



USAID
FROM THE AMERICAN PEOPLE

EDDATA II

EARLY GRADE READING ASSESSMENT WORKSHOP

Summary Notes from the Expert Workshop, Washington, DC, November 16 and 17, 2006



EARLY GRADE READING ASSESSMENT WORKSHOP

Summary Notes from the Expert Workshop, November 16 and 17, 2006

EdData II Technical and Managerial Assistance
Contract EHC-E-00-04-00004

Prepared for
Office of Education
Bureau for Economic Growth, Agriculture and Trade
United States Agency for International Development
(USAID/EGAT/ED)

Prepared by
RTI International
3040 Cornwallis Road
Post Office Box 12194
Research Triangle Park, NC 27709-2194

Photos Courtesy of Dr. Madhav Chavan, Director, Pratham (India).

Table of Contents

Introduction.....	1
Objective of the Workshop	3
Discussion Summary.....	4
Round Table 1: Analysis of Draft Outline: Discussion of Technical Issues; Review of Existing Assessment Instruments to Inform “Simple” and “Complex” Instruments.....	5
Review of Country/Language Experiences	6
Round Table 2: Discussion of Relative Content between Simple and Complex Instruments (for validity, level of complexity, and fitness for purpose).....	7
Round Table 3: Discussion of Language and Technology Issues.....	8
Round Table 4: Anticipating Possible Critiques and Building a Case for the Approach	9
Round Table 5: Approaches to Improvement: Using Assessment Results to Inform Classroom Practice in a Variety of Contexts (Examples of Good Practices)	9
Next Steps.....	10
Appendix	11
1. Expert Invitation	11
2. Agenda and Round Table Questions	12
3. Participant Bios and Contact Information.....	16



Introduction

Despite the importance of educational quality and student learning in particular, reliable data on learning remain extremely scarce, especially for developing countries. Many developing countries administer national examinations, typically at the end of a schooling cycle, in order to verify that students have gained the knowledge and skills required for graduation, and/or to control progression to the next level of schooling. But even where these exams are well designed, the results provide information against national curricular standards, which vary widely among countries in terms of specifics and overall rigor. Very few developing countries produce data on student learning against internationally comparable standards, which might permit even the broadest generalizations about how successful different countries are in equipping their children with essential skills and knowledge.

The purpose of this Task Order (TO) is to provide the U.S. Agency for International Development (USAID) with a valid set of instruments for assessing the extent to which early-grade primary-school children in USAID-presence countries are learning to read with an acceptable degree of comprehension and at an acceptable rate of fluency. The overarching objective is to provide USAID with an increased understanding of one essential dimension of education quality (reading) in its host countries, and ultimately spur more effective efforts to improve educational quality. The TO specifically addresses the achievement of Ed Data II's Result 5: International education statistics and indicators improved, promoted, and further standardized through collaboration with other USAID activities, Ministry statistics staff, multilateral and other bilateral organizations.

To this end, USAID has asked RTI to develop two reading assessment instruments: 1) an opportunity to learn assessment; and 2) a simple-screening assessment. These instruments will build upon recent experience of USAID, other donors, and country experiences, and will be designed to permit cross-country comparison of the degree of reading skill acquisition in the first few grades of the school system. These assessment instruments, or metrics, and the results obtained through field testing them in three languages—English, French, and Spanish—will serve as a basis for international discussion and will facilitate the establishment of basic standards for early-grade reading.

As outlined in RTI's proposal to USAID, the first instrument, an in-depth opportunity to learn assessment, is intended to allow for a careful diagnostic of challenges to imparting literacy in the early grades. It will be designed for maximizing reliability over simplicity. It will also allow for (a) cross-country comparison of the degree of reading-skill acquisition in the first few grades of the school system; (b) determination, in the first application, of the grade at which a country's education system is able to impart the capacity to read connected prose with fluency and understanding, for comparison purposes; and (c) identification of the specific roadblocks or areas of weakness in the progression to reading, which are correctable through feedback to teachers. This last aim is important if the assessment plans to support interventions and improvement. The second assessment instrument, a simple-screening instrument, will permit diagnostic measurement by teachers and district offices to identify children, schools, or districts with problems in imparting early grade literacy.

The in-depth assessment will focus on the opportunities children have to acquire literacy within their first few grades of schooling. The instrument will include an analysis of key learning determinants, including teacher use of time and time-on-task in the classroom, as well as the availability and quality of learning materials (against a standard metric of reading and early grade literacy expectations). The in-depth instrument will measure student knowledge and opportunity to learn against a cumulative, stair-step sequencing of literacy-acquisition measures and expectations for accomplishments as advanced by a panel of experts and developed in collaboration with RTI staff. These expectations will be linked to current research on how children learn to read, from beginning to understand the alphabetic principle, to decoding words and syllables. Sample indicators for literacy attainment within each grade could include whether a child is capable of the following:

"To prevent reading difficulties, children should be provided with:

- Opportunities to explore the various uses and functions of written language and to develop appreciation and command of them.
- Opportunities to grasp and master the use of the alphabetic principle for reading and writing.
- Opportunities to develop and enhance language and meta-cognitive skills to meet the demands of understanding printed texts.
- Opportunities to experience contexts that promote enthusiasm and success in learning to read and write, as well as learning *by* reading and writing.
- Opportunities for children *likely to experience* difficulties in becoming fluent readers to be identified and to participate in effective prevention programs.
- Opportunities for children *experiencing* difficulties in becoming fluent readers to be identified and to participate in effective intervention and remediation programs, well integrated with ongoing good classroom instruction."

Snow C.E. et al. (1998), *Preventing Reading Difficulties in Young Children* (p. 278)

- Knows the parts of a book;
- Recognizes and can name upper and lowercase letters;
- Understands that sequences of letters in written words represent sequences of sounds (phonemes) in a spoken word (the alphabetic principle);
- Can, given a spoken word, produce another word that rhymes with it;
- Uses letter-correspondence knowledge to sound out unknown words when reading;
- Notices when difficulties are encountered in understanding text or when simple texts fail to make sense;
- Engages in literary activities voluntarily;
- Represents the complete sound of a word when spelling independently;
- Accurately decodes orthographically-regular multisyllable words;
- Recalls facts and details of texts;
- Can identify words that are causing comprehension difficulties; and
- Reads aloud with accuracy and comprehension texts appropriate for their grade level (examples from Snow, C.E. et al. (1998), pp. 80-83).

