open access

UNDERSTANDING WHAT WORKS IN ORAL READING ASSESSMENTS

GLOBAL PARTNERSHIP for EDUCATION

THE WILLIAM AND FLORA HEWLETT FOUNDATION

UNESCO
United Nations Educational, Scientific and Cultural Organization

UNESCO INSTITUTE for STATISTICS

UNDERSTANDING WHAT WORKS IN ORAL READING ASSESSMENTS

GLOBAL PARTNERSHIP for EDUCATION

THE WILLIAM AND FLORA HEWLETT FOUNDATION

UNESCO
United Nations Educational, Scientific and Cultural Organization

UNESCO INSTITUTE for STATISTICS

UNESCO

The constitution of the United Nations Educational, Scientific and Cultural Organization (UNESCO) was adopted by 20 countries at the London Conference in November 1945 and entered into effect on 4 November 1946. The Organization currently has 195 Member States and 10 Associate Members.

The main objective of UNESCO is to contribute to peace and security in the world by promoting collaboration among nations through education, science, culture and communication in order to foster universal respect for justice, the rule of law, and the human rights and fundamental freedoms that are affirmed for the peoples of the world, without distinction of race, sex, language or religion, by the Charter of the United Nations.

To fulfil its mandate, UNESCO performs five principal functions: 1) prospective studies on education, science, culture and communication for tomorrow's world; 2) the advancement, transfer and sharing of knowledge through research, training and teaching activities; 3) standard-setting actions for the preparation and adoption of internal instruments and statutory recommendations; 4) expertise through technical co-operation to Member States for their development policies and projects; and 5) the exchange of specialised information.

UNESCO is headquartered in Paris, France.

UNESCO Institute for Statistics

The UNESCO Institute for Statistics (UIS) is the statistical office of UNESCO and is the UN depository for global statistics in the fields of education, science and technology, culture and communication.

The UIS was established in 1999. It was created to improve UNESCO's statistical programme and to develop and deliver the timely, accurate and policy-relevant statistics needed in today's increasingly complex and rapidly changing social, political and economic environments.

The UIS is based in Montreal, Canada.

Published in 2016 by:
UNESCO Institute for Statistics
P.O. Box 6128, Succursale Centre-Ville
Montreal, Quebec H3C 3J7
Canada
Tel: (1 514) 343-6880
Email: uis.publications@unesco.org
http://www.uis.unesco.org

©UNESCO-UIS 2016

ISBN 978-92-9189-196-2
Ref: UIS/2016/LO/TD/9
DOI: http://dx.doi.org/10.15220/978-92-9189-196-2-en

This publication is available in Open Access under the Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) license (http://creativecommons.org/licenses/by-sa/3.0/igo/). By using the content of this publication, the users accept to be bound by the terms of use of the UNESCO Open Access Repository (http://www.unesco.org/open-access/terms-use-ccbysa-en).

The designations employed and the presentation of material throughout this publication do not imply the expression of any opinion whatsoever on the part of UNESCO concerning the legal status of any country, territory, city or area or of its authorities or concerning the delimitation of its frontiers or boundaries.

The ideas and opinions expressed in this publication are those of the authors; they are not necessarily those of UNESCO and do not commit the Organization.

COVER PHOTOS: Left and bottom right, © Dana Schmidt/The William and Flora Hewlett Foundation; top right, © Margarita Montealegre, Nicaragua; centre, © Uwezo, Kenya

BACK COVER PHOTOS: Top, © Margarita Montealegre, Nicaragua; bottom © Dana Schmidt/The William and Flora Hewlett Foundation

The UNESCO Institute for Statistics (UIS) led a collaborative project to formulate recommendations to guide practitioners when selecting, conducting and using oral reading assessments. The aim is to highlight basic principles that should be applied in the different stages of oral reading assessments—from planning and design to implementation and use of the resulting data. The recommendations are drawn from a collection of articles, which can be found online in the ebook, *Understanding What Works in Oral Reading Assessments,* at http://www.uis.unesco.org

Suggested citation
UNESCO Institute for Statistics (UIS) (2016). *Understanding What Works in Oral Reading Assessments: Recommendations from Donors, Implementers and Practitioners*. Montreal: UNESCO Institute for Statistics.

Support for this initiative was generously provided by the Global Partnership for Education and the William and Flora Hewlett Foundation.

Contributors

Organization	Author
Australian Council for Educational Research (ACER)	Marion Meiers Juliette Mendelovits
ASER Centre, Pratham India	Rukmini Banerji Shaher Banu Vagh Savitri Bobde
ASER Pakistan	Sehar Saeed
Concern Worldwide	Karyn Beattie Aine Magee Homayoon Shirzad
Concern Worldwide and University College Dublin	Jenny Hobbs
Creative Associates	Joy du Plessis Fathi El-Ashry Karen Tietjen
Durham University	Christine Merrell Peter Tymms
Education Development Center	Nancy Clark-Chiarelli Nathalie Louge
Instituto para el Desarrollo de la Democracia (IPADE)	Vanessa Castro Cardenal
Juarez and Associates, USAID Lifelong Learning Project	Cristina Perdomo Ana Lucía Morales Sierra Leslie Rosales de Véliz Fernando Rubio
Laboratoire de recherche sur les transformations économiques et sociales (LARTES), Jàngandoo	Binta Aw Sall Abdou Aziz Mbodj Diéry Ba Same Bousso Meissa Bèye Diadji Niang

Ministry of Education, Guatemala	María José del Valle Catalán
Ministry of Basic and Secondary Education, The Gambia	Momodou Jeng
RTI International	Keely Alexander Margaret M. Dubeck Amber Gove Emily Kochetkova
Save the Children	Ivelina Borisova Amy Jo Dowd Elliott W. Friedlander Lauren Pisani
Twaweza East Africa	Izel Jepchirchir Kipruto John Kabutha Mugo Mary Goretti Nakabugo Lydia Nakhone Nakhone
UNICEF	Manuel Cardoso
University College London, Institute of Education	Monazza Aslam
University of British Colombia	Linda Siegel
University of Oregon	Sylvia Linan-Thompson
University of Oxford	Pei-tseng Jenny Hsieh
The William and Flora Hewlett Foundation	Patricia Scheid Dana Schmidt
Women Educational Researchers of Kenya	Joyce Kinyanjui

Table of contents

Foreword 8

Acknowledgments 9

Abbreviations 10

Chapter 1: Introduction 13

Education 2030 and data on learning outcomes 16

Overview of oral reading assessments 18

Chapter 2. Reading Assessments: Context, Content and Design 21

Home literacy environment data facilitate all children reading 22
By Amy Jo Dowd and Elliott W. Friedlander, Save the Children

Teacher quality as a mediator of student achievement 30
By Nancy Clark-Chiarelli and Nathalie Louge, Education Development Center

School-based assessments: What and how to assess reading 41
By Margaret M. Dubeck, Amber Gove and Keely Alexander, RTI International

What and how to assess reading using household-based, citizen-led assessments: Insights from the Uwezo annual learning assessment 58
By Mary Goretti Nakabugo, Twaweza East Africa

Evaluating early learning from age 3 years to Grade 3 66
By Amy Jo Dowd, Lauren Pisani and Ivelina Borisova, Save the Children

Utility of the Early Grade Reading Assessment in Maa to monitor basic reading skills: A case study of Opportunity Schools in Kenya 81
By Joyce Kinyanjui, Women Educational Researchers of Kenya

Learning-by-doing: The Early Literacy in National Language Programme in The Gambia 92
By Pei-tseng Jenny Hsieh, University of Oxford and Momodou Jeng, Ministry of Basic and Secondary Education, The Gambia

Using Literacy Boost to inform a global, household-based measure of children's reading skills 106
By Manuel Cardoso, UNICEF and Amy Jo Dowd, Save the Children

A longitudinal study of literacy development in the early years of school 118
By Marion Meiers and Juliette Mendelovits, Australian Council for Educational Research

Assessing young children: Problems and solutions 126
By Christine Merrell and Peter Tymms, Durham University

Chapter 3. Translating Reading Assessments into Practice 134

Assessing children in the household: Experiences from five citizen-led assessments 135
By John Kabutha Mugo, Izel Jepchirchir Kipruto, Lydia Nakhone Nakhone, Twaweza East Africa and Savitri Bobde, ASER Centre, Pratham India

Assessment in schools 147
By Emily Kochetkova and Margaret M. Dubeck, RTI International

Conducting an Early Grade Reading Assessment in a complex conflict environment: Is it worth it?. 157
By Karyn Beattie, Concern Worldwide and Jenny Hobbs, Concern Worldwide and University College Dublin

Administering an EGRA in a post- and an on-going conflict Afghanistan: Challenges and opportunities 170
By Homayoon Shirzad and Aine Magee, Concern Worldwide

Evaluating reading skills in the household: Insights from the Jàngandoo Barometer 177
By Diéry Ba, Meissa Bèye, Same Bousso, Abdou Aziz Mbodj, Binta Aw Sall and Diadji Niang, Laboratoire de recherche sur les transformations économiques et sociales (LARTES), Jàngandoo

Annual Status of Education Report (ASER) assessment in India: Fast, rigorous and frugal 187
By Rukmini Banerji, ASER Centre, Pratham India

Chapter 4. Using Assessment Data: Interpretation and Accountability 201

Is simple, quick and cost-effective also valid? Evaluating the ASER Hindi reading assessment in India 202
By Shaher Banu Vagh, ASER Centre, Pratham India

USAID Lifelong Learning Project: The Linguistic Profile assessment 213
By Leslie Rosales de Véliz, Ana Lucía Morales Sierra, Cristina Perdomo and Fernando Rubio, Juarez and Associates, USAID Lifelong Learning Project

Use of literacy assessment results to improve reading comprehension in Nicaragua's national reading campaign. 225
By Vanessa Castro Cardenal, Instituto para el Desarrollo de la Democracia (IPADE)

The Yemen Early Grade Reading Approach: Striving for national reform 239
By Joy du Plessis, Karen Tietjen and Fathi El-Ashry, Creative Associates

Assessing reading in the early grades in Guatemala 255
By María José del Valle Catalán, Guatemala Ministry of Education

Expanding citizen voice in education systems accountability: Evidence from the citizen-led learning assessments movement 267
By Monazza Aslam, UCL, Institute of Education, Sehar Saeed, ASER Pakistan, Patricia Scheid and Dana Schmidt, The William and Flora Hewlett Foundation

Chapter 5. Recommendations and Conclusions 280

Recommendation 1: Develop an assessment plan for comprehensive reform 281

Recommendation 2: Collect additional information to understand the context in which teaching and learning take place 289

Recommendation 3: Emphasise the relevant skills—be conscious of differences in culture and orthography of the language. 292

Recommendation 4: Properly organize the implementation of activities—logistics and monitoring 295

Recommendation 5: Cater the analysis and communication of results to the target audience . . . 297

Recommendation 6: Use the data to raise awareness and design interventions aimed at improving teaching and learning 300

Conclusion 305

Glossary 309

Foreword

With the new Sustainable Development Goal (SDG) for education, governments have pledged to ensure that every child is enrolled in school and learning by 2030. The focus in the past on access to school has given way to a clear commitment to deliver on the transformative power of education with an emphasis on learning. Thus, it is no surprise to find that five of the seven education targets highlight learning skills and outcomes of children and adults.

Reading is considered a gateway skill to all other learning. For this reason, governments are increasingly focused on assessing reading among young children—primarily through oral reading assessments, which are no longer restricted to school settings. A growing number of assessment initiatives led by citizens rather than governments are being conducted in households to help fill the gaps in delivering quality education. While there is strong and systematic support from donors for countries to measure oral reading skills, stronger advocacy and better use of resources are needed to improve learning outcomes. Additionally, further development for the generation and use of assessment data to better inform programmes and policies must be encouraged.

In response, the UNESCO Institute for Statistics (UIS) led a collaborative effort among implementers and practitioners to better understand and communicate what works when implementing oral reading assessments and why, within and across countries. The UIS brought together a diverse community of practitioners (including government officials, donors, non-governmental organizations and university researchers) to identify good practices in the design, implementation and use of oral reading assessments through the production of a series of case studies and articles. This ebook presents the complete collection of papers, recommendations and a set of concrete guidelines to improve the collection and use of oral assessment data. The contributions cover experiences in more than 60 developing countries.

By presenting a range of experiences from a collaborative but technically rigorous perspective, *Understanding What Works in Oral Reading Assessments* is uniquely designed to encourage different stakeholders to learn from each other in ways that enhance capacity, ownership and cultural sensitivity while fostering innovative forms of international collaboration.

As the SDGs become a reality, governments will need more and better data to inform policies, take corrective action and monitor progress. Early detection of learning gaps will be essential to guiding remedial action and securing the ambition of the new goal to ensure that all children are in school and learning. This publication serves as a unified voice from the community of oral reading assessment practitioners, implementers and donors on the importance of early reading skills to ensure learning for all by 2030.

Silvia Montoya
Director, UNESCO Institute for Statistics

Acknowledgments

The production of the *Understanding What Works in Oral Reading Assessments* ebook and recommendations report would not have been possible without the commitment and efforts of the authors, organizations and national governments that participated in this project. The recommendations presented here draw upon the wealth of experiences of participating authors and organizations in implementing oral reading assessments. Each article in the ebook provides critical information on good practices in the design, implementation and use of data in oral reading assessments.

The UNESCO Institute for Statistics (UIS) would like to thank all research partners for their support throughout this venture, as well as colleagues within the Global Partnership for Education and the William and Flora Hewlett Foundation who provided vital support and encouragement.

The UIS is grateful to Sylvia Linan-Thompson (University of Oregon) and Linda Siegel (University of British Colombia) for their invaluable input on technical issues.

The UIS thanks all of the authors (see list of contributors) and the peer reviewers for their careful revision: Margaret Dunlop (OISE University of Toronto) and Sheren Hamed (Jordan Education Initiative).

The UIS would also like to thank Maria Elena Brenlla, Nathalie Louge, Sara Ruto, Patricia Scheid and Hannah-May Wilson, who reviewed several articles in the ebook; and Penelope Bender, Luis Crouch and Abbie Raikes for reviewing the recommendations report.

Abbreviations

ACER	Australian Council for Educational Research
AERA	American Educational Research Association
AET	Africa Education Trust
AMCHAM	American Chamber of Commerce
app	Application
AQAP	Al-Qaeda in the Arabian Peninsula
AR	Action Research
ASER	Annual Status of Education Report
BDF	*Banco de Finanzas* (Bank of Finance)
BEACON	Basic Education for Afghanistan
CAPRI	*Centro de Apoyo a Programas y Proyectos* (Support Center for Programs and Projects)
CEC	Community Education Committees
CESESMA	*Centro de Servicios Educativos en Salud y Medio Ambiente* (Centre for Education in Health and Environment)
CETT	Centers for Excellence in Teacher Training
CLP	Community Livelihoods Project
CLSPM	Correct letter sounds per minute
CODENI	*Coordinadora de Niñez* (Childhood Coordinator)
CTT	Classical Test Theory
DIBELS	Dynamic Indicators of Basic Early Literacy Skill
DIDEDUC	Departmental Directorate of Education
DIET	District Institute of Educational Training
DIGEBI	*Dirección General de Educación Bilingüe Intercultural*
Digeduca	*Dirección General de Evaluación e Investigación Educativa*
EA	Enumeration areas
ECCD	Early childhood care and development
ECDE	Early Childhood Development and Education
EDC	Education Development Center
EDI	Early Development Instrument
EDUQUEMOS	*Foro Educativo Nicaragüense* (Nicaraguan Education Forum)
EFA	Education for All
EGMA	Early Grade Math Assessment
EGRA	Early Grade Reading Assessment
ELGI	*Evaluación de Lectura en Grados Iniciales* (Reading Assessment for Initial Grades)
ELI	*Evaluación de Lectura Inicial*
ELINL	Early Literacy in National Language
ELM	Emergent Literacy and Math
ERDC	Education Research Development Center

FAS	*Fonético Analítico Sintético*
FIOH	Future in Our Hands
FOI	Fidelity of implementation
GARA	Group Administered Reading Assessment
GATE	Gambia Association of Teaching English
GCC	Gulf Cooperation Council
GCE	Global Campaign for Education
GDP	Gross domestic product
GMR	Global Monitoring Report
GPE	Global Partnership for Education
GPI	Gender parity index
HLE	Home Literacy Environment
ICC	Intraclass correlation coefficient
IDEL	*Indicadores Dinámicos del Éxito en la Lectura* (Dynamic Indicators of Reading Success)
IDELA	International Development and Early Learning Assessment
INE	National Institute of Statistics (Spanish Acronym)
INSET	In-Service Teacher Training
IPADE	*Instituto para el Desarrollo de la Democracia* (Institute for the Development of Democracy)
iPIPS	International Performance Indicators in Primary School
IRC	International Rescue Committee
IRR	Inter-rater reliability
IRT	Item Response Theory
ITA	Idara-e-Taleen-o-Aagahi
J-PAL	Abdul Latif Jameel Poverty Action Lab
KCPE	Kenya Certificate of Primary Education
L1	First language or native speakers
LARTES	*Laboratoire de Recherche sur les Transformations Économiques et Sociales* (Research Laboratory on Economic and Social Transformations)
LEE	Evaluation of Early Reading and Writing (Spanish Acronym)
LLANS	Longitudinal Literacy and Numeracy Study
LLECE	*Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación*
MDG	Millennium Development Goals
MFC	Mother Father Council
MIA	Medición Independiente de Aprendizajes
MICS	Multiple Indicator Cluster Survey
MINED	*Ministerio de Educación Cultura y Deportes*
MOBSE	Ministry of Basic and Secondary Education
MOE	Ministry of Education
MOEST	Ministry of Education, Science and Technology
MSA	Modern Standard Arabic
NAPE	National Assessment of Progress Education
NAS	National Achievement Survey
NAT	National Assessment Test

NER	Net enrolment rate
NGO	Non-governmental organization
NL	National language
NRP	National Reading Panel
OECD	Organisation for Economic Co-operation and Development
ORF	Oral reading fluency
OSP	Opportunity Schools Programme
PAL	People's Action for Learning
PALME	*Partenariat pour l'Amélioration de la Lecture et des Mathématiques à l'École*
PAMI	Evaluation of Early Mathematics Skills (Spanish Acronym)
PASEC	Programme d'Analyse des Systèmes Educatifs de la CONFEMEN
PILNA	Pacific Islands Literacy and Numeracy Assessment
PIPS	Performance Indicators in Primary Schools
PIRLS	Progress in International Reading Literacy Study
RESP	Rural Education Support Programme
RI	Read India
RTI	Research Triangle Institute
RWC	Reading with comprehension
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SCOPE-Literacy	Standards-based Classroom Observation Protocol for Educators in Literacy
SD	Standard deviation
SDG	Sustainable Development Goal
SEGRA	Serholt Early Grade Reading Ability
SES	Socio-economic status
SIL	Summer Institute of Linguistics
SMS	Short message service
SNERS	*Système National d'Evaluation du Rendement Scolaire*
SPSS	Statistical Package for the Social Sciences
SSA	Sarva Shiksha Abhiyan
T'EGRA	Teachers' Early Grade Reading Assessment
T'EGWA	Teachers' Early Grade Writing Assessment
TIMSS	Trends in International Mathematics and Science Study
ToTs	Training of Trainers
TPOC	Teacher Performance Observation Checklist
UIS	UNESCO Institute for Statistics
UNFPA	United Nations Population Fund
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development
UVG	*Universidad del Valle de Guatemala*
WCPM	Words correct per minute
WDR	World Development Report
WERK	Women Educational Researchers of Kenya
YEGRA	Yemen Early Grade Reading Approach

© Margarita Montealegre, Nicaragua

Chapter 1
Introduction

This chapter introduces oral reading assessments and situates their importance within the Education 2030 agenda. It presents the *Understanding What Works in Oral Reading Assessments* initiative and the process to produce the ebook.

Introduction

With the Sustainable Development Goals (SDGs), the international community has pledged to ensure that every child is in school and learning by 2030. Reading is a gateway skill to all other learning, which is why governments are increasingly using oral reading assessments to evaluate and improve the skills of young children. By detecting reading weaknesses early in a child's educational experience, the resulting data can be used to better direct policies and interventions before it is too late.

To promote the use of these assessments and their results, the UNESCO Institute for Statistics (UIS) has brought together a wide range of organizations that are leading the development, implementation and financing of oral reading assessments conducted in schools and households. Through a collaborative but technically rigorous process, they have identified common practices and effective strategies to design, implement and use these tools for effective policymaking based on experiences in more than 60 developing countries. The results are presented in this ebook.

With contributions from more than 50 experts in 30 organizations, the ebook presents a series of articles highlighting good practices in executing effective oral reading assessments—from planning and design to implementation and use of the resulting data. The ebook is uniquely designed to encourage different stakeholders to learn from each other in ways that enhance capacity, ownership and cultural sensitivity, while fostering innovative forms of international collaboration.

The ebook also presents a comprehensive set of recommendations based on the experiences of the authors in non-governmental organizations, academic organizations, ministries of education, donors, international organizations and civil society groups.

THE SHIFT IN EDUCATIONAL REFORM

Over the last decades, much progress has been made toward ensuring that all children have access to quality education. Despite this progress, considerable challenges remain: 124 million children and youth are out of school (UIS database, 2016) and many more millions of children who are in school are not learning. Research studies and results from learning assessments have exposed the causes of educational failure. These include untrained teachers and absenteeism; mismatches between the language of instruction and children's mother tongue; grade repetition and dropout; children who were never enrolled in school; malnutrition; and more (Sillers, 2015). In many developing countries, a large number of children never start school or drop out, while many of those who do complete their primary education and graduate do so without acquiring the basic skills required to function in society.

