wpm for 80% comprehension of “shallow” items. For 100% comprehension, oral reading fluency scores were 60.0- 70.4 wpm. These studies as well as learning and perceptual research give some guidance regarding the reading speeds that would result in satisfactory comprehension in various languages and scripts.

A higher speed may be necessary for orthographies that omit vowels and force readers to keep alternative pronunciations of words in working memory in order to make sense of a sentence (e.g. unvoveled Arabic, Hebrew, Farsi, Pashto, Urdu). On the other hand, it may be possible to read more slowly and comprehend tonal languages, where a morpheme carries two bits of information (e. g. marked tones and short words in Lao, Vietnamese, Thai). Also, students with more limited working memory capacity perhaps ought to read faster than others in order to make sense of the messages.

Figure 1 and 2: Oral Reading Fluency and Comprehension Rates for Amharic (Sabean script) and Afan Oromo (Latin script)

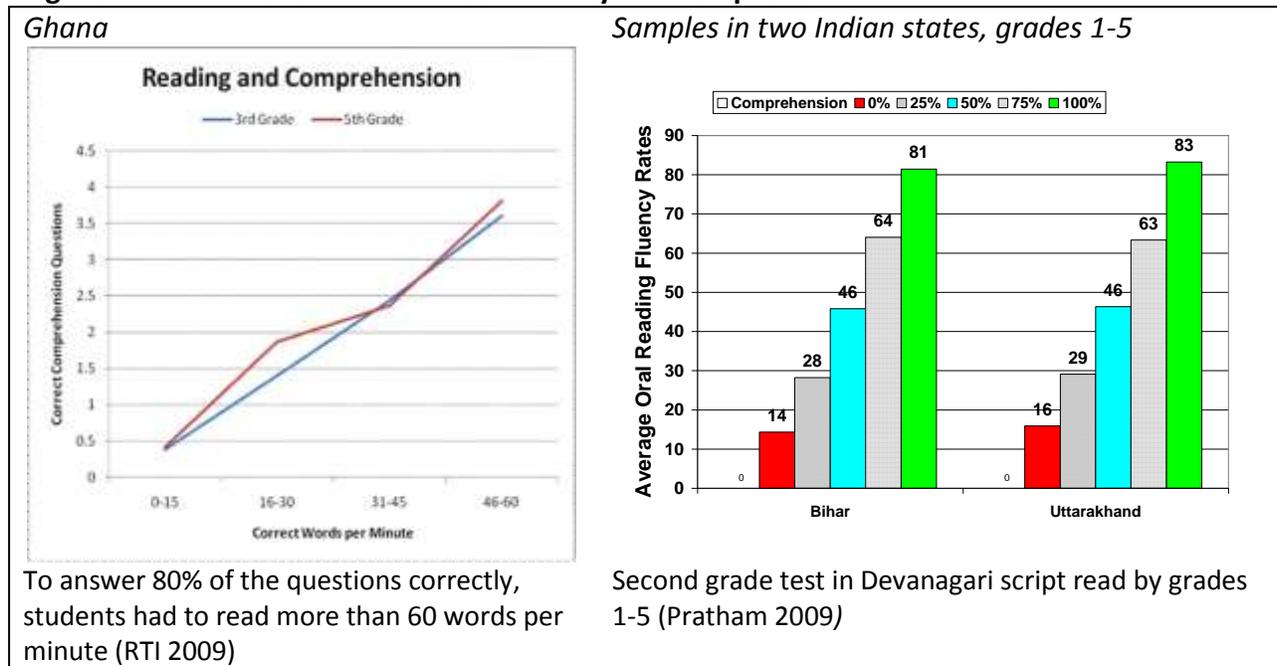


RTI 2010b, p. 41, Figure 22

One determinant of speed acquisition is visual complexity (e.g. Changizi and Shimojo 2005). The more “ink” needed to write letters and combinations the longer it takes to automatize them (Pelli et al. 2006.) Thus, Instruction in some scripts and orthographies takes much longer than others; for example, three years of instruction are required in English compared to more transparently spelled languages that may require only a few months (Share 2008). Some syllabic scripts like Kannada are nominally transparent, but their complex letter combinations also require 3-4 years to learn. The effort to read these visually complex syllabic scripts may consume working memory resources. For example, Indian students had to read about 80 words per minute in their own language (Devanagari script) to answer 80 percent of the comprehension questions (Pratham 2009; Figure 4). However, some of the effect may be due to lack of clarity in specifying comprehension items.

In some countries, average 80% comprehension was attained at higher speeds, perhaps due to vocabulary deficits or the widespread health problems. For example in Ghana, students reading in English (not their mother tongue), only answered 3 out of 5 questions correctly at 45 words per minute and needed more than 60 to answer 80 percent correctly. Thus, many children who attain the minimum reading rates may start ‘reading to learn’ only in advanced primary grades. The reasons and progression rates require more research, particularly since comprehension questions are of variable quality.

Figure 1 and 2: The relation between fluency and comprehension



The “word length effect” is another relevant phenomenon (De Luca et. al 2008). Beginning readers, take longer to read longer words. The “word length effect” gradually diminishes in silent reading after fluency and is reduced in oral reading as speed increases. For example in Italy it disappears in grade 3 (Hulme et al. 1995; Zoccolotti et al. 2005; Martelli et al. 2009).¹²

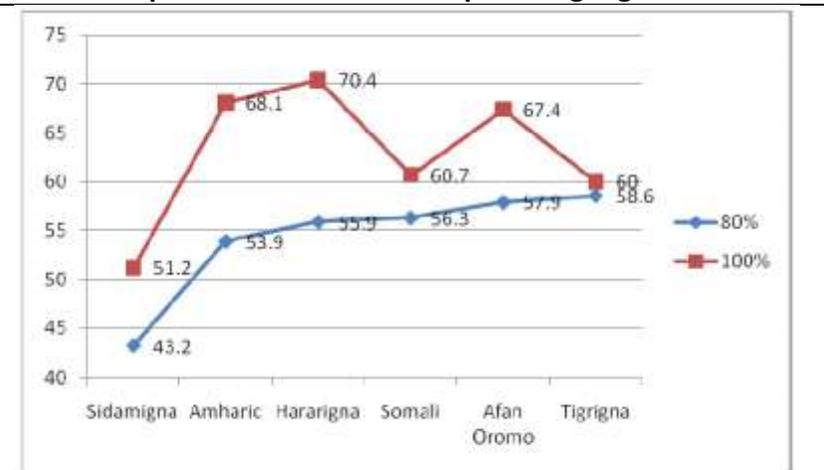
Experienced readers take about the same time to recognize longer and shorter words. Across languages, expert readers on average with about the same speed, with some exceptions. Some languages have longer words overall than others, so reading rates in the order of a few seconds and words per minute may result. For example, due to the visual complexity and cognitive demands, adults may be reading unvoweled Arabic a bit more slowly (Eviatar, Ibrahim, and

¹² Among Italian children, mean reaction times per letter were quickly reduced (they were 173 msec, 60 msec, and 36 msec per letter in grades 1-3 respectively and were further reduced in grades 3-5 Zoccolotti et al., 2005). Among adult readers, the word length effect was smaller, although detectable when using a large set of stimuli (Barca, Burani, & Arduino, 2002). For example, Bates, Burani, D’Amico, and Barca (2001) found an effect of about 9 msec per letter in naming four- to nine-letter words (annex C of Bates et al., 2001; references of these works are in Zoccolotti et al. 2005 and Spinelli 2005.)

Ganayim 2004). In laboratory conditions of rapid visual serial presentation (where eye movement is minimized) and word length is controlled, the readers of languages with longer words get lower reading speeds than the readers of generally shorter words. For example, English tends to have shorter words than Spanish and Italian, and a reader may possibly attain 500 words per minute in English but only 400 or so in Italian. The reason seems to again working memory capacity. However, these rate differences do not affect comprehension, and implications for low-level reading rates are unclear (Marialuisa Martelli, personal communication, May 17, 2010.) But the significance of 2-3 words per minute is unclear, particularly in the higher reading levels.

For research purposes, simulations or computerized methods like *rapid visual serial presentation* (e.g. Chun and Potter 1995) can disentangle the word length effect by presenting one word at a time and discounting the time that a reader takes to pronounce it. Other aspects of proficient reading could also be measured that circumvent language, such as prosody or latencies between words. However, these measurements require complex computerized procedures, and benefits are unclear for real-world reading.¹³

Figure 5. Oral Reading Fluency Scores Necessary for 80% and 100% Comprehension across Ethiopian languages



RTI 2010b, p. 43, figure 23

Overall, it seems that cross-linguistic differences are more pronounced in the early reading stages. **Cross-linguistic comparability becomes easiest at the emergence of fluency.** Regardless of the effort and time needed to learn the rules of a script, students may hold roughly comparable amounts of text in working memory after automaticity is achieved. Comparability can be enhanced by comparing curricular goals of various countries and using readability formulas (as is done in international comparison tests).