The second instrument, a simple- screening instrument, will be designed such that teachers, directors, and school-supervision personnel can apply the instrument on their own as an inexpensive and simple diagnostic of student progress in reading. Its function will be that of a

“warning light—an indicator of progress against an easy-to-understand metric of expectations for early grade reading. The simple-screening instrument will build on a correct words per

“In some countries 50 percent of fourth grade students do not understand the meaning of the texts they read (in one public school class, I found 20 non-reading students in a class of 29), but the majority of these students attend schools that cater to families in the ‘lower half of the income bracket.’ This means that 90 percent of the students in this half of the population do not understand what they read (even though many complete their primary schooling). In this situation a good literacy program (in the first two grades of primary school) can have an immense impact on the performance of the education system.”

--Ernesto Schiefelbein, Former Minister of Education, Chile

minute (cwpm) assessment RTI staff conducted in Perú in collaboration with the World Bank and DFID, under which some 35 percent of second-grade students were not able to read a single word (Abadzi, Crouch et al. 2005). Building on the standards for cwpm developed in a number of countries (including, for example, Chile, Spain and the United States), the simple-screening instrument will provide immediate feedback to teachers and administrators on how particular children or schools are performing in early-grade reading.

The assessment will be designed for use across multiple grade levels so that comparisons can be drawn between expectations and performance for students from first to third grades.

Each of these instruments will be designed to complement and inform one another. While the simple-screening instrument is akin to taking the temperature of the patient, the in-depth instrument is a comprehensive diagnostic of the environment where patients live. That is, while the simple-screening instrument will help us detect where children are generally “getting stuck” from a system-level perspective, the simple-screening assessment will provide teachers and administrators with an indication of the “trouble spots” in their particular classrooms and schools. Using information disseminated from the results of the in-depth instrument, teachers would be informed of which components of the early-grade literacy sequencing strategy are missing and require remediation. One additional role of the in-depth instrument will be to provide a basis for assessing the reliability of the simple-screening instrument. As part of the measurement strategy, the information gathered in the simple-screening instrument may overlap with or represent a subset of the information gathered using the in-depth instrument. While the in-depth instrument is designed to maximize reliability, the simple-screening instrument will be developed with simplicity in mind, such that the results of the instrument can be conveyed to community members and parents, as well as policy makers and Ministry personnel. Finally, both instruments will be designed to be adaptable across languages and cultures (to the extent possible) to allow for cross-country comparisons.

Objective of the Workshop

In order to obtain feedback on and confirm the validity of the approach outlined above, RTI convened a meeting of cognitive scientists, early-grade literacy experts, research methodologists, and assessment experts to review the proposed key components of the draft assessment instruments. During the 2-day workshop, participants were charged with bridging the gap between research and practice; that is, merging advances in the reading literature and cognitive science with assessment experiences. Researchers and practitioners presented evidence on their strategies for measuring literacy acquisition within the early primary grades. In addition, they were asked to identify the key issues to consider in designing a multi-country, multi-language early grade literacy assessment protocol. The

workshop, co-hosted by USAID, The World Bank, and RTI, included more than a dozen experts from a diverse group of countries, as well as some 14 observers from institutions such as USAID, the World Bank, the William and Flora Hewlett Foundation, George Washington University, the South Africa Ministry of Education and Plan International, among others. A detailed list of participants, including expert bios and contact information can be found in the Appendix, along with a sample invitation, workshop agenda and Round Table discussion questions.

Discussion Summary

Following opening remarks by Cheryl Kim (USAID), Robin Horn (World Bank) and Luis Crouch (RTI), RTI's Amber Gove and Kyle Snow outlined a) RTI's interest and intended approach to developing the two instruments, b) USAID's expectations for this task order and c) RTI's progress to date. In the following pages, a brief summary of the topics covered and conclusions drawn from the workshop dialogue is provided.

To begin with, part of RTI's interest in early grade reading (EGR) assessment stems from the frustration with a lack of simple, cost-effective measures of basic, early grade assessments. In the health sector, simple measures of childhood diseases and health outcomes abound and are understood in a scientific way. In education, national assessments, when they exist, tend to remain at the national level, providing little practical feedback for teachers and school administrators. In the context of increasing decentralization and devolution of the state, parents with little education are frequently asked to evaluate school quality. More often than not, they give the school very high marks as they have few concrete and simple measures on which to base their evaluation. This, coupled with the knowledge that reading is a foundation for all learning and that teaching reading *is* rocket science (but that it can be done) are the foundations for RTI's interest in pursuing this challenging endeavor.

RTI's team explained that the intention was to develop, test and vet the EGR instruments in one pilot country within the first project year. The more comprehensive (or long) instrument would be sample based, validate the simple instrument and identify key gaps in the teaching and learning process. The simple (or short) instrument would act as a screening tool to identify students and schools that are having difficulties in learning how to read. Together, these instruments are intended to be comparable across countries; identify the grade at which the education system in each country is imparting the capacity to read connected prose with fluency and understanding; and identify roadblocks and areas of weakness within the teaching of reading.

RTI accomplishments to date include a brief review of the literature, a more detailed review of assessment instruments focusing on early grade reading, development of a draft instrument outline, and calling together the expert workshop. In reviewing the literature RTI focused on the five key instructional components as outlined by the National Reading Panel: Phonemic Awareness, Phonics and Decoding, Fluency, Vocabulary and Comprehension.

Round Table 1: Analysis of Draft Outline: Discussion of Technical Issues; Review of Existing Assessment Instruments to Inform “Simple” and “Complex” Instruments

One of the first critical observations of Round Table 1 was the need for the RTI Team to identify the conceptual framework, or formulated model of reading acquisition, to guide the assessment development process. Without a model, RTI would be in danger of concluding that that which is important in reading is that which we can measure. Rather, the experts suggested, RTI should first identify that which is most important and attempt to measure those components.

It was pointed out that the “reading wars” hinge on whether we should or can assess the initial steps of reading acquisition (foundational literacy acquisition). Nonetheless, most around the room could acknowledge the fact that today there is more understanding of what constitutes reading acquisition, especially in the early grades. While not interested in “picking sides” in the reading wars, RTI explained that one of the goals of the assessment is to inform instruction and predict or identify those classrooms and students who face the greatest challenges to reading acquisition. It is also RTI’s intention to include measures that the experts can generally agree upon as being predictive of later reading ability, thereby increasing the utility of the tools for remediation and teaching purposes.

