Education For All
Fast Track Initiative

DRAFT

Working Paper Series

Reading Fluency Measurements in EFA FTI Partner Countries: Outcomes and Improvement Prospects

by **Helen Abadzi**
Senior Education Specialist
EFA FTI Secretariat

Update September 2011

© September 2011 The Education for All Fast Track Initiative Secretariat

EFA FTI Secretariat
c/o World Bank
MSN P6 - 600
1818 H Street, NW

Washington, DC 20433 USA

Web site: www.educationfasttrack.org
E-mail: info@educationfasttrack.org

The Education for All Fast Track Initiative (EFA FTI) was launched in 2002 as a global partnership between donor and partner countries to ensure accelerated progress toward the Millennium Development Goal of universal primary education by 2015. All low-income countries that demonstrate commitment to achieving universal primary education can receive support from EFA FTI.

This paper was produced by Helen Abadzi, Senior Education Specialist at the EFA FTI Secretariat. The current version is a perpetual draft, so additions and changes will be included over time.

For more information on the paper, please contact Helen Abadzi: habadzi@educationfasttrack.org or habadzi@worldbank.org.





Contents

REDO !

Figures

Figure 1 and 2: Relationships between fluency at comprehension in different countries..... 31
Figure 3 and 4: The relation between fluency and comprehension..... 36
Figure 5: Reading instruction and testing in Afghanistan..... 24
Figure 6: Fluency of FTI Partner countries in comparison to European countries at the end of grade 1 .. 14
Figure 7: Average fluency and comprehension rates in 17 selected FTI partner countries..... 15
Figure 8: Reading fluency measures and extracurricular activities in Timor Leste..... 16
Figure 9: Percent of grade 2 students who could not read a single word of a simple para, 2008-2009.... 16
Figure 10: Reading fluency test in Mali..... 20
Figure 11: Systematic Method for Reading Success in Mali..... 20
Figure 12 and 13: Outcomes of a reading fluency pilot in Mali..... 21
Figure 14: Donor support matters..... 23
Figure 15: Mexico – Proposed reading fluency standards (May 2010)..... 49

Tables

Table 1: FTI – Reading Fluency Studies in Eligible Countries (April 2010)..... 11
Table 2: Rough averages abstracted from reading fluency studies..... 14
Table 3: Summary Opportunity to Learn Indicators, sample averages..... 17
Table 4: The number of read daily multiplies into many words per year and pays off..... 19
Table 5: US Hasbrouck and Tindal norms, 2005..... 48
Table 6: Reading fluency norms from some medium and higher-income countries..... 48
Table 7: Overall consistency but also variability of scores inherent in small samples..... 48
Table 8: Reading guidelines used by the Chilean Educando Juntos NGO..... 49
Table 9: Swahili passage use in Malindi, Kenya..... 41
Table 10: Number of linguistic units in typical passages from Grade 1 reading textbook of 3 languages. 42
Table 11: Linguistic units comparison

Acronyms and Abbreviations

DI BELS	Dynamic Indicators of Basic Early Literacy Skills
EGRA	Early Grade Reading Assessment
EFA FTI	Education for All Fast Track Initiative
IEG	Independent Evaluation Group of the World Bank
NGOs	Non-Gouvernemental Organisations
PASEC	Programme d'analyse des systèmes éducatifs de la CONFEMEN
PIRLS	Progress in International Reading Literacy Study
RTI	Research Triangle Institute
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SMRS	Systematic Method for Reading Success
TIMSS	Trends in Mathematics and Science Study
USAID	United States Agency for International Development
wpm	words per minutes

Abstract

Students in lower-income countries often acquire limited literacy in school and often drop out illiterate. For those who stay, the problem is not detected until it is too late to intervene. Oral reading fluency tests given in the early grades can quickly and inexpensively assess student literacy. For this reason, one-minute reading studies have been popular. A search carried out in early 2010 showed that over 50 fluency studies have been conducted in various countries, and that norms have been established in the U.S., Mexico, and Chile. The studies often reported data in ways that were not easily comparable, and few had collected nationally representative data. However, the findings consistently showed very limited achievement. A multi-country study matching reading and instructional time data showed that the deficits are largely due to limited reading practice. The findings also suggest that few governments have taken action to improve reading outcomes on the basis of test scores. However, a number of pilot reading programs that emphasized phonics and practice were financed by donors and implemented by NGOs. These brought about substantial improvements within a few months. Their success suggests that it is eminently feasible to raise student outcomes significantly through evidence-based reading methods. Overall, the oral reading fluency tests have shown good psychometric properties, although reading achievement typically shows much variability within classes and sampling procedures could improve. Cross-linguistic comparability is rough and approximate, but overall it is possible to monitor reading outcomes across time and countries.

Executive Summary

Reading fluency is related to academic performance throughout the school years, so acquiring this skill early on is important. Curricula in all countries specify that students should learn reading in grades 1-2, but many studies have found extensive illiteracy in low-income countries, even in advanced grades. This deficiency compromises the large enrollment increases brought about worldwide through donor coordination and the Education for All Fast Track Initiative (EFA FTI). For this reason, the EFA FTI Secretariat has established quality-oriented benchmarks for Partner countries that focus on reading fluency goals for grade 2 and for the end of primary cycle.

For decades, oral reading fluency tests have been given in Europe and the US. They typically consist of one-minute reading passages and comprehension questions. Such tests can be relatively easily and inexpensively used for country- and school-level diagnosis, longitudinal monitoring, and evaluation of remedial actions. Measures of words read correctly per minute are appealing due to the simplicity and intuitive meaning of scores. They fill a gap in grades 1-3, when sample-based assessments are rarely given. Oral reading tests have a high variance, given normal student variability within classrooms. They often lack the sophistication of international comparison tests, but where psychometric properties were calculated, the tests had strong validity and reliability indicators.

The donor community and FTI Secretariat have promoted the use of reading fluency tests to monitor achievement in the early grades. In early 2010 the EFA FTI Secretariat surveyed literacy researchers and staff of various development agencies to find out how many reading studies have taken place in lower-income countries and what their outcomes were. The survey showed that over 50 reading fluency studies have been conducted worldwide. As of April 2010, one or more studies have been carried out in at least 29 of the 71 FTI-eligible countries. Seven countries had multiple undertaken studies, three had ongoing data collection in early 2010, and in 15 others, studies were being actively planned as of the same date. In 28 other countries, reading fluency assessments had not been conducted, and there were no active plans to do one. Some studies were small, but others involved from 800 to about 18,000 students. However, few have collected nationally representative data. They often centered on specific regions or central areas and tended to exclude remote areas. A few involved small convenience samples. Study reports and datasets were of variable quality.

Regardless of quality, however, the studies consistently showed very low reading speed and comprehension. They suggest that probably no low-income EFA FTI country would meet the reading indicators of the FTI indicative framework in grade 2 or at the end of primary education.

To understand a simple passage given the capacity of short-term memory, average students should read a minimum of 45-60 words per minute. Learning research and existing data suggest that this standard is possibly usable worldwide. (See details in Annex 1.) To progress in school work and attain this standard, students in probably every language or script should be expected to "crack the code" by the end of grade 1 (reading grade 1 text at around 30 words correct per minute), and to read relatively fluently by the end of grade 2 at 45-60 words per minute. With

the exception of a Nicaraguan sample of large schools, no country approached these reading speeds. Significant numbers of students in grades 1-2 (from 3 to 93 percent) in various countries did not even know the values of letters. Comprehension percentages were similarly low, and **on average, the children who were tested made little sense of the text**. To answer correctly 80 percent of comprehension questions, students in many countries had to read at higher speeds than the minimum.

Why do students score so poorly? A primary cause of illiteracy seems to be **insufficient amounts of instruction, practice, and feedback.** To become fluent and understand text better, students must be taught individual letters, study in a language that they understand, and get textbooks for practice. Instructional time needs to be spent on the pertinent tasks. However, these conditions are often not met. One USAID-financed study measured both reading fluency and the amount of instructional time spend on reading in schools of Nepal, Honduras, Guatemala, and Ethiopia. Overall fewer than half of the working days were spent on instruction. Within classrooms, less than 12 percent of the time was spent actually reading. For example, in Ethiopia, only 3 percent of the time was spent practicing the rather complex syllabic Amharic script. In a five-hour school day, 3 percent of the time would equate to *nine* minutes of reading per day (Schuh-Moore et al. 2009).

Some countries have embarked on randomized studies of improving performance, and the results are very encouraging. The few examples available worldwide suggest that specific interventions of a few months (perhaps no more than 100 instructional days) can help students learn the fundamentals of literacy. Scripted modular courses of a relatively short-term duration can help teachers stay on task where instructional time is limited, provided that materials in local languages are made available. For example the Indian non-governmental organization Pratham has developed programs that make most students literate in six weeks. In Mali, after 4 months of instruction in a local language, half the students could read 21 or more letters, compared to only 2 percent of students in public French-speaking schools after a full year of instruction.

It is reasonable to believe that low-income countries can achieve a significant surge in learning if they teach effectively their students to read in grades 1-2. The FTI Partner countries might focus on the service delivery aspects of their education plans to bring this about. Early reading instruction should use a language that children understand, with a transition into an official language (in countries where this is practiced) only after basic skills are mastered. Local education groups could support increased time on task, phonics instruction with phonological awareness, an abundance of reading materials, systematic monitoring of each child, instructional support to teachers with scripted lessons in early grades, and simple low-cost testing. Significant progress could be made in the first few months.

The EFA FTI partnership can help governments deliver improved service and instruction, and establish norms by which all countries measure and report progress regularly. Within the sector aid architecture, there ought to be an organization that would measure reading fluency every three or four years, like other international comparative tests.

1 Introduction

Reading fluency outcomes are very pertinent to the mission of the Education for All Fast Track Initiative (EFA FTI). Thus in 2009, the FTI Secretariat was empowered by its Board to promote basic skills acquisition among the 71 countries it works with. For this reason, EFA FTI revised the indicative framework of indicators to include general early-grade and end-of-cycle student performance goals (see following sections). It has also embarked on an effort to ensure consistent monitoring of early-grade reading indicators and appropriate remediation.

To compose quality standards and their rationale, the reading-related studies were extensively reviewed. Also several users, donor staff, and researchers were consulted, whose questions and insights were taken into account.[1]

This document offers the findings of a survey on reading fluency studies as well as the background research on the concepts that have been used. The main document centers on the former, whereas the annexes present the background research supporting reading fluency standards, psychometric issues, cross-linguistic comparisons, and word-counting issues.

1.5 Why Are Oral Reading Fluency Tests Important for Instruction and Monitoring?

All countries give paper-and-pencil tests that in some ways measure reading skills; some pertain to school entrance or leaving examinations, while others are standardized achievement tests developed for various countries or for the purposes of international comparisons (Annex 1). Paper and pencil tests do not easily detect whether students can read in the early grades (Annex 1). Particularly in the case of multiple-choice tests, it is unclear whether students can read; it is possible to score 25 percent just by guessing. Therefore, standardized achievement tests in lower-income countries are often administered in grade 4 or later, when most students in those grades know how to read. As a result, the illiteracy rates in the early, crucial grades are not detected until it is too late to intervene easily or prevent dropout.

Oral reading fluency tests can quickly and easily assess this skill in grades 1-2. Reading achievement can be assessed by listening to students read a simple text for about one minute and asking them 3-5 questions that link facts together. (Fluency up to about 150 words per minute can be assessed orally in higher grades through a text of about 200 words and request for a summary; e.g. Espin and Deno 1993). The questions show whether the information items stayed long enough in working memory to begin the consolidation process into long-term memory. To understand a simple passage, given the capacity of short-term memory, a student should read a minimum of 45-60 words per minute. Research suggests that aside from certain exceptions, this standard is usable worldwide. (See research evidence in Annex 1). To progress in school work and attain this standard, students should be expected to "crack the code" by the end of grade 1 (reading grade 1 text at around 30 words correct per minute), and to read relatively fluently by the end of grade 2 at 45-60 words per minute.

[1] The author is grateful to Marialuisa Martelli and Gianluigi Zocolotti (University of Rome), Andreas Karatzas (Greek Ministry of Education), Matthew Kam (Carnegie-Mellon University), Marcia Davidson (University of Utah), Tracy Alloway (Durham University) and others for their advice.