¹³ This document uses the words fluency and rate alternatively. Strictly speaking, however, fluency requires prosody. To assess prosody, examiners would describe the pitch, stress, and duration with which children express text. One study showed limited effects before grade 4 in English (Fuchs et al. 2001, Marston and Tindal 1996).

Computational and other studies show that language equivalence is better at larger blocks of text than 60-word passages. People seem to have an inherent sense of "word", and languages of longer words have shorter sentences. Furthermore, simple grade 1-2 level texts may be easier to equate. At best the word equating per minute would be rough. i.e. there may be easily a 10-15% deviation among languages, but it would be useful.

Why count words and not something else? In principle one could measure phonemes per minute, but research shows that our common unit of speech is a word. It makes sense to people. And because of working memory capacity, the 45-60 wpm help predict how well children are reading texts. The kinds of texts most comparable are the more elementary ones, where the needed background knowledge is minimal. (If educated people read complex text that requires prior knowledge, then knowledge variables get in the way.)

Nevertheless, comparing words per minute across languages seems to be a somewhat controversial topic. Some educational researchers are of the belief that languages differ fundamentally, so it is impossible to compare the reading rates of students in Khmer vs. English (e.g. RTI 2010). Some others argue against using reading rates for assessment because of possible pressures and falsification (e.g. Wagner 2011). This philosophical distrust of reading has limited comparisons to non-readers, i.e. those who read zero words correct per minute. This measure has had much impact, but has also raised questions on whether a student with 1 correct word per minute is a reader. To define a minimum lower estimate, the standard error of the estimate could instead be used.

Because many education specialists are uncertain regarding the role of working memory and cross-language comparability, there has been a tendency to ask stakeholders to formulate their own acceptable words per minute standards. Examples are provided in a Kenyan study (RTI 2010a p. 33, and 2010b, p. 45.) In the former, stakeholders defined 30-70 words per minute in Bantu languages as a "moderate" benchmark. In the latter, different groups were asked to independently determine a benchmark for the Amharic language, and then come together to agree on one figure. The groups had fundamentally different perspectives on the issue, with the first group (focused on Amhara region) preferring to set a modest fluency benchmark reachable by a large percentage of children, while the second group thought that it was more important for the benchmark to be set high. The results were a difference of 60-90 words per minute within the same language.

The experiences suggest that benchmark setting may be best viewed with respect to the percentage of students expected to attain benchmarks. Without training and understanding reading and memory functions, stakeholders may not be well positioned to state how many words per minute children should minimally read (Table 5). However, they may be able to project percentages of students who ought to attain 45-60 wpm in a period of years.

Potential Methods for Counting Words per Minute or Roughly Equating Languages

To measure reading fluency, one-minute tests inexpensive assess fluency for the early grades. Across several languages and scripts 45-60 words per minute amount to 80% comprehension when vocabulary is known and point to automaticity. Currently the consultants who count words just count the items that appear to be words. But in the multilingual societies, closer comparisons would help understand implications.

Unless sophisticated instruments are used, cross-country comparison of reading rates is possible but rather rough and approximate. It is not possible to say that language A (in which all children are reading with automaticity by grade 2) is outperforming language B (where children reach this level only by grade 3), if language A has a more transparent orthography than language B. So, it would be impossible to rank-order countries against reading fluency scores, as is done for PIRLS or TIMSS.

To count words per minute, an important issue is to determine what units constitute words in various languages. Different linguistic families or country traditions can have different writing conventions, and apparent words may really entire sentences. For example, Bantu and Turkish languages are agglutinative, while Lao and Khmer leave no spaces between words of small phrases. Arabic connects pronouns and articles, and some prepositions. Thus, similar numbers of symbols or syllables may be counted differently in different languages. For example, counting agglutinations at face value in Bantu languages may give a word count of 30 wpm, but separating words according to prominent European language conventions adds 30% more to arrive to 45. To compare reading rates across linguistic families, an agreement is needed on what particles should be counted as words. (See below). Alternatively, syllables could be counted rather than words for international comparisons, and results could be presented in different metrics nationally and internationally.

The research points to potential methods for comparing units read by students across time. Some options are below.

- **Counting actual words in connected texts or in lists.** Conventions could be developed on how words must be detected and counted. Given the similarity of grammatical parts and people's intuitive sense of what parts constitute words, words could be counted against a "reference" grammar (e.g. English that is frequent). This could give a reasonable count for many languages. Local languages written by Anglophones or Francophones probably separate words at breakpoints familiar to the researchers who wrote the languages first. But there are many situations that require different treatment. These include Bantu and Nguni languages in South Africa, also agglutinative languages in South America, Turkish etc. For example, in various languages one could count the self-standing words, but not compounds. e.g. Swahili ni-li-m-wona. (I saw him). Arabic articles and propositions could be separated, but possessive pronouns are not self standing. This method might be easiest to use within countries that have languages written down in the same system.

- **Using computational solutions to arrive at coefficients.** The equivalence across languages probably happens with texts of some length. Machine translations help establish coefficients,

and a Microsoft engineer proposed to compare counts in 35 languages. So without counting words exactly, a ratio could be used. e.g. 60 "word" text in Swahili would be equivalent to a 90 word text in English.

- **Controlling words of similar length across languages.** South Africa teaches in 11 languages. For adult literacy programs they decided to **control word length across languages**, so the Ministry of Education made up lists of 4-letter words in various languages for adult neoliterates. Then the amount of text to be processed becomes the same. For low-levels of reading there is a 'word length effect', so method keeps this variable constant. The lists would be read a bit more slowly than connected text, and comprehension is not measured. But there is much research showing a high correlation and a linear relationship at the lower levels, that it may not be necessary. The 4- or 5-letter words could be used also as text equating means, where the speed in those would be correlated to the speed of reading continuous text in various languages.

- **Measuring phonemes per minute.** In principle one could measure phonemes per minute. They make little sense to people, but they could be divided by the average letter length of words in a language. (Again, agreements are needed on what constitutes words.)

- **Rapid serial visual presentation.** There are also psychophysics solutions, such as measuring the speed whereby series of letters appear on the center of the eye (by a computer). This method bypasses language and script issues and tests the automaticity of whatever the reader sees. To deal with the visual complexity of some scripts, perhaps perimetric complexity could be taken into account. But then much programming and a computer is needed to assess students.

- There may be a need to study the **visual feature level**, given the visual-perception bottleneck. That is, how many letter features do people across scripts process per unit of time? How many are needed to make sense of text? Thus far crowding has been studied with side flankers. How about above and below?

Orthographic depth might also play a role in computations (Yap et al. 2010). Depth could be used computationally to compare languages and scripts when students are in the process of learning them. Letter to- phoneme number ratios provide a potential heuristic for scaling orthographic depth, clustering languages, and making predictions about processing demands.

Table 8 shows that word length—a marker of serial sublexical processing—predicted both lexical decision and speeded pronunciation performance far better than did word frequency in Malay. The reverse pattern was seen in English, a deep orthography, where frequency predicted recognition times better than did any other lexical variable. Length effects were also larger in speeded pronunciation, as compared with lexical decision, consistent with the pronunciation task's reliance (Yap et al. 2010). The findings suggest that reading might require more lexical support in English, German, and Dutch than in French, Malay, or Spanish, and that Finnish and Serbo-Croatian can be read aloud using only nonlexical rules. Deeper

orthographies, such as English, might be less salient for readers of shallower orthographies such as that of Malay.

Table 1: Estimating Letter-Phoneme Ratios for Nine Alphabetic Orthographies as a function of Vowels, Consonants, and All Letters (Yap et al. 2010)

Language	Ratio of Vowels/ Phonemes		Ratio of Consonants/ Phonemes		Ratio of Letters/ Phonemes	
Serbo-Croatian	5/5	1.00	25/25	1.00	30/30	1.00
Finnish	8/8	1.00	18/20	0.90	26/28	0.93
Malay	6/7	0.86	19/27	0.70	25/34	0.74
Spanish	6/7	0.86	21/18	1.17	27/25	1.08
French	12/16	0.75	21/20	1.05	33/36	0.92
Italian	5/7	0.71	18/43	0.42	23/50	0.46
Dutch	10/19	0.53	20/22	0.91	30/41	0.73
German	9/19	0.47	20/24	0.83	29/43	0.67
English	6/20	0.30	20/24	0.83	26/44	0.59
<i>M</i>	7.4/12.0	0.72	20.2/24.8	0.87	27.7/36.8	0.79
<i>SD</i>	2.5/6.3	0.24	2.1/7.4	0.22	3.1/8.3	0.20

Note—The ratio that is based on the number of vowel letters to the number of vowel phonemes is used to rank-order the languages, so that orthographically deeper languages have a lower rank.

The letter-to-phoneme ratio is at best a crude metric for quantifying orthographic depth, and more sophisticated and potentially superior approaches exist. For example, one could compute the ratio of number of graphemes to number of phonemes, or the consistency of grapheme-to-phoneme and phoneme-to-grapheme mappings. However it is unclear whether greater levels of complexity are worth the effort in reading fluency tests.