RTI presented its initial outline of the key components to be included in the short instrument, as follows: picture vocabulary, letter-sound naming, nonsense words and passage reading (the latter to incorporate rate, accuracy and comprehension). The long form was briefly reviewed as well (see table, below), before discussion focused on specific components of the instruments. The picture vocabulary component was reviewed, and initially rejected as a less than essential component of the short form (though it could possibly be included in the long form).

It was suggested that it is a challenge to find good, culturally appropriate reading passages with high frequency sight words. Instead, some oral/listening comprehension should possibly be included. Some participants recommended that we examine the possibility of including a spelling or dictation exercise, in part to separate the fluency from the comprehension task.

In response to the question of the application of the tools across multiple grades, most experts seemed to be in agreement that this was possible, though care would need to be taken in selecting appropriate reading passages, especially in selecting paragraph selections and comprehension questions that are discriminating enough.

In reference to the question regarding the challenge of false negatives (overly shy students who can read but are not willing to perform in front of a stranger) experts suggested that the training protocol include specific procedures for building rapport. Also, to the extent possible, local villagers should be recruited to conduct the assessment. Children should also be comfortable with the tasks they are to be tested on. Finally, students should be allowed to read the passages silently before doing the oral reading portion of the assessment.

It was pointed out that the RTI draft model was actually testing many of the components of the foundational stage of reading and should therefore include a measure of phonemic awareness. It was suggested that the short instrument should focus on: letter task, simple (real) word reading and simple non-word reading in order to capture three critical areas of foundational literacy. With these components, the assessment would be used to track the child's first steps in reading from which everything else follows.

It was suggested that in fact reading is culturally bound and occurs within a context. Readers employ multiple strategies for comprehending text. Thus assessments should examine not just the skills but the multiple strategies that the child uses to understand the text. While experts agreed that we all value fluency and the value of getting kids to be fluent, accurate decoders, there was a level of disagreement as to whether this should be the main focus of measurement. A question was raised about the value of sight-word recognition in the assessment and if it would be needed in more transparent languages than English. The point was raised that decoding isn't everything and if a student does not know the word they have just managed to decode then it does not do them much good.

By the end of the first day, a modest degree of consensus had been reached that the instruments should contain at a minimum the following design components:

Short Form (aka simple, subset of Long)	Long Form (aka comprehensive)
<u>Design Aspects:</u> Screening Function Non-Sample, Could be universalized Apply with little training ("popular" tool) Spotlight for calling forth support <u>Components:</u> Letter identification (name and sounds) Single-word reading Non-word reading Phoneme segmentation Passage reading (multi-level) <ul style="list-style-type: none"> - rate - accuracy - comprehension 	<u>Design Aspects:</u> Sample Based Diagnostic at the Classroom level Identifies "Causes" of Problems Expert Applied Used for Research <u>Components:</u> Repeat Short Form Components Listening Comprehension Engagement Dictation Classroom level: <ul style="list-style-type: none"> - Time on Task - Review of Lesson Plans - Curriculum Analysis

Review of Country/Language Experiences

The afternoon session, moderated and with an introductory presentation by Helen Abadzi (The World Bank), was divided among several presentations, as follows:

1. James Royer gave a brief discussion of the relationship between working memory and reading acquisition, including evidence that the brain has capacity for about 4-9 items in 10 seconds. If students do not read with sufficient speed to capture these ideas they will likely have low levels of comprehension.

2. Philip Seymour provided an overview of his recent paper comparing differences in early reading acquisition (foundational literacy) across several European languages. A copy of the notes he distributed can be found in the supplemental materials accompanying this report.
3. Malatesha Joshi shared some of his research on English and Indian languages (in particular Kannada) and hypothesized that if children are receiving good instruction in one language they are likely receiving good instruction in the other. Part of his research examined those aspects of language acquisition in Kannada that predicted performance in English.
4. Jules Kinda reviewed the results of an oral reading assessment conducted in Burkina Faso with children in both French immersion and bi-lingual schools. The initial results presented at the workshop can be found in the accompanying documentation.
5. Rima Azzam gave a brief discussion of some of the challenges of conducting assessments in Arabic, including the fact that as the texts (and language) are considered sacred, it is difficult to create exercises such as non-words (as that would entail manipulation and modification of a sacred text). Similarly, teaching and instruction methods are difficult to change, including how vowels and the 16 core shapes are taught. To further complicate the matter, the spoken language is very different than the written one (written Arabic is far more formal).
6. Madhav Chavan of Pratham shared the experience of the Annual Status of Education Report (ASER) conducted in India in which volunteers visited some 10,000 villages and conducted individually-administered basic reading and mathematics exams with more than 333,000 children. More information can be found at www.pratham.org

Round Table 2: Discussion of Relative Content between Simple and Complex Instruments (for validity, level of complexity, and fitness for purpose)

Round Table 2 examined issues of content between the two proposed instruments. Discussion began with examination of the picture vocabulary assessment as a part of the proposed tools. Opinion varied, but it was suggested that the Peabody Picture Vocabulary Test (PPVT) could be used in the long form of the assessment. Experts cautioned that without a standardized test developed and piloted it would be difficult to include a PPVT section. As this test exists in Spanish it would be possible to pilot it in Latin America as there would be only slight adaptations that would be required.

The second question posed by the RTI team was the appropriate and useful time that should be budgeted per child. Experts converged on a rough response of about 20 minutes per child, though they cautioned that examiners should not take longer or students would be fatigued. One suggestion was to break the assessment into two parts—to see a child for 10 minutes at a time, return them to the class, and then repeat another 10 minutes of assessment. Experience from the field led several experts to caution RTI regarding the training of good

enumerators—in one case the people hired were unable to follow the scripted instructions without sounding stilted and formal; to eliminate this problem, assessment organizers required the enumerators to memorize the scripts.

On a related question, the RTI team asked if responses should be open but timed (that is, given a limited number of items how long does it take for the child to complete the task) or time-limited (given 60 seconds to complete task). The general consensus was that in the interest of time and efficiency it would be better to give students 60 seconds (or another limited amount of time) to complete the required tasks.