Because reading fluency predicts subsequent academic performance, oral reading tests have been used for decades in Europe and the US, and various instruments have been developed (e.g. Brus and Voeten 1979), which take about 5minutes per child.[2] They enjoy continued use because they are easy to develop, administer, and score; results are easy to interpret. Texts are drawn directly from children's prescribed curriculum, so the tests have instructional and curricular validity for screening individual students as well as monitoring their progress (Deno, 1985; Fuchs & Fuchs, 1999; Fuchs, Fuchs, & Maxwell, 1988, Bonnet et al. 2001).

Due to problems faced by students who learn reading in English, oral fluency has received considerable attention in the U. S. One widely used test for diagnosing reading difficulties is DIBELS, the Dynamic Indicators of Basic Early Literacy Skills. In 2006, the Research Triangle Institute with USAID funding developed an internationally oriented instrument on the basis of DIBELS along with guidelines for standardized administration procedures. The result was the Early Grade Reading Assessment (EGRA).[3] EGRA is an individually administered test with 9 subtests that measure: print orientation, number of letters read per minute, common words, nonwords, phonemic awareness, connected passage of one minute, 5-7 comprehension questions, listening comprehension, and a one-line writing sample. The full EGRA battery takes about 17 minutes per child and is recommended for at least 400 students per grade in order to produce relatively reliable outcomes. In languages with consistent orthographies less testing may suffice (Share 2008; Georgiou et al. 2009).[4] Thus, the original one-minute tests with comprehension questions and letter identification may be all that is necessary for reading fluency assessment (Annex 1).

The psychometric properties of reading fluency tests suggest that they can be effective monitoring instruments. The reports produced thus far in various countries consistently show strong validity and reliability indicators. All studies that calculated internal consistency, report Cronbach alpha values above 0.9 [5] (e.g. Guyana, Kenya-Malindi, Haiti, Honduras). There is also strong evidence of concurrent validity; in Peru, a fluency test was shown to have a strong linear correlation between with a Peruvian census written test (0.56 at the school level, 0.47 at the student level; Kudo and Bazan 2008). Reading speed and comprehension often show correlations in the 0.8-0.9 range (e.g. RTI 2010b), thus validating the short-term memory rationale. (See Annex 1.) On the other hand, reading tests have a high variance, given normal student variability within schools. Performance differences between readers and non-readers are also large. More attention must be given to the consequences of the large variance, and statistical treatments may be considered.

[2] Researchers using reading speed tests include Steve Walter of Summer Institute of Linguistics Graduate School, Save the Children, Timothy Rasinski (2003), and others.

[3] Crouch and Gove 2009. An expert consultation was held at the World Bank premises in 2006, and in 2007, 8 early-grade reading studies were funded by World Bank, USAID and directly by countries. In 2008 there were 11 additional studies and a March workshop at the World Bank headquarters with 200 participants. In 2009, 27 additional studies were carried out in Latin America and Africa with EGRA or EGRA-like instruments.

[4] DIBELS has multiple subtests only in grades 1-2, and thereafter tests speed and comprehension. Even in these, grades, reading fluency rate is the most important predictor of future academic performance, and the test can be administered in 2-3 minutes (Riedel 2007). Listening comprehension seems to be weakly related to reading comprehension as shown by studies in Greek and Finnish (Diakidou et al. 2005; Müller and Brady 2001).

[5] The high internal consistency index is partly due to the fact that reading fluency is a speeded test.

Reading fluency tests could be thought of as the school entry PISA counterpart that predicts how well students are likely to learn from books. But the text choice, sampling, and administration procedures have been more informal than international comparison tests such as PIRLS, PASEC, SACMEQ (Wagner 2011).[6] Thus taking oral reading fluency to international scale has revealed some weaknesses. Certainly, many concepts can be specified better, but then these tests may become too complex to be sustainable. Can reading fluency tests remain simple enough for FTI partner countries to use easily and still give reliable monitoring information across time and countries? Time will tell.

2 The Incidence of Reading Fluency Tests in EFA FTI Partner Countries

How many reading fluency studies have been carried out in EFA FTI countries and what have been the results? One task of this study was to find out. In early 2010, the literature was searched, researchers and staff of various development agencies were contacted and asked about status and about available datasets. Also, the task managers of the World Bank were asked whether the countries of their responsibility were planning reading assessments. Also, efforts were made to suggest such assessments in the next 2-3 years.

At least 50 reading fluency studies had been done worldwide by February 2010. Many studies involved from 800 to about 3000 students, but few have collected nationally representative data. Most focused on specific regions or excluded remote areas, and a few involved small convenience samples. Of the studies, many involved EGRA or similarly detailed instruments, while others involved just passage reading and comprehension questions. One or more studies have been carried out nearly half of 71 FTI-eligible countries, but some pertain to countries that have higher incomes, such as Chile, Mexico, Paraguay, Uruguay, Spain (Andalusia), Philippines (Annex 1). Some studies focused on just one grade, and different single-grade studies may exist in one country with samples that are not comparable.

About 30 reports were obtained and studied to extract conclusions. With this information, a database was set up. For the FTI-eligible countries, information shows that around April 2010 (Table 1):

- In 29 of the 71 FTI-eligible countries, one or more studies had been carried out; most were EGRA or EGRA-like instruments, but a few were regional or informal studies conducted by researchers.
- In seven countries there were multiple studies.
- In three countries, data were being collected in early 2010,
- In 14 other countries, studies were being actively planned in 2010.
- For 28 countries there were no reading fluency assessments or active plans.

Table 1: FTI – Reading Fluency Studies in Eligible Countries (as of April 2010)

[6] SACMEQ is Programme d'analyse des systèmes éducatifs de la CONFEMEN; PIRLS is Progress in International Reading Literacy Study, and SACMEQ is Southern and Eastern Africa Consortium for Monitoring Educational Quality.

Countries	Numbers	Percentages	Actions, Outcomes
All countries in FTI database	71	100%	19 of the country results are roughly comparable
Of which fragile states	25	35%	9 countries with studies or data collection underway
Reading fluency studies conducted	29 (including 3 6 regional, 1 informal)	41%	randomized trials of treatments underway (Liberia, Mali, Niger-Plan); NGO initiatives in India, Bangladesh
Reading fluency studies planned	14	20%	Most planned for 2010
Reading fluency studies NOT planned in early 2010	28	39%	Ongoing discussions with at least two
Endorsed	41 (April 2010)	100%	16 of the country results are roughly comparable
Of which fragile states	7	17%	4 countries with studies or data collection underway
Reading fluency studies conducted	17 (5 in fragile states); 3 regional or informal	41%	randomized trials of treatments underway (Liberia, Mali, Niger-Plan)
Reading fluency studies planned	6 planned (1 fragile state) for 2010	15%	Most planned for 2010
Reading fluency studies NOT planned in early 2010	14 (5 fragile states)	37%	Ongoing discussions with at least one
Not endorsed	30	100%	4 country results are roughly comparable
Of which fragile states	18	60%	5 countries with studies or data collection underway
Reading fluency studies conducted	5 (including 4 regional)	17%	NGO initiatives in India, Bangladesh
Reading fluency studies planned	9	30%	Most planned for 2010
Reading fluency studies NOT planned in early 2010	14	47%	Ongoing discussions with at least one

Specifically among the countries endorsed by EFA FTI at some point, 17 had some reading studies, 6 are planning them, and 14 have no plans. The latter include countries like Albania, Moldova, Georgia, and Mongolia, where early-grade reading fluency has not been identified as a problem.

This substantial number of studies suggests that costs and logistics are manageable. In many, RTI has used EGRA, but also individual researchers and NGOs with more modest means have been able to give fluency tests. In countries like Tonga, where the work was carried out locally, the study preparation and data collection took place in a few weeks; in Cambodia, 40 schools with about 2400 students in six grades were assessed within a month with a reduced EGRA instrument. Larger samples collected for the purposes of monitoring of various interventions may require more time and resources, and it will be seen how sustainable the process will prove in the long run.

The reports of the studies have been of varying quality, and they present varying aspects of the data. Most studies involved grades 2-4, but a few involved grades 1 and 6-8 (Annex 2 Table 2). Some studies used the same text for all grades, while others used a higher-level text for higher grades. Some studies do not mention the time of the year when tests were administered or how schools were sampled. Sample sizes of EGRA have ranged from 800 children (in Kenya) to up to about 18,000 (in India), but some other studies had as few as 25 students. The focus on regional and more accessible areas would suggest that any conclusions *overestimate*, the populations' ability to read.

Data analysis is also of varying quality. Sometimes the number or percentage of non-readers is mentioned, and other times it is not, and in some it was uncertain whether the average words per minute included the non-readers. Sometimes the percentage of students arriving at a criterion was mentioned, other times it was not. The reports may focus on gender- age- or performance-related breakdowns, so sometimes the overall averages are shown and need to be extracted from graphs or computed. (Some oral fluency studies are not timed and were not considered, as the Aser studies in India, Pakistan, Kenya, and Tanzania; Annex 1.) Most reports discuss the reading fluency with respect to a benchmark of 60 words per minute, while a few (Honduras, Gambia) placed the numbers in the context of norms developed in Latin America and the US. It would have been useful to show percentages of students attaining various fluency rates above and below a criterion, but such data are available from only about four countries.

Most tests have been done with languages that use the Latin script. (A study has been done in Egypt, and results of studies done in Pakistan have not been disclosed.) The common testing instruments are often included in the reports. Typically texts consist of a text of 60-65 words and a set of 3-7 questions (most commonly 5). Because they pertain to beginner subject matter, they seem uniformly simple in the various languages. Sometimes they are presented as first-grade and sometimes as second-grade texts, while other times their level is unclear.

2.1 What are the Overall Findings of Various Studies?

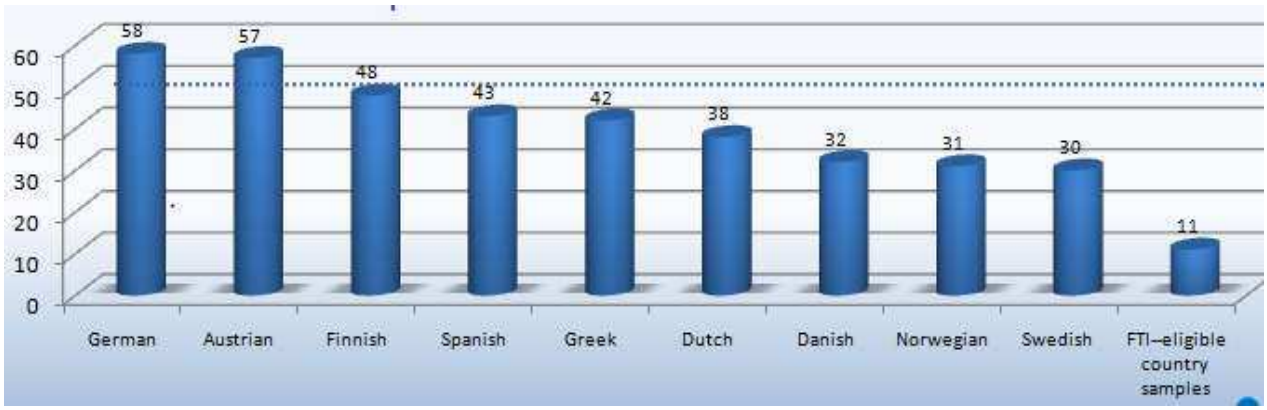
After studying the various reports, it became possible to abstract data from 17 countries that were approximately on a similar metric. These countries are *not* a representative sample of FTI-eligible countries; they merely have reported words per minute and comprehension at various grades, averaging readers and non-readers. (Sometimes comprehension is only calculated using the scores of the readers, and this point is not always clarified in the reports.)

As mentioned earlier, direct comparison among results would be inadvisable, due to differences in sampling methods, language, script, readability, and difficulty of comprehension items. Also, the percentage of non-readers in the sample varies a great deal, from 97 to 3 percent. Averages are not weighted. Nevertheless, they show trends consistent with learning research and transmit a coherent message (Table 2).