Defining and measuring comprehension – Some Issues

As mentioned above, the countries that teach reading in more than one language and more than one script (e.g. English and Swahili in Uganda, Spanish and Miskitu in Nicaragua, Hindi and English in India) require some rough comparability methods. Otherwise, students may be shown to read faster or more slowly. For example, tests in Swahili showed students reading fewer words than in English while understanding more (RTI 2007). And as shown above, there is are some complexities.

Some educational researchers have attempted to compare comprehension rather than fluency. However, comprehension presents a number of problems, which may be harder to overcome. Is it even worth measuring comprehension? The correlations between speed and comprehension are high in the lower ranges of reading. So strictly speaking, the measurement of comprehension is redundant in these ranges. This is particularly the case with native readers reading in consistently spelled languages. Even if readers are not native, a satisfactory speed implies that when they learn the language they will understand (e.g. Koranic Arabic and its use in writing local languages). However, comprehension has “face validity”. Governments must

have a sense regarding the extent to which students understand the language and can actually make sense of the text, so comprehension needs to be maintained. But it certainly needs to be clarified.

There are concerns that students may be getting significant reading percentages from background knowledge.¹⁴ One alternative to consider is retelling the text, if a reliable scoring means can be established.

How is comprehension defined?

Comprehension refers to someone's "ability to understand something, or someone's actual understanding of something."¹⁵ This means that some closely related items are linked in people's cognitive networks. This is literal comprehension: evidence that a child knows the meaning of words and how they are related (e.g. open the window vs. feed the cat). Students do not have to be taught how to comprehend literally; the mind searches for meaning within about 200 milliseconds after identifying a word (Devlin 2010).

This literal comprehension is most appropriate for assessment. The measure can inform about content that has gotten through working memory and is on the path to consolidation. Therefore to demonstrate comprehension, students should be expected to demonstrate that they understood the gist of a text. Most appropriate would be short and simple questions that sample the passage of factual information into long-term memory and require minimal prior knowledge (Georgiou et al. 2008).

However, often this is not what is used. Reading specialists of higher-income countries typically use the word "comprehension" for higher-order functions, such as predictions, inferences, summarizing etc. Comprehension questions may then ask for these more complex functions that require prior knowledge, transformations, need to examine more material into working memory. And, tests in various countries and within the same country have often used various types of questions. Some may be simple recall questions, while others (as in Mexico) may involve inference as well. EGRA has a mixture of 3 literal and 2 inferential questions. Also, test instructions have differed. In some tests, students were scored only on questions they could answer during the time they read, while in others students got extra time to read the entire text before answering, and in some they were even permitted to re-read a text while responding.

This varying definition of comprehension creates validity and reliability problems. Test questions for grade 2 could be relatively simple, but this variability clarifies that they are not comparable. To compare comprehension questions across tests such as various

¹⁴ About 45% scores can be predicted from background knowledge (Linda Farrell, Readsters, private communication, July 26, 2012)

¹⁵ www.yourdictionary.com

administrations of EGRA (even within the same language and country), scaling and equating procedures are needed. This would require pretesting and introduce complexity to tests meant to be simple.¹⁶ It is not being done, and studies in the international early-grade community probably do not exist. So the magnitude of variation from the usual hypotheses is unknown. Efforts have been made to set a common metric of comprehension. Out of 5 questions, answering 4 could show fairly high comprehension. If the questions were simple and shallow, perhaps the item difficulties would not be very different for students of the same ability. Indeed informal reading inventories tend to use 80% or 4 out of 5 correct answers to comprehension questions without item analyses (Marcia Davidson, personal communication 4-29-2010).

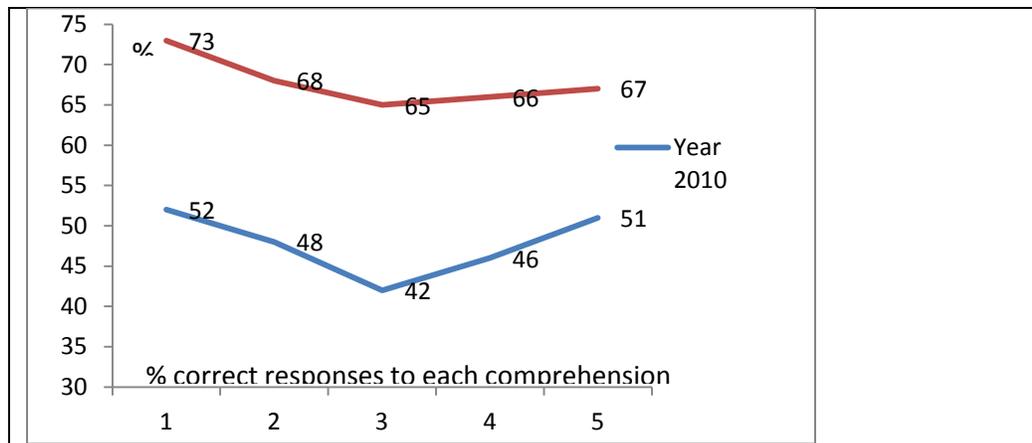
Questions are used because they are easy to score, but a retelling of the gist is also used. DIBELS uses a "retell fluency" test in lieu of comprehension questions. This is administered only to students who read more than 40 words per minute (www.dibels.uoregon.edu). In that subtest about 50% of the number of relevant words used in retelling in one minute is considered satisfactory, but the rationale for this number is not clear. Unfortunately there is research indicating poor interrater reliability in scoring the retelling.

Although the 80% standard has been used, it is really a convenience number of getting 4 questions out of 5. The international community has given little thinking on what constitutes sufficient comprehension to understand a text relatively comfortably. Missing information strains working memory, as students must bring in prior knowledge while keeping material in it at the same time. There is a great deal of research on comprehension in English, but it is beyond the purview of this review. (Clearly a review is needed with an eye on inferences for one-minute tests.)

Another issue that complicates comprehension rate as a criterion is the serial effects of working memory (Gupta et al. 2005). Questions about items that are at the beginning and end of the text may achieve higher responses (see figure).

Figure 6: Cambodian EGRA, comprehension of connected text, grades 1-6

¹⁶ Some academics have suggested using a common story in multiple languages and scoring comprehension on it, but it is unclear how that metric would be set. Furthermore, testing security issues would arise about that story.



However, the finding from RTI datasets that 53-60 wpm are read on average at 80% comprehension opens the way to some assumptions. It implies that at that speed average students are reading fast enough to retain a text in working memory long enough to make sense out of it (about 60 for grade 2 text, higher wpm for more complex text). This would be achieved only with the easiest items. Items demanding more complex thinking would result in a lower comprehension score or require faster reading. Thus 80% comprehension seems like an upper boundary, attained only with the easiest items.

Often one hears from education specialists who have been in the field that some students read fast but do not understand. (This argument is sometimes offered to suggest that working memory does not work as expected.) From one simple test it is impossible determine many issues. About 10% of the children may have an abnormally limited working memory and may not be able to hold the material long enough in their minds. But more likely culprits are limited vocabulary knowledge, complex comprehension standards, or other sources of measurement error.

Given the various doubts regarding the link between fluency and comprehension, as well as a reluctance to compare words per minute across languages has created some difficult situations for officials of some governments. They are often asked to set their own standards on how much comprehension they ought to expect from their students. But they are not presented the basics of working memory to understand the concept. Working memory is a relatively fixed window, and to make sense of the text. Students must throw material into that space as fast and as effortlessly as possible in order to have time to understand its greater meaning. Due to a limited understanding of memory functions by the educational community, the GPE benchmarks only say that students will “read with comprehension”, but this may prove misleading to officials. Comprehension is also a score of 20%. The serial position effects of working memory mean that students barely reading within the confines of this mechanism may remember only the beginning and the end of a story. Thus, the hopes of officials for high comprehension may amount in reality to 20-40%.

It makes more sense to explain to officials the need for high comprehension (at least 80%) and encourage them to set percentages of students that ought to attain those comprehension levels over time.

One example of the difficulties encountered in asking officials to set their own standards comes from a Kenyan study (RTI 2010a p. 33, and 2010b, p. 45.) In the former, stakeholders defined 30-70 words per minute in Bantu languages as a “moderate” benchmark. In the latter, different groups were asked to independently determine a benchmark for the Amharic language, and then come together to agree on one figure. The groups had fundamentally different perspectives on the issue, with the first group (focused on Amhara region) preferring to set a modest fluency benchmark reachable by a large percentage of children, while the second group thought that it was more important for the benchmark to be set high. The results were a difference of 60-90 words per minute within the same language. The experiences suggest that benchmark setting may be best viewed with respect to the percentage of students expected to attain benchmarks. Without training and understanding reading and memory functions, stakeholders may not be well positioned to state how many words per minute children should minimally read.