On the issue of identifying classrooms or students based on a simple dichotomous indicator (can they read or not) it was suggested that for the foundation literacy portions of the assessment, a metric of 75% correct would be adequate to state that the child had achieved foundation literacy. Those students who had achieved 25% or less would be considered non-readers.

Sample instruments, developed in Ghana by CAL (Jim Bauman) and in Afghanistan by Colette Chabbott, were shared and discussed with the group. The Ghana instrument in particular uses basic elements of phonemic awareness to check for decoding of words, includes a picture dictionary for sounds of letters and uses visual and verbal clues to ensure students comprehend the words and text.

Round Table 3: Discussion of Language and Technology Issues

Mike Royer began the language and technology session with a demonstration of the software program his team has designed, entitled Cognitive Aptitude Assessment Software. The system enables test administrators and researchers to record student words and responses and assess the number of words (or letters or syllables) correct per minute. Additional information can be found at www.educationalhelp.com.

Marilyn Adams then presented the read-aloud voice recognition and comprehension software her team developed at Soliloquy Learning (with 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. A tutorial and video are available at www.soliloquylearning.com.

The discussion on technology varied from feasibility of using computer assisted instruction in developing countries to cultural implications of new and foreign technologies (and a brief digression as to the merits of the \$100 laptop). One expert pointed out that collection of data through electronic mechanisms (e.g. digital recorders) for later coding would limit the amount of local capacity building that should go into these types of assessment efforts. Implications for validity and reliability were also raised—it would not be possible to construct a test for one medium and then change the delivery medium. Scoring redesign (from paper to electronic form) would not, however, generate problems in this respect. Experts also cautioned that it is better to have well-trained enumerators who score results in the field than to have enumerators analyze results at a central location as subsequent evaluation tends to “over-think” the results.

As a summary question regarding language issues, RTI posed the question of using a translation activity for assessment (that is, testing two languages at the same time using translation from one language to the other). Experts suggested that instead two parallel forms of the assessment should be developed for testing each language individually as this would reduce confusion and confounding effects.

Round Table 4: Anticipating Possible Critiques and Building a Case for the Approach

A brief video developed with the World Bank and RTI's Luis Crouch was shown to participants to kick-off the discussion on policy dialogue. The video, available [here](#), documents results from a pilot assessment in 22 schools, while transmitting a simple message of quality learning standards, accountability and support for reform. The video set the stage for further discussion of advocacy, education reform support and parent and teacher involvement in the quality reform process.

Following discussion of the video, Sylvia Linan-Thompson described a professional development model for early literacy in grade 1-3 classrooms, under expansion as part of USAID's Centers for Excellence in Teacher Training (CETT). The program uses classroom coaching and scripted instruction for teachers in several countries in Latin America. Additional information on CETT can be found at <http://www.readingforallchildren.org/>

Several experts suggested that teachers be involved in the test development and application process as strategy for minimizing their resistance and increasing teacher support for the initiative. Drawing lessons from Pratham's experience, Mr. Chavan argues that the point of the effort is to bring reading to the forefront of the issues people think about. The challenge is to make the outcomes measurable and negotiate some starting point to be used as an indicator of the challenges in the sector.

It was pointed out that measurement is (and should be) used to indicate where teachers need assistance in their instruction methods. Additionally, participants were reminded that one of the objectives of the assessment is to provide simple tools for improving reading instruction. One of the controversial points raised is that many educators do not wish to make education a measurable process.

Round Table 5: Approaches to Improvement: Using Assessment Results to Inform Classroom Practice in a Variety of Contexts (Examples of Good Practices)

The final Round Table session included an open discussion with observers. Experts called for the need to build teacher knowledge about reading, strengthen skills to impart that knowledge and provide guided practice and support for teachers. Part of RTI's assessment development process should be to provide a form of give-back to teachers and schools through the development of the simple instrument.

Methods for returning assessment results back to teachers were discussed, with the example of CETT put forward as a model of accompaniment and instructional coaching.

The Pratham model was discussed in the context of providing easily understood and measurable goals for teachers through charts that document student progress over time. Such tools are posted in the classroom enabling parents and visitors to check on student progress.

The discussion provided a broad framework for development of the instruments but did not include a procedure for going about it, which may require further discussion. The question of how the materials are selected remains to be resolved, as well as the precise procedure to be followed. Also to be verified and discussed further are the issues of validity and reliability as well as use of results.

Next Steps

Based on the results of the workshop, RTI will develop draft protocols for the in-depth and simple-screening assessments. Our team will review the comments, submitted and proposed materials and lessons learned and will develop draft assessment instruments, sampling protocols, and enumerator training manuals for the piloting of each of the draft in-depth and simple-screening instruments. The meeting generated considerable interest and a rich discussion; RTI intends to continue the dialogue with the workshop participants. RTI thanks USAID and the World Bank for its support of and interest in this effort.

Appendix

1. Expert Invitation



November 8, 2006

Dear Dr. Adger,

Please join us on November 16–17, 2006 at the World Bank Headquarters in Washington, D.C., for an important technical meeting to identify critical components of a methodology to be used worldwide to assess early grade reading proficiency. The meeting is being sponsored by USAID, the World Bank, and the Research Triangle Institute. (A draft agenda is attached and additional logistical information will be forthcoming).

RTI has been contracted by USAID to develop two assessment instruments to measure the extent to which school children in early primary grades are learning to read with an acceptable degree of comprehension and rate of fluency. The project is being funded under USAID's EdData II program. The first instrument will be a simple one that can be applied by teachers, local officials, or parents to determine if a child reads fluently and understands basic text. The second will be more comprehensive and include classroom observations and measures of how instructional time is used in class (opportunity-to-learn assessment). The second instrument will be used to validate and calibrate the first instrument as well as for deeper diagnosis of reading problems.

We hope you will be able to attend the meeting and share your insights on measurement of early grade reading acquisition. There is no need to prepare a paper or presentation. We will send you some points for reflection in advance, and ask that you bring your favorite tools and share your knowledge of measurement-based interventions or improvement strategies for early grade reading.

We would be grateful if you could respond by November 10 whether you will be able to participate in the meeting. Please communicate your decision to Amber Gove, RTI Senior Research Analyst, at agove@rti.org.

We look forward to hearing from you and working with you on this important education issue.