In the last 15 years, the focus of educational reform has been gradually shifting from increasing school attendance to improving the quality of education. The shift in focus to instructional quality has been driven in large part by learning assessment results. Although large-scale international and regional assessments have demonstrated for years that children in developing countries were not learning at the same rate as their counterparts in Western countries, the recent move to assess reading skills in primary school has helped mobilise reform efforts. Since 2009, the number of countries around the world that have collected assessment data to measure early reading skills

has increased exponentially through assessments with non-representative sample sizes (studies, impact evaluations, project benchmarks) and those administered at the system-level (examinations, participation in regional or cross-national initiatives and implementing a national learning assessment).

INTRODUCING ORAL ASSESSMENTS

Although there are many types of learning assessments, this report focuses on standardised measures that are designed, administered and scored in a consistent manner and are criterion referenced. In essence, they measure what children are expected to know and be able to do. The assessments are individually administered one child at a time and are direct assessments of foundational skills for learning. We refer to them as oral assessments because children respond orally—usually to written stimuli. Administering an assessment orally is more inclusive as this method allows all children to participate—even those who are not literate. Governments do not necessarily organize the administration of the assessments; generally, there are many partners involved in the different stages of the assessment process.

Although the assessments are not explicitly based on the education curriculum in particular countries, they are often compatible with the curriculum as they measure key components of reading and/or numeracy skills acquisition. This report focuses on oral reading assessments.

The use of oral assessments to measure children's reading development has been instrumental in shifting the focus of educational reform to one that emphasises system accountability, improved instruction and the identification of student learning needs. Unlike international (e.g. PIRLS) and regional assessments (LLECE, PASEC, PILNA, SACMEQ), oral assessments can be—relative to policy impact—smaller, quicker and cheaper (Wagner, 2011) to design and administer in local languages. These are critical features in settings where children enter school speaking a number of different languages and funds for conducting assessments may be limited. Further, results are actionable, targeted to early reading and are usually available for dissemination in a shorter timeframe compared to regional or international assessments. It is these last three characteristics that have contributed to the impetus needed to change the focus of educational reform from access to education to quality of instruction and student learning outcomes.

It is important, however, to recognise the limitations of oral reading assessments. First, they are resource intensive in terms of staff required to complete the process. Second, they are time consuming as they involve training several groups of individuals to perform the various tasks required. Third, the reading comprehension measures are limited and may not discriminate among students for several reasons: there are few items; the test generally allows lookbacks; and the questions included are typically explicit and inferential so do not involve interpreting, integrating ideas and information, or evaluating and critiquing content.

Education 2030 and data on learning outcomes

With the adoption of the Sustainable Development Goals (SDGs), governments have pledged to ensure that every child is enrolled in school and learning by 2030. The focus on learning outcomes is a shift from the Millennium Development Goals (MDGs), which focused on ensuring access to, participation in and completion of formal primary education (UNESCO-TAG, 2015).

Policymakers at the global and national levels clearly recognise the importance of determining whether the quality of education is improving and the role that the monitoring of learning outcomes plays in achieving this end. It is not enough to know how many children are enrolled in school or how many teachers are hired to reach the SDGs. They need to know whether children possess the basic reading and mathematics skills essential to future learning. They need to know what children can and cannot do early on to ensure that there are policies and practices in place to support early intervention and remediation. Waiting until the end of primary education to ascertain learning levels will be too late for many children.

To help transform this promise into action, governments will need more and better data to identify areas of improvement, install change and monitor progress. The good news is that through household surveys, learning assessments and research studies, educators, administrators and other stakeholders have been engaged in answering questions, such as: What are children learning? Where are they learning? And who is being left behind?

The ability to read is essential for progress in the education system. Having relevant, high-quality early grade literacy data is a crucial step in attaining this goal. Although assessment is vital to guiding government policy and changes to instruction, it alone is not enough. Data should be analysed and governments should continuously evaluate their policy agendas, school-level implementation and progress through the use of assessments and their results to ensure that all children are learning.

A FOCUS ON READING

The SDG for education calls for monitoring learning outcomes, and several indicators in the Education 2030 Framework for Action specifically refer to reading. Reading is considered a gateway skill to all other learning. Children who fail to develop appropriate reading skills in the first few years of schooling are likely to continue to lag behind their peers (Juel, 1988). In low income countries, these children often drop out of school before completing primary education. Thus, ensuring that all children learn to read has served as the impetus for assessing reading in the early years of schooling—primarily, through oral reading assessments. Although there is consensus that reading is an important skill, there is, however, less agreement on what skills should be assessed and how they should be assessed.

SHARING EXPERIENCES TO UNDERSTAND WHAT WORKS IN ORAL READING ASSESSMENTS

Given the focus on reading and on trying to guarantee early success as a contribution to primary school completion, many organizations have started using one-on-one oral assessments that involve printed stimuli. The rationale for using oral assessments as opposed to written assessments will be described throughout this report. Of these, a few warrant mentioning at the outset.

First, participation in most pencil and paper assessments requires some word reading ability so if many children are not able to respond, there will be very low discriminant capacity at the lower end of the scale. Also, given the relative informality of many school settings, it is possible that in group assessments, especially if teachers are present or the assessment content is leaked, children may be coached or even helped during a group-administered, pencil-and-paper test. Assessments that are orally administered, one-on-one, by individuals who are from outside the school, help circumvent some of these problems. In addition, oral assessments can assess very basic oral skills such as phonological awareness and basic literacy skills, such as letter knowledge.

For these reasons, the use of oral assessments has become relatively widespread. Despite some commonalities among the instruments used, there are also differences in the purpose, design and administration of these assessments. Given the wide array of assessments available to practitioners, the UNESCO Institute for Statistics (UIS) led a collaborative effort with organizations that have been actively financing, designing and implementing oral assessments *(see **Box 1**).* Representatives from these organizations were asked to submit case studies and position papers that exemplify good practices. The information from these papers was then synthesised and used to derive the resulting recommendations.

It is hoped that these recommendations will provide the field with a set of concrete guidelines to improve data collection and their use.

Box 1. Collaborators of *Understanding What Works in Oral Reading Assessments*

- 30 organizations
- 50 contributors
- Combined experiences from more than 60 developing countries

The methodology of this collaborative exercise drew on the following principles:

1. *Moving towards consensus.* Being a consensus-building exercise, the organizations' own know-how served as the starting point. Experiences were shared and different perspectives were compared.

2. *Focus on identifying balance between cultural specificity and global applicability.* Maintaining equilibrium between these two principles and addressing the challenge of identifying the culturally specific lessons that apply only to certain regional, linguistic or cultural contexts was deemed important. Equally important is the goal to identify overall principles that may apply to a wide variety of developing contexts.

3. *Parsimony.* It was key to emphasise the importance of streamlining and simplifying assessment instruments and methodologies without incurring a loss of precision and explanatory power as these are relevant to policymaking.

The 20-month process that culminated in the development of these recommendations can be summarised in **Figure 1**.

Figure 1. Development phases of the oral reading assessments recommendations

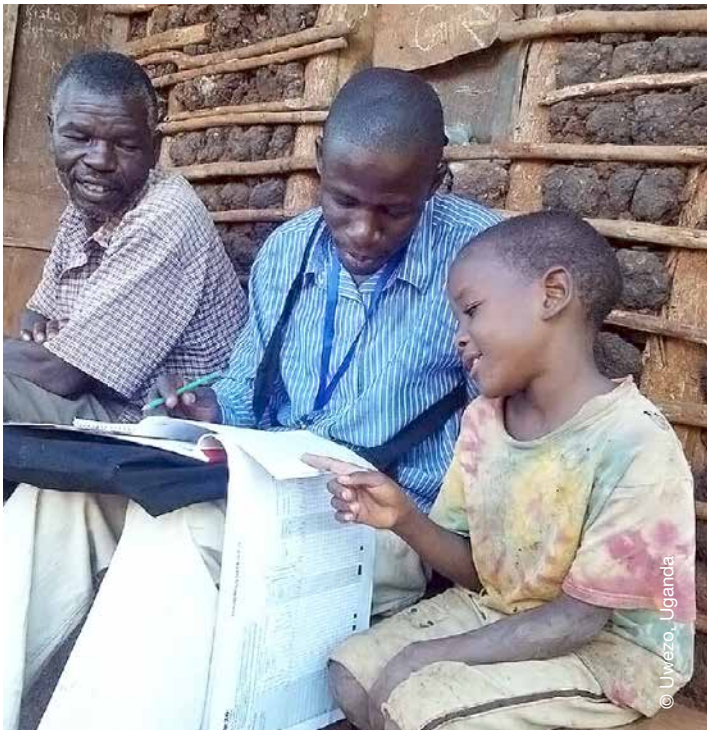
Overview of oral reading assessments

Assessment, in educational contexts, refers to a variety of methods and tools that can be used to evaluate, measure and document learning progress and skill acquisition *(see **Box 2**).* In addition to providing information on current student achievement, empirical data can be used to determine teaching instruction quality, identify students' learning needs or evaluate language ability. The most common use of oral reading assessments is to determine students' current level of performance. These data often serve as a baseline for specific interventions or generalised reform efforts.

Box 2. Commonalities among oral reading assessments

Although oral reading assessments are designed for different purposes, they share some characteristics. Any given assessment is typically a standardised measure that is designed, administered and scored in a consistent manner and is criterion referenced. The assessments measure what children are expected to know and be able to do. They are individually administered, direct assessments of key components of reading skills acquisition. Most often, these are assessments of learning (i.e. they are designed to inform stakeholders and not teachers).

Once a need for reform has been established and an intervention is implemented, oral assessments can serve as an outcome measure to determine the effect of the intervention. When assessments are used to determine the effect of an intervention, it serves as an evaluation tool. According to Fenton (1996), 'evaluation is the application of a standard and a decision-making system to assessment data to produce judgments about the amount and adequacy of the learning that has taken place'. Essential to this process is the availability of


© Uwezo, Uganda

standard or normative scores that provide parents, educators, administrators and donors with an index by which to judge whether learning progress is meaningful. This section will provide an overview of the different types of oral assessments.

ACCOUNTABILITY ASSESSMENTS

Accountability assessments are used to report to the public and other stakeholders on educational trends and to demonstrate the effectiveness of the education system in serving children and in meeting the needs of the community and state.

Citizen-led assessments

Citizen-led assessments are generally those that are led by citizens or civil society organizations

TABLE **1**

Citizen-led assessments

Citizen-led assessment	Country	Target population (assessed children)	Year initiative was launched
ASER	India	5—16 years old	2005
ASER	Pakistan	5—16 years old	2008
Beekunko	Mali	6—14 years old	2011
Jàngandoo	Senegal	6—14 years old	2012
Uwezo	Kenya	6—16 years old	2009
Uwezo	Uganda	6—16 years old	2009
Uwezo	United Republic of Tanzania	7—16 years old	2009

Note: Other citizen-led assessments include Medición Independiente de Aprendizajes (MIA) launched in Mexico in 2014 and LEARNigeria in Nigeria in 2015; the assessments target children aged 5-16 years and 5-15 years, respectively. LEARNigeria, similarly to ASER-India, also surveys all children aged 3-15 years yet only those aged 5 or older are assessed. Both MIA and LEARNigeria are not yet administered to a nationally-representative sample of children.

Source: adapted from (Aslam et al., 2016) and the UIS Catalogue of Learning Assessments, 2016

rather than by governments *(see **Table 1**).* They are conducted in households rather than in schools and measure basic reading and numeracy skills. Citizen-led assessments can provide recurrent estimates of children's basic learning levels and (so far) tend to be similar in design and administration. Citizen-led assessments are a different model of assessment. Rather than being in the hands of a limited number of professionals, the community has a stake in administering and interpreting these assessments.

Volunteers administer the measurement tools that assess children's reading skills in homes or communities. Children's reading levels are typically characterised as being either at the letter, word or passage level (often two passages with varying levels of difficulty are included in the assessment). This approach allows stakeholders to track changes in the number of students at each level over time.

The results from citizen-led assessments are used for accountability and advocacy by (see article by Aslam et al.):

- generating nationally representative and locally owned data on acquisition of foundational skills that are not dependent on school access;
- helping re-orient the debate from school access to improved learning for all;
- creating new opportunities for citizens to better understand the status of their children's learning so that they can decide for themselves whether governments are delivering on promises related to equity and quality in education;
- promoting new mechanisms for evidence-based policy, proven programme interventions and actions to improve learning.
- creating a sense of community and shared purpose.

There are two further points worth noting about citizen-led assessments. First, while citizen-led assessments have mostly been used for generating accountability pressure, it typically has not been a high-stakes accountability pressure tied to particular teachers or schools. Rather, their main purpose has usually focused on education-system accountability or overall community-based accountability. In addition, they have also been used in the classroom to group children by skill and to place them at the right level, rather than based on grade or age or curricular expectations. The approach of teaching at the right level is currently gaining some traction among educators in developing countries.

School-based assessments

A second type of accountability assessment is the school-based oral assessment. The most commonly used is the Early Grade Reading Assessment, which has also been used settings other than in schools. Other widely used school administered

assessments include the Initial Grades Reading Evaluation (EGLI in Spanish) and Literacy Boost. These assessments are administered in schools and results are often used to advocate for educational reform. In the reform process, stakeholders use data from these assessments to make decisions on the use and effectiveness of resources, personnel and institutions. Reform efforts initiated after data have been collected on a national sample often include changes in instructional approaches and curriculum, textbook development and resource allocation. Although one could classify these assessments as driving accountability, it is important to note that the accountability sought here is at the level of the teacher support system, the system that provides learning materials to learners and the overall policy. Few, if any, of these assessments are used to assess individual teachers and as a matter of fact, they are designed to be sample-based assessments that do not identify individual teachers or learners.

Since literacy begins before formal schooling, assessments, such as the International Development and Early Learning Assessment (IDELA) (see article by Dowd et al.) and the Performance Indicators in Primary School (PIPS) (see article by Merrel and Tymms), seek to identify which skills children possess prior to beginning formal primary education. Like the measures used with school-aged children, results from these assessments provide data on children's level of skill acquisition and can be used to improve early childhood programmes. Measures designed to assess children's knowledge and skills at school entry can also provide Grade 1 teachers with information on children's relative learning performance that can be used to plan instruction to support all learners.

School-based oral reading assessments have also been used as outcome measures in the evaluation of intervention projects in a variety of contexts. The data are collected at two or three points during the span of a project. When used as a formative measure, students are assessed while the intervention is being implemented and results are used to make programmatic changes. The use of data to make decisions is critical when implementing a new instructional approach. However, at the end of the project, results of the summative assessment are used to determine the effect of the intervention or reform effort. Literacy Boost, for instance, a well-known reading intervention, has been implemented in a number of countries and in 35 languages. Results from the Literacy Boost assessments are used to shape and evaluate the implementation of Literacy Boost programmes.

REFERENCES

Juel, C. (1988). "Learning to read and write: A longitudinal study of children in first and second grade". *Journal of Educational Psychology*. Vol. 80, pp. 437-447.

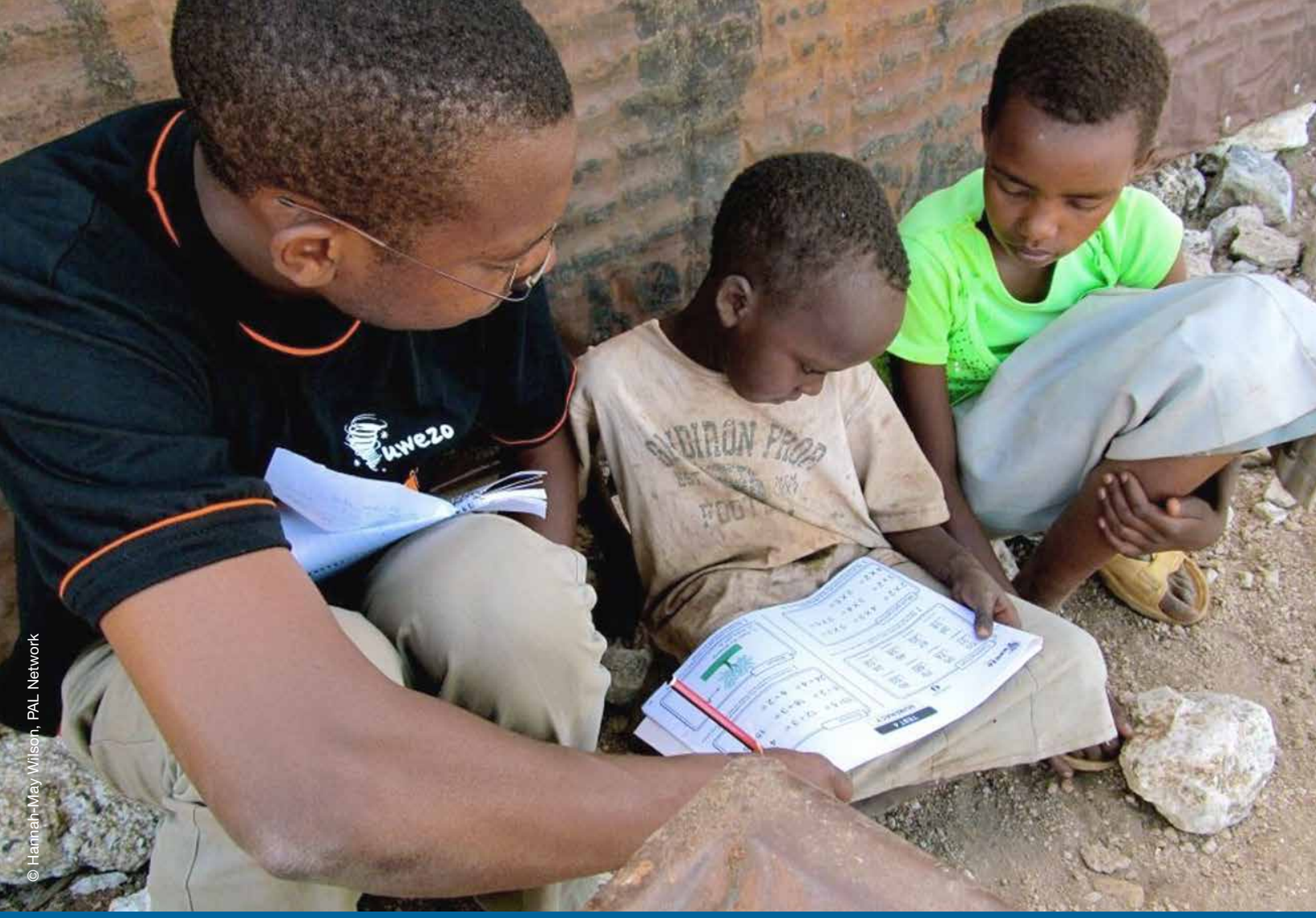
Sillers, D. (2015). USAID presentation at the 2015 Global Education Summit. https://www.usaid.gov/sites/default/files/documents/1865/Sillers.pdf. (Accessed January 2016).

UNESCO Institute for Statistics Catalogue of Learning Assessments. http://www.uis.unesco.org/nada/en/index.php/catalogue/learning_assessments. (Accessed January 2016).

UNESCO Institute for Statistics Database. http://www.uis.unesco.org/. (Accessed January 2016).

UNESCO TAG (2015). *Technical Advisory Group Proposal: Thematic Indicators to Monitor the Education 2030 Agenda*. Paris: UNESCO. http://www.uis.unesco.org/Education/Documents/43-indicators-to-monitor-education2030.pdf

Wagner, D.A. (2011). "Smaller, Quicker, Cheaper: Improving Learning Assessments to Developing Countries". Paris: UNESCO-IIEP. http://www.literacy.org/sites/literacy.org/files/publications/213663e.pdf



© Hannah-May Wilson, PAL Network

Chapter 2
Reading Assessments: Context, Content and Design

The articles in this chapter describe the types of assessments used to measure early reading skills. The advantages and challenges of using various techniques are described. Suggested strategies to collect additional information alongside reading assessments are provided.

Home Literacy Environment Data Facilitate All Children Reading

AMY JO DOWD AND ELLIOTT W. FRIEDLANDER
Save the Children

ABBREVIATIONS

HLE	Home Literacy Environment
SES	Socio-economic status

1. INTRODUCTION

Jolly is an 8-year old girl who is completing her first year of primary school in Rwanda. When Jolly returns home from school each day, her mother makes sure she completes her homework, and her father and Jolly read together. When there is free time, Jolly sings and plays cards with her six brothers and sisters. Flora is a 9-year old girl also completing her first year of school. She lives in the same district of Rwanda as Jolly. When Flora gets home, she first fetches water, then collects kindling, then cooks dinner for her family. No shared reading occurs because, according to her father, there is nothing in the house to read. Even if there were, Flora's life is so busy that she only completes her homework with friends while walking to school (Tusiime et al., 2014).

Despite living close to one another, being of the same age and grade, speaking the same language, reading the same textbooks and being taught by similar teachers, Flora and Jolly will have drastically different experiences at school. Regardless of what curricula are used or which skills are emphasized in the classroom, the daily experiences that Flora and Jolly have at home and in the community will affect their motivation, learning and development.

As we gather oral reading assessment data to better understand how to help learners, it is critical to collect data on the learning environment. A thorough mapping of children's learning environment—both inside and outside schools—provides an empirical foundation for building better learning interventions. With greater insight into the supports and obstacles to learning that children experience throughout their lives, we can design and improve programmes that can meet the diverse needs of all learners.

In this article, we offer a field-tested method to add learning environment data to improve the quality and utility of oral reading assessment data collection and analysis. We do so by first defining the Home Literacy Environment (HLE) and briefly reviewing its empirical relationship to learning, with special attention to studies in both developed and developing world contexts. Then, we describe how we measure the HLE in developing world contexts and how we analyse these data to inform efforts to improve learning in the developing world.

2. WHAT IS HLE?

Hess and Holloway (1984) define the HLE in five dimensions: 1) the value placed on reading, 2) the press for achievement, 3) the availability of reading materials, 4) reading to children, and 5) opportunities for verbal interaction. While this definition of the HLE as a predictor of reading skills in children prevails in the developed world where plentiful print and readers are common, it lacks two things: consideration of children's interest in

and motivation to read as well as the roles that neighbors, extended family and community may play in providing opportunities to read and be read to. These determinants of opportunities to read and amounts of reading practice may be particularly salient characteristics of a literacy environment in the developing world. Thus, while Hess and Holloway's HLE framework is a central feature of Save the Children's best practice, we also acknowledge that this frame is improved by capturing children's interest in and motivation to read as well as accounting for the varied places and people with whom opportunities to learn occur beyond the school walls (Dowd, 2014; Friedlander et al., 2016).