Table 2: Draft Oral Reading Fluency Benchmarks by Ethiopian Language

	Proposed Benchmark (words per minute)	Percentage of children at benchmark (2010)	Target percentage of children at benchmark by 2015
Amhara region	60 wpm	4.3%	80%
Amharic (Group 2)	90 wpm	1.0%	50%
Afan Oromo	70 wpm	4.2%	60%
Tigrigna	60 wpm	0.4%	45%
Sidaamu Afoo	75 wpm	0.1%	70%
Hararigna	60 wpm	0.6%	50%

Source: RTI 2010b, p. 45

International Comparability: How to Count Words for Reading Fluency Tests?

The fact remains that comparability is at the paragraph or larger volume level rather than at the word level, which seems most useful for instructional work.

The length of words as written on paper is to a considerable extent arbitrary. In many languages words spaced apart represent linguistic units, and in others they do not. Some scripts, like Lao and Thai, break up at the phrase rather than at the word level. On the other hand, Arabic includes gaps within words. Agglutinative languages like Turkish, Quechua, or Bantu languages may write words in long concatenations. In these, many words are connected that are counted separately in English or other European languages.

As mentioned elsewhere, humans are born with knowledge of certain syntactical rules and intuitively identify which parts of speech constitute “words” (e.g. Culbertson and Legendre 2010). To develop international comparability, words must be counted in consistent ways. This issue deserves more consideration, because words have been counted on the basis of common Indoeuropean grammatical structures (given studies in English, Dutch, other European languages). However, other language families may result in agglutinations of various sizes. For Example, Arabic attaches articles and pronouns to nouns or verbs. Some decisions must be made about which units to count as words. Perhaps syllables ought to be counted rather than words, but these are not as intuitive. Also, different conventions could be used for reporting in a specific country, and adjustments could be made on the basis of algorithms for international comparison purposes. (e.g. counting syllables and dividing by the average number of syllables in a word of a specific language.) Information is available in linguistic databases with respect to the statistical characteristics of a language, such as word frequency or the consistency of the relationships between orthography and phonology.¹⁷

Table 3: Swahili passage use in Malindi, Kenya

Jumamosi iliyopita Katana na dada zake, Kadzo na Fatuma, Walienda kuogelea baharíni. Kabla ya kuondoka walibeba mahamri, maembe, samaki na maji ya machungwa. Walibeba pia nguo zao za kuogelea. Wote waliingia kwenye matatu kuelekea huko. Walipofika baharini waliona watu wengi sana. Katana alikuwa na hamu sana ya kuogelea. Maskini Katana, aliingia baharíni bila kubadili nguo zake! Dada zake walimcheka sana.	Jumamosi i\li\yo\pita Katana na dada zake, Kadzo na Fatuma, Wa\ li\enda ku\ogelea baharí\ ni. Kabla ya ku\ondoka wa\ li\ beba mahamri, maembe, samaki na maji ya machungwa. Wa\ li\beba pia nguo zao za ku\ogelea. Wote wa\li\ingia kwenye matatu ku\elekea huko. Wa\li\po\fika bahari\ ni wa\li\ona watu wengi sana. Katana a\li\kuwa na hamu sana ya ku\ogelea. Maskini Katana, a\li\ingia baharí\ ni bila ku\badili nguo zake! Dada zake wa\li\m\cheka sana.
Word count: 60	Potential word count: 92

RTI: EQUIP2 data (RTI 2010a)

For example, some tests have shown students in Swahili fewer words than in English. In the 2007 EGRA administration in Malindi (RTI 2007), Kenyan students read on average 10.2 correct words per minute in Swahili and 11.4 words per minute in English. (Reading comprehension was the same.) However, when Swahili is broken down into words which exist independently in English, the number of words increases by about 30 percent, and the speed difference

¹⁷ Linguistic databases compiling quantitative and objective estimates about the principal variables that affect reading and writing acquisition. (Peerman, Lete, and Sprenger-Charolles 2007).

disappears. As Table 7 shows, Swahili passages may have 10-20 percent more syllables than English but still have 30 percent fewer countable words.

Yemen connected text and questions

الان سوف أسالك عدة اسئلة عن القطعة التي قرأتها، حاول الإجابة عن الأسئلة بأفضل ما تستطيع .	
أين ذهب الراعي مع أغنامه؟ "إلى المرعى"	7
لماذا ذهب الخروف الصغير إلى النهر؟ "لأنه شعر بالعطش"	19
بماذا إتهم الذئب الخروف؟ "بأنه أساء إليه قبل عام"	31
كم كان عمر الخروف؟ "ستة أشهر"	40
ماذا قرر الذئب أن يفعل بالخروف؟ "أن يأكله"	49
لماذا هرب الذئب؟ "لأنه خاف من قدوم الراعي والكلب"	58

The Yemen EGRA text has 9 cases of an attached article and 6 cases of a preposition that in other languages is often detached (in, then). Thus the text could be considered to have either 9 or 15 words more, i.e. 67 or 73 words total.

Egypt EGRA connected text

صَبَاحُ الْخَيْرِ. أَنَا إِسْمِي أَمِيرَةٌ ، وَعُمْرِي تِسْعُ سِنَوَاتٍ

وَأَنَا فِي الصَّفِّ الثَّالِثِ ، وَلِي أُخْتُ صَغِيرَةٌ اسْمُهَا مَرِيْمٌ وَعُمْرُهَا خَمْسُ سِنَوَاتٍ

عِنْدَمَا أَلْعَبُ مَعَ صَدِيقَاتِي تُحِبُّ مَرِيْمٌ أَنْ تَلْعَبَ مَعَنَا.

جَلَسْتُ مَعَ صَدِيقَاتِي لِأَصْنَعُ طَائِرَةً وَرَقِيَّةً

جَاءَتْ أُخْتُ الصُّغْرَى مَرِيْمٌ وَقَالَتْ لِي: أُرِيدُ أَنْ أَلْعَبَ مَعَكَ

قَلْتُ لَهَا : لَا أَنْتِ صَغِيرَةٌ، اجْلِسِي وَأَنْظُرِي.

The Egyptian EGRA has 4 cases of an article and 6 cases of attached prepositions (and, then)

To avoid complex solutions it is relatively easy to compare the number of phonemic units rather than words. Syllables are more amenable to comparisons. For example Italian, second graders read at a rate of 137 syllables per minute or 65 words per minute, and third graders read at a rate of 181 syllables per minute or 86 words per minute (P. Tressoldi, personal communication to M. Jukes). However, syllables lack the intuitive sense of words.

Table 4: Number of linguistic units in typical passages from Grade 1 reading textbook of 3 languages (Compiled by M. Jukes).

	Words	Syllables	Phonemes	Sentences	Syllables/Word	Phonemes/Syllable
English B	61	90	220	10	1.5	2.4
English C	63	85	221	8	1.3	2.6
Swahili D	41	100	192	9	2.4	1.9
Swahili E	41	130	240	10	3.2	1.8
Spanish	63	99	219	9	1.6	2.2
Variability*	21.8%	17.3%	7.8%			

The agglutinative language, Swahili, has significantly more syllables per written word.

Standardized Achievement Tests in Reading

Donor policy influence and financing has resulted in the construction and administration of sample-based standardized achievement tests for most low-income countries. The tests are typically multiple-choice and require reading for the assessment of language, math, and other subjects. There are also regional standardized tests, such as PASEC¹⁸ for Francophone Africa and SACMEQ for Anglophone Africa,¹⁹ which compare language outcomes for various countries.

The more recent versions use item response theory, which tracks item performance across ability levels and makes it easier to detect illiteracy. However, all tests are affected by reading fluency. This is also a problem that the International Association for the Evaluation of Educational Achievement (IEA) has tried to confront in its TIMSS (Trends in Mathematics and Science Study) and PIRLS (Progress in International Reading Literacy Study) that are given in grades 4 and 8 every 4 and 5 years respectively. PIRLS tests whether students have acquired the curriculum intended in schools. It assesses student proficiency in reading comprehension and focuses on the transition from learning to read to “reading to learn”. (It collects extensive data on the context for learning to read and includes 10 passages, 5 literary and 5 information in 126 items for 167 score points). The test has items that are multiple choice and also constructed response questions.

Due to limited literacy among many students participating in TIMSS and PIRLS, these tests are given in grades 4-6. To assess literacy more directly, IEA is preparing a pre-PIRLS test, which will be tested in six countries along with PIRLS in grade 4 in 2011, with data available in 2013. It will be a basic inferential diagnostic testing. Pre-PIRLS may be administered in grade 3, but still it will not reach grades 1 and 2.

Why should fluency measurements be made through texts read aloud rather than silently? In principle, it would save time to read silently, but research has shown lower correlations between fluency and comprehension when students indicate the amount of text read in one minute. Fuchs et al. (2000) found that for 4th graders reading silently, the correlation with the questions answered on the passage was .38, and 0.47 with the Iowa Test of Basic Skills. For oral reading by comparison the rate correlated 0.84 with the passage and 0.80 with the Iowa Test. Perhaps “recycling” a passage read aloud may help students reading more slowly retain it into working memory, though that same function may inhibit retention in higher reading rates. The Chilean NGO Educando Juntos (see below) recommends repeated silent readings and averages of purported rates, but it is unclear how much the correlation would improve.