Regards,

Joseph Carney
Director
Office of Education
USAID

Robin S. Horn
Education Sector Manager
Human Development Network
The World Bank

Luis Crouch
Research Vice President
RTI International

2. Agenda and Round Table Questions

Agenda: Early Grade Reading Assessment Workshop

Thursday, November 16, 2006

9:00–9:30	Introduction, Opening Remarks	Cheryl Kim, USAID Robin Horn, World Bank Luis Crouch, RTI
9:30–11:00	<ul style="list-style-type: none"> Discuss RTI Approach to Designing “Simple” and “Complex” Instruments Presentation of Draft Outline for “Simple” and “Complex” Instruments 	Amber Gove, RTI Kyle Snow, RTI
11:00–12:30	Round Table 1: <ul style="list-style-type: none"> Comments/Analysis of Draft Outline: Discussion of Technical Issues Review of Existing Assessment Instruments to Inform “Simple” and “Complex” Instruments 	Expert Round Table Luis Crouch, Moderator
12:30–1:30	Lunch Provided	
1:30–5:30	Review of Country/Language Experiences <ol style="list-style-type: none"> Memory, Speed and Comprehension Lessons from Multiple Language Research Orthography and Literacy Burkina Faso (measurement) Arabic (script/measurement) Pratham/India (remediation and measurement) Peru (dissemination and policy) Implications and Lessons	Helen Abadzi, Moderator <ol style="list-style-type: none"> James Royer, UMass, USA Philip Seymour, Univ Dundee, UK Malatesha Joshi, Texas A&M, USA Jules Kinda, Burkina Faso Rima Azzam, AIR, USA Madhav Chavan, Pratham, India Luis Crouch, RTI Expert and Observer Group Discussion

Friday, November 17, 2006

9:00–9:15	Review of Discussion	Amber Gove
9:15–11:00	Round Table 2: Discussion of Relative Content between Simple and Complex Instruments (for validity, level of complexity, and fitness for purpose)	Expert Round Table Kyle Snow, Moderator
11:00–12:30	Round Table 3: Discussion of Language and Technology Issues	Expert Round Table Luis Crouch, Moderator
12:30–1:30	Lunch Provided	
1:30–3:30	Round Table 4: Anticipating Possible Critiques and Building a Case for the Approach	Expert Round Table Amber Gove, Moderator
3:45–5:00	Round Table 5: Approaches to Improvement: Using Assessment Results to Inform Classroom Practice in a Variety of Contexts (Examples of Good Practices)	Expert Round Table Luis Crouch, Moderator
5:00–6:00	Closing Discussion – Review, Lessons and Next Steps	Expert and Observer Group Discussion

Questions for Round Table Discussions

Round Table 1:

Comments/Analysis of Draft Outline: Discussion of Technical Issues; Review of Existing Assessment Instruments to Inform “Simple” and “Complex” Instruments

1. What essential components are missing from our draft instrument outlines? What components could be eliminated?
2. Our intention is to apply the same instrument, or at least the same key components of the instrument, to students in first few grades of primary school (Grades 1-3). We are doing this purposively, so we can draw “curves” of increased learning and to see where a “breakthrough” is made against a passage of given difficulty. What are the pros and cons of this? There will be a ceiling effect, but in effect that is what we are looking for.
3. To what extent are the essential components (and less essential components) identified in Question 1 age and/or grade invariant? Given our intent to use the same pair of measures in grades 1-3, it is important to anticipate any developmental changes that emerge over time in the relationships among components of reading being assessed.
4. What assessments have been particularly good at understanding the components of student reading? Should we ensure that we include issues related to, say, phonemic manipulation in the more complex one? Should we test, for example, “pure” listening and sound recognition in the more complex instrument? Please share specific examples from assessment instruments you have used, and their value in establishing the “cause” of problems that then become visible “later” as fluency problems.
5. What are some of the benefits of using these sorts of items? What are some of the challenges of using these sorts of items?
6. There seems to be an issue in oral applications in that some children may be able to read but are too shy to do so in the presence of a stranger. What are recommended techniques for avoiding false negatives? What was the experience of, for example, Pratham? Does starting with letters and then going on to paragraphs help avoid this (note Pratham started with paragraphs and then did letters).
7. Several of the assessments we reviewed include pictures (Peabody Picture Vocabulary Test, DIBELS, etc.). What is the rationale for using pictures versus text (assessing pre-reading vocabulary?), and what are the tradeoffs to presenting pictures?

Round Table 2:

Discussion of Relative Content between Simple and Complex Instruments (for validity, level of complexity, and fitness for purpose)

1. Our intent is to develop two tools. A simpler one is intended for “universal” application by teachers, perhaps by a literate person in the community, or by district officials. It can provide a simple way to screen or pinpoint teachers or schools having problems, but is not meant to diagnose the nature of the problems in much depth. The more complex instrument would eventually be applied on a sample basis by experts. Having a complex instrument would serve one purpose during the pilot stage and a different “eventual” purpose. During the pilot stage the complex instrument is used to establish validity and reliability. Second, the more complex instrument could diagnose in more depth the simpler one. In principle, in the pilot stage one could simply apply only a complex instrument, and then *select out* fewer and simpler items to then constitute the simple instrument. Thus, in the pilot stage there need not be two instruments as such, but the idea is to develop two. Is this a sound general strategy? For example, our informing belief is that in most countries the more basic or causal problems (e.g., lack of phonemic awareness) will tend to be fairly common across schools, and thus a sample-based approach, to detect that teachers are not working on this,

or that this is a problem, will suffice. A simpler instrument will help schools and districts assess which schools are having problems.

2. What is the minimum number of components for the simple assessment instrument to ensure reliability and validity? Does the simplest instrument that we can design have acceptable reliability and validity? Does reliability matter as much in oral instruments as it does in written instruments?
3. What is the maximum number of components we should include in the complex instrument?
4. On a scale of 1-10, how simple should the simple instrument be in comparison to the complex instrument?
5. What should be the maximum amount of time per student allotted to the simple instrument? What about to the complex instrument?
6. What is the most feasible use of the simple instrument: screening children with reading problems (i.e., diagnostic), or providing a general view of reading proficiency across skills levels (descriptive), and are these mutually exclusive uses?