3. IS THE HLE RELATED TO LEARNING?

The preponderance of evidence that proves the relationship between the HLE and children's academic achievement comes from developed country settings (Hess and Holloway, 1984; Snow et al., 1998). There are, however, developing country studies that verify this link. In this section, we review the strong evidence of the relationship between the HLE and learning in developed world contexts and the emerging evidence of its different yet nonetheless positive association with learning in developing world contexts.

The links between language and literacy in the home and a child's school performance and reading achievement in particular is well documented in developed world contexts (Bradley et al., 2001; Hart and Risley, 1995). Across samples of different ages, socio-economic statuses, languages, and many different measures of literacy-related skills and abilities, the trend is clear: the more supportive the HLE, the better the child's reading achievement (Bus et al., 2000; Snow et al., 1998). In fact, Taylor (1983) even challenged "whether we can seriously expect children who have never experienced or have limited experience of reading and writing as complex cultural activities to successfully learn to read and write from the narrowly defined pedagogical practices in our schools". This question posed in the

United States more than three decades ago remains relevant today across the globe.

In the developing world, several studies find empirical links between the HLE and learning (Chansa-Kabali and Westerholm, 2014; Kabarere et al., 2013; Kalia and Reese, 2009; Wagner and Spratt, 1988). In addition, studies reporting on students' motivation and voluntary reading at home also found positive links between HLE and reading achievement (Abeberese et al., 2014; Elley, 1992). Studies by Save the Children conducted largely in rural areas of developing countries measured the HLE as books in the home, verbal interactions, models of independent reading, shared child-family reading, and help or encouragement to study. Analyses found generally consistent links between reading skills and student reported measures of the HLE (Dowd and Pisani, 2013; Friedlander et al., 2012). Additional indicators of motivation and reading skills used in Malawi significantly predicted all reading skills even when controlling for socio-

economic status, gender, repetition and age (Save the Children Malawi, 2013).

The evidence suggests that to better understand the development of reading skills in developing world contexts, it is necessary to collect and analyse data that represent the five dimensions of the HLE, children's motivation to read and children's opportunities for reading practice inside and outside both the home and the school. Including these elements will provide us with a better understanding of and a broader evidence base that more appropriately represents the rich variety of learning environments in different languages, cultures, physical environments and living situations around the world. Measuring the HLE and children's interest and motivation will help us investigate and eventually improve our definitions of 'best practices' to support reading achievement.

4. HOW TO COLLECT HLE ALONGSIDE ORAL READING ASSESSMENT DATA

Save the Children began collecting oral reading assessment data in 2007. Since that time, HLE data collection shifted as we discovered the need for a broader framework in developing world contexts. In 2009, we merely asked children whether or not there were books at home and whether any reading occurred at home. From the resulting data, we saw strong associations between the presence of books and readers and reading achievement. We next added questions on book variety and whether there were individuals who could read at home, and in 2011, we began to collect data that specifically mapped onto all of Hess and Holloway's five dimensions. With each increase in the level of HLE specification, our understanding of its links to the variation in children's reading skills grew. In 2013, further exploration beyond the HLE, namely motivation and use of skills beyond school walls, demonstrated the need for greater information on literacy in the lives of children. Current Save the Children's best practice in collecting HLE data uses a survey of family members and activities as well as follow up questions to capture information on the motivation for reading and literacy use outside the home.

To collect the data described above, an assessor first establishes a friendly rapport with the sampled child and collects informed assent to participate in the study. Following this, the assessor asks the child background questions, including what types of books are found in their home. Country teams develop a list of relevant types of reading materials on which to inquire in a given context, which generally includes textbooks, newspapers, magazines, religious books, storybooks, coloring books and comics. Then, the assessor asks the child, 'Who do you live with?' As the child responds, the assessor fills in the boxes in the matrix shown in **Figure 1**. For each person the child names, the assessor asks whether the child saw the person reading during the last week, whether the person told them or helped them to study in the last week, etc. As the child responds, the assessor records '1' for yes and a '0' for no in the matrix.

Over time, we have determined that questioning no more than eight family members sufficiently captures the majority of families in contexts where we work—an extremely small percent of sampled children live in homes with more than eight members. Our field teams have conveyed the two ways to collect this data efficiently. The first is to fill in the data column by column while the second is to fill in the first column, then ask about the literacy habits of each member (e.g. 'Do you see Mom read? Does Mom read to you?'). Depending on the number of family members a child has and the rapidity with which a child responds, collecting this data adds an additional five to seven minutes to our oral reading assessments.

5. HOW TO ANALYSE HLE AND ORAL READING ASSESSMENT DATA

The data collected from the matrix and questions listed in the last section enable several different types of analyses. First, we can very simply model the relationship between binary descriptions of the HLE and reading achievement. This allows us to answer questions such as 'What is the relationship between

Figure 1. HLE survey matrix

Name/initials	Relationship 1-Mom, 2=Dad, 3=Sister, 4=Brother, 5=Grandma, 6=Grandpa, 7=Other Female, 8=Other Male	Seen reading 1=YES, 0=NO	Told/helped you to study 1=YES, 0=NO	Read to you 1=YES, 0=NO	Told you a story 1=YES, 0=NO

Other than at school, did anyone outside your home read to you last week?	__No (0)	__Yes (1)
Other than school, did you read to anyone outside your home last week?	__No (0)	__Yes (1)
Other than at school, did you read alone last week?	__No (0)	__Yes (1)
In the last week, did you use your reading skills outside of school?	__No (0)	__Yes (1)
If yes, where? _____		__Yes (1)
In the last week, have you helped anyone using your reading skills?	__No (0)	__Yes (1)

reading to children and children's reading abilities?' or 'How are the presence/absence of books at home associated with reading achievement?'

The family member-specific data also enable more sophisticated analyses that investigate how reading achievement is predicted by the number of readers at home, the amount of reading a child is exposed to, the saturation of literacy habits in the home or even patterns of who reads related to reading achievement. Also, data on the types of books at home enables investigation into how different materials may or may not predict reading achievement. For instance, we can consider the presence of child-appropriate materials only (e.g. storybooks, comics) as an interesting subset linked to learning to read. When examining the overall relationship between the HLE and reading, we often combine the variables relating to the HLE into one or two sub-indices representing materials/activities and motivation/use. Collecting data on all of these aspects of children's literacy environments outside of school offers rich possibilities to advance our understanding of children's reading development and our efforts to improve reading globally.

6. HOW CAN HLE DATA INFORM EFFORTS TO IMPROVE LEARNING?

Given the importance of the HLE in the development of children's reading skills, Save the Children always measures it alongside reading skills when a study intends to consider influential factors or the impact of specific factors. This is important because it helps define baseline opportunities and challenges, enables accurate estimates of factors that influence learning and facilitates analyses of equity in impact.

HLE data in and of itself provides key details into the access to reading materials and opportunities to read that children have at baseline. This insight

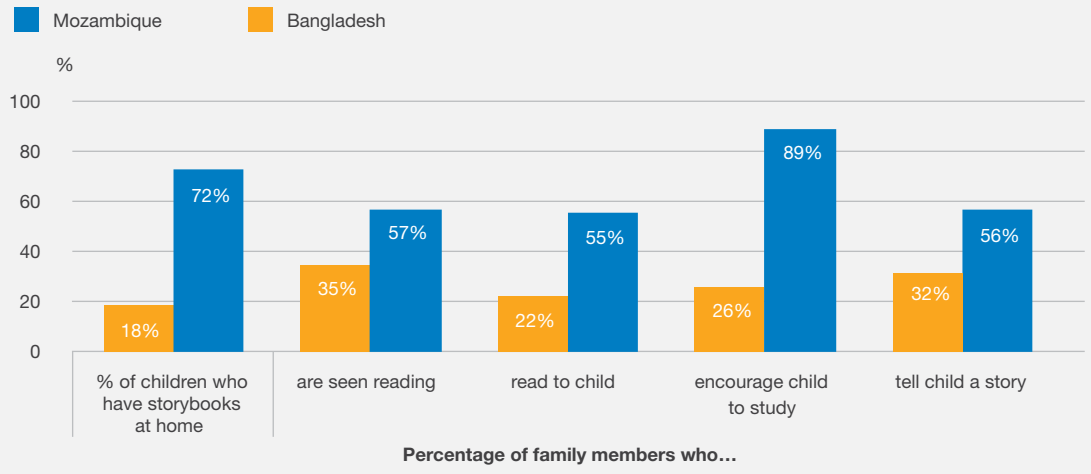
can illuminate challenges that might otherwise be overlooked. For example, if there are few or no children's books or very few people seen reading at home, interventions can shift focus to provide more books or to identify community reading mentors to support children who come from poor HLEs. HLE data can also reveal opportunities such as a setting in which most children report their parents already read to them. **Figure 2** shows simple HLE profiles that set the stage for interventions in Nacala, Mozambique and Meherpur, Bangladesh. Comparing these two very different contexts for learning to read reveal the different challenges that children may face.

The significantly (p=0.001) higher percentage of families that engage in these practices in Bangladesh signals more opportunities to build from in Meherpur. Conversely, there are greater challenges in Nacala, Mozambique but there is also greater room for growth. These HLE data set the stage for interventions by teams in each country by indicating the level of learning environment support outside the school walls.

Including HLE data in analyses also clarifies the relationship between other student-level characteristics and reading factors. For example, if, we find that girls in Region X have significantly higher average reading skills than boys in the same region, we may draw the wrong conclusions if HLE data is not included. Perhaps the relationship may be explained by the fact that girls in Region X have more access to books and readers in the home. Without HLE data in analyses, we might mistakenly conclude that there is something intrinsic about girls or about the way people treat girls that makes them read better. We would also miss the fact that students with poor HLEs are not receiving the support they need to succeed. When we account for access to books and readers in statistical models of reading achievement, it enhances our contextual understanding of the supportive culture for learning to read. It further helps identify potential important target groups for intervention as well as possible remedies to help struggling readers outside of the school.

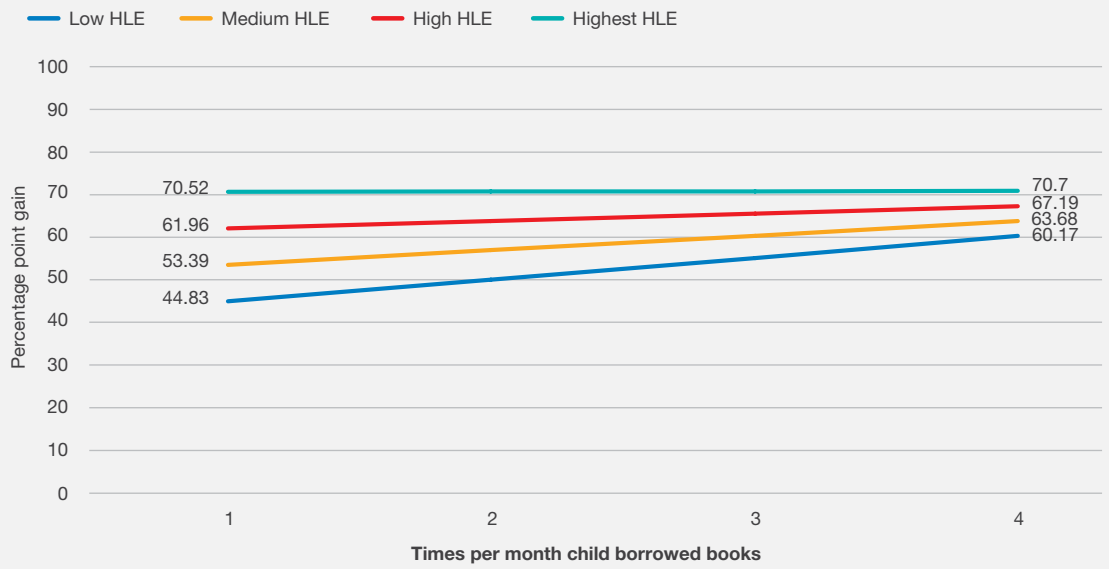
It is very common to collect information on sex and socio-economic status (SES). These two characteristics are often considered key factors that influence learning. In our experience, the HLE is also a key predictor of reading. A seven-country study of factors that influence reading found that, controlling for age, sex, SES and HLE as well as early childhood participation and chore workload (as available), the HLE significantly predicted reading skills in nearly a

Figure 2. HLE materials and activities in Bangladesh and Mozambique



Source: Save the Children Bangladesh (2013) and Mozambique (2014)

Figure 3. Gain in Pashto reading accuracy by HLE and book borrowing frequency in Pakistan



Source: Mithani et al., 2011

third of the 43 multivariate models fitted, while SES did so in 16% and sex in only 4% (Dowd et al., 2013).

Finally, without HLE data, we miss the opportunity to understand impact and equity. For example, the chance to determine if an intervention helped all children equally and not just those with supportive HLE backgrounds. Even more important, if children from deprived HLE backgrounds had lower average scores before an intervention or policy change, we could determine if the shift closed that gap and if not, what else might be needed to achieve this goal. **Figure 3** displays regression results of the statistically significant ($p<0.05$) relationship between gains in Pashto reading accuracy by HLE and book borrowing frequency in Pakistan.

The more often children borrowed books, the closer the average predicted gains of children with low HLE (blue) are to those of classmates with higher HLE. For children with the highest HLE at baseline in this context (green), the impact of borrowing books is minimal.

Accepting that HLE plays an important role in children's reading skill development makes it imperative that we collect these data alongside oral reading assessments to better understand the context, accurately estimate effects and critically analyse remedies.

7. LIMITATIONS

There are several limitations to collecting HLE data using the field-tested method outlined here. First, it does not enable a view of the quality of the interactions that occur with reading materials and oral language in the home. Second, it is student reported and therefore susceptible to social desirability bias. Third, there can be varied interpretations and understandings of what it means to 'read to children' that can affect how staff adapt and pilot the questions, how assessors pose these questions to children and how children answer them. These limitations would theoretically make relationships harder to discern in the data. However, in our data collections, we consistently see the same relationships, indicating that the data have reasonable reliability and validity. Even considering these limitations, collecting data on the HLE and grappling with how to address the limitations mentioned here serves to support our understanding and progress towards ensuring that all children are reading.

8. CONCLUSION

Flora and Jolly attend relatively similar schools and have access to virtually the same school-bound resources. Data that only represents limited background characteristics such as sex would miss the broader picture of how and why Flora and Jolly are developing as readers. Baseline HLE data collection allows us to identify extra supports that Flora might need as well as identify stronger reading families like Jolly's that may serve as a resource to other struggling readers. Continually collecting data on these factors allows better accounting for different aspects of equitable attainment.

Save the Children recommends collecting data on the five dimensions of literacy—1) the value placed on reading, 2) the press for achievement, 3) the availability of reading materials, 4) reading to children, and 5) opportunities for verbal interaction—alongside information about children's interest in reading and the varied places and people with whom opportunities to learn occur beyond the school walls. Collecting these HLE data alongside oral reading assessment scores will further enhance our global evidence base as well as our store of tested solutions to ensure that basic skills are acquired.

REFERENCES

Abeberese, A. B., Kumler, T. J., and Linden, L.L. (2014). "Improving Reading Skills by Encouraging Children to Read in School: A Randomized Evaluation of the Sa Aklat Sisikat Reading Program in the Philippines". *Journal of Human Resources*, Vol. 49, No. 3, p.p. 611–633. http://www.nber.org/papers/w17185.pdf

Bradley, R. H., Corwyn, R. F., McAdoo, H. P. and Coll, C.G. (2001). "The Home Environments of Children in the United States Part I: Variations by Age, Ethnicity, and Poverty Status". *Child Development*, Vol. *72*. No. 6, p.p. 1844-67. http://www.ncbi.nlm.nih.gov/pubmed/11768149

Bus, A., Leseman, P. and Keultjes, P. (2000). "Joint book reading across cultures: A comparison of Surinamese-Dutch, Turkish-Dutch, and Dutch parent-child dyads". *Journal of Literacy Research*, Vol. *32*, No. 1,p.p. 53-76. http://jlr.sagepub.com/content/32/1/53

Chansa-Kabali, T. and Westerholm, J. (2014). "The Role of Family on Pathways To Acquiring Early Reading Skills in Lusaka's Low-Income Communities". *An Interdisciplinary Journal on Humans in ICT Environments*, Vol. 10, p.p. 5-21.

Dowd, A.J. (2014). *Practice, opportunity to learn and reading: parent and social factors in literacy acquisition.* Paper presented at the CIES Annual Conference, Toronto, Canada.

Dowd, A.J., Friedlander, E., Guajardo, J., Mann, N. and Pisani, L. (2013) *Literacy Boost Cross Country Analysis Results.* Washington, DC: Save the Children.

Dowd, A. J. and Pisani, L. (2013). "Two Wheels are Better than One: the importance of capturing the home literacy environment in large-scale assessments of reading". *Research in Comparative and International Education*, Vol. 8, No. 3, p.p. 359-372.

Elley, W. B. (1992). *How in the world do students read? IEA Study of Reading Literacy*. Report for the International Association for the Evaluation of Educational Achievement. The Hague: Institute of Education Sciences.

Friedlander, E., Dowd, A.J., Borisova, I. and Guajardo, J. (2012). *Life-wide learning: Supporting all children to enjoy quality education*. New York: UN Women and UNICEF. http://www.worldwewant2015.org/node/283236

Friedlander, E., Dowd, A., Guajardo, J. and Pisani, L. (2016). "Education for All or *Literacy* for All? Evaluating Student Outcomes from Save the Children's Literacy Boost Program in Sub-Saharan Africa". In A. Abubakar, and F. van de Vijver (eds.), *Handbook of Applied Developmental Science in Sub-Saharan Africa.* New York, NY: Springer.

Hart, B. and Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H Brookes Publishing.

Hess, R. D. and Holloway, S. D. (1984). "Family and School as Educational Institutions". *Review of Child Development Research, 7*, 179–222.

Kabarere, V., Muchee, T., Makewa, L. N., & Role, E. (2013). "Parental Involvement in High and Low Performing Schools in Gasabo District, Rwanda". *International Journal about Parents in Education*, Vol. 7, No. 1, p.p. 30-42.

Kalia, V. and Reese, E. (2009). "Relations Between Indian Children's Home Literacy Environment and Their English Oral Language and Literacy Skills". *Scientific Studies of Reading*, Vol. 13, No. 2, p.p. 122-145. http://www.tandfonline.com/doi/abs/10.1080/10888430902769517

Mithani, S., Alam, I., Babar, J. A., Dowd, A. J., Hanson, J. and Ochoa, C. (2011). *Literacy Boost Pakistan: Year 1 Report*. Washington D.C.: Save the Children.

Save the Children Malawi (2013). *Save the Children International Basic Education Program All Children Reading TIANA Project 2013 Endline Report*. Blantyre: Malawi.

Snow, C. E., Burns, M. S. and Griffin, P. (1998). *Preventing Reading Difficulties in Young Children*. National Research Council.

Taylor, D. (1983). *Family literacy: Young children learning to read and write.* Portsmouth, NH: Heinemann.

Tusiime, M., Friedlander, E. and Malik, M. (2014). *Literacy Boost Rwanda: Literacy Ethnography Report.* Save the Children, Stanford University and the Rwanda Education Board.

Wagner, D. A. and Spratt, J. E. (1988). "Intergenerational Literacy: Effects of Parental Literacy and Attitudes on Children's Reading Achievement in Morocco". *Human Development*, Vol. 31, No. 6, p.p. 359-369.

Teacher Quality as a Mediator of Student Achievement

NANCY CLARK-CHIARELLI AND NATHALIE LOUGE[1]
Education Development Center

ABBREVIATIONS

EDC	Education Development Center
NRP	National Reading Panel
SCOPE-Literacy	Standards Based Classroom Observation Protocol for Educators in Literacy
USAID	United States Agency for International Development

1. INTRODUCTION

In recent years, standardised primary grade reading assessments have revealed disturbingly low levels of primary grade student achievement in reading and math in many countries around the world. As organizations and governments strive to improve primary grade learning outcomes, understanding which factors account most for the dramatic differences in student achievement will increase the likelihood of the success of interventions.

Through sample-based national testing using the Early Grade Reading Assessment (EGRA) or other similar tools, we now know a lot about where students stand in relation to the competencies necessary for reading with comprehension. However, we still do not know enough about teachers who are entrusted with ensuring that students attain the necessary reading and math skills by the end of the primary grades. Similarly, although more information is now available about primary school access, we still do not know what instruction looks like and what competencies are being taught. In addition, in the field of education research, reliable measurement of teacher quality is still at the "comparatively early stages of development" (Centre for Education Statistics and Evaluation, 2013).

Efforts to address teacher skill gaps and improve the quality of teaching is likely to fail without information on teacher skills since any professional development programme is only successful if it builds on existing knowledge and skills. So far, only a limited number of attempts have been made in developing nations to investigate the teaching cadre and better understand the content of instructional practice in a systematic way. Without such information, student achievement data provides only half of the story—it identifies the problems, but not the opportunities for solutions that may lie in improving teaching quality.

This paper addresses two key areas in literacy assessment that focuses on teacher quality:

1. How can we assess teacher quality in primary grade literacy instruction that goes beyond credentials or content knowledge? What tools help assess the quality of classroom literacy instruction?

1 This publication is made possible by the generous support of the American people through the United States Agency for International Development (USAID) under the Basa Pilipinas Project and the Philippines Department of Education. The contents of this publication are the sole responsibility of Education Development Center, Inc. (EDC), and do not necessarily reflect the views of USAID or the United States Government.

2. What are the implications for building professional development in primary grade literacy that provides support for change in instruction?