Table 2: Rough averages abstracted from reading fluency studies

Grade	Average words per minute	Average comprehension % questions answered	Average % non-readers (if available)	Number of countries where these grades were tested
1	12	23	53	8
2	23	32	35	13
3	38	41	24	13
4	62	72	10	5
5	70	63	12	3
6	56	73	0	2

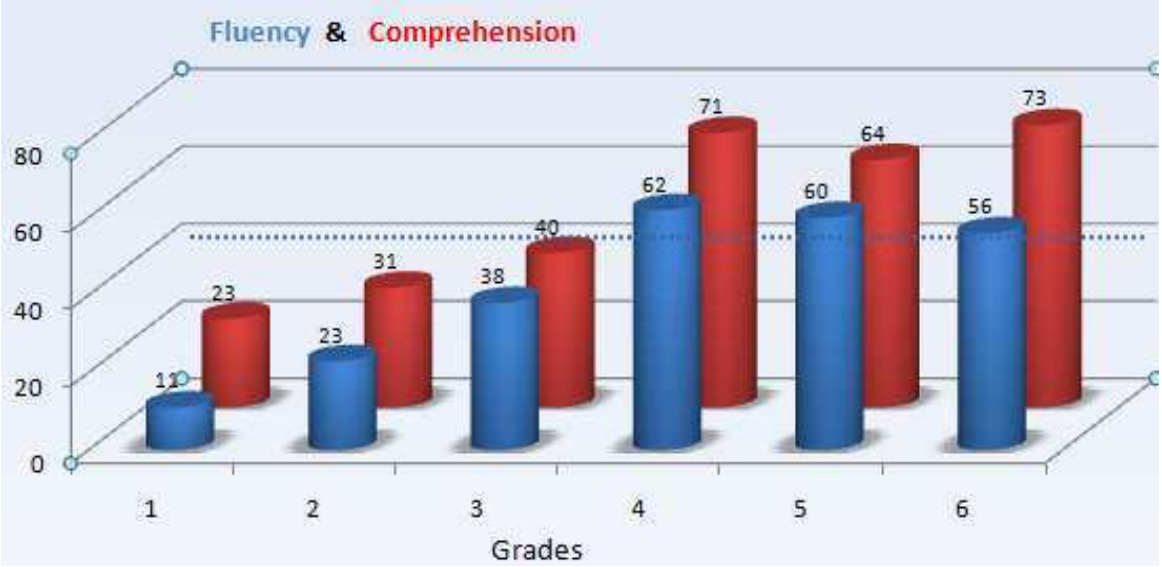
Figure 1: Fluency of FTI Partner countries in comparison to European countries at the end of grade 1



Source: Seymour et al. 2003

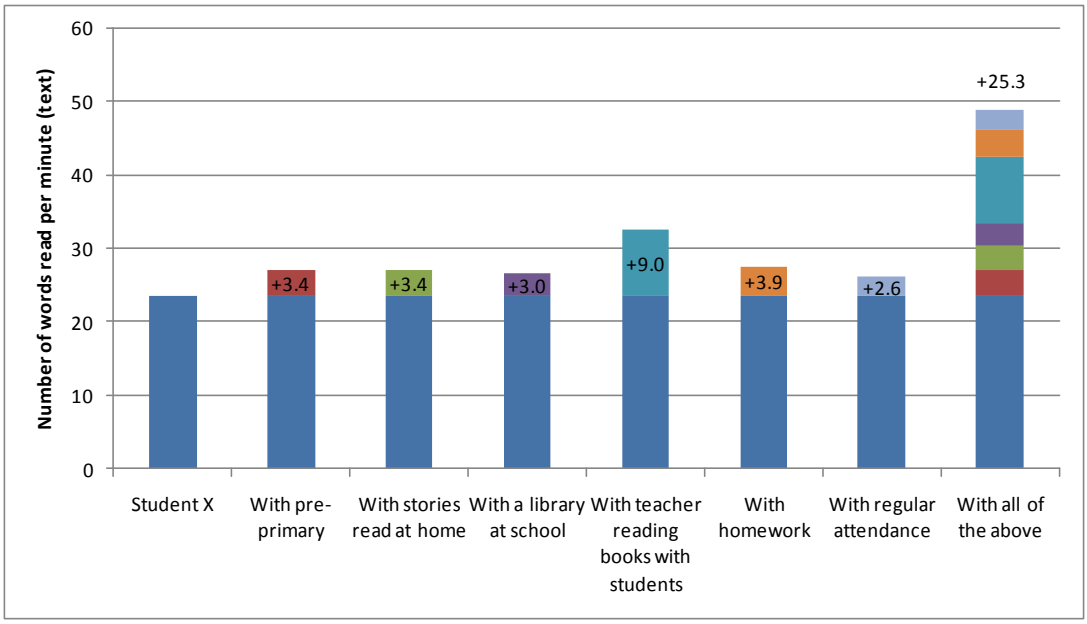
Overall, the data produced on these 17 countries do not approximate 30 words per minute in grade 1 or 45-60 words in grade 2. The only exception is a sample of larger schools averaging 67 words per minute in Nicaragua, a country that has greatly emphasized literacy since the Sandinista rule. Students in higher grades also read slowly, though they tend to be the better students who have not dropped out. US and Latin American norms would suggest a speed of 109-139 words for grade 5, but the two countries which offer this grade show only 60. Comprehension percentages are similarly low though they increase across grades as speed increases. The percentages suggest that at least the younger children who were tested made little sense of the text; reports show that significant numbers of students in grades 1-2 did not even know the values of letters.

Figure 2: Average fluency and comprehension rates in 17 selected FTI partner countries



In some countries, students were tested in official languages that have complex spelling rules (English and French) rather than the local languages of a country (e.g. Zerma in Niger). Also in Kenya some students were tested in both local and official languages. Remarkably, local language did not seem to give students much decoding advantage. When languages use the same script, students who learn to read in one language can apply the rules to another. Vocabulary limitations may reduce comprehension scores, and this was evident in some countries, such as Senegal. But overall, the research showed a strong relationship between fluent reading and comprehension, thus confirming predictions of memory research. Thus, the primary determinant may be whether students have been taught to read and practiced sufficiently. Under such circumstances they may read both languages, though vocabulary limitations my reduce comprehension scores.

Figure 3: Reading fluency measures and extracurricular activities in Timor Leste



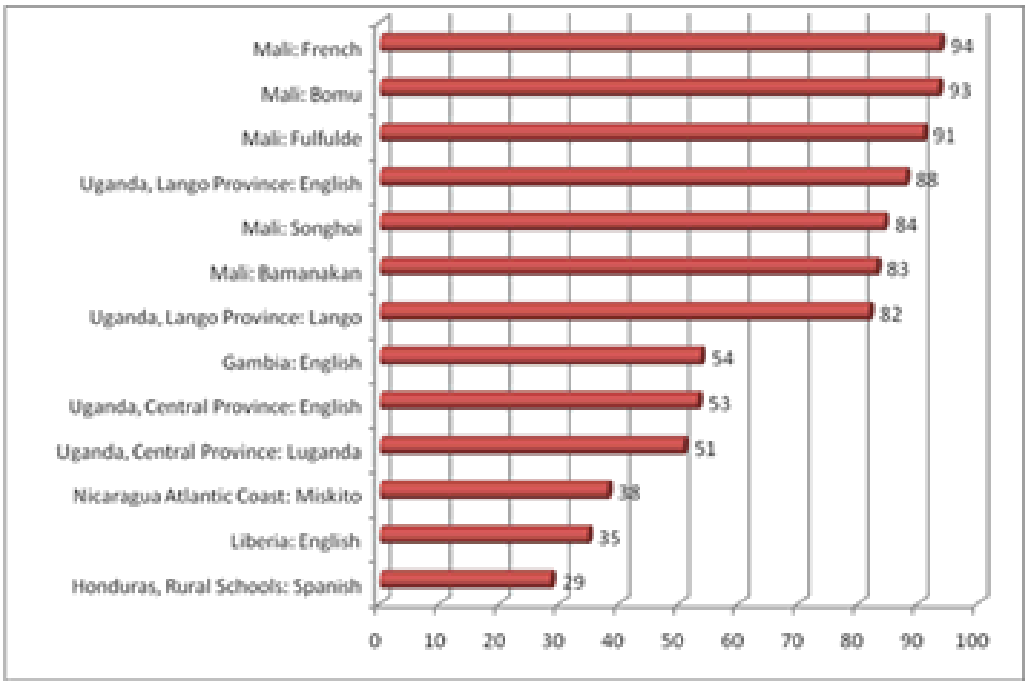
Behind gross averages in every country there is a lot variability. Some countries have several schools that perform very well, while other countries have few high performers. For example, in some schools sampled by the 2007 Kenyan study, only 7-9 percent of Standard 2-3 students read at 45 words per minute and could answer 3 of 5 comprehension questions; in others, 31 percent of standard 2 students met this low-level criterion at the end of the year.[7] In the Senegalese sample, performance appeared to be better; 11 percent of second graders and 48 percent of third graders tested in French read at 45 words per minute (with comprehension rates of 62 and 66 percent respectively). Also 28 percent of third graders tested in Wolof read at 45 words per minute, and scored 77 percent in comprehension questions. The study noted that many of the students who met reading criteria tended to have books at home and/or had attended preschool.[8]

Overall, these rough data suggest that **probably no FTI country would meet the qualitative indicators of the FTI indicative framework, either in grade 2 or at the end of primary education**. Even the countries that have no tests yet (e.g. Angola) have service delivery conditions that would predict outcomes similar to those of Mozambique (IEG 2007). Exceptions may be the lower-middle-income countries that benefited from FTI (Moldova, Albania). The limited data which exist suggest initial success with literacy but limited progress later on. For example, an Albanian sample of about 1416 students in the beginning of grade 5 read on average 90 words per minute. And only 45-77% of the students answered all comprehension questions correctly (Llambiri 2004).

[7] 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.

[8] Gove 2009b. Beyond Access: Effective Reading for All and the Early Grade Reading Assessment. Presentation at the World Bank Institute, April 30, 2009, RTI.

Figure 4: Percent of grade 2 students who could not read a single word of a simple paragraph, 2008-2009



Source: Results of end of grade two early grades reading assessments. Reports for each country available at www.eddataglobal.org

2.2 What Are Likely Reasons for the Low Student Performance?

Most reports do not extensively discuss the reasons for student performance, and some point to socioeconomic variables that cannot be easily improved by schools. One USAID-financed study (EQUIP2), however, investigated instructional issues in greater detail. In 2008, the Academy of Educational Development (AED) took measures not only of reading but also of instructional time use and textbook availability. The study involved Guatemala, Ethiopia, Nepal, and Honduras.

Table 3: Summary Opportunity to Learn Indicators, sample averages

	1	2	3	4	5	6	7	8	9	10	11	12
Country	% Open	Teacher attendance rate	Student attendance rate	% of day used	% time on task	% Instructional days	% w/ text	% obs using text	% obs read	oral read wpm	Class size	# supp visits (yr)

						equiv						
Guatemala	97%	88%	92%	72%	59	33%	63%	3%	11%	46	27	7
Honduras	93%	94%	97%	82%	56	30%	58%[9]	22%	21%	73	29	5
Ethiopia	93%	89%	97%		41%	33%	83%	4%	3%	18	44	18
Nepal	90%	91%		92%	60%	45%	84%	14%	12%	26	44	3

Note: data in EQUIP2 Study (Table III draft), AED 2009.

All four countries have above-average total intended hours of instruction when compared to the regional averages of 789 (Latin America), 665 (South and West Asia), and 809 (sub-Saharan Africa) respectively for grades 1-3.[10] However, schools in all four countries were found to use less than half of students' opportunity to learn. On average about 56 out of 180 days were used for instruction in Guatemala, 60 out of 200 in Honduras, 69 out of 203 in Ethiopia, and 87 out of 190 in Nepal. Also instructional time in class (time on task) was limited; it ranged from 60 percent in Nepal to 41 percent in Ethiopia. Furthermore, less than 12 percent of the time was spent actually reading overall, and this included reading any materials (including from the blackboard): 21 percent in Honduras, 12 percent in Nepal, 11 percent in Guatemala, and only 3 percent in Ethiopia. Textbooks were part of the problem; 63-84 percent of students had them, but in some schools textbooks often sat on shelves. Obviously, opportunity to practice reading should include more than such a small fraction of each school day; in a five-hour school day, 3 percent of the time would equate to 9 minutes of reading per day. To automatize reading extensive instruction and practice with books is needed, and students in this sample did not get it.