Round Table 3:

Discussion of Language and Technology Issues

1. What steps should be taken to ensure comparability in English, French and Spanish? Has research identified “families” languages for which reading acquisition is general internally consistent, yet perhaps different than as occurs in other “families”?
2. Is the best approach one where an instrument is developed in one language and the translated into others, resulting in different versions of the same instrument, or one where instruments are developed following the same general framework (e.g., components measured) even if the items are not related?
3. Assuming there is interest and support for expanding into other languages (Arabic, Bengali. . .) what are the challenges in terms of comparability in language complexity, regular vs. irregular language, etc.? What experiences/challenges have you had in multiple language studies?
4. Our experience in countries with home languages different from the dominant national or international language is that, regardless of policy, early grade instruction in reading really takes place in both languages simultaneously, depending on individual teacher ability, predilection, and so on. Thus we see a value, in those situations, in testing in both languages. This will not only test, but will help establish whether the issue is proficiency in teaching reading or the inherent difficulties created by complex bilingual environments. That is, what experience have you had in using instruments that assess student skills in two languages simultaneously?
5. In these situations, is translation an intuitively appealing way to test comprehension? We tried this in Kenya between Kiswahili and English, using a small sample, as well as in Uganda between Luero and English, and found a rather good correlation between “comprehension” measured in this way and fluency, and “comprehension” measured via more traditional techniques. Is this worth trying in heavily bilingual situations?
6. What issues that should be taken into account if we do want to design instruments that test simultaneously in languages with both transparent and deeper orthographies, ? For instance, in languages with transparent orthographies should one even bother with words?
7. What do we know (and can we tell from research and experience) about the ability of children at the end of second grade to read fluently in: north and south Indian languages, Amharic, voweled Arabic, mainly unvoweled Urdu, voweled Arabic (but specifically for the Maghreb)? What do we know about the effects of removing the vowels from Arabic?
8. Is it practical/feasible/cost-effective to use software to score instruments (e.g. CoolEdit, Adobe Audition) and what experiences have you had in doing that? Similarly, should technology be developed to provide reports (even if scores are input by a user after computed locally)?
9. What inexpensive ways have other studies used to record and score student answers to oral exams?

Round Table 4:

Anticipating Possible Critiques and Building a Case for the Approach

1. In developing instruments that are “just simple enough” one may face critiques on two levels. First, that one has oversimplified the complex process of assessing children’s reading, that reading is a “holistic” process that cannot be assessed as consisting of separate and sequential or consecutive skills, and so on. While this sort of debate may seem somewhat outdated in the US and other developed countries, we find it alive and well in the developing world. In Peru, for example, the officials in the Ministry of Education do not want to have anything to do, or so it seems to us, with assessing fluency, much less assessing things such as letter recognition, as this seems to “debase” the notion of reading. What strategies would you recommend for overcoming these beliefs?
2. Second, in producing materials that are simple and short there may be reliability issues. Yet we believe that there are considerable social mobilization and accountability advantages in having simple and short-enough measures of literacy that can indeed be assessed by a community or non-experts. What are some of the tradeoffs here? Do we have any priors on what is “simple enough” to fulfill a social mobilization and simple accountability role (“can your child read? – here is how you can tell?”) yet not so simple that it becomes nonsense?
3. What are strategies that others have used to defend their assessment models?
4. Many current international assessments use paper and pencil exams for students age 9 or 4th grade and above. Other countries (Peru) are using group pencil-and-paper tests as early as Grade 2. Our intention to work with younger students is based on the idea that 4th grade is too little, too late in terms of remediation strategies, and to test orally on the assumption that pencil and paper group instruments cannot detect some of the early precursor problems (especially in the bottom of the distribution). What possible critiques is one likely to face in testing younger students, and testing (mostly or partially) orally?
5. Our general intention, certainly for the simpler version, is to use “text only” assessments without multiple choice. We find many assessments use a picture-based approach and/or multiple-choice. We suspect the latter is perceived as being more “holistic” perhaps because there is a belief that it can test decoding along with comprehension, and there is a belief that even having items that test “pure” decoding goes against the curricular ideology. Are there deeper reasons for preferring picture-based assessments? What are some of the tradeoffs here? Can we use our simple-complex design to see whether using pictures adds meaningful knowledge, by establishing useful correlations?

Round Table 5:

Approaches to Improvement: Using Assessment Results to Inform Classroom Practice in a Variety of Contexts (Examples of Good Practices)

1. How can results be used to inform pre-service and in-service training for teachers?
2. What experiences and strategies have been particularly successful in linking assessment results to teacher training efforts? How can we best set up the assessment process so as to facilitate acceptance of the results by teachers and educators?
3. In using assessment results to inform practice, is there a danger that such methods are accused of being “too instrumentalist” or that they come too close to “teaching to the test?” But, is the teaching of basic reading so basic indeed that “teaching to the test” (or, at least, teaching to the underlying constructs) is really not a problem?
4. It is a very powerful and simple policy message (facilitating accountability and school support measures) to state that 80 percent of students in School X can read, while in School Y, only 10 percent can read. This requires classification of students along a dichotomous (0/1) variable, at least at the school level. Do the policy benefits to this approach outweigh the validity issues for classifying students into these categories? What other approaches have been used to translate complex assessment ideas into simple/powerful measures that parents, teachers and others can understand?