2. EDUCATIONAL CONTEXT

2.1 The link between teacher quality and student achievement

In the U.S., research has shown that teachers have a substantial impact on student learning. One recent meta-analysis of over 2,000 research studies of teacher quality found that the effect size of teacher quality on student achievement averages .50 (after controlling for student characteristics), which translates into more than a half of a school year of achievement gains (Hattie, 2009). Although individual student background is usually found to explain much of the variance in student scores, some studies have shown that high quality instruction throughout primary grades can substantially offset the disadvantages associated with poverty (Darling-Hammond, 2000). A study by Rowe (2003) found that:

> "…whereas students' literacy skills, general academic achievements, attitudes, behaviors and experiences of schooling are influenced by their background and intake characteristics, the magnitude of these effects pale into insignificance compared with class/teacher effects. That is, the quality of teaching and learning provision are by far the most salient influences on students' cognitive, affective, and behavioral outcomes of schooling—regardless of their gender or backgrounds. Indeed, findings from the related local and international evidence-based research indicate that 'what matters most' is quality teachers and teaching, supported by strategic teacher professional development".

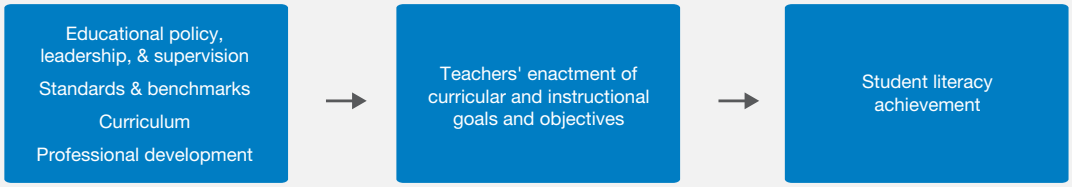
Moreover, there is evidence that the effects of teacher quality on student performance are cumulative. Students who are assigned to several ineffective teachers in a row have significantly lower achievement and educational gains than those who are assigned to several highly effective teachers in sequence (Sanders and Rivers,1996). This research holds a lot of promise for promoting education in developing countries.

Underlying the hypothesis that teachers are a key mediator in influencing student achievement is a conceptual theory of change. **Figure 1** articulates this process associated with improvement in literacy instruction and the ultimate goal of positive changes in students' literacy achievement. Moving from left to right, this diagram identifies the inputs and processes that provide sources of support for student literacy achievement. Undergirding classroom instruction are inputs that are controlled at the macro-level of the school system—be it at the national, regional or district level. These are inputs over which most teachers usually have less control. These include 1) educational policy, leadership and supervision; 2) standards and benchmarks; 3) curriculum; and 4) opportunities for professional development. The mediator between these macro-level policies, structures and increases in students' literacy achievement is ultimately the actual instruction that teachers deliver and students receive. It is in the classroom and daily instruction where teachers enact curricular and instructional goals and objectives. Also, it is the quality of this enactment that is associated with student gains. A similar model has been described by Desimone (2011) in her research on effective professional development in which she posits a change theory including the following steps: 1) professional development experience for teachers; 2) professional development increases knowledge and skills and influences attitudes and/or beliefs; 3) improvement in content and pedagogy of instruction; and 4) gains in student learning.

2.2 Indicators of classroom quality in literacy instruction

As many have described, reading is comprised of a set of components which must be taught in order for students to read but which if presented discretely are not necessarily sufficient in order for them to become a skilled reader (Comings, 2014).

Figure 1. Theory of change to produce student literacy achievement



Source: EDC Philippines, 2013/2014

Students need instruction and practice on individual components as well as time reading connected text in which the components seamlessly work together (Snow et al., 1998; National Reading Panel, 2000). Perhaps the most succinct report on the vital components of reading development was expressed in the 2008 U.S. report of the National Reading Panel (NRP). In this seminal report, the panel identified five key components: phonemic awareness, phonics, fluency, vocabulary and comprehension. Thus, any tool designed to assess quality of literacy instruction must minimally assess these five components as well as dimensions of instruction that are more holistic in how these components are combined in effective literacy instruction (e.g. level of classroom discourse, effective management of instructional time). To this end, the question must be asked, How can we assess teacher quality in primary grade literacy instruction?

3. SCOPE-LITERACY ASSESSMENT

To address the need for information on the quality of teacher instruction in literacy, the Standards Based Classroom Observation Protocol for Educators in Literacy (SCOPE-Literacy) was developed for use in international settings. The SCOPE-Literacy is a classroom observation tool and is founded on a research-based set of guiding principles aligned with Education Development Center's reading model, Read Right Now! (Education Development Center, 2013). These guiding principles identify teaching strategies that are most effective in developing competent readers and writers—strategies consistent with what students should be able to achieve in language and literacy. The SCOPE-Literacy's guiding principles and teacher observation protocols are as follows:

- ***Teacher builds a supportive learning environment that provides the foundation for student participation and risk taking.*** Rules and routines provide efficient use of class time and allow students to engage in purposeful activity. Particularly for young learners and for those learning new languages, a risk-free environment must be created through the teacher's skillful use of modeling and reframing of responses when students make errors. Teacher intervention when conflicts or student non-compliance occurs is calm, non-threatening and effective. The ultimate goal is to facilitate independent, productive problem-solving strategies among learners.

- ***Teacher uses effective grouping strategies to support learner participation and language and literacy learning.*** The use of a variety of grouping strategies (i.e. whole group, small group, pairs) supports high collaboration and cooperation. Smaller groups also support language development among students as each student is given more time to verbally interact with others than in traditional large groupings.

- ***Teacher ensures full participation of all learners regardless of their gender, special needs or other differences.*** The teacher orchestrates the class such that prior knowledge and personal interests are used as the basis

© Liang Qiang / World Bank

for conversations, activities and learning experiences. Individual differences are valued and specific strategies are used to engage all learners. Effective use of 'wait time' promotes the participation and risk-taking of students.

- ***Teacher and students have access to classroom materials.*** High quality materials are available and in sufficient quantity for the number of students in the classroom. Books support instructional goals and student learning.

- ***Teacher manages instructional time effectively.*** The teacher has a lesson plan and follows it. There is evidence that lesson plans build on one another and support mastery of reading and writing competencies. Clear instructions about what students are expected to do are appropriately brief.

- ***Teacher builds students' oral language skills.*** The teacher provides learners with rich and meaningful lessons in oral language development and models the use of appropriate language structures, vocabulary and pronunciation throughout instruction. The teacher may often need to intentionally bridge between a familiar language and one that students are learning. In turn, students are given opportunities to express themselves, use new vocabulary and practice new language structures.

- ***Teacher provides opportunities for meaningful reading activities.*** The teacher matches texts to learners' reading levels and interests. Students are given an opportunity to choose reading material. Time is given for learners to read authentic texts and engage in meaningful reading tasks in a variety of ways (e.g. silent reading, paired reading, reading aloud, choral reading).

- ***Teacher provides opportunities for learning, for word identification and spelling.***

Figure 2. SCOPE-LITERACY dimensions and indicators

Section I. Classroom structure

1. Supportive learning environment
   - Understanding of rules and routines
   - Environment supports student language and literacy learning
   - Teacher management of conflicts and non-compliance
2. Effective grouping strategies
   - Grouping strategies
   - Learner participation
   - Learner cooperation and collaboration
3. Participation of all learners
   - Learners prior knowledge and interests
   - Strategies that support learner inclusion
   - Practice that provides learner access to learning
4. Opportunities for reflection
   - Opportunities to self-assess reading and writing
   - Tools to support learner reflection and self-assessment
   - Ongoing assessment
5. Classroom materials
   - Print-rich environment
   - Classroom materials to support literacy learning
   - Use of books in instruction
6. Manages reading and writing instruction
   - Lesson planning
   - Patterns of instruction
   - Directions to support learner

Section II. Language and literacy instruction

7. Opportunities for oral language development
   - Learner talk
   - Teacher language
   - Direct instruction
   - Discussion
8. Opportunities for meaningful reading
   - Text choice
   - Opportunity to read individually
   - Print resources
9. Opportunities for learning to decode and spell words
   - Direct instruction
   - Adaptations for individuals
   - Strategies for decoding
10. Develops reading fluency
    - Modeling fluency
    - Varied instructional strategies
    - Activities to build automaticity
11. Opportunities for developing vocabulary
    - Teacher modeling
    - Vocabulary selection
    - Varied approaches to vocabulary instruction
    - Strategies for learning word meanings independently
12. Opportunities for developing reading comprehension
    - Learner thinking
    - Instructional strategies
    - Questioning
13. Writing instruction
    - Opportunities for self-expression
    - Writing process
    - Direct instruction

Instruction in phonemic awareness, word identification and phonics occurs in short episodes of direction instruction. The teacher clearly and succinctly explains specific principles and provides engaging activities for practice. Opportunity to apply spelling principles is guided by the teacher and specific strategies are provided to foster learner independence.

- ***Teacher provides opportunities for developing fluent reading.*** The teacher models fluent reading and draws students' attention to specific features of fluency. Teacher engages readers in enjoyable and motivational reading games and activities that increase automatic word recognition and smooth reading.

- ***Teacher provides opportunities for vocabulary development.*** The teacher exposes students to new words and models use of sophisticated vocabulary. Teacher teaches specific word meanings from books or words/concepts important to the curriculum. Words are studied in depth and are used in multiple contexts.

- ***Teacher builds students' comprehension in texts they listen to and read themselves.*** The teacher poses a variety of questions that provide opportunities for literal comprehension as well as inferential and higher-level thinking. The teacher models and explains 'thinking' strategies to help students understand text (e.g. summarisation, predicting).

- ***Teacher provides opportunities for systematic writing instruction that supports students' expressions of their own thoughts and ideas.*** Students engage in authentic writing using a multi-step process (plan, draft, revise, edit and publish). Teacher provides brief, focused lessons

that may include process, mechanics, genres or techniques (e.g. dialogue).

Using these guiding principles as the foundation, the SCOPE-Literacy assesses classroom reading and writing instruction along thirteen dimensions of practice and is organized into two major sections: classroom structure and language literacy instruction. The thirteen dimensions of literacy practice and indicators reflecting the dimensions are displayed in **Figure 2**.

Based on the observation of an instructional session on literacy, each item is scored on a scale from 1 to 5 with 1 being on the lower end and 5 being on higher end of performance. In addition to the numerical rating, there is a statement that accompanies each score to further guide the assessment of each dimension.

Rating 1 Deficient	There is minimal or no evidence of the practice.
Rating 2 Inadequate	There is limited evidence of the practice.
Rating 3 Basic	There is some evidence of the practice.
Rating 4 Strong	There is ample evidence of the practice.
Rating 5 Exemplary	There is compelling evidence of the practice.

A reliability analysis of the SCOPE-Literacy found that the internal consistency of items was coefficient alpha = .891.

In order to provide a contextualisation for the use of the SCOPE-Literacy, an example of its use in Basa Pilipinas is provided. Basa Pilipinas is USAID/Philippines' flagship basic education project in support of the Philippine Government's early grade reading programme. Basa is implemented in close collaboration with the Department of Education and aims to improve the reading skills for at least one million early grade students in Filipino, English and selected mother tongues. These goals will be achieved by improving reading instruction, reading delivery systems and access to quality reading materials. The project commenced in January 2013 and will continue through December 2016.

The SCOPE-Literacy observation tool can be accessed here.

4. DISCUSSION OF SCOPE LANGUAGE AND LITERACY INSTRUCTION FINDINGS

The literacy practices of Grade 2 teachers from a sample of schools participating in the Basa project within two regions of the country were observed on two occasions—November 2013 and December 2014. In Grade 2, students are learning to read in their mother tongue, Filipino and English. The mother tongue within the two regions observed differs. The SCOPE-Literacy observations were conducted during the Filipino language portion of the day that occurs for 50 minutes daily.

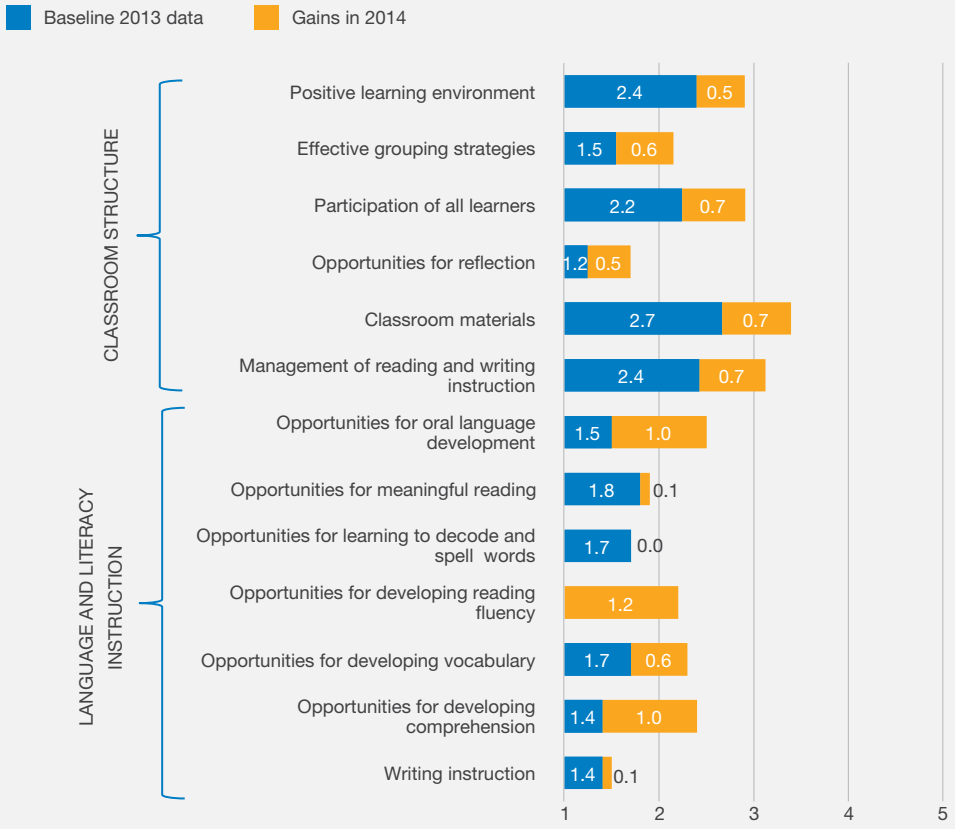
A sample of 33 Grade 2 teachers were observed using the SCOPE-Literacy tool in November and December 2013 and again in December 2014 to measure changes in teaching practices as a result of the Basa intervention. As **Figure 3** demonstrates, teachers started out with very low scores at baseline in 2013, ranging between 'deficient' and 'inadequate'. By the end of 2014, teacher practices showed a broader range of scores with more teachers performing at the 'basic' level. This suggests some improvement from 2013 to 2014, indicating that teachers were starting to apply new teaching practices. Ratings of 4 and 5 or 'strong' and 'exemplary' are not easy to attain. However, a 'basic' rating of 3 is quite positive in the context of introducing new literacy instruction techniques.

By December 2014, improvements were seen in all teaching practices observed. However, the largest gains were seen in the 'language and literacy instruction' domain. In the 'classroom structure' domain, teachers saw the largest improvements in the "ensuring participation of all learners", "ensuring accessible classroom materials" and "effective management of reading and writing

instruction" items of the SCOPE-Literacy. In fact, in 2014, nearly half of observed teachers scored 'strong' in 'classroom materials' and "management of reading and writing instruction". This is important given that Basa has introduced a large number of new reading materials for teachers to manage in the classroom. Ensuring access to the materials is key for student learning as is equitable student participation in the classroom. Teachers didn't score as highly in the areas of effective grouping strategies and opportunities for reflection—both items that require more advanced classroom management skills. Teachers who can effectively group students are better at providing differentiated learning opportunities as well as opportunities for reflection that can deepen students' understanding of text.

While baseline scores were lower overall in the domain of 'language and literacy', this is also where teachers showed the most improvement. This is not unexpected as teachers may not have had much exposure to teaching reading prior to the Basa intervention. For the 'language and literacy instruction' domain, teachers largely improved literacy instruction in the areas of oral language development, developing reading fluency and developing comprehension. The improvement in opportunities for developing reading fluency was particularly striking as we saw almost no evidence of this practice in the first observation. Oral language development is also a key skill for teachers, particularly in a multi-lingual context where bridging opportunities from one language to another needs to be intentionally planned by teachers.

Figure 3. Change in average scores for SCOPE-Literacy, 2013 to 2014 (n=33)



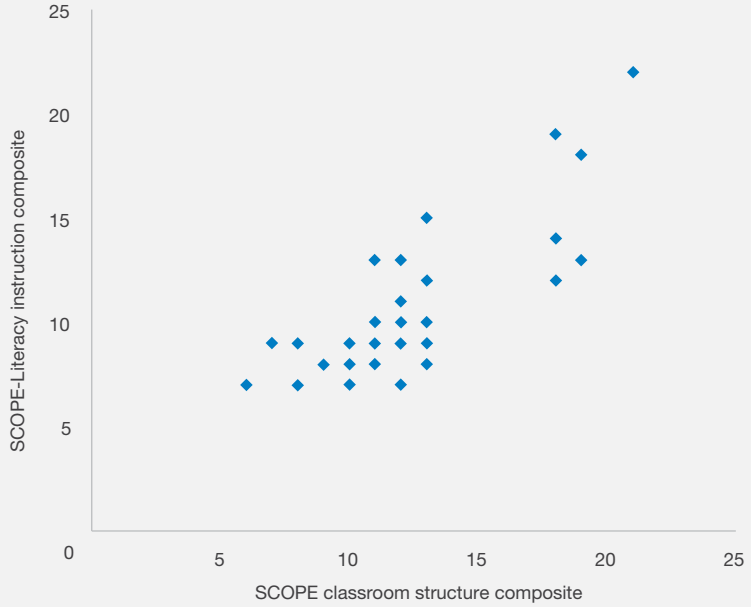
Note: Observations of 33 Grade 2 teachers in the Philippines using the SCOPE-Literacy tool in November and December 2013 and again in December 2014 to measure changes in teaching practices as a result of the Basa intervention.
Source: EDC, Philippines, 2013/2014

There is one general caveat to consider. Basa teachers follow an instructional sequence in which all 14 domains of the K-12 curriculum are not taught everyday but over a period of five days for Filipino and ten days for English. This is by design to allow adequate time for pupils to complete skill-related tasks in their second and third language. Depending on the lesson plan for the day, it would not be expected that teachers teach all domains. In addition, domains such as phonics take on additional meaning in a language such as English, which has an opaque orthography versus Filipino, a syllabic language that has a transparent orthography. Since the teachers were observed during their Filipino class, one possible reason for no increase in the score for "opportunities for learning to decode and spell words" is because by Grade 2 in the third quarter, a majority of the students have already learned to decode and spell in Filipino.

These results indicate that teachers who have stronger practice in classroom structure also have more advanced practices of teaching literacy. The structures around literacy learning support the more nuanced implementation of instructional strategies and the tailoring of instruction to the needs of particular students.

A correlation (*see **Figure 4***) between the two sections of the SCOPE-Literacy results was found in 2014. The scatterplot shows that the relationship between the two components of the SCOPE tool appears to be linear. The coefficient of the correlation between the two sections of the SCOPE-Literacy was statistically significant (Pearson's r=.946; Kandall's tau = .820 and Spearman's rho = .905, all three significant at $p<0.001$ level). These results suggest that there is a strong link between classroom structure and more advanced practices of teaching literacy.

Figure 4. Correlation between two components of SCOPE-Literacy, 2014 (n=33)

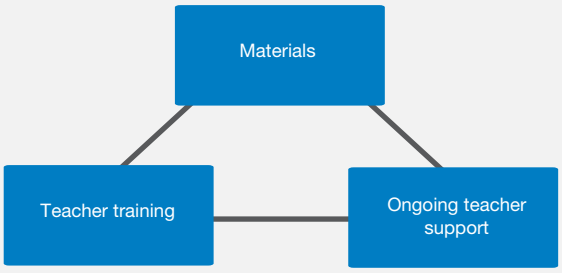


Note: Observations of 33 Grade 2 teachers in the Philippines using the SCOPE-Literacy tool in November and December 2013 and again in December 2014 to measure changes in teaching practices as a result of the Basa intervention. The SCOPE classroom structure composite is comprised of six dimensions on the SCOPE-Literacy tool: 1) supportive learning environment; 2) effective grouping strategies; 3) participation of all learners; 4) opportunities for reflection; 5) classroom materials; and 6) manages reading and writing instruction. The SCOPE-Literacy instruction composite is comprised of seven dimensions on the SCOPE-Literacy tool: 1) opportunities for oral language development; 2) opportunities for meaningful reading; 3) opportunities for learning to decode and spell words; 4) develops reading fluency; 5) opportunities for developing vocabulary; 6) opportunities for developing reading comprehension; and 7) writing instruction.
Source: EDC, Philippines, 2013/2014

5. IMPLICATIONS FOR BUILDING PROFESSIONAL DEVELOPMENT IN PRIMARY GRADE LITERACY THAT PROVIDES SUPPORT FOR CHANGE IN INSTRUCTION

Basa Pilipinas embodies a professional development model designed to be embedded (integrated into teacher practice) and comprehensive. The framework of the professional development can be thought of as a three-legged stool as the professional development is the most stable or robust if all three 'legs' are present—materials, teacher training and on-going support for teachers (*see **Figure 5***). These three areas in the form of communities of practice, work together to provide teachers with the knowledge and understanding of what to teach, how to teach it and why teach it.

Figure 5. Basa's model of comprehensive professional development



5.1 Key features of effective literacy professional development

Previous research supports Basa's approach to professional development in literacy and has identified five key features that can be articulated in terms of literacy: content focus, active learning, coherence, duration and collective participation (Desimone, 2010-2011; Desimone et al., 2002).