The AED study assessed oral reading fluency in the third grade. The averages in Ethiopia and Nepal (18 and 26 words per minute (wpm) respectively) were affected by the significant percentages of third-grade students who could not read at all: 44 percent in Nepal and 36 percent in Ethiopia. In Nepal only 26 percent read at 41-60 wpm, and in Ethiopia (Afar language) few read more than 45 wpm. The visual complexity of the syllabic scripts in these countries may create delays, given the limited instructional time. By contrast, in Honduras, more time was devoted to reading practice, and this Honduran sample averaged 73 wpm. (The EGRA study gave similar results.) Over 60 percent of the Honduran children could read at least 70 wpm, and 35 percent could read above 90 wpm. Regression analyses were not carried out, but a visual inspection of the data suggests that time for instruction and practice is a primary

[9] Grade 3 only

[10] Benavot 2004, cited in AED 2009. The official school year in Guatemala is 180 days (900 intended instructional hours), in Honduras 200 days (1050 intended instructional hours), in Ethiopia 203 days (812 intended instructional hours), and 192 days (1152 intended instructional hours) in Nepal.

variable that affects reading fluency. These findings resonate with earlier instructional time and reading research (Abadzi 2007). The more time students spend reading, the more practice they get and the more fluent they become (Table 4).

Table 4: The number of words read daily multiplies into many words per year and pays off

Test Score – Percentile	Minutes Reading Per day	Words Read per Year
98	67.3	4,733,000
90	33.4	2,358,000
70	16.9	1,168,000
50	92	601,000
30	4.3	251,000
10	1	51,000
2	0	-

Source: Anderson, Fielding, and Wilson 1988

2.3 What Actions Have Various Countries Undertaken on the Basis of Reading Fluency Tests?

In many countries, the reading fluency results were the subject of considerable communication. RTI and large NGOs conducted workshops and urged follow-up. The Gambian government sought innovative English programs and later decided to institute local-language instruction. Aside from the Gambia, however, few governments seem to have taken actions to improve instruction. To spur progress, RTI and some other organizations have worked to find funding and to convince NGOs to pilot evidence-based reading programs, often in local languages.

Donor support has been instrumental in improving reading fluency in countries where this goal was specifically targeted (Figure 8). Some examples are below.

Mali and Niger. Over 83 percent of second grades and 63 percent of 4th graders could not read a single word. The EGRA results showed that only 2.2 percent of students in Mali and almost no students in Niger could read at 45 words per minute. The Hewlett Foundation financed certain NGOs to implement the Systematic Method for Reading Success (SMRS), a three-month instruction program for grade 1 classrooms in 25 primary schools in Mali and in 16 second chance schools of Niger (ages 9-15 years). In Mali it was taught in Bamanankan and in Niger in Zerma.

Students received 30 minutes of reading instruction daily built around two booklets of 40 lessons each. These consisted of:

- Phonological awareness (recognition of sounds that make up words);
- Phonics; initial introduction of the name and sound of each letter, then blending letters into words.

- Sight words (high frequency words to be recognized on sight)
- Words of only 2-3 letters; Use of pictures in lieu of longer words or letters not taught.
- Listening to teacher read aloud followed by vocabulary and comprehension questions
- Re-reading successfully at students' own pace

Figure 4: Reading fluency test in Mali



EGRA testing, 2009

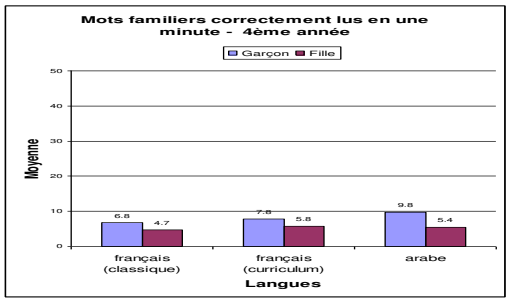
Figure 5: Systematic Method for Reading Success in Mali

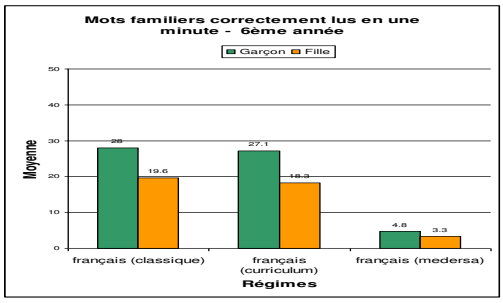


Pictures are substituted for longer words that students have not yet studied

The method has complexities such as sight words or short words that are more appropriate for English (Share 2008). It also requires the production and teacher training on many sketches, whose processing by beginning students is unclear. However, practice and good time use may offset the complex tasks. At the end of the instructional period (approximately 6 months in Mali and 4 months in Niger), the Ministry of Education randomly sampled six SMRS schools (104 students) and six national curriculum or bilingual schools (121 students). In Mali, after about four months of instruction, 49 percent of students could read 21 or more letters, versus 2 percent in national schools after a full year of instruction; (baseline had been almost zero). Also, 42 percent of SMRS students could read more than 50 percent of a word list compared to 2 percent of students in national schools; 89.7 percent of the students in the SMRS schools could read relevant text fluently with good comprehension, compared to 41.1 percent of the students in national curriculum (or bilingual) schools after a full year of instruction. In Niger more than 2/3 of students tested were able to correctly identify the Zarma alphabet, 45 percent recognized commonly utilized words and 22 percent could read and accurately comprehend a text (Plan International 2008; Alfari 2009). To some extent the efforts were compromised by strikes and limited teacher training, but they established that better reading outcomes are possible in these very low-performing countries.

Figure 6 and 7: Outcomes of a reading fluency pilot in Mali





Mali – Plan International: grades 4 and 6, familiar words read correctly per minute

Liberia. A phonics-based, scripted reading program of about 90 lessons was developed for English foundation skills in grades 2-3. (It was financed by USAID and administered by RTI.) Many publicly available materials were downloaded from the internet and printed at no cost. A randomized trial in 2010 compared the fully specified instructional approach to a control group and a "light" treatment consisting of parental communication of results. At midterm, and despite only 3.5 months of intervention, the full treatment group showed considerable fluency gains. Compared against baseline, full treatment children increased the number of words read correctly by 51.2 percent, while "light treatment" schools increased by 28.9 percent. Thus, full treatment schools increased their number of words read from 19.4 to 29.5 words per minute, while "light treatment" schools increased from 21.0 to 27.1. So, at mid-term, children in full treatment schools were reading 7.2 words more per minute than those in control (29.5 compared to 21.0). The gap was also large for light treatment schools, with a difference of 6.1 words (27.1 compared to 21.0; Piper, RTI draft report, October 2009). At the end of the year, "light treatment" outcomes did not significantly differ from control schools, but full-treatment schools clearly had superior outcomes.

Kenya. Lesson plans were developed to link teacher daily practice to national curriculum goals. Teachers were trained to focus teaching on scope (what), sequence (when) and instructional model (how). They practiced phonological awareness, phonics, reading fluency, comprehension and vocabulary. They used flashcards, word charts and puzzles, along with books. [11] Teachers received monthly visits for pedagogical support and they conducted continuous assessment. Randomized field trials showed an 80 percent increase over baseline in most reading tasks after 8 months. Non-readers dropped from 54 to 34 percent in the treatment sample. However, control schools also improved their performance, possibly because they were located near experimental schools, and the teachers became interested in the methods (RTI-USAID 2008).

[11] USAID's Centers for Excellence in Teacher Training (CETT) has a professional development model for early literacy in grade 1-3 classrooms. The program uses classroom coaching and scripted instruction for teachers in several countries in Latin America. (www.readingforallchildren.org/)

India-Pratham NGO. Around 2005, Pratham became widely known because it developed a 45-lesson program, to be delivered in 6 weeks, consisting of two hours of reading and one hour of math daily. Also tutoring activities were aimed at teaching Indian scripts as well as English to children who had reached Grades 3 and 4 without mastering Grade 1 and 2 competencies. For half of each four-hour school day, children who had been identified by the teacher as lagging behind their peers were pulled out of the regular class and sent for remedial instruction in groups of 20. The tutor (*balsakhi*, a child's friend) was typically a young woman with secondary education hired from the local community, who received two weeks of training and was paid 500-750 rupees (about US$10-15) per month. Randomized trials showed moderate results. In the second year (when everyone was familiar with the program), 8 to 13 percent more of the particularly marginalized children were able to carry out the skills mentioned above. However, the intervention was considered cost-effective. The program cost just US$2.25 per year for every child who received the remedial assistance compared to the average expenditure on education in India of US$78 per year (J-PAL 2006).

Also in India, *'remedial reading camps'* have achieved impressive improvements in the Jaunpur district of Uttar Pradesh, where 60 percent of 7- to 14-year-olds could not read a first-grade story. A randomized evaluation examined the impact of three interventions: informing the community about the status of local schools, outlining to village education committees their roles and responsibilities; training community members to assess children's learning and presenting these findings at village meetings; training local volunteers in simple techniques for teaching children to read and introducing reading classes after school. The evaluation found that *information-sharing did not improve reading skills but that extra classes after school had a big impact*. The evaluation showed that it was possible to improve fluency through an additional instruction of two hours a day over three months (Banerjee et al. 2008 in the 2010 Global Monitoring Report, p. 113).

Bangladesh. Following reading tests that showed modest outcomes, the BRAC NGO conducted a 4-week dialogic reading intervention with the intention of increasing preschoolers' expressive vocabulary. Eighty rural preschoolers randomly selected from five preschools participated in the 4-week program. Their expressive vocabulary, measured in terms of definitions, was tested on 170 challenging words before and after the program and compared with that of control children who participated in the regular language program. Both groups were read eight children's storybooks in Bangla with illustrations; but the dialogic reading teacher was given a set of ''wh'' and definitional questions to enhance children's verbal participation. In four weeks, the mean vocabulary scores of dialogic program children increased from 26% to 54% whereas the control children remained at the same level (BRAC 2009). Vocabulary is very important for comprehension, but it improves fluency mainly for deep orthographies (Share 2008). The role of improved vocabulary in the transparent Bangla orthography is unclear.

3 How to Improve Reading Fluency: Guidance from Pilot Studies and Research

What is needed to achieve minimal reading fluency and text comprehension? The prerequisite for reducing reaction time to letters and activating the visual word form is practice (Annex 1). In grade 2, it is possible for the poor to catch up during the regular school day, evenings, and

vacations. The above examples and other research show that quick and sizeable gains are possible. Research suggests that low-income students would benefit from:

- Mother tongue instruction or at least support for second language learners, partly due to the facility of transparently spelled languages (Georgiou et al. 2009)
- Relatively short and defined modules for reading fundamentals, with specific speed and accuracy objectives (e.g. for 6 weeks or as needed for a language and script; Pratham methodology, proximal goals help teachers feel effective – Bandura 1997).
- Use of instructional time for reading and other curricular activities (Abadzi 2007);
- Scripted lesson plans and materials that teachers of limited education can easily follow; Intensive supervision and systematic class visits, informal reading assessments (Liberia and Mali pilots);
- Instruction through phonological awareness, synthetic phonics, decoding, analogies in letter combinations; an initial focus on visually simpler and more common patterns (Pelli et al. 2006).
- Corrective feedback (individual reading for a minute per day), offered to everyone systematically as reinforcement (Salamone and Correa 2002, Gee 2007). Use of the better students as monitors of others.
- Plentiful amounts of materials for students to practice and take home;
- Community awareness of reading needs and monitoring (as in Liberia; Crouch and Kodra 2008)
- Placement in the lower grades of teachers better able to control classes and maximize instructional time.

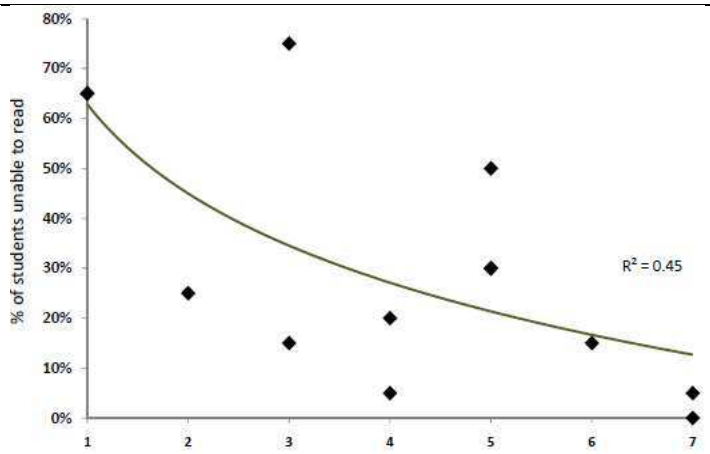
Communities can also monitor fluency, as was shown to be possible in India, Kenya, and Tanzania. Even illiterate parents can distinguish whether children read fluently. It is important, therefore, that parents learn to expect the acquisition of fluency in grades 1 and 2 and demand that schools prepare their children for this task. Public information videos (such as one developed in Perú) can help parents and educators understand these standards and their rationale.