The Aser (Uwezo) Reading Test

¹⁸ Programme d’analyse des systèmes éducatifs de la CONFEMEN, where CONFEMEN stands for “Conférence des ministres de l’éducation des pays ayant le français en partage

¹⁹ Southern and Eastern Africa Consortium for Monitoring Educational Quality

A different type of oral fluency test is given in India by the Pratham NGO, and various areas are surveyed on a regular basis. According to the amount of print students can read within a few minutes, they are classified as reading at the letter, word, sentence, or paragraph level. (Aser 2009; Aser means “effect” in Urdu.) This concept was employed before words per minute became widely known. Pratham has been collecting monitoring data using this concept, and in 2010 Kenya also conducted an Aser assessment. (See www.uwezo.net)

The Aser concept makes intuitive sense to people and has face validity. In Tanzania in 2009 22,000 households were tested by volunteers, covering all children in them in grades 1-8. In Tanzania about 90 percent of the second graders could not read a grade 2 story in Kiswahili, and even half of the older children failed to do so.

For statistical uses however, the concept has limited utility. The scale of measurement is at best ordinal. The progression from letters to words, sentences, paragraphs is discontinuous with unclear classifications in between. If viewed as a single scale, it is unknown whether it has a linear trajectory, and whether the intervals between its benchmarks are equal. Clearly sentences and paragraphs vary a lot in length in real life, and it’s unclear how long the prototypical concept would look like. For these reasons, the results are in the form of frequency distributions and amenable mainly to nonparametric statistics. By contrast, words per minute is a ratio scale, and a child reading 40 words reads twice as fast as child reading 20 words of the same text.) Nevertheless, the Pratham scale indirectly represents the timeframe in which a person can process a message. For this reason, testing with one-minute reading and with the Aser test has revealed close relationships.

Table 5: India – Aser Results at letter, word, paragraph, story levels

Reading Level Results For Enrolled Children - All India - 2009						
Class	Nothing	Letter	Word	Para	Story	Total
All	7.58%	16.78%	16.66%	18.89%	40.09%	100.00%
1	31.24%	44.58%	16.10%	4.76%	3.31%	100.00%
2	11.20%	33.63%	31.91%	14.87%	8.40%	100.00%
3	5.31%	19.41%	28.68%	26.83%	19.78%	100.00%
4	2.63%	10.51%	19.48%	29.80%	37.58%	100.00%
5	1.82%	6.69%	13.16%	25.51%	52.82%	100.00%
6	1.04%	3.86%	8.06%	20.75%	66.29%	100.00%
7	0.85%	2.80%	5.32%	15.64%	75.38%	100.00%
8	0.57%	1.85%	3.35%	11.76%	82.46%	100.00%

Psychometric and Sampling Issues of Oral Reading Fluency Tests

To be useful longitudinally as monitoring tools, fluency measures should be consistent over large samples. Some studies have tested **test-retest reliability** or **alternate-form reliability**. The former could be compromised by repeated reading of the same passage, and the latter would require tests of the same readability level. Multiple forms in Kenya by M. Jukes gave similar results. An Aser study in India showed statistically non-significant changes in test-retest and alternate form reliability (Aser 2009). Overall, it is unknown how repeated large-scale administrations in the same country would differ in results. To minimize error variance if reading rates are monitored on a regular basis, testing should be carried out at the same time each year. Timing matters, because students learn more every day. With the information known at this time, it seems that one-minute connected reading tests can be relatively easily and inexpensively used for country and school-level diagnosis, monitoring, and evaluation of remedial actions.

High reliability is possible because of large differences between fluent and non-fluent students. However, speed tests tend to have high reliability. The psychometric characteristics of reading fluency tests can also be studied by considering each word as an item, i.e. have a 60-item test. In many languages and scripts the “items” would be highly correlated, in others not. Highly correlated items would produce inflated coefficients, but others not.

Creating a monitoring baseline may seem in principle easy but to estimate correctly changes across time, very precise sampling procedures are needed. These may complicate the inherently simple nature of the test. Thus there are statistical precision choices that must be made. A large measurement error signifies broad confidence intervals, which may make the tests less useful for monitoring purposes. It may be difficult to document changes over time if the distributions from one year to the next overlap. Minimizing measurement error necessitates more complex tests. However, it is possible to study the distribution around the mean and pay less attention to statistical significance and effect size.

There are other requirements for the use of reading fluency tests as a monitoring tool. Students progress rapidly, and testing should be done early in the year to minimize school effects, and at the same time every year. However, the poorer students tend to lose skills over the summer, so perhaps testing should be done at the end of the year or a month after the school year starts and students have recovered.

The sample size is also something to monitor for programming purposes; 400 students per grade are needed to show growth from the baseline, more if other variables are involved (e.g. rural-urban differences). Small samples and incidental samples may provide a lot of one-time information. But their utility for monitoring purposes is limited because they are likely to have significant variability from year to year. Repeated measures from Mexico and Andalusia demonstrate this over time. Timing must also be considered, and comprehension questions must be comparable from year to year.

Potential for group testing. Reading fluency tests have been administered individually worldwide, but logistics may become complex in schools, where students should not hear other

students reading. If some group paper-and-pencil tests existed that closely tracked the oral tests, this would be useful. Such a possibility exists with the test Wordchains. Students are asked to mark the ends of connected words within three minutes (e.g. “houseboxminebicycle”). The test will not work with scripts that have final letters, but it works at least with the Latin script. Experimentation should be carried out on its use in various languages. Other EGRA subtests could also be studied for adaptation to a group setting, such as the dictation test and potentially the listening comprehension test. However, the pros and cons must be considered, even if tests can be given in groups; for example, collecting and scoring group tests may cancel out time gains from individual administrations.

In addition, there is software to measure speed and accuracy (e.g. www.soliloquylearning.com). It has read-aloud voice recognition and comprehension (developed through the support of NICHD). The software records student voices, corrects the student when words are read incorrectly, includes comprehension questions and scores the students for future analysis. Another system simply enables test administrators and researchers to record student words and responses and assess the number of words (or letters or syllables) correct per minute (www.educationalhelp.com). Such software can also measure silent student reading.

Norms and Benchmarks Pertaining to Oral Reading Fluency Tests

In middle-income countries, there is little need to think of words per minute or fluency standards for normal populations. Some students may acquire fluency early and some others later on, but the vast majority of students acquire basic fluency in grades 1-2. In fact, 45-60 words per minute constitutes minimal achievement, often acquired by the end of grade 1; 45 words per minute corresponds to the 18th and 25th percentile of US norms in the spring of grade 2, and similarly of Latin American norms (Tables 1-5). In many low-income countries, however, students may drop out early and fail to attain fluency by the time they graduate or drop out. Simple numerical benchmarks may focus governments and educators on attainable goals that can also be intuitively understood.

For monitoring reasons, a number of countries have developed norms. There are English, Spanish, and Italian norms, and they show considerable consistency. The most prominent are:

US norms (Hasbrouck and Tindal 2006) were developed after large-scale testing of representative samples (Table 2). English has complex spelling, and several U.S. students are poor. Other countries have lower-income populations but simpler orthographies. Since detailed norms have not been developed in many other countries, a number of reading studies make reference to the U.S. norms (Tables 3-5, Figure 1).

In **Chile**, the teacher-to-teacher network program of the Ministry of Education set goals for grades 1 and 2 at 30 and 70 words per minute, while the Chilean non-governmental organization (NGO) Educando Juntos has proposed goals around 34 and 64 words per minute for grades 1 and 2 respectively (www.educandojuntos.cl). Studies in Mexico, Uruguay, and Paraguay also offered similar benchmarks.