1. Content focus

Professional development must be structured around the key components and dimensions of literacy instruction and classroom management. This may mark a major departure from traditional methods of literacy instruction. Moreover, while many cognitive processes involved in reading and writing may also apply across subject areas, professional development that is focused specifically on literacy is more effective. For example, teacher communities of practice should devote specific time to literacy as a focus rather than as a more generic discussion on 'questioning'. If the content is embedded in a teacher's daily instruction or current curriculum, the new information or learning is more meaningful (Knowles, 1980). Regarding classroom management, teachers need to adopt classroom management strategies to foster successful teaching and learning. For example, it is essential that teachers learn to group students according to reading level to provide differentiated and targeted instruction and to keep children on task at all times (Baker, 2007).

2. Active learning

While well-designed lectures may be critical in exposing teachers to new information on reading development and literacy instruction, professional development needs to foster interaction among teachers and with the facilitator. Teacher reflection on their own literacy practices and sharing in small groups, video analysis and action planning are all activities that may be effective in applying new information in an interactive manner.

3. Coherence

The link between the literacy materials and curriculum teachers are using in the classroom, training sessions and on-going support must be aligned. This demands that district leaders, supervisors and school heads are well versed on the content and methods of the professional development teachers are receiving in language and literacy. As instructional leaders, school heads should play an instrumental role in on-going professional development designed to foster better literacy instruction.

4. Duration

Brief episodes of training for teachers are not effective. According to Desimone (2011), a minimum of 20 hours of professional development

is necessary to be effective. Face-to-face training must be bolstered by 'back at home' on-going support that encourages teacher reflection and sharing on their implementation of new teaching strategies. Teacher sharing of their own practice and students' learning is reported to be a meaningful and motivating experience. It should be noted that Basa teachers receive a minimum of 30 hours of professional development in a semester during their first year in the programme.

5. Collective participation

The strength of all teachers within a grade level sharing in professional development is important. Coherence between grades is also critical. Building a coherent plan for literacy instruction in the early grades in literacy should ideally take place at the district, division or regional level.

5.2 Potential limitations of the study

While the SCOPE-Literacy provides classroom assessors with a range of scores, the use of the full scale may be limited in developing contexts. If few teachers provide instruction at the higher level of the scores, then it often results in a more limited range of scores—as we see in this study. This could result in an inflated estimate of inter-item reliability.

On the other hand, it is also reasonable to hypothesize that high inter-item reliability may be an indication of the reciprocal relationship between 'classroom structure' and 'literacy instruction' as better instruction relates to better features of classroom structure and vice versa.

6. SUMMARY AND RECOMMENDATIONS

SCOPE-Literacy is a useful tool to evaluate the quality of teacher practice. The data presented in this paper are based on a small sample of teachers and classrooms but the results indicate their potential usefulness in better understanding the quality of classroom instruction. As a mediator of student achievement, knowledge of teacher practices will inform teachers, school leaders and other administrators in the shaping of policy and supervision strategies. Moreover, while the tool may be used as an overall measure of classroom quality in language and literacy, it has great promise in the area of professional development. Based on an initial assessment of teachers' needs, specific dimensions of the SCOPE-Literacy may be used as a way to plan and monitor professional development in literacy. A cohesive professional development plan for schools, districts and countries is critical to the ultimate goal of student achievement in literacy.

REFERENCES

Baker, R. S. (2007). "Modeling and understanding students' off-task behavior in intelligent tutoring systems". *Proceedings of ACMCHI 2007: Computer-Human Interaction*, pp. 1059-1068.

Centre for Education Statistics and Evaluation (2013). *Great Teaching, Inspired Learning*: *What does the evidence tell us about effective teaching?* Sydney, Australia: NSW Department of Education and Communities. http://www.dec.nsw.gov.au.

Comings, J.P. (2014). "An evidence-based model for early-grade reading programmes". *Prospects*, Vol. 45, No. 2, pp. 167-180. http://link.springer.com/article/10.1007/s11125-014-9335-9.

Darling-Hammond, L. (2000). "Teacher quality and student achievement: A review of state policy evidence". *Education Policy Analysis Archives*, Vol. 8, No. 1, pp. 1-44.

Desimone, L. M (2011). "A primer on effective professional development". *Phi Delta Kappan,* Vol. 92, No. 6,pp. 68-71.

Desimone, L., M., Porter, A.C., Garet, M.S., Yoon, K.S. and Birman, B.F. (2002). "Effects of professional development on teachers' instruction: Results from a three-year longitudinal study". *Educational Evaluation and Policy Analysis*, Vol. 24, No. 2,pp. 81-112.

Education Development Center (2013). *Read Right Now!* Waltham: MA.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Knowles, M. (1980). *The modern practice of adult education: Andragogy versus pedagogy.* Rev. and updated ed. Englewood Cliffs, NJ: Cambridge Adult Education.

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literacy on reading and its implications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development.

Rowe, K. (2003). *The importance of teacher quality as a key determinant of students' experiences and outcomes of schooling*. Paper presented at the Australian Council for Educational Research Conference, 19-21 October. http://research.acer.edu.au/research_conference_2003/3.

Sanders, W.L. and Rivers, J.C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Research Progress Report. Knoxville: University of Tennessee Value-Added Research and Assessment Center. http://www.cgp.upenn.edu/pdf/Sanders_Rivers-TVASS_teacher%20effects.pdf.

Snow, C.E., Burns, M.S. and Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

School-based Assessments: What and How to Assess Reading

MARGARET M. DUBECK, AMBER GOVE, KEELY ALEXANDER
RTI International[1]

ABBREVIATIONS

AERA	American Educational Research Association
EGRA	Early Grade Reading Assessment
RTI	Research Triangle Institute

1. INTRODUCTION

The goal of this article is to respond to two questions regarding the Early Grade Reading Assessment (EGRA): what reading skills are selected for inclusion in the EGRA and the how the survey generates results that are valid and reliable. The EGRA is a reliable and valid measure of skills that contribute to reading development, typically administered to students in the first few grades of primary school to inform system and school-level improvement.

As stated in the assessment's inception documents (RTI International, 2009; RTI International and International Rescue Committee, 2011), the EGRA is not intended to be a high-stakes accountability measure to determine whether a child should advance to the next grade—nor should it be used to evaluate individual teachers. Rather, the subtasks included in the EGRA can be used to inform the focus of instruction. As a formative assessment, the EGRA in its entirety or select subtasks can be used to monitor progress, determine trends in performance and adapt instruction to meet children's instructional needs.

2. WHAT SHOULD BE ASSESSED IN READING

2.1 What is reading?

Reading is a cognitive process encompassing the concepts that print is speech in a written form and that the ultimate goal is to read with understanding. Various models aim to explain the reading process—among them are the construction-integration model (Kintsch, 1998), the dual coding model (Paivio, 1971) and the transactional model (Rosenblatt, 1978). As another example, Snow and the RAND Reading Study Group (2002) described the internal text model as a complex combination of extraction and construction of meaning identifying the role of linguistic knowledge, cognitive capacities, vocabulary, background knowledge, motivation and strategy knowledge.

Less complex is the Simple View of Reading (Gough and Tumner, 1986; Hoover and Gough, 1990). In this model, reading with understanding is the product of decoding and language (i.e. expressed as a mathematical equation: decoding x language = reading comprehension). According to this model, when one of the two factors is lacking, reading comprehension does not occur. The Simple View implicitly includes other constructs but highlights two essential factors to inform an instructional response. It values the contribution of both decoding

1 The authors would like to acknowledge the energy, efforts and resources of the students, teachers, ministry and donor agency staff—principally those from the United States Agency for International Development (USAID)—of the more than 70 countries that have conducted EGRAs to date.

skills and language skills to reading comprehension. Therefore, for readers who can decode, their reading comprehension is commensurate with their language comprehension skills, and readers must have content knowledge in a variety of domains to support their language comprehension abilities.

Using a 'simple view of reading', a response to a reader (or an entire education system) that is struggling is to identify which factors need addressing: decoding skills, language skills or both? Explicit and systematic instruction in decoding is necessary to improve the accuracy and automaticity of word recognition, which will support reading comprehension. A response that develops knowledge or comprehension benefits struggling readers or children learning to read in a nonnative language. The 'simple view of reading' is a useful way to consider the process of learning to read and the EGRA utilises this framework.

Learning to read is a process that develops in a predictable manner but is influenced by individual differences and contexts (e.g. the pedagogy, language). First, an emergent reader develops a basic understanding of the connections between spoken and written words. For example, a child may recognise the logo of a mobile phone company that is posted throughout the community by its colour or shape of the letters—but this recognition is not reading. Children also develop phonological knowledge, which supports manipulating word parts and sounds. This is followed closely by developing print knowledge, such as learning the relationships between individual letters and sounds. Thereafter, readers develop their orthographic knowledge to learn to encode (spell) or decode (read) meaningful or frequently occurring parts in written words.

The time required to master these basic reading skills varies by language and context. Among other factors, the nature of a language's writing system has been shown to influence the rate at which familiar word reading skills are acquired. Moreover, Seymour et al. (2003) showed that the shallow orthographies (consistent sound-symbol correspondences) of languages such as Finnish or Greek contribute to nearly perfect word accuracy after a year of schooling. Conversely, in opaque orthographies such as English, with complex graphemes, contextual variations and irregularities interfere with word recognition and learning to read takes longer. For example, Seymour et al. (2003) found that after a year of schooling, children learning to read in English recognised only a third of the words they attempted to read.

From the earliest phases, word recognition relies on oral language skills such as vocabulary (Oullette, 2006). For example, when a reader knows a word's meaning, it provides a means for a self-check that the sound that was uttered (i.e. read) is the correct word. Yet, considering that the ultimate goal of reading is to read with understanding, as basic reading skills progress beyond the word-recognition phase, reading proficiency depends less on basic reading skills and more on vocabulary and prior knowledge (August and Shanahan, 2006; Hoover and Gough, 1990; Vellutino et al., 2007). Reading proficiency also corresponds to the increasing demands of the texts readers are expected to understand.

2.2 What is practical to both assess and improve?

The EGRA battery is a template for developing individualised, locally tailored assessments for each country and language. The definition of what skills to assess is also based on a practical calculation of what skills would benefit most easily from intervention. Research from various contexts suggest which literacy skills can be reliably measured and are predictive of later reading success (August and Shanahan 2006; National Early Literacy Panel, 2008; National Institute of Child Health and Human Development, 2000). The skills, which are practical to assess and improve, are divided into three domains: phonological awareness, print knowledge and orthographic knowledge. The EGRA measures these domains:

i. Phonological awareness

Phonological awareness is a collection of skills defined as a sensitivity to language at the

phonological level. Many studies have supported its role in predicting early word reading and spelling in both shallow and opaque languages (Badian, 2001; Denton et al., 2000; McBride-Chang and Kail, 2002; Muter et al., 2004; Wang et al., 2006).

ii. Print knowledge

Print knowledge includes an understanding of the orthographic system and the written language. Through a learner's investigations, print knowledge advances in a hierarchical yet recursive way, implying that one print knowledge component is a prerequisite for another component but that skills are not necessarily mastered before new learning commences. Print concepts include book orientation, directionality and a purpose for reading.

Understanding the distinctive symbols and names of alphabet letters also falls under the domain of print knowledge. Besides letter recognition, alphabet knowledge also encompasses knowledge of letter names and their corresponding sounds. Letter knowledge has been consistently shown to be a strong predictor of early word reading and spelling (Adams, 1990; Ehri and Wilce, 1985; Piper and Korda, 2010; RTI International, 2013; Wagner et al., 1994).

iii. Orthographic knowledge

Orthographic knowledge is an understanding of words in their written form. It includes the knowledge that certain sequences of letters compose words that represent spoken sounds. Applying this knowledge helps learners identify familiar words, decode unfamiliar words in isolation and read connected texts, such as a sentence or a story.

2.3 What does the EGRA assess in reading?

The EGRA is a collection of subtasks that measure skills needed for the acquisition of reading. From 14 existing subtasks (*outlined in* ***Table 1***), users can select the ones that align with their research question and the particular stage(s) of literacy development of interest. Researchers interested in capturing a range of abilities can select the subtasks that are expected to reflect student performance depending on the context of which phase of development the assessment is administered.

The EGRA has been used mostly to understand primary school children's reading abilities. This corresponds to the period where instruction progresses from playing with language via songs and rhymes to learning the alphabet to exploring how to combine letters to read and spell individual words, and ultimately, to using that knowledge to read connected text. The EGRA would also be appropriate for measuring the progress of older children or young adults who are in the early stages of learning to read.

3. HOW DOES THE ASSESSMENT GENERATE RESULTS THAT ARE RELIABLE AND VALID?

The process for designing and developing an early reading assessment begins first and foremost with an understanding of the purpose of the study or data collection opportunity. As outlined in another article in this volume (see Kochetkova and Dubeck), early in the process, stakeholders should come together to define how assessment results will be used and whether the proposed assessment and associated survey tools will contribute to the desired result.

The *Early Grade Reading Assessment Toolkit*—first developed in 2009 (RTI International, 2009) and updated in 2015 (RTI International, 2015)—provides detailed guidance on how to develop and adapt[2] an EGRA. As stated previously, the EGRA has been found to be a valid and reliable tool for understanding students' early literacy skills. Validity is the degree to which theory and evidence

2 When creating an EGRA, some adapters evaluate the pilot data using item response methodology to determine what modification might be needed prior to finalising the instrument. For example, the RTI International regularly uses Rasch measurement methodology to examine item functioning for EGRA subtasks. This analysis evaluates the effectiveness of each item such as an individual word within a reading passage, and assesses if the item (word) is producing expected responses. Rasch measurement is based on a probabilistic model where the likelihood of a student responding correctly to an item is a function of the student's skill (or ability) and the item's difficulty.

TABLE **1**

Description of EGRA subtasks

Subtask name	Purpose and procedures	Phase(s) of development
Orientation to print	Measures knowledge of early print concepts, such as a word, a letter and directionality. It is untimed and does not have a discontinuation rule.	Pre-alphabetic
Letter name identification	Measures knowledge of letter names. A hundred letters are presented in random order in both upper and lower case. It is timed to 60 seconds and is discontinued if none of the letters in the first line (i.e. first 10 letters) is read correctly.	Partial alphabetic
Letter sound identification*	Measures knowledge of letter–sound correspondence. A hundred letters are presented in random order in both upper and lower case. It is timed to 60 seconds and is discontinued if none of the sounds in the first line (i.e. first 10 sounds) is produced correctly.	Partial alphabetic
Initial sound discrimination	Measures the ability to discriminate beginning sounds. Three words are presented and the aim is to identify the word that begins with a different sound from the other two. It is oral and has 10 sets of words. It is discontinued if no points are earned in the first five items.	Pre-alphabetic Partial alphabetic
Segmentation (phoneme or syllables)	Measures the ability to segment a word into individual phonemes or syllables. This subtask is oral and has 10 items. It is discontinued if no points are earned in the first five items.	Pre-alphabetic Partial alphabetic
Syllable identification	Measures the ability to read individual syllables. Fifty syllables are presented. It is timed to 60 seconds and is discontinued if none of the first five syllables is read correctly.	Partial alphabetic
Familiar word reading	Measures the ability to read individual grade-level words. Fifty words are presented. It is timed to 60 seconds and is discontinued if none of the words in the first line (i.e. first five words) is read correctly.	Partial alphabetic Alphabetic
Nonword reading*	Measures the ability to decode individual nonwords that follow common orthographic structure. Fifty nonwords are presented. It is timed to 60 seconds and is discontinued if none of the words in the first line (i.e. first five nonwords) is read correctly.	Partial alphabetic Alphabetic
Oral reading fluency*	Measures the ability to read a grade-level passage of approximately 60 words. It is scored for accuracy and rate. It is timed to 60 seconds and is discontinued if none of the words in the first line (i.e. about 10 words) is read correctly.	Consolidated-alphabetic
Reading comprehension (with or without lookbacks)*	Measures the ability to answer questions on the grade-level passage. Questions include explicit and inferential types; Lookbacks (i.e. referencing the passage for the answer) can be used if appropriate.	Consolidated-alphabetic Automatic
Cloze	Measures sentence-level comprehension. Several words are presented to complete the sentence. It is untimed and does not have a discontinuation rule.	Consolidated-alphabetic Automatic
Listening comprehension*	Measures receptive language of an orally read passage with both explicit and inferential questions. It is untimed and does not have a discontinuation rule.	Used diagnostically across various phrases
Vocabulary	Measures receptive language skills of individual words and phrases related to body parts, common objects and spatial relationships. It is untimed and does not have a discontinuation rule. Written assessment (government developed) and EGRA oral reading fluency (Grade 4).	Used diagnostically across various phrases
Dictation	Measures the ability to spell and to apply writing conventions in a grade-level sentence. Words can be scored for partial representation.	Partial alphabetic Alphabetic Consolidated-alphabetic
Interview	Gathers information about the child that is related to literacy and language development (e.g. first language, access to print). It is self-reported by the child.	Any phase of interest

Note: * Denotes the subtasks that are considered core for most contexts.

© Dana Schmidt/The William and Flora Hewlett Foundation

support the testing approach and interpretation of the results. Reliability is the overall consistency of a measure—i.e. whether the measure generates similar results under consistent conditions either within a sample of like learners or across repeated measures (American Educational Research Association [AERA] et al., 2014; Glaser et al., 2001). To borrow an example from the health sector, a blood-pressure cuff is a valid way of measuring blood pressure. It is not a valid way of assessing how much an individual weighs. The blood-pressure cuff is a reliable measure if it consistently reports similar or same results for the same individual or like individuals under similar conditions. If it were to give wildly different results for an individual in two applications several minutes apart, it would not be considered a reliable measure.

The validity of the EGRA is tied to (1) the conceptual underpinning of the tool (and its inclusion of valid subtasks of early reading skills) and (2) the usefulness of the results in reporting on student performance in early reading. Overall validity is sometimes characterised using four key aspects: construct, content, concurrent, and predictive validity (AERA et al., 2014). The second edition of the *EGRA Toolkit* (2015) provides a substantial list of references for each of the subtasks that support the construct and content validity of the approach in English (the French and Spanish versions of the toolkit provide additional information for the validity of the measures in those languages) (Sprenger-Charolles, 2009; Jiménez, 2009). For more details regarding these references, see **Appendix I**.

The EGRA Tooklit, Second Edition can be accessed here

The EGRA has also been a part of several concurrent-validity studies. Concurrent validity is shown when an assessment correlates with another assessment that has been previously validated. This gives confidence that the assessments are measuring the same construct and the results are valid. In this stratagem, researchers simultaneously (or concurrently) administer two assessments (usually of the same or similar construct and content) to the same student, then compare the results. Studies attempting to validate the EGRA against other validated assessments have been conducted in Peru (Kudo and Bazan, 2009), Honduras and Nicaragua (Bazan and Gove, 2010). The Kudo and Bazan study (2009) in particular comprehensively reviewed oral and written assessment studies of concurrent validity. Studies seeking to validate other studies against the EGRA include assessments conducted in India (Vagh, 2012) and Kenya (ACER, 2015). For additional information, please see also Vagh's article on concurrent validity in this volume. The correlations comparing the EGRA and other assessments ranged from .41 to .98 and are summarised in **Table 2**. The correlation coefficient provides the strength of the linear relationship between two variables, the closer to 1 the more confidence one has that the assessments are measuring the same construct. The high correlations (.9–.97) are with other oral reading assessments from India and Kenya. The Spanish assessments listed were written tests, which accounts for medium-sized correlations to the EGRA (.41–.47).

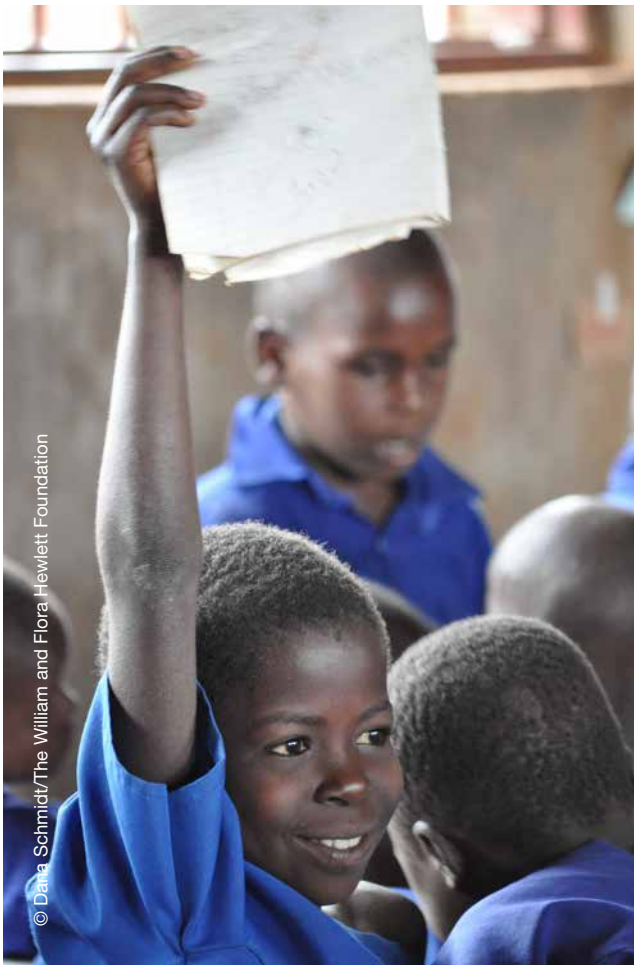
For details on how the studies were conducted, please see the original reports (links available in the reference list at the end of this article).

TABLE **2**

Summary of concurrent validity results

Country	Assessments	Language(s)	Correlation results*	Citations
India	Fluency battery (EGRA adaptation composite score) and Annual Status of Education Reports (ASER)	Hindi	0.9 to 0.94 (depending on ASER classification of student skill level) using Spearman correlation coefficients n varies from 256 to 8,092 depending on round of data collection	Vagh (2012)
Kenya	The EGRA composite score and Twaweza's Uwezo initiative assessments	English Kiswahili	0.961 0.977 n = 1,207 total, approximately 400 for each assessment domain	ACER (2015)
Peru	Written assessment (government administered) and the EGRA	Spanish	0.47 n = 475	Kudo and Bazan (2009)
Honduras	Written assessment (government administered) and the EGRA oral reading fluency (Grade 3)	Spanish	0.42 n = 213	Bazan and Gove (2010)
Nicaragua	Written assessment (government developed) and the EGRA oral reading fluency (Grade 4)	Spanish	0.41 n = 374	Bazan and Gove (2010)

Note: *Pearson's r correlation coefficients stated unless otherwise noted.