Figure 8: Donor support matters

	Percentage of children unable to read by years of Save the Children support

Years of Support	Average Words Per Minute
7	31
7	20
6	20
5	23
5	23
5	25
4	15
4	19
3	15
3	5
2	41
new	7
new	13

Note: Each row represents one Save the Children-supported school.

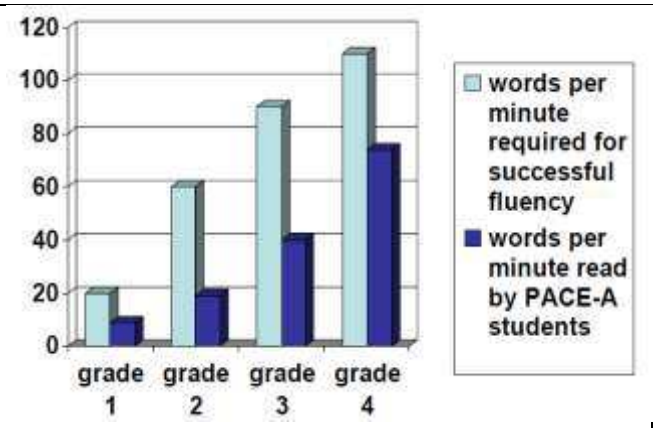


Source: Save the Children support of 13 schools in Ethiopia

Figure 9: Reading instruction and testing in Afghanistan

a





PACE-A program of Afghanistan

Although all emphasis has been on the early grades, reading speed must rise consistently through primary and secondary school. This means that students need reading materials in sufficient quantity for home and class study. Students in high-income countries may reach 150 words per minute in grades 4-6, but in poorer circumstances, students may permanently read below that level during secondary school or beyond. And fluency tests are not only necessary for the low-income countries. For the higher-income countries among the FTI partners (e.g. Albania or Moldova), students may perform well in the early grades but fail to attain the fluency

needed to read comfortably the textbooks of later grades. (Research in Albania points to "hidden dropout" resulting from such gradually deteriorating ability to follow the class; Llambiri 2004.) Therefore, vigilance is needed until older students become expert readers.

4 Measurement Issues for Future Monitoring

Worldwide there is an interest in doing assessments that are quicker and cheaper (Wagner 2003). The studies produced thus far in various countries consistently show strong validity and reliability indicators, and the tests have proved resilient to statistical and methodological mishaps. Reading fluency tests fill a gap in grades 1-3, when sample-based assessments are rarely given. Though exact costs have not been estimated, the number of existing studies suggests that costs and logistics are manageable. Nevertheless, they are not as technically sophisticated as the international comparison tests. Increasing the technical demands for reading fluency tests may create complexities that will reduce their appeal. Tradeoffs must be carefully considered.

With the information known at this time, it seems two or three of the EGRA subtests (letter and one-minute reading with comprehension questions) can relatively easily and inexpensively be used for country and school-level diagnosis, longitudinal monitoring, and evaluation of remedial actions. Specificity is needed to include the time of the year that fluency tests are given, calculations providing fluency and comprehension percentages by readers and non-readers, the percentages of students who could not read a single word. The reports could also clarify if letters have been assessed on the basis of sound or names, since these two may produce different results. Also, the psychometric qualities of comprehension questions should be better specified. For example, questions should merely sample the content of students' working memory rather than test linkages to previous knowledge. The exact procedure for obtaining comprehension data is important, because on some occasions measurement error may arise and create interpretation difficulties (e.g. Dowd 2011).

The use of regional or convenience samples was a good start, but these probably overestimated student achievement and are hard to use as baseline measures. For performance monitoring purposes, there is a need for more demographically representative samples. The sample sizes per grade recommended on the basis of confidence intervals have been manageable for implementation, but there is a need to understand empirically if the error variance generated from these can sufficiently track reading progress over time.

5 Potential Reading Benchmarks for Low-Income Countries

For future monitoring, efforts should be made to communicate a need for common procedures and metrics. Though languages and scripts create some variability, the universal characteristics of human cognition also set the ground for some uniformity. Several studies show that languages have common structures and conform to working memory capacity (e.g. Miestamo et al. 2008, p. 46). Common standards may help track performance improvement internationally and compare findings effectively.

Defining standards as research was being interpreted and reading studies conducted, has been a work in progress. In 2009, the EFA FTI Secretariat developed tentative minimal standards for the evaluation of the partnership performance and of country-level performance. The EFA FTI partner countries would be expected to establish effective reading programs and increase every year the percentage of students who attain the benchmarks. The 2009 benchmarks expected partner countries to determine numerical goals for the end of grade 2 and the end of primary (grade 5 or 6):

- early grade performance: % students who, after two years of primary school, demonstrate reading fluency and comprehension to 'read to learn'; and
- end of cycle performance: % students who are able to read with comprehension, according to their countries' curricular goals, by end of primary school.

Contacts with FTI partner countries revealed the need to make standards clearer and easier to communicate from central educational authorities to teachers and parents. All terms in the above goals required operational definitions. Educators ought to know how to estimate fluency, comprehension, and "reading to learn."[12] Decisionmakers also ought to have some criteria by which to decide what percentages of students could realistically attain the benchmarks every year, given interventions in schools. Of course benchmarks also ought to permit countries to set higher or lower fluency goals if desired.

The initial benchmarks were to be reviewed in 2010, so efforts to specify them further were made in 2010-11. The task proved more complex than expected due to a need for specification of multiple variables, limited technical knowledge among many staff in the education sector, and concerns regarding political reactions. As this document approached publication, final decisions had not been made.

Likely benchmarks would include statements aimed at defining words per minute as well as illiteracy (reading almost 0 words per minute). Learning research would be used as well as the studies which show satisfactory comprehension (80% correct responses) in many countries at 45-60 words per minute. (Annex 1). The use of 45 words per minute as a lower boundary would be close to U.S. and Latin American cutoff scores for identifying students at risk. In addition to specifying average speed and comprehension rates, it would be important to reduce the percentage of students who are illiterate, i.e. read 0 words per minute. [13]

For some languages, what constitutes words would have to would have to be specified given agglutinations, and some studies may be needed to determine language- or script-specific reading benchmarks (for example tonal languages and non-semitic languages written in unvoweled Arabic). However, humans seem to have a sense of syntax and what constitutes words, and there are various possible methods for this (Annex 1).

[12] "Reading to learn" requires higher than minimal reading speeds to permit more thinking time in working memory, so the grade 2 goal refers to the percentage of students who can carry out a grade 4 skill. (See Annex 1.)
[13] It has been difficult to define illiteracy as zero words per minute, since 1-2 words many also signify the same result. One means to do so may be to use the standard error of measurement of various reading tests, which is around 5 words per minute.

Illustrative benchmarks, shown below, are likely to refer to the above variables.

- the percentage of children who after two years of school read fluently enough to answer at least 80% of comprehension questions (about 45-60 words per minute according to research and existing data),
- the percentage of children reading fluently enough to answer at least 80% of comprehension questions from texts at their grade level (at least 90 words per minute) at the end of the primary cycle.
- % of children who are illiterate at the end of grade 2 (i.e. scoring 0-5 words per minute, special ed. excluded);
- % of children who are illiterate at the end of primary;(i.e. no children scoring 0-5 words per minute, special ed. excluded);

For the FTI evaluation framework, the benchmarks could refer to the % of countries on track to eliminate child illiteracy on the basis of:

- Reading at least 45 words per minute by the end of grade 2, unless a country specifies a different reading rate;
- Demonstrating satisfactory levels of comprehension (e.g. 80%) as measured by simple recall questions on grade 1-2 texts.
- Reading at least 90-120 words per minute by the end of primary, unless a country specifies a different reading rate
- Demonstrating satisfactory levels of comprehension (e.g. 80%) as measured by simple recall questions on grade-level texts.
- Eliminating cases where students read near zero words per minute in grade 2 or end of primary (special education excluded).

6 Conclusions and Future Prospects

The reading fluency studies located up to mid-2010 are heterogeneous and can be compared only roughly, but they offer a picture that seems consistent with various studies and observations. Data on 19 countries make it likely that on average, students in most FTI-eligible countries read at speeds that are too slow to permit sufficient comprehension at their grade level. Without fluency and comprehension students cannot “read to learn”, as they should be doing after grade at the latest.

To improve learning outcomes, students need instruction and practice on the relevant tasks. Textbook procurement has been a big problem in lower-income countries, but every effort must be made to give texts to every student for home use. To maximize reading time, a specific reading curriculum may be needed rather than reading instruction during ‘language arts’. FTI countries must focus the service delivery aspects of their plans to bring this about. FTI staff must also advise on these variables when they help countries formulate sector plans. To do so, they need more training on the instructional conditions that bring about fluency. Local Education Groups and Ministry staff may also receive training and information through audiovisual presentations, local and GDLN conferences, and e-learning courses. A program of measuring reading fluency at all

primary (and also secondary) grades may be instituted to monitor improvements and to ensure that by 2015, all who are in school know at the very least how to read.

EFA FTI will encourage its partners to fulfill the new quality-oriented goals of reading fluency. The large number of reading fluency tests carried out signifies that international consensus has been forming. However, the rationale needs to be disseminated more widely. Thus far, many of the studies have been managed by international consultant firms. Costs and sustainability concerns however, dictate that measurements must become country driven. Responsibility for could be given to a local organization with sufficient implementation and analytical capacity: pedagogical institutes, evaluation agencies, and education or psychology departments of universities. These may carry out the work inexpensively as well. In the short run, the new FTI funding vehicle as well as bilateral or other multilateral support could finance the activities.

EFA FTI partner countries should also develop a regular program for measuring reading fluency in all grades on a regular basis at the end of the cycle. National-level representative samples should be used, at a schedule that countries may determine (for example every two years). Different grades may be assessed different years and in different schools, but it will be important to obtain longitudinal comparable data. International comparison tests and public examinations could be given in tandem, and thus the results of these tests could be informed by students' reading performance.

It is reasonable to believe that low-income countries can achieve a significant surge in learning if they teach effectively their students to read in grades 1-2. The FTI Partner countries might focus on the service delivery aspects of their education plans to bring this about. Early reading instruction should use a language that children understand, with a transition into an official language (in countries where this is practiced) only after basic skills are mastered. Local education groups could support increased time on task, phonics instruction with phonological awareness, an abundance of reading materials, systematic monitoring of each child, instructional support to teachers with scripted lessons in early grades, and simple low-cost testing. Significant progress could be made in the first few months.

The EFA FTI partnership can help governments deliver improved service and instruction, and establish norms by which all countries measure and report progress regularly. Within the sector aid architecture, there ought to be an organization that would measure reading fluency every three or four years, like other international comparative tests. Examples to emulate would be the OECD PISA or the International Evaluation Association's TIMSS and PIRLS tests.

Annexes

Annex 1: Memory, Measurement, and Cross-Cultural-Comparability

Reading fluency and development of benchmarks require some knowledge of memory research and understanding of its implications. This domain of knowledge is rarely taught in faculties of education or economics. This annex explains in simple language the essential concepts, issues, and implications. 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?

Even among middle-income countries reading fluency in the early grades determines student achievement all the way to the university. 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. So, 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.[14] 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.

Worldwide, curricula prescribe that students must learn to read in grades 1-2 and perform various tasks related to reading content. Therefore fluency is expected even when this goal is not explicitly stated.[15] Research in European languages and middle-income environments

[14] 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).

[15] 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.

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.

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.

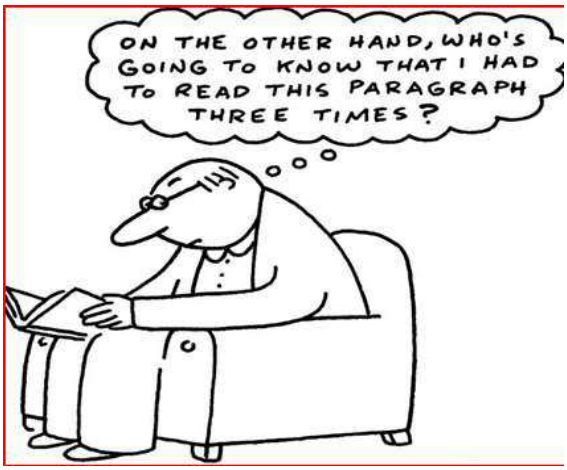
Educated people in high-income countries usually process text fast and within the capacity of their working memory, so its limitations are not obvious. But in low-income countries students may recognize text slowly, if at all, and they may have little knowledge to retrieve quickly. Thus, working memory 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. Illustratively, we must read a simple sentence of seven known words within 12 seconds in order to make sense of it. Dividing 7 words into 12 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).[16]

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

[16] 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 with shorter words are averaged. 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).