By Bisquera	40.70	70.71	101.4	120.22	136.18	152.88	167.44	177.70
General-1998	54.10	64.29	89.3	101.03	102.97	111.91	124.48	157.75
General-1999	59.28	77.75	95.97	108.67	122.73	119.54	157.63	157.75
2002	51.43	70.24	70.67	98.18	91.43	109.38		138.50

Note: Equipo de Orientacion de Marbella, Andalusia, Spain 2002. E.g. N=92

Table 9: Reading guidelines used by the Chilean Educando Juntos NGO

Correct words per minute	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Very fast	56	84	112	140	168	196	214	214
Fast	47-55	74-83	100-111	125-139	150-167	178-195	194-213	194-213
High average	38-46	64-73	88-99	111-124	136-149	161-177	174-193	174-193
Low average	29-37	54-63	76-87	97-110	120-135	143-160	154-173	154-173
Slow	22-28	43-53	64-75	85-96	104-119	125-142	135-153	135-153
Very slow	21	42	63	84	103	124	134	134

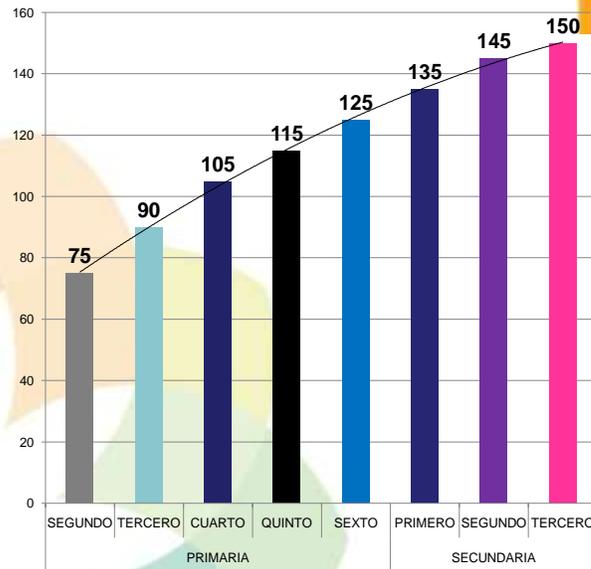
Figure 7: Mexico – Reading Speed Standards (May 2010)

Estándares acordados para Velocidad Lectora



La propuesta está basada en la distribución de las velocidades lectoras en cada uno de los grados. Esta información se ha combinado con la velocidad promedio de los alumnos encuestados en marzo de 2010, que en cada grado en la prueba ENLACE 2009 están incluidos en los niveles de logro bueno y excelente.

Asimismo considera la evidencia proporcionada por la investigación internacional que indican que los incrementos en Velocidad lectora decrecen o se van estabilizando al aumentar el grado escolar



4

References

- Abadzi, H. 2006. Efficient learning for the poor. Washington, DC: The World Bank.
- Abadzi, H. 2007. Absenteeism and beyond: Instructional Time Loss and Consequences. World Bank: Policy Research Working Paper 4376.
- Abadzi, H., Crouch, L., Echegaray, M., Pasco, C., & Sampe, J. 2005. "Monitoring basic skills acquisition through rapid learning assessments: A case study from Peru." *Prospects*, 35(2), 137-156.
- Abdul Latif Jameel Poverty Action Lab (J-PAL), Pratham, & ASER (2009). Evaluating READ INDIA: The development of tools for assessing Hindi reading and writing ability and math skills of rural Indian children in Grades 1-5. Unpubl. Draft. Chennai, India: J-PAL.
- Abdul Latif Jameel Poverty Action Lab (J-PAL), 2006. Making Schools Work for Marginalized Children: Evidence from an inexpensive and effective program in India. Abdul Lateef Jamil Poverty Action Lab Briefcase no. 2, November 2006 Massachusetts Institute of Technology.
- Alcock, K J., Nokes, K; Ngowi, F; Musabi, C; Mbise, A; Mandali, R; Bundy, D; Baddeley, A.; Kiswahili reading tests: The development of reading tests for use in a regularly spelt language. Partnership for Child Development, Wellcome Centre for the Epidemiology of Infectious Disease, University of Oxford
- Alfari, Djibo. 2009. Panel: Success in Early Grade Reading Project in Mali and Niger. In G. Brozo and E. Sturtevant. *Beyond Access: Effective Reading for All*. Report from the 6th Annual Global Perspectives on Literacy. International Reading Association.
- Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuo-spatial short-term and working memory in children: are they separable? *Child Development*, 77, 1698-1716.

- Anderson, R., Fielding, L., & Wilson, P. 1988. Growth in reading and how children spend their time outside of school. *Reading Research Quarterly*, 23, 285-303.
- Annie E. Casey Foundation . 2010. Early Warning: Why Reading by the End of Third Grade. Baltimore, www.aecf.org
- ASER 2009. Evaluating the reliability and validity of the ASER testing tools. Unpublished draft. New Delhi (www.asercentre.org).
- Atkinson, R. C. & Shiffrin, R.M. 1968. Human memory: A proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), *The Psychology of Learning and Motivation*, Vol 2. New York: Academic Press.
- Baddley, A. 1999. *Essentials of Human Memory*. East Sussex, UK: Psychology Press.
- Bandura, Al. 1997. *Self-efficacy: The exercise of control*. New York: Worth Publishers
- Barr, R., C. Blachowicz, C. Katz, and B. Kaufman. 2002. *Reading Diagnosis for Teachers: An Instructional Approach* (4th ed.). Boston, MA: Allyn and Bacon.
- Beaulieu, C., C. Plewes, L.A. Paulson, D. Roy, L. Snook, L. Concha, and L. Phillips. 2005. "Imaging Brain Connectivity in Children with Diverse Reading Ability." *NeuroImage* 25: 1266–1271.
- Bolger, D. J., C. A. Perfetti, and W. Schneider. 2005. Cross-cultural effect on the brain revisited: Universal structures plus writing system variation. *Human Brain Mapping*, 25, 92.
- Bonnet G., Braxmeyer N., Horner S., Lappalainen H.P., Levasseur J., Nardi E., Rémond M., Vrignaud P. et White J. (2001), *The Use of National Reading Tests for International Comparisons : Ways of Overcoming Cultural Bias*, Paris: ministère de l'Éducation nationale-DPD, , n° 04.07.
- Bonnet, G, Daems, F, Glopper, C M de, Horner, S, Lappalainen, H P, Nardi, E, Remond, M, Robin, I, Rosen, M, Solheim, R G, Tonnesen, F E, Vertecchi, B, Vrignaud, P, Wagner, A K H, White, J. 2003. *Culturally balanced assessment of reading [c-bar]*, Opdrachtgever: European network policy makers evaluation of educational systems, Ministère de la Jeunesse, de l'Éducation Nationale et de la Recherche, DEP Éd., Paris.
- BRAC. 2009. Dialogic teacher. Unpublished abstract. Dhaka, BRAC.
- Breznitz, Z. 1997a. Effects of Accelerated Reading Rate on Memory for Text among Dyslexic Readers. *Journal of Educational Psychology* 89: 289–297.
- . 1997b. Enhancing the Reading of Dyslexic Children by Reading Acceleration and Auditory Masking. *Journal of Educational Psychology* 89: 103–113.
- Brus, B. T. and Voeten, M. J. M. 1979. *Een-minuut-test*. [One minute test]. Nijmegen, the Netherlands: Berkhout.
- Campbell, J.R., Kelly, D.L., Mullis, I.V.S., Martin, M.O., & Sainsbury, M. 2001. *Framework and Specifications for PIRLS Assessment 2001—2nd Edition*. Chestnut Hill, MA: Boston College.
- CFIE-IPN. 2006. *Fluidez lectora por grado escolar en una muestra de niños mexicanos 1er Congreso de Innovación Educativa*.
<http://148.204.73.101:8008/jspui/handle/123456789/457>.
- Chall, J. S. 1982. *Reading: The Latest: PsycCRITIQUES Vol 27 (6)*.
- Changizi M.A. and S. Shimojo. 2005. Character complexity and redundancy in writing systems over human history. *Proceedings of the Royal Society-Biology*, ;272(1560):267-75.

- Chen, R., & Vellutino, F. R. 1997. Prediction of reading ability: A cross-validation study of the simple view of reading. *Journal of Literacy Research*, 29, 1–24.
- Chun, M. M. and M. C. Potter. 1995. A Two-Stage Model for Multiple Target Detection in Rapid Serial Visual Presentation. *J Exp Psychol: HPP* 21(1): 109-127
- Cowan, N. 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87-185.
- Cowan, N. 2005. Working-memory capacity limits in a theoretical context. In C. Izawa & N. Ohta (eds.), *Human learning and memory: Advances In theory and applications*. The 4th Tsukuba international conference on memory. Erlbaum. (pp. 155-175)
- Culbertson, Jennifer, and Geraldine Legendre. 2010. Investigating the evolution of agreement systems using an artificial language learning paradigm. In *Proceedings of the 2010 Western Conference on Linguistics*. Department of Linguistics, California State University, Fresno. (Artificial grammar reveals inborn language sense, study shows. ScienceDaily. Retrieved May 13, 2011, from <http://www.sciencedaily.com/releases/2011/05/110513112256.htm>)
- Crouch, L. and M. Korda. 2008. EGRA Liberia: Baseline Assessment of Reading Levels and Associated Factors. Report prepared under contract to the World Bank.
- Crouch, Luis and Amber Gove Leaps or One Step at Time: Skirting or Helping Engage the Debate? The Case of Reading. Draft : 19 November 2009
- Daneman, Meredyth and Patricia A. Carpenter. 1980. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Davidson, M., and J. Towner. 2005. Evaluating An Oral Reading Fluency Cut-Score to Identify Second Grade Children Who Are At-Risk for Reading Failure. Paper submitted to Educational Measurement.
- de la Colina, M. G., R. Parker, J. Hasbrouck, and R. Lara-Alecio. 2001. Intensive Intervention in Reading Fluency for At-Risk Beginning Spanish Readers. *BilingualResearch Journal* 25(4): 417–452.
- De Luca, M., L. Barca, C. Burani, P. Zoccolotti, Pierluigi. 2008. The Effect of Word Length and Other Sublexical, Lexical, and Semantic Variables on Developmental Reading Deficits. *Cognitive & Behavioral Neurology*. 21, 4, 227-235.
- Dehaene , Stanislas and Cohen, Laurent (2010) The unique role of the visual word form area in reading. *Trends in Cognitive Science*, 959.
- Dehaene, S., F. Pagado, L. Braga, P. Ventura, G. Nunes Filho, A. Jobert, G. Dehaene-Lambert, R. Kolinsky, J. Morais and L. Cohen. 2010. How learning to read changes the cortical networks for vision and language. *Science*, 330, 6009, 1359-1364.
- Devlin, Keith. 2010. The mathematical brain. In D. De Souza (ed.) *Mind, Brain, and Education*. Bloomington, IN: Solution Tree Press, p. 169.
- Deno, S. L. 1985. Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 2, 219-232.
- Diakidou, I. N., Stilianou, P., Karefillidou, C., & Papageorgiou, P. 2005. The relationship between listening and reading comprehension of different types of text at increasing grade levels. *ReadingPsychology*, 26, 55–80.
- Dowd, Amy Jo and Elliott Friedlander. 2009. Bangladesh Program: Emergent and early grades reading assessment validation study results. Save the Children (draft) 4/7/09.