Reliability or the consistency of results for a population is estimated on a scale of 0 to 1 (1 being perfect consistency) through several means, including test-retest (where the same individual repeats the assessment usually within a week of the first assessment application) or coefficient alpha, a Classical Test Theory (Cueto and León, 2012) measure that examines the contribution of each subtask to the overall consistency of the instrument. Few results for test-retest reliability of the EGRA could be found, likely because test-retest approaches require identification and tracking of individual students, which can be quite challenging in low-resource settings. One study from India did report test-retest results for a Hindi adaptation of the EGRA with coefficients ranging from 0.83 to 0.98. (Vagh, 2012). Coefficient alpha results—generated using the summary results from each subtask (such as oral reading fluency)—for several studies can be found in **Table 3**. For academic assessments (i.e., with minimal learning between time 1 and time 2) above 0.8 are generally considered acceptable for research purposes; all results in Table 3 are above 0.8.

TABLE **3**

Coefficient alpha results of the EGRA administered in Grade 2 by country and language of instrument

County	Subtasks	Language	n*	Coefficient alpha
Ghana	· Letter sound fluency · Nonword fluency · Oral reading fluency · Reading comprehension	English Akuapem Asante Twi Dagaare Dagbani Dangme Ewe Fante Ga Gonja Kasem Nzema	7,915 687 1,633 541 431 447 492 692 430 423 439 442	0.89 0.9 0.9 0.89 0.89 0.92 0.93 0.86 0.89 0.93 0.88 0.89
Indonesia	· Letter sound fluency · Phonemic awareness initial sound · Nonword fluency · Oral reading fluency · Reading comprehension · Dictation	Bahasa Indonesia	4,812	0.89
Jordan	· Letter sound fluency · Syllable sound fluency · Nonword fluency · Oral reading fluency · Reading comprehension · Dictation	Arabic	1,447	0.9
Kenya	· Letter sound fluency · Syllable sound fluency · Nonword fluency · Oral reading fluency · Reading comprehension	Kiswahili	2,112	0.91
Liberia	· Letter sound fluency · Familiar word fluency · Nonword fluency · Oral reading fluency · Reading comprehension	English	1,249	0.87
Malawi	· Letter sound fluency · Syllable sound fluency · Familiar word fluency · Nonword fluency · Oral reading fluency · Reading comprehension	Chichewa	3,360	0.97
Nigeria	· Letter sound fluency · Nonword fluency · Oral reading fluency · Reading comprehension	Hausa	1,271	0.89
Philippines	· Letter sound fluency · Familiar word fluency · Nonword fluency · Oral reading fluency · Reading comprehension	Cebuano Ilokano Hiligaynon Maguindanaoan	415 399 392 397	0.93 0.94 0.94 0.94
Tanzania	· Syllable sound fluency · Familiar word fluency · Nonword fluency · Oral reading fluency · Reading comprehension · Dictation word score · Dictation punctuation score · Dictation sentence word Score · Dictation sentence score	Kiswahili	2,152	0.96

Note: **n* is recorded for the subtask with the lowest *n* (highest number of missing data).
Source: Authors' calculations from EGRA data sets.

4. WHAT INFORMATION DO PRACTITIONERS AND POLICYMAKERS NEED TO MAKE IMPROVEMENTS IN LEARNING?

The EGRA is almost always accompanied by context questionnaires, classroom and school inventories, observation tools and other instruments that can help contextualise and inform the student assessment results. These instruments provide critical information on a child's home language, human and physical resources in the school, availability of textbooks and reading materials. They serve to link the EGRA results to various components or characteristics of the education system. **Table 4** is an overview of how the EGRA results have been used to inform the sector, drawing on the '5Ts' (test, teach, tongue, text and time) framework put forth in Gove and Cvelich (2011).

General impact evaluations (which may draw on multiple dimensions) using both the EGRA and other school-based assessments of early reading skills similar to the EGRA are included by country in **Table 5**. Many of these impact evaluations have been published through the websites of non-governmental organizations (NGOs) or international donors while a few have made it into the peer-

© Dana Schmidt/The William and Flora Hewlett Foundation

TABLE **4**

Summary review of literature using EGRA results, by topic

Education System Dimension	Assessments
Test: Use of assessment for system-level improvement, global monitoring or classroom-based assessment.	Crouch and Gove (2011) Davidson, Korda and Collins (2011) Dubeck and Gove (2015)* Gove et al. (2013)* Gove et al. (2015) Jiménez et al. (2014)* Wagner et al. (2012)*
Teach: Instructional practices, coaching	Nielsen (2013) Piper and Zuilkowski (2015)*
Tongue: Language-of-instruction policies and language use within the classroom.	Piper (2010) Piper and Miksic (2011) Piper et al. (2015a)* Trudell et al. (2012) Trudell and Piper (2014)*
Text: Availability of materials, use of student and teacher materials.	Ministry of Basic and Secondary Education, Republic of The Gambia (2009) RTI International (2015)
Time: Time on task, instructional time.	Adelman et al. (2015) Moore et al. (2011)* Moore et al. (2012)*

Note: *Denotes peer-reviewed articles or book.

TABLE **5**

Impact evaluations using the EGRA and similar school-based oral reading assessments, by country (endline results only)

Country	Programme	Citation
Afghanistan	Literacy Boost	Azami and Para (2014)
Bangladesh	Literacy Boost	Guajardo et al. (2013); Jonason et al. (2014)
Burundi	Literacy Boost	Rosenkrantz et al. (2014)
Democratic Republic of Congo	Healing Classrooms	Aber et al. (under review)
Egypt	Girls' Improved Learning Outcomes (GILO)	RTI International (2014)
El Salvador	Literacy Boost	Pisani and Alvarado (2014)
Ethiopia	Literacy Boost	Friedlander et al. (2012); Gebreanenia et al. (2014); Jonason & Solomon (2014)
Haiti	Literacy Boost Tout Timoun Ap Li (ToTAL)	Save the Children (2013); RTI International (2015)
Indonesia	Literacy Boost	Brown (2013); Pisani et al. (2014)
Kenya	Primary Math and Reading (PRIMR) Initiative	Piper et al. (2014)*; Piper et al. (2015b)*
Liberia	EGRA Plus EGRA Plus Liberia Teacher Training Program 2 (LTTP2)	Davidson and Hobbs (2013)* Piper and Korda (2010) King et al. (2015)
Malawi	Literacy Boost Malawi Teacher Professional Development Support (MTPDS) Malawi Teacher Professional Development Support (MTPDS)-Reading Intervention	Dowd and Mabeti (2011); Save the Children (2013, 2014); Pouezevara et al. (2013); Pouezevara et al. (2013)
Mali	Programme Harmonisé d'Appui au Renforcement de l'Education (PHARE) Institute pour l'Education Populaire (IEP)	Ralaingita and Wetterberg (2011)*; Spratt (2014)
Mozambique	Literacy Boost	Mungoi et al. (2011)
Nepal	Literacy Boost	Karki and Dowd (2013); Pinto (2010);
Pakistan	Literacy Boost	Mithani et al. (2011); Moulvi and Pava (2014)
Philippines	Literacy Boost Basa Pilipinas	Badiable et al. (2013); Dunlop (2015) Education Development Center (2015a)
Rwanda	Literacy, Learning and Leadership (L3)	Education Development Center (2015b)
Senegal	Harnessing Youth Volunteers as Literacy Leaders (HYVALL)	Education Development Center (2014)
South Africa	District Development Support Program (DDSP)	Ralaingita and Wetterberg (2011)*
Sri Lanka	Literacy Boost	Wickramsekara et al. (2014)
Zimbabwe	Literacy Boost	Pisani and Chinyama (2013)

Note: *Denotes peer-reviewed articles or book chapters.

reviewed literature (those marked with an asterisk). There is growing awareness on the part of the education community, however, of the need to publish results—in particular, impact evaluation results—in the peer-reviewed literature.

5. CONCLUSIONS

The development and adaptation of reliable and valid approaches to understanding early reading skills is a complex process but one that is informed by a considerable body of research—both in high-income contexts and increasingly, in low- and middle-income countries. This article has provided an overview of what is assessed and how the survey generates results that are valid and reliable. The EGRA relies on a proven set of subtasks for understanding key foundational skills in reading drawn from across student assessment approaches. Development and adaptation of the instrument to each new context requires careful adherence to the guidance and recommendations. When adapted correctly and applied using proper survey techniques, the EGRA and its results are reliable and valid depictions of student skills.

APPENDIX **1**

Research supporting the EGRA approach and subtasks

Foundational skill	Measured by EGRA subtask(s):	Supporting research—direct excerpt from the *EGRA Toolkit*, 2nd edition (RTI International, 2015)
Phonological awareness	Initial sound identification; initial sound discrimination Optional: phoneme (or syllable) segmentation	Phonological awareness has been shown across numerous studies in multiple languages to be predictive of later reading achievement (Badian, 2001; Denton et al., 2000; Goikoetxea, 2005; McBride-Chang and Kail, 2002; Muter et al., 2004; Wang, Park and Lee, 2006).
Alphabetic principle and alphabet knowledge	Letter identification (either by letter names or letter sounds); nonword reading; familiar word reading; dictation Optional: syllable identification	Research has shown alphabet knowledge to be a strong early predictor of later reading achievement (Adams, 1990; Ehri and Wilce, 1985; Piper and Korda, 2010; Wagner et al.,, 1994; Yesil-Dağli, 2011) for both native and nonnative speakers of a language (Chiappe et al., 2002; McBride-Chang and Ho, 2005; Manis et al., 2004; Marsick and Watkins, 2001). One of the main differences between successful readers and struggling readers is the ability of successful readers to use the letter-sound correspondence to decode new words they encounter in text and to encode (spell) the words they write (Juel, 1991).
Vocabulary and oral language	None directly but indirectly measured by listening comprehension Optional: vocabulary (untimed)	Reading experts have suggested that vocabulary knowledge of between 90 and 95% of the words in a text is required for comprehension (Nagy and Scott, 2000). It is not surprising then, that in longitudinal studies, vocabulary has repeatedly been shown to influence and be predictive of later reading comprehension (Muter et al., 2004; Roth et al., 2002; Share and Leiken, 2004).
Fluency	Oral reading fluency with comprehension Timed and scored for speed and accuracy: letter name identification, letter sound identification, nonword reading, and familiar word reading	Numerous studies have found that reading comprehension has a relationship to fluency, especially in the early stages (Fuchs et al., 2001). For example, tests of oral reading fluency, as measured by timed assessments of correct words per minute, have been shown to have a strong correlation (0.91) with the reading comprehension subtest of the Stanford Achievement Test (Fuchs et al., 2001). The importance of fluency as a predictive measure does, however, decline in the later stages. As students become more proficient and automatic readers, vocabulary becomes a more important predictor of later academic success (Yovanoff et al., 2005).
Comprehension	Reading comprehension Optional: maze; cloze	Research has not yet produced a proven means to consistently and thoroughly test the higher-level and more nuanced comprehension skills in a standardised way that could be accepted as valid and reliable. However, options are under consideration and it is hoped that the measurement of comprehension will continue to improve as this skill is one of the most important measures of reading success.

REFERENCES

Aber, J.L. (under review). "Impacts of "Healing Classrooms" on children's reading and math skills in DRC". *Journal of Research on Educational Effectiveness.*

ACER (2015). *Report on the concurrent validity and inter-rate reliability studies of Uwezo.* Washington, DC: Results for Development Institute (R4D).

Adams, M.J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Adelman, M., Baron, J.D., Blimpo, M., Evans, D.K., Simbou, A. and Yarrow, N. (2015). *Why do students learn so little? Seeking answers inside Haiti's classrooms.* World Bank Working Paper 96500. Washington, DC: World Bank. Retrieved from https://openknowledge.worldbank.org/bitstream/handle/10986/22064/Why0Do0student0e0Haiti0s0classrooms.pdf?sequence=1

American Educational Research Association (AERA), American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing [US]. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

August, D. and Shanahan, T. (eds) (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth.* Mahwah, NJ: Erlbaum.

Azami, S. and Pava, C. (2014). *Literacy Boost Afghanistan Year 1, October 2014.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-afghanistan-year-1-october-2014

Badiable, A., Guajardo, J., Fermin, R. and Robis, E.J.C. (2013). *Literacy Boost Metro Manila endline report.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-metro-manila-endline-report-2013

Badian, N. (2001). "Phonological and orthographic processing: Their roles in reading prediction". *Annals of Dyslexia*, Vol. 51, pp. 179-202.

Bazan, J. and Gove, A. (2010). *Análisis psicométrico de EGRA y su validez concurrente con otras evaluaciones de desempeño en lectura: caso Honduras y Nicaragua.* Prepared for USAID under the Education Data for Decision Making (EdData II) project. Research Triangle Park, NC: RTI International. http://pdf.usaid.gov/pdf_docs/PNAEA399.pdf

Brown, C. (2013). *Literacy Boost Indonesia: Endline report 2013.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-indonesia-endline-report-2013

Chiappe, P., Siegel, L. and Wade-Woolley, L. (2002). "Linguistic diversity and the development of reading skills: A longitudinal study". *Scientific Studies of Reading,* Vol. 6, No. 4, pp. 369-400.

Crouch, L. and Gove, A. (2011). "Leaps or one step at a time: Skirting or helping engage the debate? The case of reading". W.J. Jacob and J.N. Hawkins (eds), *Policy debates in comparative, international and development education.* New York: Palgrave MacMillan. pp. 120-151. http://www.palgraveconnect.com/pc/doifinder/10.1057/9780230339361.0018

Cueto, S. and León, J. (2012). *Psychometric characteristics of cognitive development and achievement instruments in round 3 of Young Lives.* Young Lives Technical Note 25. Oxford, UK: Young Lives, Oxford Department of International Development (ODID), University of Oxford. http://www.grade.org.pe/upload/publicaciones/Archivo/download/pubs/NDMtn25.pdf

Davidson, M. and Hobbs, J. (2013). "Delivering reading intervention to the poorest children: The case of Liberia and EGRA-Plus, a primary grade reading assessment and intervention". *International Journal of Educational Development*, Vol. *33*, No. 3,pp. 283-293.

Davidson, M., Korda, M. and Collins, O.W. (2011). "Teachers' use of EGRA for continuous assessment: The case of EGRA plus: Liberia". A. Gove and A. Wetterberg (eds), *The Early Grade Reading Assessment: Applications and interventions to improve basic literacy*. Research Triangle Park, NC: RTI Press, pp. 113-138. http://www.rti.org/pubs/bk-0007-1109-wetterberg.pdf

Denton, C.A., Hasbrouck, J.E., Weaver, L.R. and Riccio, C.A. (2000). "What do we know about phonological awareness in Spanish?" *Reading Psychology*, Vol. 21, No. 4, pp. 335-352.

Dowd, A.J. and Mabeti, F. (2011). *Literacy Boost—2 year report Malawi.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-2-year-report-malawi

Dubeck, M.M. and Gove, A. (2015). "The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations". *International Journal of Educational Development*, Vol. 40, pp. 315-322.

Dunlop, M. (2015). *Literacy Boost Metro Manila endline report, March 2015.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-metro-manila-endline-report-march-2015

Education Development Center, Inc. (2014). *Harnessing Youth Volunteers as Literacy Leaders (HYVALL): Endline student assessment report.* Washington, DC: EDC.

Education Development Center, Inc. (2015a). *USAID/Philippines Basa Pilipinas Program: Evaluation report for school years 2013/2014 and 2014/2015*. Washington, DC: EDC.

Education Development Center, Inc. (2015b). *Rwanda National Reading and Mathematics Assessment: Midline report*. Washington, DC: EDC.

Ehri, L.C. and Wilce, L.S. (1985). "Movement into reading: Is the first stage of printed word learning visual or phonetic?" *Reading Research Quarterly*, Vol. 20, pp. 163-179.

Friedlander, E., Hordofa, T., Diyana, F., Hassen, S., Mohammed, O. and Dowd, A.J. (2012). *Literacy Boost. Dendi, Ethiopia. Endline II—final report.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-dendi-ethiopia-endline-ii-final-report

Fuchs, L., Fuchs, D., Hosp, M.K., and Jenkins, J. (2001). "Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis". *Scientific Studies of Reading,* Vol. 5, No. 3, pp. 239-256.

Gebreanenia, Z., Sorissa, M., Takele, M., Yenew, A. and Guajardo, J. (2014). *Literacy Boost Tigray: Ethiopia endline evaluation report.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-tigray-ethiopia-endline-evaluation-report

Glaser, R., Chudowsky, N. and Pellegrino, J.W. (eds). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.

Goikoetxea, E. (2005). "Levels of phonological awareness in preliterate and literate Spanish-speaking children". *Reading and Writing,* Vol. 18, pp. 51-79.

Gough, P.B. and Tumner, W.E. (1986). "Decoding, reading, and reading disability." *Remedial and Special Education,* Vol. 7, pp. 6-10.

Gove, A., Chabbott, C., Dick, A. DeStefano, J., King, S., Mejia, J. and Piper, B. (2015). *Early learning assessments: A retrospective.* Background Paper for the *Education for All Global Monitoring Report 2015.* Paris, France: UNESCO. http://unesdoc.unesco.org/images/0023/002324/232419e.pdf

Gove, A. and Cvelich, P. (2011). *Early reading: Igniting education for all. A report by the Early Grade Learning Community of Practice.* Rev. edn, Research Triangle Park, NC: Research Triangle Institute. http://spectra.rti.org/pubs/early-reading-report_gove_cvelich.pdf

Gove, A., Habib, S., Piper, B. and Ralaingita, W. (2013). "Classroom-up policy change: Early reading and math assessments at work". *Research in Comparative and International Education*, Vol. 8, No. 3, pp. 373-386.

Guajardo, J., Hossain, M., Nath, B.K.D. and Dowd, A.J. (2013). *Literacy Boost Bangladesh endline report.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-bangladesh-endline-report-2013

Hoover, W. and Gough, P. (1990). "The simple view of reading". *Reading and Writing: An Interdisciplinary Journal,* Vol. 2, pp. 127-160.

Jiménez, J. (2009). *Manual para la evaluación inicial de la lectura en niños de educación primaria.* Prepared for USAID under the Education Data for Decision Making (EdData II) project. Research Triangle Park, NC: RTI International. http://pdf.usaid.gov/pdf_docs/PNADS441.pdf

Jiménez, J. E., Gove, A., Crouch, L. and Rodríguez, C. (2014). "Internal structure and standardized scores of the Spanish adaptation of the EGRA (Early Grade Reading Assessment) for early reading assessment". *Psicothema*, Vol. 26, No. 4, pp. 531-537.

Jonason, C., Guajardo, J., Nath, B.D. and Hossain, M. (2014). *Literacy & Numeracy Boost Bangladesh endline, April 2014.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-numeracy-boost-bangladesh-endline-april-2014

Jonason, C. and Solomon, S. (2014). *Literacy Boost Somali region, Ethiopia: Endline report 2014.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-somali-region-ethiopia-endline-report-2014

Juel, C. (1991). "Beginning reading". R. Barr, M. L. Kamil, P. Mosenthal, and P.D. Pearson (eds.), *Handbook of reading research,* New York: Longman, pp. 759-788.

Karki, V. and Dowd, A. J. (2013). *Literacy Boost Kapilvastu, Nepal: Year 1 report, 2013.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-kapilvastu-nepal-year-1-report-2013

King, S., Korda, M., Nordstrom, L. and Edwards, S. (2015). *Liberia Teacher Training Program: Endline Assessment of the Impact of Early Grade Reading and Mathematics Interventions.* Prepared for USAID/Liberia, Ministry of Education: Republic of Liberia, and FHI 360. Research Triangle Park, NC: RTI International.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge, UK: Cambridge University Press.

Kudo, I. and Bazan, J. (2009). *Measuring beginner reading skills: An empirical evaluation of alternative instruments and their potential use for policymaking and accountability in Peru.* Policy Research Working Paper 4812. Washington, DC: World Bank.

Manis, F.R., Lindsey, K.A. and Bailey, C.E. (2004). "Development of reading in grades K-2 in Spanish-speaking English language learners". *Learning Disabilities Research and Practice,* Vol. 19, No. 4, pp. 214-224.

Marsick, V.J. and Watkins, K.E. (2001). "Informal and incidental learning". *New Directions for Adult and Continuing Education,* Vol. 89, pp. 25-34.

McBride-Chang, C. and Ho, C. S.-H. (2005). "Predictors of beginning reading in Chinese and English: A 2-year longitudinal study of Chinese kindergarteners". *Scientific Studies of Reading*, Vol. 9, pp. 117-144.

McBride-Chang, C. and Kail, R.V. (2002). "Cross-cultural similarities in the predictors of reading acquisition". *Child Development,* Vol. 73, pp. 1392-1407.

Ministry of Basic and Secondary Education, Republic of The Gambia. (2009). *Report on impact assessment of interventions on early grade reading ability (EGRA) in schools.* https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=270

Mithani, S., Alam, I., Babar, J.A., Dowd, A.J. and Ochoa, C. (2011). *Literacy Boost Pakistan: Year 1 report*. Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/pepas-literacy-boost-pakistan-endline-report-january-2014

Moore, A.M.S., DeStefano, J. and Adelman, E. (2011). "Time misspent, opportunities lost: Use of time in school and learning". W.J. Jacob and J.N. Hawkins (eds.), *Policy debates in comparative, international and development education,* New York: Palgrave MacMillan, pp. 241-258. http://www.palgraveconnect.com/pc/doifinder/10.1057/9780230339361.0018

Moore, A.M.S., Smiley, A., DeStefano, J. and Adelman, E. (2012). "The right to quality education: How use of time and the language of instruction impact the rights of students". *World Studies in Education*, Vol. 13, No. 2, pp. 67-86.