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.



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)

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.[17] 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.

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; Annex 1 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 12-second working memory model taught for decades in psychology courses has become more nuanced over the years. [18] At the same time, research has shown the increased

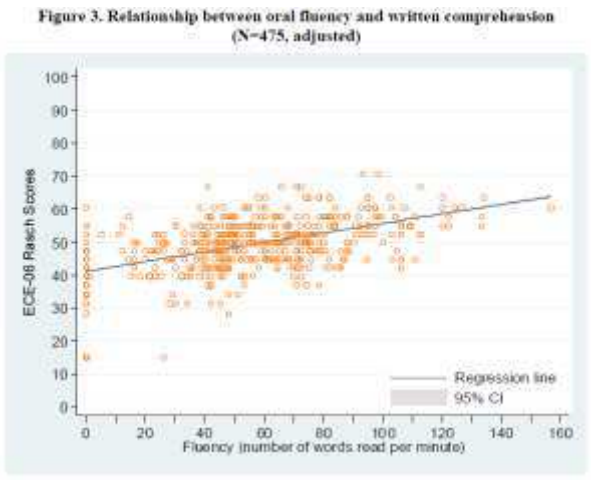
[17] Brain imaging methods, such as functional magnetic resonance imaging, are expensive, but event-related potentials could be easily used in low-income countries. The perceptual aspects of reading are also assessed by studying eye movements through computerized displays (e.g. Sereno and Rayner 2003).

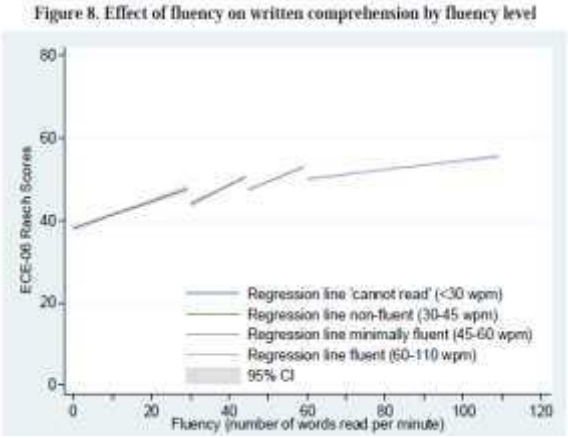
[18] The 7 items in working memory were based on the recall of English-language digits (Miller 1956), but neurological research suggests greater generalizeability (Miglore et al. 2008). The 12 seconds resulted from reading of trigrams (Peterson and Peterson 1959). Working memory studies have produced variable results, depending to some extent on the paradigms chosen. The older studies and models used protocols similar to fluency tests of beginning students and still seem appropriate due to greater ecological validity (e.g. Daneman and Carpenter 1980). By contrast, 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 at al. 2006). However, the ratio between storage capacity and time remains about the same. Overall, newer models of working memory emphasize the dynamic exchange between cognitive networks and the central executive function; large amounts of information can be retrieved in milliseconds, obscuring the size of the verbal store. But still at some point the capacity limits are surpassed, and cognitive overload may result.

commonalities among languages due to the underlying neurological mechanisms (Bolger et al. 2005).

It would be ideal if studies were carried out in all countries, languages, and writing systems in order to determine parameters of the visual word form activation and relate these to reading speed and comprehension. However, studies require considerable money and time and could divert donor funds from more important purposes. 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. Studies could be used sparingly, for example for tonal languages.

Annex 1 Figure 1 and 2: Relationships between fluency at comprehension in different countries





Peru, grade 3 in Spanish; oral fluency in comparison to a written comprehension test (Kudo and Bazan 2008)

Why does comprehension result from fluency?

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. Many studies have demonstrated a close relationship between speed and comprehension (review in Fuchs et al. 2001).[19] The

[19] 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

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).

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).[20]

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 (Valencia et al. 2010).[21] In lower-income countries, however, upper-primary students often read slowly, and the simple reading rate measure over one minute may provide sufficient information quickly and cheaply.

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 and short-term 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. [22]

---

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).

[20] 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).

[21] For students in the US who read more than about 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).

[22] 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

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.[23]

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).[24] 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).

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% of 100% comprehension scores, oral reading fluency levels need to be somewhere between 50 and 80

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").

[23] 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).

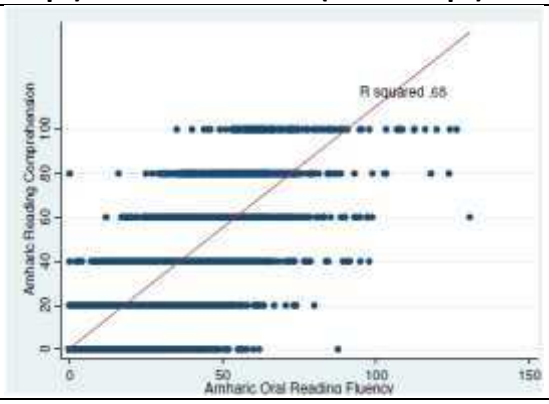
[24] A study of word lists showed that the Chinese who have short words could remember about nine and the Welsh who have long words could remember about five (Ellis 1980). However, sentences even out the variability because some expressions are longer than others.

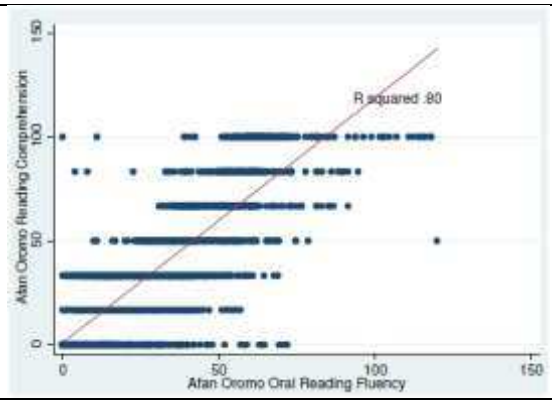
wpm (RTI 2010). Regressions showed that the range of oral reading fluency needed for 80% to 100% comprehension was between 53 wpm to 59 wpm for 80% comprehension of "shallow" items. For 100% comprehension, oral reading fluency scores were 60.0- 70.4 wpm.

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. unvoweled 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.

Annex 1 Figure 5 and 6: Oral Reading Fluency and Comprehension Rates for Amharic (Sabean script) and Afan Oromo (Latin script)





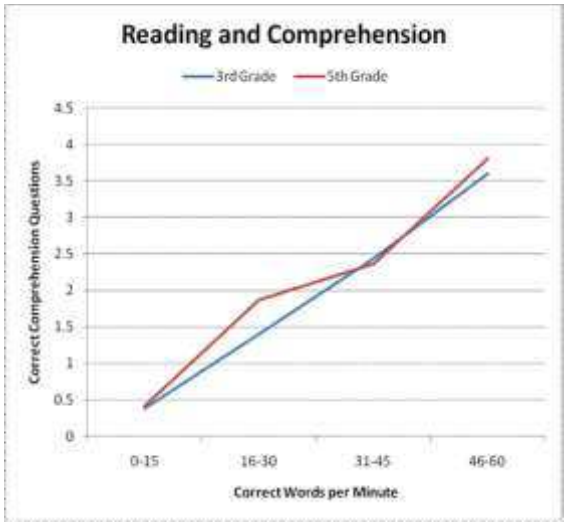
RTI 2010b, p. 41, Figure 22

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; Annex 1 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.

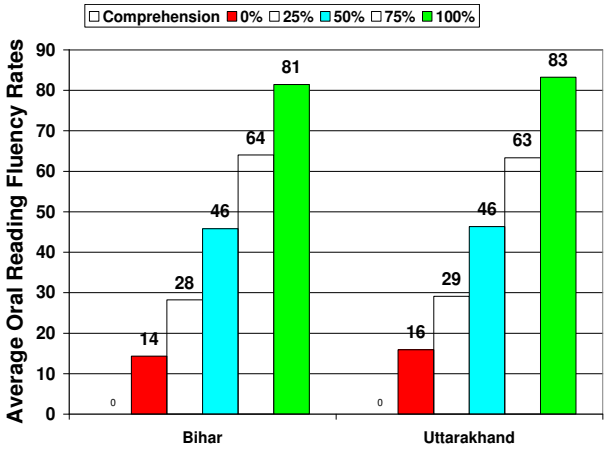
Annex1 Figure 3 and 4: The relation between fluency and comprehension

Ghana



To answer 80% of the questions correctly, students had to read more than 60 words per minute (RTI 2009)

Samples in two Indian states, grades 1-5



Second grade test in Devanagari script read by grades 1-5 (Pratham 2009*)*

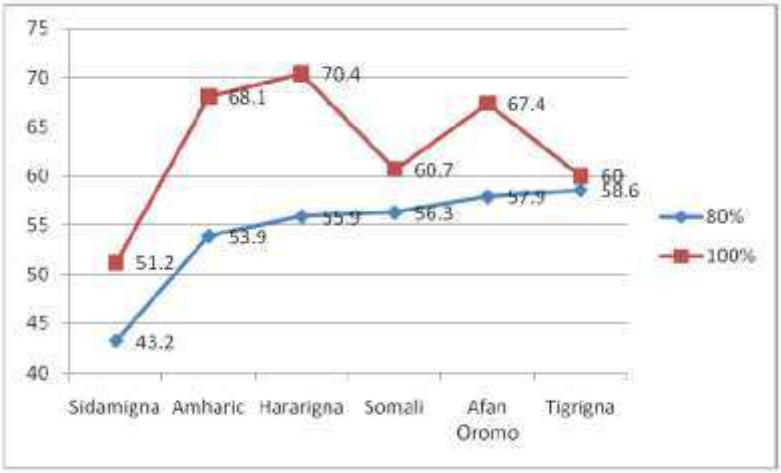
The "word length effect" is another relevant phenomenon (De Luca et. al 2008). Beginning readers, take longer to read longer words. The "word length effect" gradually diminishes in silent reading after fluency and is reduced in oral reading as speed increases. For example in Italy it disappears in grade 3 (Hulme et al. 1995; Zoccolotti et al. 2005; Martelli et al. 2009). [25] 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

[25] 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.)

and words per minute may result. But the significance of 2-3 words per minute is unclear, [26] 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.[27]

Annex 1 Figure 7. 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. The clearest point of cross-linguistic comparability is at the emergence of fluency. Regardless of the effort and time needed to acquire a script, students may hold comparable amounts of text in working memory after automaticity is achieved. Comparability is necessarily approximate, but it can be enhanced by comparing curricular goals of various countries and using readability formulas (as is done in international comparison tests).

[26] Due to the visual complexity and cognitive demands, adults may be reading unvoweled Arabic a bit more slowly (Eviatar, Ibrahim, and 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.)

[27] 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).

To count words per minute, an important issue is to determine what units constitute words in various languages. Different linguistic families 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. 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 ads 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. In conclusion, 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.

It can be argued that it is unnecessary to compare performance across languages or countries. However, it is important to discuss significant performance differences across countries or languages. Means must be determined for grouping them and deciding the magnitude of differences that might be considered worthwhile. For example, use of orthographic depth computations could be used (Yap et al. 2010; Annex 4). As discussed earlier, however, the benefits and costs of the added complexity would have to be assessed.

Some academics have objected to any setting of standards, which raise the prospect of pressures, falsification, and emphasis on speed to the neglect of comprehension (Wagner 2011). Some others have offered the opinion that stakeholders should formulate their own 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 (Annex 1 Table 5).

Defining and measuring comparative speed and comprehension

As mentioned above, many countries 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). Without efforts to create comparability, 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). Thus, there is a need to create conventions for rough comparisons to derive findings of broad applicability. This section briefly presents the research highlights related to this issue.

The question is how to specify indicators for this relationship, either from the speed or from the comprehension side. Both variables have complexities. From the speed side, conventions are needed to define words per minute across languages. From the comprehension side, psychometric comparability is a problem.

The correlations between speed and comprehension are high in the lower ranges of reading. So strictly speaking, the measurement of comprehension is redundant. 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, governments must have a sense regarding the extent to which students understand the language and can actually makes sense of the text, so the assessment of comprehension is important.