- Dowd, Amy Jo. 2011. Fluency, accuracy and comprehension across the globe: Unpacking the utility of 60 words per minute. Paper presented at annual meeting of the Comparative and International Education Society, Montreal, Canada, May 2, 2011.
- Dowrick, P. W. and Weol Soon. Kim-Rupnow, and Thomas J. Power. 2006. Video Feedforward for Reading. *The Journal of Special Education*, 39/NO, 194-207.
- Ellis, N. C. and R. A. Hennelly. A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology*, 71:43–51, 1980.
- Equipo de Orientación Educativa de Marbella. 2003. Evaluación de la velocidad lectora oral y análisis de la correlación de esta variable con la nota global de junio. Consejería de Educación y Ciencia. Junta de Andalucía.
- EQUIP2. 2010. Using Opportunity to Learn and Early Grade Reading Fluency to Measure School Effectiveness in Woliso, Ethiopia. (USAID).
- Espin, C.A., & Deno, S.L. (1993). Performance in reading from content-area text as an indicator of achievement. *Remedial and Special Education* 14 (6), 47-59.
- Eviatar, Zohar; Ibrahim, Raphiq; Ganayim, Deia. 2004. Orthography and the Hemispheres: Visual and Linguistic Aspects of Letter Processing. *Neuropsychology*. Vol 18(1), Jan 2004, 174-184.
- Fuchs, L. S., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review*, 28(4), 659-671.
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5(3), 239-256.
- Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education*, 9(2), 20-28.
- Fuchs, L. S., Fuchs, D., Eaton, S., & Hamlett, C. L. 2000. Relation between reading fluency and reading comprehension as a function of silent versus oral reading mode. Unpublished data.
- Gee, J. P. 2007. What video games have to teach us about learning and literacy (2nd ed) New York: Palgrave MacMillan.
- Georgiou, G., Das, J. P., & Hayward, D. 2008. Comparing the contribution of two tasks of working memory to reading in relation to phonological awareness and rapid naming speed. *Journal of Research in Reading*, 31, 302-318.
- Georgiou, G., Parrila, R., & Papadopoulos, T. 2008. Predictors of word decoding and reading fluency in English and Greek: A cross-linguistic comparison. *Journal of Educational Psychology*, 100, 566-580.
- Georgiou, G., Das, J. P., & Hayward, D. 2009. Revisiting the “simple view of reading” in a group of children with poor reading comprehension. *Journal of Learning Disabilities*, 42, 76-84.
- Geva, E., & Siegel, L. S. 2000. Orthographic and cognitive factors in the concurrent development of basic reading skills in two languages. *Reading and Writing: An Interdisciplinary Journal*, 12, 1-30.
- Good, R.H., III, Simmons, D.C., & Smith, S.B. 1998. Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School of Psychology Review*, 27(1), 45-56.

- Gove, A. 2009. Early Grade Reading Assessment Toolkit. RTI International, USAID and the World Bank. Available: www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=149
- Gove 2009b. Beyond Access: Effective Reading for All and the Early Grade Reading Assessment. Presentation at the World Bank Institute, April 30, 2009, RTI.
- Gupta, P. Lipinski, J., Abbs, B., and Lin, P-H. (2005). Serial Position Effects in Nonword Repetition. *Journal of Memory and Language*, 53, 141-162
- Hasbrouck and Tindal. 2006. Hasbrouck, J., and G.A. Tindal. 2006. "Oral Reading Fluency Norms: A Valuable Assessment Tool for Reading Teachers." *The Reading Teacher* 59: 636–644.
- Hulme, Charles Roodenrys, Steven Brown, Gordon Mercer et.al. 1995. The role of long-term memory mechanisms in memory span. *British Journal of Psychology*; Nov 1, 1995.
- IEG. 2007. Mozambique: Project Performance Assessment Review
- Issé, Mahdi M. 2009. Djibouti. Ministère De L'Éducation Nationale Et De L'Enseignement Supérieur Evaluation en lecture des élèves de 2^{ème} année Rapport d'analyse Mai 2009. MENESUP
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grade. *Journal of Educational Psychology*, 80(4), 437–447.
- Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review*, 103(4), 773-780.
- Karatzas, Andreas. 2005. The acquisition of orthographic competency. (Η μάθηση της ορθογραφικής δεξιότητας.) Athens: Grigoris Publishers.
- Kudo, Inez and Jorge Bazan. 2008. Measuring beginner reading skills: An empirical evaluation of alternative instruments and their potential use for policymaking and accountability in Peru. World Bank: Policy Research Working Paper Series No 4812.
- LaBerge, D., and S. J. Samuels. 1974. Toward a theory of automatic information processing in reading. *Cognitive Psychology* 6: 293–323.
- Llambiri, Stavri. 2004. "Braktisja e fshehtë", "Objektivat minimale të domosdoshme të nxënësve", Tirana, Shtëpia Botuese "Albas".
- Liao, C. H., Georgiou, G., & Parrila, R. (2008). Rapid naming speed and Chinese character recognition. *Reading and Writing: An Interdisciplinary Journal*, 21, 231-253.
- Marston, D., & Tindal, G. 1996. Technical adequacy of alternative reading measures as performance assessments. *Exceptionality*, 6, 201–230.
- McCandliss, B.D., Cohen, L., & Dehaene, S., (2003). The Visual Word Form Area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, 7, 293-299.
- Martelli, M., Di Filippo, G., Spinelli, D., & Zoccolotti, P. 2009. Crowding, reading, and developmental dyslexia. *Journal of Vision*, 9(4):14, 1-18.
- Miestamo, Matti, Kaius Sinnemäki, Fred Karlsson (eds). 2008. Language Complexity: Typology, Contact, Change. Language companion series 94. John Benjamins Publishing co.
- Mitton, Giselle. 2008. Success in Early Reading Pilot project in Mali and Niger Implementation report GAD MLI0080 & NER064. Plan International.
- Migliore, Michele, Gaspae Novara, and Domenico Tegolo. 2008. Single neuron binding properties and the magic number 7. *Hippocampus*, 18, 1122-1130.
- Miller, George A. 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, 63: 81-97.