Moulvi, Z.F. and Pava, C. (2014). *Literacy Boost Quetta, Pakistan, Year 2, November 2014.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-quetta-pakistan-year-2-november-2014

Mungoi, D., Mandlante, N., Nhatuve, I., Mahangue, D., Fonseca, J. and Dowd, A.J. (2011). *Endline report of early literacy among pre-school and primary school children in Gaza Province, Mozambique*. Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/endline-report-early-literacy-among-pre-school-and-primary-school-children-gaza-province

Muter, V., Hulme, C., Snowling, M.J. and Stevenson, J. (2004). "Phonemes, rimes, vocabulary, and grammatical skills as foundation of early reading development: Evidence from a longitudinal study". *Developmental Psychology*, Vol. 40, pp. 665-681.

Nagy, W. E. and Scott, J. (2000). "Vocabulary processes". M.E.A. Kamil, P.B. Mosenthal, P.D. Pearson and R. Barr, (eds.), *Handbook of reading research*, Vol. III, Mahwah, NJ: Erlbaum, pp. 269-284.

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel.* Washington, DC: National Institute for Literacy.

National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: U.S. Government Printing Office. https://www.nichd.nih.gov/publications/pubs/nrp/Pages/smallbook.aspx

Nielsen, H.D. (2013). *Going to scale: The Early Grade Reading Program in Egypt: 2008-2012.* Prepared for USAID under the Education Data for Decision Making (EdData II) project, Data for Education Programming in Asia and the Middle East (DEP-ASIA/ME). Research Triangle Park, NC: Research Triangle Institute. https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=606

Oullette, G.P. (2006). "What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension". *Journal of Educational Psychology,* Vol. 98, pp. 554-566.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.

Pinto, C. (2010). *Impact of Literacy Boost in Kailali, Nepal 2009-2010: Year 1 report*. Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-kailali-nepal-year-1-report

Piper, B. (2010). *Uganda Early Grade Reading Assessment findings report: Literacy acquisition and mother tongue.* Prepared for the William and Flora Hewlett Foundation. Research Triangle Park, NC: RTI International and Makerere University Institute for Social Research. https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=293

Piper, B., Jepkemei, E. and Kibukho, K. (2015b). "Pro-poor PRIMR: Improving early literacy skills for children from low-income families in Kenya". *Africa Education Review*, Vol. 12, No. 1, pp. 67-87.

Piper, B. and Korda, M. (2010). *EGRA Plus: Liberia. Program evaluation report.* Prepared for USAID/Liberia under the Education Data for Decision Making (EdData II) project, Early Grade Reading Assessment (EGRA): Plus Project. Research Triangle Park, NC: RTI International. http://pdf.usaid.gov/pdf_docs/pdacr618.pdf

Piper, B. and Miksic, E. (2011). "Mother tongue and reading: Using early grade reading assessments to investigate language-of-instruction policy in East Africa". A. Gove and A. Wetterberg (eds.), *The Early Grade Reading Assessment: Applications and interventions to improve basic literacy*. Research Triangle Park, NC: RTI Press, pp. 139-182. http://www.rti.org/pubs/bk-0007-1109-wetterberg.pdf

Piper, B., Schroeder, L. and Trudell, B. (2015a). "Oral reading fluency and comprehension in Kenya: reading acquisition in a multilingual environment". *Journal of Research in Reading.* http://dx.doi.org/10.1111/1467-9817.12052

Piper, B. and Zuilkowski, S.S. (2015). "Teacher coaching in Kenya: Examining instructional support in public and nonformal schools". *Teaching and Teacher Education*, Vol. 47, pp. 173-183.

Piper, B., Zuilkowski, S.S., and Mugenda, A. (2014). "Improving reading outcomes in Kenya: First-year effects of the PRIMR Initiative". *International Journal of Educational Development*, Vol. 37, pp. 11-21.

Pisani, L. and Alvarado, M. (2014). *Literacy Boost El Salvador endline report, November 2014.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-el-salvador-endline-report-november-2014

Pisani, L. and Chinyama, A. (2013). *Literacy Boost Zimbabwe baseline report, 2012*. Washington, DC: Save the Children.

Pisani, L., Puta, S., Ni, L., Giri, B., Alesbury, C. and de Fretes, M. (2014). *Literacy Boost Belajar Indonesia midline and endline report.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-belajar-indonesia-midline-endline-report

Pouezevara, S., Costello, M. and Banda, O. (2013). *Malawi reading intervention: Early grade reading assessment, final assessment—2012.* Prepared for USAID under the Malawi Teacher Professional Development Support Program. Washington, DC: Creative Associates. http://pdf.usaid.gov/pdf_docs/pa00jqj4.pdf

Ralaingita, W. and Wetterberg, A. (2011). "Gauging program effectiveness with EGRA: Impact evaluations in South Africa and Mali". A. Gove and A. Wetterberg (eds.), *The Early Grade Reading Assessment: Applications and interventions to improve basic literacy,* Research Triangle Park, NC: RTI Press, pp. 83-112). http://www.rti.org/pubs/bk-0007-1109-wetterberg.pdf

Rosenblatt, L. (1978). *The reader, the text, the poem: The transactional theory of the literary work.* Carbondale, IL: Southern Illinois University Press.

Rosenkranz, E., Jonason, C. and Kajangwa, D. (2014). *Literacy Boost Burundi endline report, August 2014.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-burundi-endline-report-august-2014

Roth, F.P., Speece, D.L. and Cooper, D.H. (2002). "A longitudinal analysis of the connection between oral language and early reading". *Journal of Educational Research,* Vol. 95, pp. 259-272.

RTI International. (2009). *Early grade reading assessment toolkit.* Research Triangle Park, NC: RTI International.

RTI International. (2013). *The Primary Math and Reading (PRIMR) Initiative: Annual report, 1 October 2012–31 September 2013.* Prepared under the USAID Education Data for Decision Making (EdData II) Project. Research Triangle Park, NC: RTI. http://pdf.usaid.gov/pdf_docs/PA00K262.pdf

RTI International. (2014). *Girls' Improved Learning Outcomes: Final report*. Research Triangle Park, NC: RTI International. http://pdf.usaid.gov/pdf_docs/pa00jtbc.pdf

RTI International. (2015). *Early Grade Reading Assessment toolkit,* Second Edition, prepared for USAID under the Education Data for Decision Making (EdData II) project, Research Triangle Park, NC: RTI. http://static1.squarespace.com/static/55c4e56fe4b0852b09fa2f29/t/56e0633545bf213c2b5269e5/1457546040584/EGRA+Toolkit+Second+Edition_March_8_2016+.pdf

RTI International (2015). *Tout Timoun Ap Li - ToTAL (All Children Reading) final report, REVISED, 2 August 2012–5 December 2014* [Haiti]. Prepared under the Education Data for Decision Making (EdData II) project. http://pdf.usaid.gov/pdf_docs/PA00K911.pdf

RTI International and International Rescue Committee (IRC). (2011). *Guidance notes for planning and implementing early grade reading assessments.* Washington, DC: RTI and IRC. https://www.eddataglobal.org/reading/index.cfm?fuseaction=pubDetail&id=318

Save the Children (2013). *Reading is the future: Lekti se lavni: baseline-endline assessments report* [Haiti]. Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/reading-future-lekti-se-lavni-baseline-endline-assessments-report

Save the Children (2013). *Save the Children International Basic Education Program: TiANA project endline report, 2013* [Malawi]. Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/save-children-international-basic-education-program-tiana-project-endline-report-2013

Save the Children (2014). *TiANA final evaluation report for ST Anthony and Namadidi education zone, Zormba rural, Malawi, September 2014.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/tiana-final-evaluation-report-st-anthony-and-namadidi-education-zones-zormba-rural-malawi

Seymour, P.H., Aro, M. and Erskine, J.M. (2003). "Foundation literacy acquisition in European orthographies". *British Journal of psychology*, Vol. 94, No. 2, pp. 143-174.

Share, D.L. and Leikin, M. (2004). "Language impairment at school entry and later reading disability: Connections at lexical versus supralexical levels of reading". *Scientific Studies of Reading,* Vol. 8, pp. 87-110.

Snow, C. and the RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Research prepared for the Office of Educational Research and Improvement (OERI), U.S. Department of Education. Santa Monica, CA: RAND Corporation.

Spratt, J., King, S., and Bulat, J. (2013). *Independent Evaluation of the Effectivenss of Institut pour l'Education Poulaire's Read-Learn-Learn (RLL) Program in Mali.* Prepared for the William and Flora Hewlett Foundation under Grant #2008-3229. Research Triangle Park, NC: RTI International.

Sprenger-Charolles, L. (2009). *Manuel pour l'evaluation des competences fondamentales en lecture.* Prepared for USAID under the Education Data for Decision Making (EdData II) project. Research Triangle Park, NC: RTI International. http://pdf.usaid.gov/pdf_docs/PNADQ182.pdf

Trudell, B., Dowd, A.J., Piper, B. and Bloch, C. (2012). *Early grade literacy in African classrooms: Lessons learned and future directions.* Conference paper for Triennial on Education and Training in Africa, African Development Education Association. http://www.adeanet.org/triennale/Triennalestudies/subtheme1/1_5_04_TRUDELL_en.pdf

Trudell, B. and Piper, B. (2014). "Whatever the law says: Language policy implementation and early-grade literacy achievement in Kenya." *Current Issues in Language Planning*, Vol. 15, No. 1, pp. 4-1.

Vagh, S.B. (2012). *Validating the ASER testing tools: Comparisons with reading fluency measures and the Read India measures.* Unpublished report. http://img.asercentre.org/docs/Aser%20survey/Tools%20validating_the_aser_testing_tools__oct_2012__2.pdf

Vellutino, F.R., Tunmer, W.E., Jaccard, J.J. and Chen, R. (2007). "Components of reading ability: Multivariate evidence for a convergent skills model of reading development". *Scientific Studies of Reading*, Vol. 11, pp. 3-32.

Wagner, D.A., Lockheed, M., Mullis, I., Martin, M.O., Kanjee, A., Gove, A. and Dowd, A.J. (2012). "The debate on learning assessments in developing countries". *Compare: A Journal of Comparative and International Education*, Vol. 42, No. 3, pp. 509-545.

Wagner, R.K., Torgesen, J.K. and Rashotte, C.A. (1994). "The development of reading-related phonological processing abilities: New evidence of bi-directional causality from a latent variable longitudinal study". *Developmental Psychology*, Vol. 30, pp. 73-87.

Wang, M., Park, Y. and Lee, K.R. (2006). "Korean-English biliteracy acquisition: Cross-language phonological and orthographic transfer". *Journal of Educational Psychology*, Vol. 98, pp. 148-158.

Wickramesekara, P., Navaratnam, S. and Guajardo, J. (2014). *Literacy Boost, Gampaha District Sri Lanka country office endline report—December 2014.* Washington, DC: Save the Children. http://resourcecentre.savethechildren.se/library/literacy-boost-gampaha-district-sri-lanka-country-office-endline-report-december-2014

Yesil-Dağli, Ü. (2011). "Predicting ELL students' beginning first grade English oral reading fluency from initial kindergarten vocabulary, letter naming, and phonological awareness skills". *Early Childhood Research Quarterly,* Vol. 26, No. 1, pp. 15-29.

Yovanoff, P., Duesbery, L., Alonzo, J. and Tindall, G. (2005). "Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency". *Educational Measurement,* Vol. Fall, pp. 4-12.

What and How to Assess Reading Using Household-Based, Citizen-Led Assessments: Insights from the Uwezo Annual Learning Assessment

MARY GORETTI NAKABUGO
Twaweza East Africa

ABBREVIATIONS

ASER	Annual Status of Education Report
EA	Enumeration areas
EGRA	Early Grade Reading Assessment
NAPE	National Assessment of Progress in Education
MDGs	Millennium Development Goals
SDGs	Sustainable Development Goals
UIS	UNESCO Institute for Statistics

1. INTRODUCTION

Global efforts to achieve universal education for all have given rise to a movement of individuals and institutions committed to measuring learning attained out of the formal education settings, such as classrooms and schools. Assessment in formal school settings does not fully account for the learning of children who may be out of school on a given day when the assessment is undertaken. Conversely, assessments outside of the school setting would allow access to children irrespective of their participation in school (i.e. irrespective of attendance, enrolment and school choice). What has emerged from this movement is a shift of focus from measuring success in education based on inputs such as gross enrolment ratios, presence of infrastructure such as classrooms, textbooks, and teachers to a focus on actual learning outcomes of all children. It is now widely accepted that education cannot be truly universal until every child who enrols learns (Filmer et al., 2006). However, the challenge has been developing a universal understanding and measure of what constitutes learning (Pritchett et al., 2013).

As a result of attempts to address this learning measurement challenge, a number of learning assessment systems have been developed, one of which is the household-based assessment (also commonly referred to as citizen-led assessment). The household-based assessment is a learning assessment that is done at the household level when children are at home away from the formal school setting. The assessment is undertaken by trained citizen volunteers and is designed in ways that enable all children within a selected age-group (some of whom may be out of school for specific reasons) to be assessed. In addition, the household-based learning assessment engages children in the presence of their parents/guardians so that instant feedback on the children's learning levels is provided to facilitate the sharing of information and to help create awareness.

The household-based, citizen-led assessment was mainly popularised by the Annual Status of Education Report (ASER) in India in 2005 and has now taken root in Pakistan, Mali, Senegal and East Africa (Kenya, the United Republic of Tanzania and Uganda). Mexico and Nigeria have since come on board as well. The East-African household-based assessment initiative is code-named 'Uwezo', a Kiswahili word that means 'capability'. The Uwezo initiative is a programme of Twaweza East Africa, an organization that helps enable children to learn, citizens to exercise agency and governments to be

more open and responsive in the United Republic of Tanzania, Kenya and Uganda. The Uwezo assessment, like other citizen-led assessments, is a simple household-based tool to measure basic literacy and numeracy among children aged 6-16 years.

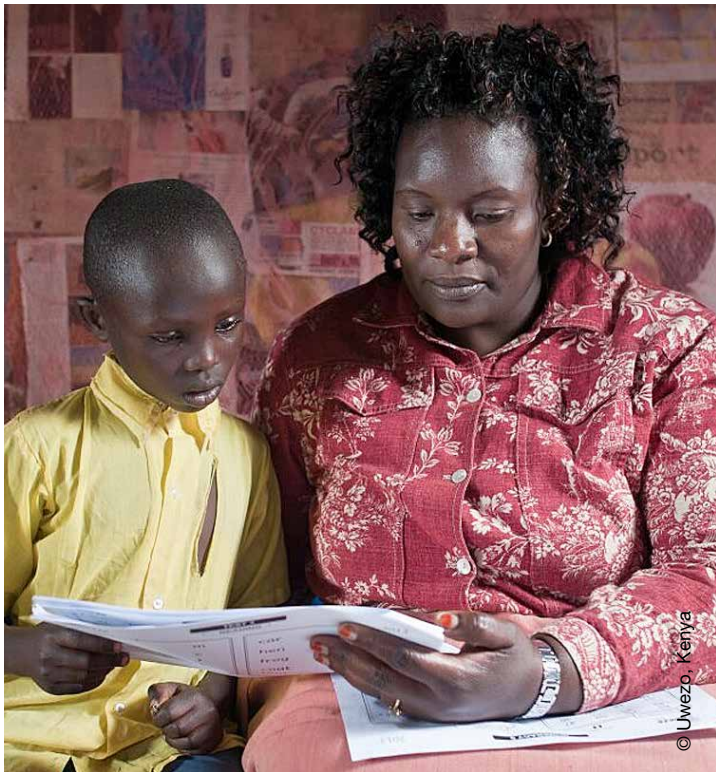
This article gives an account of how the Uwezo household-based, citizen-led assessment of oral reading is undertaken by addressing the following questions:

- What to measure? Who to measure?
- How to administer?
- Where to measure?
- When to measure?
- Who measures?
- Why measure using the household assessment?

The following sections address each of these key questions.

2. WHAT AND WHO DOES THE UWEZO HOUSEHOLD-BASED, CITIZEN-LED ASSESSMENT MEASURE?

The Uwezo household-based, citizen-led assessment measures the basics of oral reading mainly based on a phonics approach. Phonics denotes a systematic teaching of sound-symbol relationships to decode words. The phonic approach is based on the thinking that learning to read is incremental in nature. For children to be able to read words, they need to know letter sounds or letter names as the ability to read whole sentences is dependent on the ability to read words. Thus, the Uwezo household-based, citizen-led assessment works on the assumption that without learning basic competencies such as how to sound out letters, a child cannot progress to reading extended texts or move up in the school system. In addition—in keeping with common beliefs among citizen-led assessment practitioners—the Uwezo assessment developers believe that basic competencies in oral reading need to be measured and built irrespective of grade. This belief is contrary to conventional


© Uwezo, Kenya

assessments and examinations used in East Africa, which are usually based on content from specific grades.

Instead of grade-level assessments tied to grade-level content, the Uwezo assessment measures basic reading competencies of children aged 6-16 years.[1] The rationale is that it cannot be assumed that children possess these basics so a one-on-one oral assessment of these competencies is essential.

To this end, the Uwezo assessment uses a simple authentic tool that can be understood, implemented and the results interpreted by non-education professionals and ordinary citizens. In summary, the competencies that are assessed by the Uwezo tool are the basics of reading based on a phonics approach that each individual in the target population group ought to have. Therefore, using the tool can help indicate possession or lack of the basic skills. An example of the Uwezo oral reading assessment tool is presented in **Figure 1**.

1 In Kenya and Uganda, the Uwezo learning assessment is administered to children aged 6-16 years and in the United Republic of Tanzania with those aged 7-16 years. In the United Republic of Tanzania, children begin primary school at age 7 years.

Figure 1. Example of the Uwezo oral reading test: Uwezo Kenya 2013 test

LITERACY TEST

- Letter/sound
- Word
- Paragraph
- Story
- Comprehension

ENGLISH READING TEST

LETTER

e	x
d	w
k	c
h	b
j	a

WORDS

room	face
table	dog
desk	pen
ear	fish
bean	man

PARAGRAPH

My mother works in Lamu.
Lamu is a busy town.
The people there are good.
They are very kind.

STORY

Juma reads to us a story from his book everyday. He reads the story aloud in class. We enjoy listening to the stories. Yesterday he read about the sun and the wind. Both of them lived in the same sky. The wind did not like the sun. It wanted to be the head of the sky. One day, the wind chased the sun away. It told the sun to go to another sky. The sun did not go. The next morning, the wind ran after the sun. The sun fell down and started crying. That is how it began to rain. We clapped for Juma.

Q1. What does Juma do every day?

Q2. How did rain begin?

Source: extracted from Uwezo Kenya 2013 National Learning Assessment Test Booklet

3. HOW IS THE MEASUREMENT OF ORAL READING COMPETENCIES DONE IN THE UWEZO ASSESSMENT?

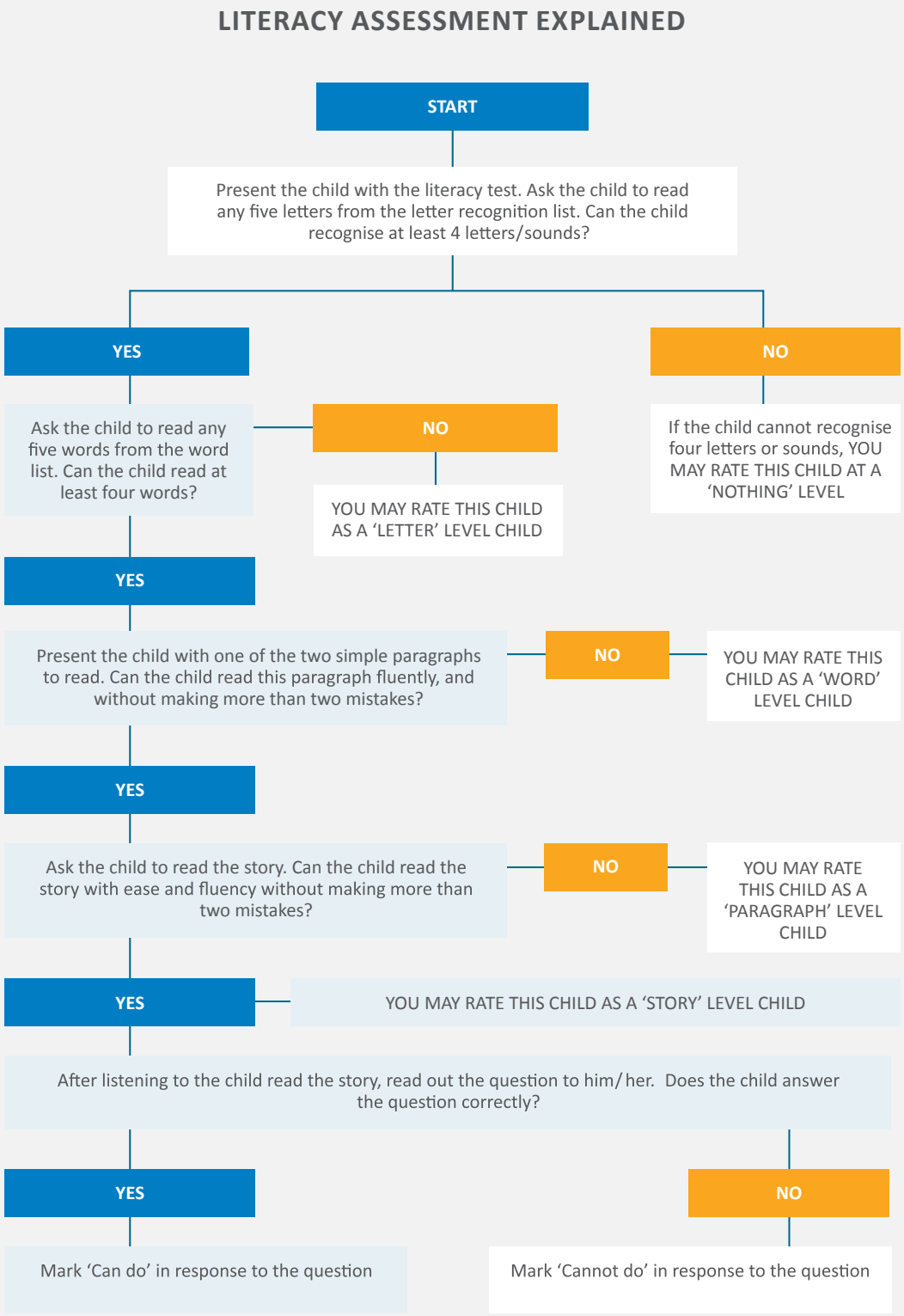
While many learning assessments in East Africa are pen and paper based, the Uwezo household-based citizen-led assessment is administered orally. This is because the Uwezo assessments are undertaken in countries where most children—even those in Grade 2 and above—cannot read so pen and paper tests are not an appropriate method of assessment. As such, alternative methods and mechanisms that suit these contexts are needed. Oral assessments are also necessary to measure basic sounding out skills.