To demonstrate comprehension, students should be expected to demonstrate that they understood the gist of a text. Most appropriate would be short and questions that sample the passage of factual information into long-term memory (Georgiou et al. 2008). However, tests in various countries and within the same country have often used various types of questions. Some were simple recall questions, while others (as in Mexico) involved inference as well. Questions for grade 2 text would probably be relatively simple, but they have different psychometric properties and strictly speaking are not comparable.[28] And the need to retain material long enough in working memory would get confounded with vocabulary knowledge of non-native speakers. Normally, scaling and equating procedures are needed to compare test scores across countries. This would require pretesting and introduce complexity to tests meant to be simple.[29]

In principle it would be desirable to set a common metric of comprehension rather than count words per minute. For example, 80% comprehension scores could be expected across countries. But it has been hard to determine precisely what constitutes sufficient comprehension. In principle students should show that they understood 100% of a simple text,

[28] 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.

[29] 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.

but there is a need to allow for measurement error. Thus, many oral fluency tests put a comprehension benchmark as 80 percent.[30]

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.

Annex 1 Table 6: 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 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

[30] 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. Informal reading inventories tend to use 80% or 4 out of 5 correct answers to comprehension questions (Marcia Davidson, personal communication 4-29-2010). Due to serial effects of memory, questions querying items at the beginning and end of the text may achieve higher responses (Gupta et al. 2005).

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.[31]

Annex 1 Table 7: 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 disappears. As Annex Table 7 shows, Swahili passages may have 10-20 percent more syllables than English but still have 30 percent fewer countable words.

One potential means to compare reading rates more precisely may be to compare the number of linguistic 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

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

third graders read at a rate of 181 syllables per minute or 86 words per minute (P. Tressoldi, personal communication to M. Jukes).

Annex 1 Table 8: 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.

Also, more can be understood with respect to the effort needed to learn fluent reading in a certain language given orthographic depth. Letter to- phoneme number ratios provide a potential heuristic for scaling orthographic depth, clustering languages, and making predictions about processing demands. Annex 1 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.

Annex 1 Table 9: 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**
M	7.4/12.0	**0.72**	20.2/24.8	**0.87**	27.7/36.8	0.79
SD	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.

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 PASEC[32] for Francophone Africa and SACMEQ for Anglophone Africa,[33] 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.

[32] 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

[33] Southern and Eastern Africa Consortium for Monitoring Educational Quality

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 was 0.38 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

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 ahd 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 40 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.

Annex 1 Table 1: 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 of 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 (Annex 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 (Annex 1 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 (Annex Tables 3-5, Figure 1).

In **Chile**, the teacher-to-teacher network program of the Ministry of Education set goals for grades 1 and 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.

A **Cuban** researcher suggested 30 words per minute at the end of grade 1 as a reading speed for a "normal" child. (Perez Villar 1996).

A study from **Spain** reported averages for the first and second grades about 50–55 and about 75 words per minute respectively (Equipo de Orientación Educativa de Marbella 2003). Among the low-income Spanish-speakers in the U.S., a speed of just 30–60 words per minute in Spanish in Grades 1 and 2 is used as an index of disadvantage (de la Colina 2001). Similarly, a cutoff score of 40 words per minute places second graders in the U.S. at risk (Davidson and Towner 2005).

In **Mexico** norms were established in May 2010 after much deliberation and study of norms from other countries and with technical assistance by the World Bank. Students are expected to read 75 words per minute at the end of grade 2 (Annex 1 Figure 8).

Annex 1 Table 2: US Hasbrouck and Tindal norms, 2005

Grade	50th % ile	25th % ile	18th %ile [derived]	10th % ile
1	53	28	21	15
2	89	61	45	31
3	107	78	63	48
4	123	98	85	72
5	139	109	90	83
6	150	122	108	93
7	150	123	110	98
8	150	124	110	97

Annex 1 Table 3: Reading fluency norms from some medium and higher-income countries

Grade	Cuba 1996	Chile MOE 2005	Chile-Educando Juntos	USA (Hasbrouk & Tindal)	Paraguay 2005	Mexico 2006	Andalusia 2002
1	30	35	30	53	50	49	51.43
2	40	70	60	89	60-70	70	70.24
3	60	100	80	107	70-80	80	70.67
4	80	120	110	123	100-120	97	98.18
5	100	160	130	139	120	112	91.43
6	120-140+	200	160	150	120+	111	109.38

Annex 1 Table 4: Overall consistency but also variability of scores inherent in small samples

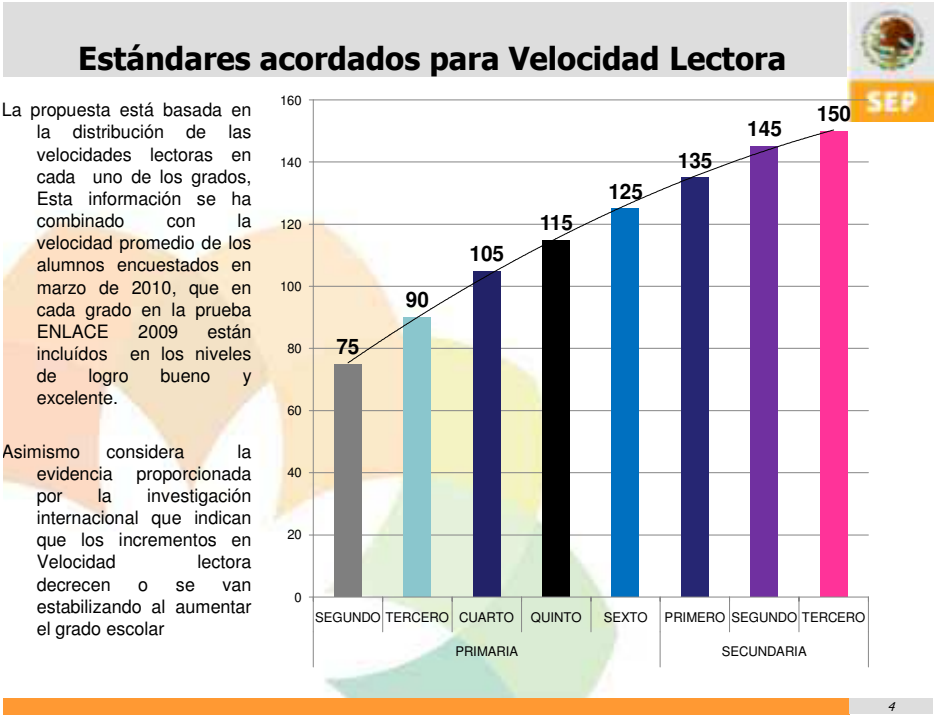
Means obtained	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
By Bisquera	40.70	70.71	101.4	120.22	136.18	152.88	167.44	177.70
General-1998	54.10	64.29	89.3	101.03	102.97	111.91	124.48	157.75
General-1999	59.28	77.75	95.97	108.67	122.73	119.54	157.63	157.75
2002	51.43	70.24	70.67	98.18	91.43	109.38		138.50

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

Annex 1 Table 5: 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

Annex 1 Figure 8: Mexico – Reading Speed Standards (May 2010)



Annex 2 Table 1: Database of countries and reading tests conducted

Country	FTI-endorsed	Status (April 2010)	Languages	Funding Agency	Implementing Agency	Scope (sample size when available)	Month	Year
Afghanistan	No	Completed	Dari, Pashto	USAID/Afghanistan	CARE with 3 partners (IRC, AKF, CRS)	1000 learners in PACE-A classes	Nov each year	2007, 2008, 2009
Albania	yes	selected testing	Albanian	Unicef	Education and Development NGO	1416 students beginning grade 5 in pilot schools where "hidden dropout" was being remedied		approx 2003
Angola	no	In planning	Portuguese	World Bank	World Bank/Consultants	In planning	Sept	2010
Argentina	ineligible	Completed	Spanish	Centro de Estudios de Politicas Publicas (CEPP)	Centro de Estudios de Politicas Publicas (CEPP)	Reduced version adapted for use as classroom supervision and teacher tool for San Luis province schools;		2009
Bangladesh	no	Completed	Bangla	Bangladesh Rural Advancement Committee (BRAC)	BRAC-IED (Institute for Education Development)	Small pilot sample, 100 students		2007
		Completed	Bangla	Save the Children	Save the Children	School based sample of 6 schools, grades K, 1 and 2	Nov	2008
Benin	yes	no plans						
Bhutan	yes	no plans						
Burkina Faso	yes		French, Moore	BNPP (Bank trust fund)	individual researcher	informal sample of about 35 students		2006
Brazil	ineligible	Completed	Portuguese	Graded School, São Paulo, Brazil	Graded School, São Paulo, Brazil	Adapt for use with teachers in early primary grades in Brazil		2009
Burundi	no	In planning	Kirundi	World Bank	World Bank/Consultants	In planning		2010

Country	FTI-endorsed	Status (April 2010)	Languages	Funding Agency	Implementing Agency	Scope (sample size when available)	Month	Year
Cambodia	yes	data analysis	Khmer	FTI	Government	National sample, no incomplete schools, about 2400	March	2010
Cameroon	yes	completed	French, Kom	BNPP (Bank trust fund)	individual researcher	about 300 regional		2007
Cape Verde	no	no plans						
Central African Republic	yes	no plans						
Chad	no	no plans						
Congo, Democratic Republic of the	no	In process	French	USAID/DRC	RTI International	In process	May	2010
Congo, Republic	no	planned 2010						
Cote d'Ivoire	no	no plans						
Djibouti	yes	completed	French	Bank	government	Results reported in incompatible ways		2009
Egypt	ineligible	Completed	Arabic	USAID/Egypt	GILO (Girls' Improved Learning Outcomes)/RTI International	Pilot with 100 students in 3 suburban Cairo schools; full application in 60 schools (half treatment, half control) in upper Egypt in 2009, plans for 2011	May Jan, Oct	2008, 2009, 2010
Eritrea	no	planned 2010		nonprofit	nonprofit	regional		unknown
Ethiopia	yes	Completed	Oromiffa	USAID/Washington EQUIP2 and Save the Children	AED and Save the Children	Regional application in 26 schools, 20 third grade students in each school; Classroom observation of time on task in grades 1-3;	Sept	2008

						Teacher and Principal interviews		
		Completed		DFID	University of Oxford Department for International Development	2000 students	Aug	2009
		In process	Amharic, Afan Oromo, Tigrigna, Somali, Sidama, Harari	USAID/Washington, USAID/Ethiopia	RTI International/EdData	Regionally representative samples in 8 regions. Grades 2 and 3 were assessed, for a total of more than 8000 children.	June	2010
Gambia, The	yes	Completed	English	World Bank	RTI International	Nationally representative sample of 1200 grade 1-3 students in 40 schools	May	2007
		Completed	English	World Bank	World Bank/Consultants	8959 students in grade 3 and 5, 2696 of them received an oral literacy test in addition.	June	2008
		Completed	English	World Bank	World Bank/Consultants	5660 students in grade 4 and 6, 1755 of them received an oral literacy test in addition.	June	2009
		Completed; Final report available at www.eddataglobal.org	English	World Bank	The Gambia Ministry of Education	Nationally representative sample of 1200 grade 1-3 students in 40 schools	May	2009
Georgia	yes	no plans						

Country	FTI-endorsed	Status (April 2010)	Languages	Funding Agency	Implementing Agency	Scope (sample size when available)	Month	Year
Ghana	yes	Completed	English	World Bank	World Bank/Consultants	Nationally representative sample of 1700 grade 3 and 5 students in 50 schools	June	2009
		In planning	English	USAID/Washington	RTI International/EdData	Small demonstration sample in English		2010
Guatemala	ineligible	Completed	Spanish, Mam, K'iche, Ixil	USAID/Washington	Equip2/AED and Save the Children	Application in 23 schools, 20 third grade students in each school; Classroom observation of time on task in grades 1-3; Teacher and Principal interviews		2008
		Completed	Spanish	Amigos de Patzun	Amigos de Patzun	Regional sample of 2nd and 3rd grade students in 20 treatment and 20 control schools		2009
Guinea	yes	no plans						
Guinea-Bissau	no	no plans						
Guyana	yes	Completed	English	World Bank	RTI International	Nationally representative sample of 2699 grade 1-3 students in 40 schools	Oct	2008
Haiti	yes	Completed	Haitian Creole, French	USAID/Haiti	AIR	Regional sample of 3161 Grade 1-3 students in 223 schools		2009
		Completed	Haitian Creole, French	World Bank	RTI International	Regional sample of 3000 grade 2-4 students in 80 schools	Mar	2009
Honduras	yes	Completed	Spanish	World Bank	RTI International, CIASES	Program sample of 2226 students in grade 2-4 in 46 rural, multi-grade schools (ProHeco program)	Oct	2008