3. Participant Bios and Contact Information

Invited Experts

1. **Marilyn Jager Adams** holds a Ph.D. from Brown University in cognitive psychology and developmental psychology and is internationally regarded for her research and applied work in the area of cognition and education. Recipient of the American Educational Research Association's Sylvia Scribner Award for outstanding educational research, Dr. Adams's contributions include the landmark book, [*Beginning to Read: Thinking and Learning About Print*](#) (MIT Press). On the applied side, she is Senior Literacy Advisor for Instruction for PBS's [*Between the Lions*](#), as well as senior author of [*Fox in a Box*](#), an award-winning, standards-based literacy assessment kit. Dr. Adams has also written/designed three empirically proven instructional programs. These include [*Odyssey: A Curriculum for Thinking*](#), which was originally developed for barrio students in Venezuela; [*Phonemic Awareness in Young Children*](#) on linguistic awareness for emergent readers and special needs students; and Open Court's 1995 edition, [*Collection for Young Scholars*](#), a program for reading, writing, and literacy development for elementary school students. She is currently a Visiting Professor in the Cognitive and Linguistic Sciences Department at Brown University and Chief Scientist at Soliloquy Learning, Inc., where she has been working on developing the potential of speech-recognition technology as a medium for supporting and understanding reading development. marilyn.adams@verizon.net
2. **Carolyn Temple Adger**, Ph.D. sociolinguistics, is director of the Language in Society Division and the Language Education and Academic Development Division at the Center for Applied Linguistics in Washington, DC. Her research has centered on language in culturally diverse educational situations. At present, she is leading USAID-supported technical assistance activities in Ghana that involve development and implementation of literacy standards for primary school, creating a culture of literacy in schools, and introducing English in non-formal education.
3. **Rima Azzam**, a Principal Research Analyst with AIR, has a multidisciplinary background. She received her B.Sc. in Psychology from the University of Surrey in England, her M.Sc. in Educational Psychology from the University of London, and her M.Ed. in Neuropsychology and Ed.D. in Special Education from Teachers College, Columbia University. She has a solid foundation in research with extensive academic training in research methodologies and strong management skills. Dr. Azzam has worked with a variety of populations including mentally and emotionally handicapped children and youth. She has extensive hand-on experiences in assessing, teaching and tutoring children and adults with learning difficulties. Her experience also includes teacher training, curriculum development, and test development in different cultures and languages. Her work has involved managing, designing and conducting research in US Federal as well as international education and mental health projects. razzam@air.org
4. **Jim Bauman** is a Senior Associate and Program Manager in the Language Testing Division (LTD) at the Center for Applied Linguistics. He is the program manager for the WIDA Consortium's effort to build and maintain comprehensive standards based tests of English language proficiency, oral and written, for English language learners in grades K-12. He also serves as principal investigator for a USAID funded program to the Education Development Center to advise the Ghanaian government on educational development. Specifically, he spearheads an effort to develop assessments for early literacy achievement in local Ghanaian languages and English. He has also been involved in building online tests of reading and listening proficiency in Russian and Arabic and spearheaded a successful effort to train prospective raters of oral language proficiency tests using the ACTFL scale by means of an online course in Blackboard. Finally, he acts as program manager for a large-scale test development effort funded by the WIDA Consortium and the U.S. Department of Education to create an alternative content assessment in English language arts, math, and science for low proficient English language learners in elementary and high school. jim@cal.org

5. **Madhav Chavan** is a member of India's National Advisory Council, an apex advisory body constituted as an interface with civil society, with regard to the implementation of the National Common Minimum Program (NCMP) of the Government of India. He is one of the founders and the Director of Programs of Pratham (www.pratham.org), an organization that has been working for universalization of elementary education in India since 1994. A Chemist by training, he got his doctorate in Chemistry from the Ohio State University in 1984. After a stint as a post-doctoral fellow and a Visiting Assistant Professor at the University of Houston he returned to India and was appointed Reader in Physical Chemistry at the University Department of Chemical Technology, University of Mumbai in 1987. In 1989 he got involved in the adult literacy work in the slums of Mumbai and founded the Committee of Resource Organization (CORO) for Literacy and began to move away from Chemistry. He was involved in the social organization aspects of the Integrated Waste Management Project of the Department of Science and Technology, Government of India between 1992-94. Dr. Chavan has been a creative contributor in the field of adult literacy, elementary education, and the development sector in general. His contributions range from scripting and presenting a unique prime-time Mumbai Doordarshan serial "Akshardhara" (1990-91) on adult literacy, penning the Government of Maharashtra's theme song for girls' education, and the most recent 'learning to read' technique to the innovation of replicable and scalable programs for education. Dr. Chavan has been a visiting fellow to the Chinese Academy of Sciences (1986), Queensland University of Technology (1999) and has lectured on issues in education at various universities in the United States. madhavchavan@vsnl.com

6. **Jeff Davis** is the lead assessment specialist in the International Development Program at AIR. He is the project manager for a standards and assessment project in Honduras, and provides technical assistance for projects in Egypt, Malawi, Namibia, Pakistan, and Zambia. Dr. Davis coordinates a team of US-based specialists who collaborate with groups of host country counterparts in assessment-related areas, including research design, curriculum analysis, survey methods, item development, test construction, standardized field administration, statistical analysis, and reporting. He is responsible for creating and maintaining high-quality assessments in various countries. JCDavis@air.org

7. **R. Malatesha Joshi** is a professor of literacy education at Texas A & M University, where he teaches and conducts research relating to language and literacy, assessment and intervention of reading and spelling problems. He is the editor of 'Reading and Writing: An Interdisciplinary Journal' which is ranked as one of the top ten journals in education and educational research by in-cites, a branch of ISI web of knowledge. He also serves on the editorial board of five other journals. He has received funding from NATO to direct institutes on language and literacy during the past twenty years, the most recent one being in Italy in 2001. He is the coauthor/co-editor of over dozen volumes; the most recent is the "Handbook of orthography and literacy" which examines the nature of writing systems in over 25 languages and its influence on literacy acquisition. His papers have appeared in Journal of Learning Disabilities, Journal of Attention Disorders, Neuropsychologica, School Psychology review, and Psychological Reports. mjoshi@tamu.edu

8. **Jules Kinda** is a professor of linguistics specialized in African linguistics and in the terminology and teaching methodology of African languages at the University of Ouagadougou, Burkina Faso. Dr. Kinda also works with OSEO, an NGO that provides bilingual education to rural schools. He has published instructional manuals and dictionaries in the Moore language of Burkina Faso for children and adult learners. He is also a trainer of trainers in adult learning methods. Dr. Kinda has carried out numerous studies in adults' and children's literacy acquisition and measurement in Burkina Faso. jul_kind@yahoo.fr

9. **Peggy McCardle, Ph.D., M.P.H.**, is the Chief of CDBB and directs the Language, Bilingualism, and Bilingual Development and Disorders Program and the Reading, Writing and Related Learning Disabilities research programs. Dr. McCardle holds a bachelor's degree in French, a Ph.D. in linguistics, and a master's degree in public health. She has been a classroom teacher and a speech-language pathologist, and has held university faculty positions at South Carolina State College, the University of Mississippi, the University of Maryland, and the Uniformed Services University of the Health Sciences, and hospital-based clinical

positions at Womack Army Community Hospital, Ft. Bragg, North Carolina, and at Walter Reed Army Medical Center, Washington, D.C. Her publications address various aspects of public health and developmental psycholinguistics (e.g., language development, bilingualism, and reading). She was the NICHD liaison to the National Reading Panel, currently serves as liaison to the National Institute for Literacy, and leads or serves on various inter-agency working groups. She co-edited *The Voice of Evidence in Reading Research* (2004, Brookes Publishing), which presents information about reading research and its findings, for educators, administrators, and others concerned with getting research results into the classroom, and *Childhood Bilingualism* (2006, Multilingual Matters), which addresses research issues in the development of bilingual language abilities, as well as various thematic journal issues on these and related topics. mccardlp@mail.nih.gov