- Müller, K., & Brady, S. (2001). Correlates of early reading performance in a transparent orthography. *Reading and Writing: An Interdisciplinary Journal*, 14, 757-799.
- OED 2005. "Niger: First Education Project (Credit 1151-NIR), Primary Education Development Project (Credit 1740-NIR), Basic Education Sector Project (hybrid)--(Credit 2618-NIR)." Project performance assessment report. Washington, DC: World Bank: Independent Evaluation Group.
- Ouellette, G. 2006. What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98, 554-566
- Peereman, Ronald, Bernard Lété, Liliane Sprenger-Charolles. 2007. Manulex-infra: Distributional characteristics of grapheme–phoneme mappings, and infralexical and lexical units in child-directed written material. *Behavior Research Methods* 2007, 39 (3), 593-603
- Pelli, D. G., C. W. Burns, B. Farell, and D. C. Moore-Page. 2006. Feature detection and letter identification. *Vision Research* 46(28): 4646–4674.
- Pérez Villar, J.. 1996. ¿Cómo lee mi paciente?: Contribución a la metodología del examen directo en psiquiatría de niños. *Revista Cubana de Pediatría* 68(3): 201–210.
- Peterson, L.R. and M. J. Peterson. 1959. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193-198.
- Pikulski, J. J., & Chard, D. J. 2005. Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*, 58(6), 510-519.
- Piper, B. 2009. "Impact Study of SMRS Using Early Grade Reading Assessment in Three Provinces in South Africa." Integrated Education Programme, RTI International.
- Piper, B. 2009. Early Grade Reading Assessment (EGRA) Plus Liberia: Project Overview. EdData II Technical and Managerial Assistance, Task Number 6 Contract Number EHC-E-00-04-00004-00 Strategic Objective 3, December 2009.
- Pratham. 2009. Fluency and comprehension in Bihar and Uttarakhand. Unpublished report.
- Riedel, Brant. 2007. The Relation Between DIBELS, Reading Comprehension, and Vocabulary in Urban First-Grade Students. *Reading Research Quarterly* 42: 546–567.
- Rasinski, T., 2003. *The Fluent Reader: Oral Reading Strategies for Building Word Recognition*. United States: Scholastic.
- Rasinsky, Timothy V. 2003. One-Minute Reading Probe for Assessing Fluency de Timothy V. Rasinsky (p 159).
- Riedel, Brant W. (2007). The relationship between DIBELS, reading comprehension, and vocabulary in urban first grade students. *Reading Research Quarterly*, Vol 42, No. 4, 546-562.
- Research Triangle Institute (RTI) 2007. Results of EGRA design workshop. Sponsored by EdData II Project and Aga Khan Foundation EMACK II Project. Presentation in Mombasa , 23-27 April 2007
- _____. 2007. Early Grade Reading Kenya Baseline Assessment: Analyses and Implications for Teaching Interventions Design July 18-26, 2007 Malindi, Kenya Prepared by RTI International for the Aga Khan Foundation.
- _____. 2007. Senegal Early Grade Reading Assessment (EGRA) Results from Senegalese Primary School Students. Learning to Read in French and in Wolof—Report for the World Bank

- _____. 2008. Early Grade Reading Kenya: Baseline Assessment EdData II Technical and Managerial Assistance, Task Number 4 Contract Number EHC-E-01-04-00004-00 Strategic Objective 3, January 2008
- _____. 2009. EGRA Plus: Liberia Data Analytic Report: EGRA Plus: Liberia Mid-Term Assessment Early Grade Reading Assessment (EGRA) Plus: Liberia EdData II Task Number 6 usaid Contract Number EHC-E-06-04-00004-00. Strategic Objective 3 October 31, 2009.
- _____. 2010a. "Early Literacy: Igniting education for all". RTI Publication.
- _____. 2010b. Ethiopia Early Grade Reading Assessment. Data Analytic Report: Language and Early Learning. EdData II, Ed Data II Task Number 7 and Ed Data II Task Number 9, October 31, 2010
- Royer, M., Abadzi, H., and Kinda, J. 2004. The Impact of Phonological-Awareness and Rapid-Reading Training on the Reading Skills of Adolescent and Adult Neoliterates. *International Review of Education*, 1: 53-71.
- Salamone, J. D. and Correa, M. 2002. Motivational views of reinforcement. Implications for understanding the behavioral functions of the nucleus accumbens dopamine. *Behavioral Brain Research*, 137, 3-25.
- Schilling, S. G., Carlisle, J. F., Scott, S. E., & Zeng, J. 2007. Are fluency measures accurate predictors of reading achievement? *Elementary School Journal*, 107(5), 429-448.
- Schuh Moore, Audrey-marie, Joseph DeStefano and Elizabeth Adelman, 2009. Opportunity to Learn as a Measure of School Effectiveness in Guatemala, Honduras, Ethiopia, and Nepal. Educational Quality Improvement Program (EQUIP2). Academy for Educational Development.
- Seymour, P., H. K. M. Aro, and J. M. Erskine. 2003. Foundation Literacy Acquisition in European Orthographies. *British Journal of Psychology* 94(2): 143–174.
- Shaywitz, Sally. 2003. *Overcoming Dyslexia*. New York: Alfred Knopf.
- Share, D.L. 2008. On the Anglocentricities of current reading research and practice: the perils of overreliance on an "outlier" orthography. *Psychological Bulletin*, 134, 584-615.
- Share, D. L. 1999. Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, 72, 95–129.
- Shaw, R. & Shaw, D. 2002. DIBELS oral reading fluency-based indicators of third grade reading skills for Colorado State Assessment Program (CSAP). Technical report. Eugene, OR: University of Oregon.
- Simeon, D.T., and S. Grantham-McGregor. 1989. "Effects of Missing Breakfast on Cognitive Functions of School Children of Different Nutritional Status." *American Journal of Clinical Nutrition* 49: 646–653.
- Snow, C. E., M. Susan Burns, and P. Griffin (eds.). 1998. *Preventing Reading Difficulties in Young Children*. Committee on the Prevention of Reading Difficulties in Young Children. Washington DC: National Research Council.
- Spinelli, Donatella. 2005. Length Effect in Word Naming in Reading: Role of Reading Experience and Reading Deficit in Italian Readers. *Developmental Neuropsychology*, 27(2), 217–235.
- Sprenger-Charolles, L. 2008. The Gambia Early Grade Reading Assessment: Results from 1200 Gambian primary students learning to read in English. *Journal*. Retrieved from <https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=116>.

- Swanson, Lee and T. Alloway. 2010 (upcoming). APA Educational Psychology Handbook, Eds (Harris, K., Urdan, T., & Graham, S.) Vol 1 - Educational Psychology: Contributions to Education.
- Sereno, Sara C. and Keith Rayner. 2003. Measuring word recognition in reading: eye movements and event-related potentials. *TRENDS in Cognitive Sciences* Vol.7 No.11.
- Tan, A., and T. Nicholson. 1997. Flashcards Revisited: Training Poor Readers to Read Words Faster Improves their Comprehension of Text. *Journal of Educational Psychology* 89: 276–288.
- Tindal, G., Hasbrouck, J., & Jones., C. (2005). Oral reading fluency: 90 years of measurement. Eugene, OR: Behavioral Research and Teaching, Technical report #33; http://brt.uoregon.edu/techreports/ORF_90Yrs_Intro_TechRpt33.pdf
- Torres, Ximena. 2007 Metodologia para la evaluacion de la velocidad lectora. www.educandojuntos.cl
- Tressoldi, P. E., Stella, G., & Faggella, M. (2001). The development of reading speed in Italians with dyslexia: A longitudinal study. *Journal of Learning Disabilities*, 34(5), 414-417.
- Vander Meer, C. D., Lentz, F. E., & Stollar, S. (2005). The relationship between oral reading fluency and Ohio proficiency testing in reading. Technical Report. Eugene, OR: University of Oregon.
- Wagner, D. A. 2003. "Smaller, quicker, cheaper: Alternative strategies for literacy assessment in the UN Literacy Decade." *International Journal of Educational Research*, 39.
- Wagner D.A. 2011. Smaller, Quicker, Cheaper. Improving Learning Assessments for Developing Countries. Education for All Fast Track Initiative Secretariat.
- Warmington, M. and C. Hulme. 2012. Phoneme awareness, visual verbal-paired associate learning, and rapid automatized naming as predictors of individual differences in reading ability. *Scientific Studies of Reading*, 16 (1), 45-62.
- Wilson, J. 2005. The relationship of Dynamic Indicators of Basic Early Literacy Skills [DIBELS] oral reading fluency to performance on Arizona Instrument to Measure Standard: (AIMS). Technical Report. Tempe School District No. 3, Tempe, Arizona.
- Willms, J. D. 2008. Learning Divides: Ten Policy Questions About The Performance And Equity Of Schools And Schooling Systems. Montreal: Unesco Institute of Statistics.
- Wimmer, H., K. Landerl, and U. Frith. 1999. "Learning to Read German." In *Learning to Read and Write: A Cross-Linguistic Perspective*, M. Harris and G. Hatano, eds. Cambridge, U.K.:Cambridge University Press, pp. 34–50.
- Wood, D. E. 2006. Modeling the relationship between oral reading fluency and performance on a statewide reading test. *Educational Assessment*, 11 (2), 85–104.
- Vagh, Shaher Banu. 2009. Evaluating the Reliability and Validity of the ASER Testing Tools. ASER Centre: www.asercentre.org
- Valencia, S. W. , A. T. Smith, A. M. Reese, M. Li. 2010. Oral Reading Fluency Assessment: Issues of Construct, Criterion, and Consequential Validity. *Reading Research Quarterly*. 45(3), 270–291.
- Vukovic, R. K and Linda Siegel. 2006. The role of working memory in specific reading comprehension difficulties. In Alloway, T. and Gathercole, S. (eds.) *Working Memory and Neurodevelopmental Disorders*. New York: Psychology Press.

- Yap, Melvin J. Susan J. Rickard Liow, Sajlia Binte Jalil, Siti Syuhada Binte Faizal. 2010. The Malay Lexicon Project: A database of lexical statistics for 9,592 words. *Behavior Research Methods*, 2010, 42 (4), 992-1003.
- Zoccolotti, P., M. De Luca, E. Di Pace, F. Gasperini, A. Judica, D. Spinelli. 2005. Word length effect in early reading and in developmental dyslexia. *Brain and Language*, 93, 3, 369-373.