		Completed	Spanish	USAID/Washington	Equip2/AED and CARE	One time application in 35 schools, 20 third grade students from each school, (depending on enrollment rates); classroom observations of time on task in grades 1-3; Teacher and principal interviews.		2009
India	ineligible	Completed		DFID	University of Oxford Depart for International Development	2000 students in Andhra Pradesh	Aug	2009
Indonesia	ineligible	no plans						
Jamaica	ineligible	Completed	English	USAID/Washington	RTI International/EdData; University of the West Indies	384 grade 2 and 3 students in 48 schools; EGRA reduced version conducted as part of SSME (see SSME tab)	May	2007
		In planning	English	USAID/Jamaica				2010
Kenya	yes	Completed	English, Kiswahili, Kikuyu, Luo	The William and Flora Hewlett Foundation	RTI International	Regionally representative sample. Total of 2000 children and 4000 assessments in Standard 3. Each student assessed in English and mother tongue, and some in mother tongue, English and Kiswahili.	Sept	2009
		Completed	English, Kiswahili	USAID/Kenya and USAID/Washington	RTI International/EdData	Project based sample; grade 2 students in 20 treatment and 20 control schools	October	2007, 2008
		completed	English, Kiswahili	nonprofit	individual researcher	convenience sample, about 100 students		2007

Country	FTI endorsed	Status (April 2010)	Languages	Funding Agency	Implementing Agency	Scope (sample size when available)	Month	Year
Kiribati	no	no plans						
Kyrgyzstan	yes	no plans						
Liberia	yes	Completed	English	World Bank	RTI International	National sample of grade 2 and grade 3 students; total of 47 schools sampled	Jun	2008
		Completed	English	World Bank	World Bank/Consultants	Nationally representative sample of 780 grade 3 and 5 students in 26 schools	June	2009
		Completed	English	USAID/Liberia	RTI International/EdData	Project sample of 180 schools 2 levels of treatment; 2957 students in Grades 2 and 3 at baseline	Nov, Jun, Jun	2008, 2009, 2010
		In process	English	USAID/Liberia	Education Development Center, RTI International	Sample of youth and adults; total of 95 schools sampled (1700 observations)	Jun	2009, 2010, 2011
Madagascar	yes	Completed	Malagasy	World Bank	World Bank/Consultants			2009
Malawi	yes	Completed	Chichewa, English	Save the Children	Save the Children	20 schools, grade 2 and 4	Feb, Nov	2009, 2010
Mali	yes	Completed; Final report available at www.eddataglobal.org	Bamanakan	PLAN International	Institut pour l'Education Populaire (IEP)	Apply EGRA as a baseline in treatment and control schools and develop/test instructional approaches		2007
		Completed	Bamankan, Bomu, Fulfulde, Songhoy	The William and Flora Hewlett Foundation	RTI International	20 grade 2 students in 25 schools in each language	Apr	2009
		Completed	French, Arabic	USAID/Mali	RTI International/PHARE	French grades 2, 4 and 6; Arabic grades 2 and 4	Apr	2009

		Completed	Bamankan, Bomu, Fulfulde, Songhoy	The William and Flora Hewlett Foundation	RTI International	Application in 4 Malian languages in grades 1, 2 and 3 (50 program and 50 control schools)	Apr	2009
		Completed	French, Shenara, Bamanakan	Save the Children	Save the Children	60 schools, grade 3	April	2010
		In process	French; Bamananka n; Shenara	Save the Children	Save the Children	Approximately 1,200 grade 3 children from 60 schools	April	2010
Mexico	ineligible	completed	Spanish	Government	Government	3 samples, regional		2009
Mozambique	yes	In planning	Portuguese	World Bank	World Bank			2010
			Portuguese	USAID/Washington EQUIP	AED, Aga Khan Foundation			2010
Myanmar	no	no plans						
Namibia	ineligible	In planning	English	European Commission		National level sample for benchmarking	Apr	2011
Nepal	yes	Completed	Nepalese, Tharu, Rana Tharu, Doteli, Magar, Bajang	USAID/Washington	Equip2/AED and Save the Children	One time application in 23 schools, 20 third grade students in each school; observation of time on task in grades 1-3; Teacher, Principal interviews	Aug	2009
		Completed	Nepali (Tharu)	Save the Children	Save the Children	20 schools, grade 2 and 4	Aug, Mar	2009, 2010
Nicaragua	yes	Completed	Spanish	USAID/Nicaragua	RTI International/EdData , CIASES	National level sample of 6649 students in grades 2-4	Oct	2008
		Completed	Spanish, Miskito	USAID/Washington	RTI International/EdData , CIASES	Instrument pilot in 60 schools	Apr	2008
Niger	yes	Completed	Zarma	PLAN International	VIA	Apply EGRA as a baseline in treatment and control schools and develop/test instructional approaches		2009

Country	FTI endorsed	Status (April 2010)	Languages	Funding Agency	Implementing Agency	Scope (sample size when available)	Month	Year
Nigeria	ineligible	Completed	English	DFID	ESSPIN	20 grade 2 students, 30 grade 4 students in 12 schools in 2 states: Kano and Lagos	Nov	2009
		In process	English	USAID/Nigeria	Creative Associates International, RTI International, Johns Hopkins U.- Center for Communication Programs, School to School	Grade 4	June	2010
		In planning	English, Hausa	USAID/Nigeria	Creative Associates International, RTI International, Johns Hopkins U.- Center for Communication Programs, School to School	Grade 2	Sept	2010
Pakistan	no	Completed	Urdu and Pashto	Embassy of the Kingdom of the Netherlands	Save the Children	243 grade 2 students from 15 schools in Allai Tehsil (mix of GPS and GGPS)	Dec	2009
Papua New Guinea	yes	In process		World Bank	World Bank	Grades 1,2,3 - 400 children per grade	Nov	2010
Paraguay	ineligible	completed		government	government	unkown		2006
Peru	ineligible	Completed	Spanish, Quechua	World Bank	World Bank, local NGOs, technical advisor Luis Crouch	Small non-representative sample in 20 schools for policy dialogue		2006
		Completed	Spanish	USAID/Washington	RTI International/EdData , Fundacion de Desarollo Agrario (FDA)	512 students in grades 2 and 3 in 64 schools; EGRA conducted as part of SSME		2007

		In process	Spanish	DFID	University of Oxford Department for International Development	2000 students	Aug	2009
Philippines	ineligible	Completed	Tagalog, English, Magindanoan, Ilongo, T'boli	Save the Children	Save the Children	1426 grade 1 and grade 3 students from 39 schools in Metro Manila (Taguig and Paranaque) and South Central Mindanao (Lake Sebu, T'boli in South Cotabato, Lutayan and Bagumbayan, Sultan Kudarat)	Nov	
		completed	Tagalog,	nonprofit	individual researcher	small regional sample		2006
Rwanda	yes	Completed	English	World Bank	World Bank/Consultants	Nationally representative sample of 2000 grade 3 and 5 students in 60 schools	June	2009
Senegal	yes	Completed	French, Wolof	World Bank	RTI International	French: 502 grade 1, 2 and 3 students; Wolof: 186 grade 1 and 3 students in 32 schools	May	2007
		Completed	French, Wolof, Pulaar	The William and Flora Hewlett Foundation	RTI International	French: 700 Grade 2 students in 50 schools; Wolof and Pulaar: Instrument validation samples (75 students each)	May	2009
		In process	French	World Bank	World Bank/Consultants		Sept	2010
Sierra Leone	yes	Completed	English	World Bank	World Bank/Consultants	2900 students in grade 3 and 4	June	2008
Solomon Islands	no	planned 2010						
South Africa	ineligible	Completed	English, IsiXhosa, IsiZulu, Sepedi, Tshivenda, Xitsonga	Government of South Africa	Government of South Africa	Measuremet to track national progress; USAID provided basic technical assistance; South African government is continuing		2007

		Completed	isiZulu, Setswana, Sepedi	USAID/South Africa	RTI International; Molteno Institute	650 students in 45 schools, 1196 observations		2009
Somalia	no	no plans						
Sri Lanka	no	no plans						
Sudan	no	planned 2010						
South Sudan	no	planned 2010						
Tanzania	no	In planning	Kiswahili					2010
Timor-Leste	yes	Completed	Portuguese, Tetum	World Bank	World Bank	484 grade 1-3 students in Portuguese, 460 grade 1-3 students in Tetum	June	2009
Togo	yes	planned						
Tonga	no	Completed	Tongan	World Bank	World Bank	Grades 1,2,3 - 400 children per grade	Nov	2009
Uganda	no	Completed	English, Luganda, Lango	The William and Flora Hewlett Foundation	RTI International	2000 students in 50 schools, each student assessed in English and mother tongue, P2 and P3	Oct	2009
Uzbekistan	no	no plans						
Vanuatu	no	In process	English	World Bank	World Bank		May	2010
Vietnam	yes	Completed	Vietnamese	DFID	University of Oxford Department for International Development	2000 students	Aug	2009
		In process	Vietnamese	JSDF	Save the Children	60 schools, grade 3	Apr	2010
Yemen	yes	In process	Arabic	USAID/Washington	RTI International/EdData	To be determined		2010
Zambia	yes	no plans						
Zimbabwe	no	no plans						

Note: Information current as of mid-2010. More updates available from RTI www.eddata.org

Annex 2 Table 2: Some fluency and comprehension rates reported in existing studies of low-income countries

		Grade 1		Grade 2		Grade 3		Grade 4		Grade 5		Grade 6	
Country	**Language**	**Words correct per minute**	**Comprehension %**	**Words correct per minute**	**Comprehension %**	**Words correct per minute**	**Comprehension %**	**Words correct per minute**	**Comprehension %**	**Words correct per minute**	**Comprehension %**	**Words correct per minute**	**Comprehension %**
Afghanistan	Pashtu, Dari	9	40	13	40	35	60	78	100				
Albania	Albanian									90	61		
Bangladesh	Bangla	17.4	47	37.4	79								
Burkina Faso	Mooré			23.2	48	31.2	36	56	71				
Burkina Faso	French	0	0	35.4	36	31	9	75	49				
Ethiopia	Afar					18							
Gambia	English	2.2	8.6	4	10.9	9.2	25						
Ghana	English					28	36.6			43	58		
Honduras	Spanish			36.2	42.8	60.1	63.8	96.7	79.4				
Honduras	Spanish					73							
India	Hindi	34.5	40	40.5	45	49.5	52	62	64	76.5	69	81.5	73
Kenya	English			10.2	9								
Kenya	Swahili			11.4	6								
Kenya	English					30.4	5.4						
Kenya	Swahili					21.2	9.3						
Liberia	English			19.6	33.5	19.6	33.5						
Mali	multiple			1.4	0			4				30	
Niger	Zerma			4	2.2								
Nepal	Nepali					27							
Nicaragua	Spanish, Miskitu	35.15	73	67.35	82	91.5	86						
Senegal	French	8	9	29	30	35	50						

Senegal	Wolof	2	5			29	52						
Timor Leste	Portuguese	4	3	19	17	45	40						
Timor Leste	Tetum	5	7	20	28	45	55						
Guatemala	Spanish					50							
	Averages	12	23	23	32	38	41	62	72	70	63	56	73
	Notes: Readers and non-readers are averaged; some languages are listed separately, others averaged in studies, some data were derived from estimates and extrapolations; averages rounded.												

6 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 Massachussetts 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. Sturtevart. 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 artficial 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 aná lisis 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 McMillan.

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. Ppresentation 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. Ministere De L‟Education Nationale Et De L‟Enseignement Superieur Evaluation en lecture des élèves de 2ème 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. Congnitive Psychology 6: 293–323.

Llambiri, Stavri. 2004. "Braktisja e fshehtë", "Objektivat minimale të domosdoshme të të nxënit", 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?: Contribució n 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 Assesing 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 Internaitonal 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 Assesssment: 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. Eugenet, 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 – Revisited. A Review of Learning Indicators in Developing Countries, with a Focus on Reading Assessments. Draft report to FTI Secretariat.

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.*

Zambia DHS EdData Survey 2002 Education Data for Decision-making Central Statistical Office. Lusaka, Zambia ORC Macro Calverton, Maryland, USA April 2003.

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.