10. **James M. Royer** is a Research Professor of Psychology at the University of Massachusetts, Amherst. He received his PhD from the University of Illinois. He has authored 4 books and 70 journal articles on cognitive approaches to assessment and instruction with a specific focus on reading, mathematical cognition, and the assessment and remediation of students with learning difficulties. He has worked as a consultant on international literacy projects in Grenada, Guatemala, Haiti, Burkina Faso, Nigeria, India, Pakistan and Indonesia. Dr. Royer and his colleagues have also created educational software that is designed to identify and strengthen weak academic skills. royer@psych.umass.edu
11. **Philip Seymour**'s early interest was in experimental cognitive psychology, especially the role of semantic processes in linking verbal and pictorial information (Seymour, P.H.K. *Human Visual Cognition*, West Drayton: Collier Macmillan, 1979). He then specialized in the application of cognitive neuropsychological methods in the description of disturbed reading and spelling processes in childhood dyslexia (Seymour, P.H.K. *Cognitive Analysis of Dyslexia*. London: Routledge & Kegan Paul, 1986). Dr. Seymour's current work focuses on reading acquisition, particularly the role of awareness of large linguistic units (syllables, rimes) and small units (phonemes). He is developing a cross-linguistic approach to these issues in which learning to read and dyslexia in English is contrasted with development in a range of European languages which differ in syllabic complexity and in the depths of their orthographies. phks@edenfield65.freeseerve.co.uk
12. **Sylvia Linan-Thompson**, is an associate professor, Fellow in the Mollie V. Davis Professorship in Learning Disabilities at The University of Texas, Austin, and director of the Vaughn Gross Center of Reading and Language Arts. She develops and examines reading interventions for struggling readers who are monolingual English speakers, English language learners and bilingual students acquiring Spanish literacy. She has authored articles, chapters and a book on these topics and has developed instructional guides. Dr. Linan-Thompson is currently co-principal investigator of studies examining the oral language and literacy development in English and Spanish of Spanish speaking children, the efficacy of a 3-tiered model of reading intervention in general education classrooms and in bilingual classrooms. Dr. Linan-Thompson is associate director of the National Research and Development Center on English Language Learners that is examining the effect of instructional practices that enhance vocabulary and comprehension for middle school English language learners in content areas. sylvialt@mail.utexas.edu
13. **James Wile** is the Director of International Development at the International Reading Association, a non-profit professional network of over 350,000 teachers, teacher educators, librarians, publishers, school administrators around the world. Along with other departments of the Association, the International Development Division creates powerful, capacity-building programs for classroom teachers and policymakers that match the needs of the economically developing countries with the vast professional expertise and information resources of national, regional, and international affiliates of the International Reading Association. Dr. Wile works to develop relationships with professional partners, works with local reading associations in Africa, Europe, Latin America, North America, Oceania, and Asia to develop support systems for literacy activities in developing countries, and obtains and administers grant funding for specific literacy projects. Dr. Wile is responsible for IRA's current teacher education programs in Pakistan,

Macedonia, Ghana, Kenya, and Tanzania. In his role as division director he has traveled extensively to meet with educators, government officials, and classroom teachers for the purpose of creating education alliances and programs that advance and sustain the goal of global literacy. He holds a B.A. from the University of Michigan in Journalism, an M.A. from the University of Michigan in Reading Curriculum, and a Ph.D. from the Ohio State University in Language and Literacy. His professional articles have appeared in *The Reading Instruction Journal*, *The Ohio School Psychologist*, *Language Arts*, *The Teacher Educator*, *Action in Teacher Education*, the *Mediterranean Journal of Education Studies*, and the *Thinking Classroom*.
jwile@reading.org

Observers

1. Rebecca Adams, Education Officer, Office of Education, USAID, RebeccaAdams@usaid.gov
2. Penelope Bender, Senior Technical Advisor, Plan International, penelope.bender@planusa.org
3. Sandra Bertoli, Education Data Analyst, USAID, sbertoli@usaid.gov
4. Barbara Bruns, Lead Economist, Human Development Network, The World Bank, bbruns@worldbank.org
5. Colette Chabbott, Adjunct Faculty, International Education Program, Graduate School of Education & Human Development George Washington University, chabbott@gwu.edu
6. Deon Filmer, Senior Economist, Development Research Group, The World Bank, dfilmer@worldbank.org
7. Cornelia Jesse, Human Development, Africa Region, The World Bank, cjesse@worldbank.org
8. Cheryl Kim, Education Team Leader, Latin America and Caribbean Bureau, USAID, ckim@usaid.gov
9. William Lorie, Human Development Network, The World Bank, wlorie@worldbank.org
10. Catherine Powell-Miles, Africa Bureau, Education Division, USAID, Cmiles@usaid.gov
11. Meshack Moloi, Systemic Evaluation, General Education, South Africa, Moloi.Q@doe.gov.za
12. Dana Schmidt, Quality Education for Developing Countries Initiative, The William and Flora Hewlett Foundation, DSchmidt@hewlett.org
13. Don Sillers, Economist, Office of Poverty Reduction, USAID, Dsillers@usaid.gov
14. Jim Stevens, Senior Operations Officer, Human Development Network, The World Bank, jstevens2@worldbank.org

Organizers/Hosts

1. USAID: Joseph Carney, Director, Office of Education, USAID, jcarney@usaid.gov
2. World Bank:
 - Robin Horn, Education Sector Manager, Human Development Network rhorn@worldbank.org
 - Helen Abadzi, Senior Evaluation Officer, Independent Evaluation Group habadzi@worldbank.org
3. RTI International:
 - Luis Crouch, Research Vice President, International Development Group lcrouch@rti.org
 - Amber Gove, Senior Research Analyst, Education Policy and Systems, International Development Group, agove@rti.org
 - Kyle Snow, Early Childhood Education Program Director, ksnow@rti.org