Furthermore, the Uwezo assessment of oral reading competencies is administered in the household and is undertaken in a non-threatening way—the test is not timed, no mark is awarded and qualitative feedback on learning achievement of the assessed child is given to the child and parent/guardian instantly. In this scenario, anyone within the target population group can attempt the test at his/her own pace. The assessment is terminated at the point where the child being assessed exhibits the highest reading competency level (*refer to **Figure 2** for the Uwezo literacy assessment flow chart*).

Rather than award marks or a grade that says little about the reading competencies of the child, the household assessment administrator/volunteer ticks the highest level of competency the child achieved. As noted earlier, the assessment tool is based on the natural progression in reading ability, starting from letter sounds and reading words to reading and comprehending extended text at the highest foundational level. If this assessment was administered in a school context, the practice of recording the highest competency level would be useful in clustering children according to the current reading competencies and teaching them at the right level. **Figure 3** is an extract from the Uwezo assessment tool where assessment results for each child are recorded by ticking the highest level of reading competency demonstrated.

For example, if the child reads up to word level and is not able to proceed to paragraph level, a tick is inserted under 'word' level. This automatically implies that the child is also able to sound/name letters but is not able to read sentences at paragraph and story levels.

Figure 2. Uwezo process of assessing basic oral reading competencies



Source: extracted from Uwezo 2013 volunteer manual

Figure 3. Template for recording children's highest competencies in literacy (and numeracy)

H1400. **Basic Learning Levels**

Test all Children between the ages of 6 - 16 only and tick the highest level attained

English Literacy **(Tick the highest level)**					Comprehension For those reading story only				Local language ..... **(Tick the highest level)**					Comprehension For those reading story only				Numeracy **(Tick the highest level)**							Bonus **(For all children 6 - 16 yrs)**				***(For all children*** 6- 16 yrs) **Visual acuity** H1500. Can the child point to the direction where figure E is facing from 20ft?			
					Q1		Q2							Q1		Q2									Q1		Q2		Right Eye		Left Eye	
Nothing	Letter	Word	Para	Story	Can do	Cannot do	Can do	Cannot do	Nothing	Syllables	Word	Para	Story	Can do	Cannot do	Can do	Cannot do	Nothing	Counting	Num Rec. 10 - 99	Addition	Subtraction	Multiplication	Division	Can do	Cannot do	Can do	Cannot do	Yes	No	Yes	No

Source: Uwezo survey booklet

4. WHERE DOES THE UWEZO ASSESSMENT TAKE PLACE?

Unlike other learning assessments that are conducted at school level, such as the international Early Grade Reading Assessment (EGRA) and national assessments such as the National Assessment of Progress in Education (NAPE) of Uganda, the Uwezo assessment is implemented at the household level. In-school assessments are most appropriate in developed countries where "typically all children are in school, and all schools are listed and fall under the jurisdiction of some national or provincial authority" (ASER 2014: 2015). In the context of the developed world where all schools and children are accounted for, it is then possible to draw a representative sample of children either by age or grade nationally or regionally.

The Uwezo survey booklet can be accessed here.

In less developed countries, such as those where the Uwezo assessment and other household-based assessments are undertaken, accurate data on all children and schools (private and government) may not be readily available nor reliable if they are available. Also, school attendance varies with an estimated 20% of primary school children absent from school on any given day in Kenya, The United Republic of Tanzania and Uganda (Uwezo, 2014). In addition, there is a sizeable percentage of children that fall within the target age bracket who are out of school (*please refer to* ***Table 1***). In such a context, drawing a representative sample for in-school assessment becomes a challenge and the best place to find all children is in the household. Indeed, if the commitment is to ensure "learning for ALL children", then measurements should also be representative of all children.

In order to provide an equal chance for all children aged 6-16 years to be assessed in each household, a two-stage stratified sampling design is embraced in each country. In the first stage, 30 enumeration

TABLE 1

Rate of out-of-school children			
Country	**Target population**	**Rate of out-of-school children of primary-school age**	**Rate of out-of-school children of lower secondary-school age**
Kenya	6–16 years old	15.1	—
Uganda	6–16 years old	11.9	23.1
United Republic of Tanzania	7–16 years old	15.5	41.6

Source: Uwezo and the Unesco Institute for Statistics

areas (EAs)/villages are selected from each of the districts using probability proportional to size—a sampling procedure where the selection probability for each EA is set to be proportional to the number of households within the EA. This means that EAs with higher numbers of households have a higher chance of being selected. The second and ultimate stage involves randomly selecting 20 households from each of the 30 EAs, resulting in 600 randomly selected households per district.

Another characteristic to note about the household-based assessment is that the communication of results is an important element. The main rationale for the household approach is to assess children in the presence of their parents/guardians so that instant feedback on the children's learning levels is provided to inspire the parents to take action to support learning. Thus, a unique characteristic of household-based assessments like the Uwezo assessment is that they are undertaken in communities so the assessment tends to influence communities to take action to improve learning outcomes. Feedback is instantly given to the citizens at the household level and regularly to policymakers through national report launches and media coverage. "Assessment administration in the school does not allow for awareness that might lead to such action (in fact, in many assessments administered by schools, the school staff does not even see the results, not to mention children or parents)" (Results for Development Institute, 2013). Prior to leaving the home, the assessor presents the family with a calendar, which includes written messages of what the parents can do to improve their children's learning (e.g. "encourage your child to read at home").

5. WHEN TO MEASURE USING HOUSEHOLD ASSESSMENTS?

There are different types of learning assessments in East Africa: examinations, national assessments and the household-based, citizen-led assessments. It is common among countries in East Africa to conduct public examinations at the end of primary education to establish if students have mastered the curriculum contents at the end of the cycle. For example, the Primary Leaving Examination is administered at the end of Grade 7 (the end of primary education) in Uganda; the Kenya Certificate of Primary Education is administered in Grade 8; and in the United Republic of Tanzania, the National Primary Examination is administered in Grade 7. These examinations are high-stakes. That is, they are used to select students who qualify to enter secondary education programmes (UIS Database of Learning Assessments).

National learning assessments in East Africa—such as the National Assessment of Progress in Education (NAPE) administered in Grades 3, 6 and 9 in Uganda and the National Assessment for Monitoring Learning Achievement administered in Grade 3 in Kenya—are conducted irregularly and data from such measurements are not released instantly after the data collection. This makes acting on the data to promote improvement in learning difficult as feedback is very delayed. In contrast, the Uwezo assessments are administered annually to assess basic reading competencies of children in the target age group. The frequency of this household-based, citizen-led assessment is meant to "act as a thermometer to measure the 'health' of learning and provide evidence on the action that is urgently needed" to drive the attention of decision makers, policy planners and implementers towards improving learning outcomes (Banerji, 2014).

6. WHO MEASURES?

The Uwezo learning assessments are conducted by citizen volunteers who are recruited from

the assessment villages and trained thoroughly on the processes of administering the reading assessments and recording the results. Volunteers are monitored as they administer the assessment. As described earlier, the assessment tools are simple and straightforward, designed so they can be easily understood and administered by non-teachers. The minimum qualification for a volunteer is the possession of a lower secondary certificate, to be a person of integrity and to possess the ability to speak and read the local language used in the assessed community. Two volunteers are recruited in each assessed village through a district-based civil society organization. The volunteers then undergo a two-day residential training by Uwezo-trained trainers[2] on the entire process of administering the assessment before they commence the activity.

This practice of working with citizen-volunteers to administer reading assessments is contrary to common practice in most countries where governments work with professionals to administer the assessment. Working with citizen-volunteers allows ordinary citizens to engage in understanding the status of schooling and learning in their own areas. This engagement and understanding has the potential to lead to the next step—community action to improve learning outcomes. The concept of citizen-volunteers performing the basic measurement of the learning among children and doing something to improve it is equivalent to Village Health Teams in the health sector in Uganda who are trained to diagnose and administer basic medication for basic illnesses (Uganda Village Project, 2009).

7. WHY MEASURE USING A HOUSEHOLD-BASED ASSESSMENT?

Household assessments play multiple roles and provide evidence for use by multiple audiences. At the national level, due to their annual occurrence, they provide regular data on the performance of the education system and what action might be taken to improve learning.

At the school level, household assessments play a role in accountability. The data collected on school-going children can be linked to the school they attend (private or government). It is possible for the information collected on children's competencies to be used to hold individual schools accountable to stakeholders if a large enough sample is drawn from an individual school. The information collected on out-of-school children can be used to measure gaps in early reading (an important skill for lifelong learning).

Further, collecting data in the household allows practitioners to obtain household level data which when analysed alongside the evidence on literacy and numeracy, can help confirm certain assumptions. For instance, the data have shown the association between socio-economic levels and competency levels. This type of data cannot be captured using school-based assessments.

Most importantly, the main reason that household assessments are undertaken, as the title 'household' suggests, is to empower ordinary citizens and communities to play a key role in the education of their children. The fact that assessment is conducted at home as opposed to at school where formal teaching takes place helps contradict the thinking that the role of education is the responsibility of schools and teachers alone. Parents have a major role to play in education too and this role is strengthened by the knowledge of how the education system is functioning. Without this knowledge, parents have no evidence to work with and may assume that the mere fact that their children are in school equals learning. In recent years, the Uwezo assessment team has also worked closely with local council leaders in selected districts to communicate findings from the Uwezo assessment in the communities they lead. Although, currently, community-specific data are not collected, district data can help provide a general overview of children's learning levels.

2 Uwezo implements a cascade model of training, which includes training of trainers in a refresher course followed by regional and national level training sessions before they train the volunteers who administer the learning assessment at the household level.

8. CONCLUSION

New Sustainable Development Goals (SDGs) have been developed to replace the Millennium Development Goals (MDGs). The fourth SDG focuses on ensuring "inclusive and equitable quality education and promote life-long learning opportunities for all" (UNSD, 2015). Just as quality education remains a core part of the new set of goals so should assessment of and for learning. There will be a need for assessment approaches that not only measure if learning has been attained but that also lend themselves easily to action for the improvement of learning. Since the target is to have all children learn well, there will be a need to assess at scale and support each school age child to acquire the basic foundations of reading, writing, numeracy and other higher level sets of learning skill sets, such as critical thinking. Household assessments have the potential to engage and build local resources and a nationwide capacity to ensure that all children achieve quality learning.

REFERENCES

ASER (2015). *Annual Status of Education Report (Rural)* 2014. New Delhi: ASER Centre.

Banerji, R. (2014). "Citizen-led assessments: Nuts and bolts". Presentation at the CIES conference, Toronto, Canada.

Filmer, D., Hasan, A., and Pritchett, L. 2006. "A Millennium Learning Goal: Measuring Real Progress in Education". *The Center for Global Development and The World Bank Working Paper No. 97*.

Pritchett, L., Banerji, R. and Kenny, C. (2013). *Schooling is Not Education! Using Assessment to Change the Politics of Non-Learning.* Washington D.C.: Centre for Global Development.

Results for Development Institute (2013). *Assessing Learning, Achieving Impact.* Washington D.C.: Results for Development Institute.

Uganda Village Project (2009). Village Health Teams. http://www.ugandavillageproject.org/what-we-do/healthy-villages/village-health-teams/ (Accessed on 6 July 2015).

UNESCO Institute for Statistics Database of Learning Assessments. http://www.uis.unesco.org/Education/Pages/learning-assessments-database.aspx

United Nations Sustainable Development (UNSD). Knowledge Platform. https://sustainabledevelopment.un.org/sdgsproposal (Accessed 16 September 2015).

Uwezo (2014) *Are our Children learning? Literacy and Numeracy in East Africa*. Nairobi: Uwezo East Africa.

Uwezo (2013). *Uwezo Uganda 2013 Volunteer Workbook*. Kampala: Uwezo Uganda.

Uwezo Kenya (2013). *Are our children learning? Annual Learning Assessment Report*. Nairobi: Twaweza East Africa.

Evaluating Early Learning from Age 3 Years to Grade 3

AMY JO DOWD, LAUREN PISANI AND IVELINA BORISOVA
Save the Children

ABBREVIATIONS

ECCD	Early childhood care and development
EDI	Early Development Instrument
ELM	Emergent Literary and Math
HLE	Home learning environment
IDELA	International Development and Early Learning Assessment
ICC	Intraclass correlation coefficient
MICS	Multiple Indicator Cluster Surveys
USAID	United States Agency for International Development
WCPM	Words correct per minute

1. INTRODUCTION

In 2007, Save the Children began building a system and tools for measuring learning outcomes. Pursuing an agenda of evidence-based programming for children age 3 years to Grade 3, we have encountered many challenges. However, over time and with the support of partners all over the world, we have developed learning measurement resources that cover the continuum of child development from age 3 years to Grade 3. In this article, we aim to share some of the main challenges faced in measuring learning outcomes for evidence-based programming as well as their resolution in hopes that it helps others working in this complex arena.

From the outset, our most foundational question has been how to most effectively measure learning for ALL the children we support through our community-based programmes. The answers we have found have driven the development of Save the Children's learning measurement approach. Although primary school oral assessments have been the main focus globally, it became clear early on that a more holistic perspective using an assessment that captures pre-primary learning as well would better and more sustainably serve the measurement of learning improvements. This is because a child's readiness to learn (or school readiness) as they enter formal education systems enables their success in learning to read and achieving academic success later on.

The phrase "readiness to learn" in itself is deceiving. It implies that certain things need to happen before a child is learning. However, from the time children are born, they are constantly learning. In fact, children's learning trajectories are steepest at earlier ages and skills developed at that time literally pave the way to later school achievement (Thompson and Nelson, 2001). Research shows the significance of emergent literacy skills developed during the preschool period for reading achievement in the primary grades (Scarborough, 1998; Lonigan et al., 2008). Oral language, phonological awareness, alphabet knowledge and print awareness are strong and independent predictors of how quickly and how well children will read once they are exposed to formal reading instruction in Grades 1, 2 or 3 (Lonigan et al., 2000; Lonigan et al., 2008; Wagner et al., 1997).

Measuring both school readiness skills and reading can help us strengthen interventions and system efficiency in supporting learning. The artificial silos

that put children into 'preschool' and 'primary school' obscure the fact that children are always learning. Early experiences build the foundation for later learning but the path is not linear and it is multifaceted (Learning Metrics Task Force, 2013). Thus, Save the Children decided that in order to address learning in the developing world and in the lives of young children, we needed rigorous measurement of learning earlier in the age spectrum than the middle of primary school.

It is vital to collect data to better understand children's development during these formative, foundational years and into the first years of schooling. Skills developed early both overlap with and are important precursor skills to later reading achievement. Improving our understanding of children's readiness for school will help us contextualise the learning crisis by enhancing our view of where children begin developing key academic and life skills. An earlier assessment allows for a more concrete identification of the skills on which interventions or systemic change can build and estimate the impact of improvements. Further, being able to identify characteristics of groups of children whose learning is not progressing on par with their peers during the early childhood period can provide information to empower parents, communities, schools and ministry officials working to support the development of all children.

Large numbers of assessments have been performed across the world in the past decade. The international community has a large population of capable assessors and assessment-savvy educational professionals. We now need to turn their attention to adding information from children of younger ages into the dialogue on the learning crisis and including early childhood interventions in the debate on viable solutions. Doing so is the most cost effective investment in education and creating a world of readers and learners (Gertler et al., 2014; Heckman et al., 2010). Measuring and addressing learning from age 3 years to Grade 3 will help ensure that all children who complete primary schooling will be able to make meaning of text.

This article addresses the question of how to quickly, feasibly and reliably gather information on all children's early learning status and their progress toward making meaning of text. These are not individual diagnostic tools but simple yet rigorous methods for estimating the skill levels of children in groups (schools, centres, districts, provinces or nations) to learn how to shape innovations that build upon existing skills to maximise children's potential.

2. ASSESSING A CONTINUUM OF LEARNING

Save the Children uses the Literacy Boost oral reading assessment among children in Grades 1 to 3 as well as a holistic oral assessment of school readiness called the International Development and Early Learning Assessment (IDELA) with children aged 3 to 6 years old before they enter primary school or just as they enter Grade 1. The Literacy Boost assessment includes reading mechanics such as letter knowledge, decoding, fluency and accuracy as well as comprehension. IDELA incorporates measures of emergent language and literacy; early numeracy and problem solving; motor; and social-emotional skills, as well as approaches to learning and executive function (short-term memory and inhibitory control). Together these two assessments capture a fluid continuum of language and literacy skill development—starting with foundational and moving to more advanced skills.

Over eight years, Save the Children has conducted oral reading assessments with thousands of children as part of its Literacy Boost program in 24 countries and in more than 35 languages. Across the dozens of learning assessments of children in Grades 1 through 4, there is at least one consistent finding: a diverse range of skills and capabilities exist in any age group. Save the Children's oral assessments (IDELA and Literacy Boost) aim to detect skill variations to allow us to better understand where children are prior to learning interventions, and how their skills grow and develop over the course of a programme. Since assessments are used to shape and evaluate programmatic action, it is

critical that they provide opportunities for children to demonstrate which skills they have mastered. The assessments therefore include a range of foundational and higher order skills. **Table 1** lists the skills measured in the IDELA language and literacy domain alongside those measured in the Literacy Boost assessment.

As local context is critical, country teams may add additional lower or higher order subtests or administer test component(s) in multiple languages. The range of assessed skills allows intervention teams to learn about strengths and gaps in learning as well as facilitates measuring the progress of both children with limited skills as well as those with more advanced skills. Letter knowledge skill represents overlap between IDELA and Literacy Boost (see Table 1). In IDELA, it is among the highest order skills assessed while in the Literacy Boost assessment when it is used in primary schools, it is considered 'foundational'. IDELA also offers linkages to precursor skills, such as expressive vocabulary, emergent writing skills and exposure to print. These skills underlie and support the acquisition of more advanced reading and writing skills as a child learns to move from oral to written language. Further, IDELA measures a child's ability to stay on task and to persist when faced with a challenging activity, which also relates to how that child is likely to respond to a difficult reading passage she/he encounters later on.

2.1 Bridging silos of education to investigate early learning

Save the Children often segments work into silos or sectors of pre-primary, primary and secondary education as do many organizations, ministries and institutions involved in children's education. Unfortunately, measurements and assessments are also similarly structured, creating the notion that certain skills are the end result of a stage. However, as **Table 2** shows, using the example of writing skills, children's learning is much more detailed and fluid within and between stages.

Oral reading assessments can help mitigate this issue as they are administered in the middle of primary schooling instead of at the end, which allows for course correction before too much time has gone by. Because children don't learn in silos, the oral assessments developed by Save the Children can help bridge the gap between pre-primary and the early grades by providing a rigorous tool to measure the skills underlying oral reading success. These oral assessments enable programme teams to consider what skills and abilities children take into primary school.

TABLE **1**

Emergent language and literacy skills by instrument

Skills	IDELA: emergent language and literacy domain	Literacy Boost assessment
Expressive vocabulary	✓	
Phonological awareness	✓	
Emergent writing	✓	
Concepts about print	✓	
Listening comprehension	✓	
Letter knowledge	✓	✓
Single word reading		✓
Decoding		✓
Fluency		✓
Accuracy		✓
Reading comprehension		✓

TABLE **2**

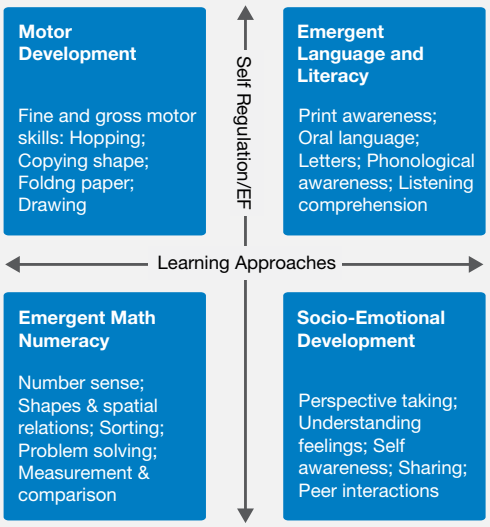
Assessment silos and children's learning by level of education

Level of education	Pre-primary	Primary	Secondary
Assessment silo	School readiness	End of primary exam	End of secondary exam
Children's learning—writing as an example	Scribble Write letters Write words	Write words Write sentences Write paragraphs	Write paragraphs Write essays Write reports

IDELA development began in 2011 with the identification of prioritised constructs to measure learning and a review of existing tools. Importantly and appropriately, this early childhood assessment tool is more holistic than a primary grade reading assessment composed of language, math, socio-emotional and motor skills. After a two-country pilot that year, Save the Children invested three more years in refining the adaptation and administration of its items and retested them in 12 more countries. A study of 11 field trials with over 5,000 children found that the 22-item core tool has an internal consistency of .90 out of 1 (Pisani et al., 2015).

IDELA measures four domains alongside approaches to learning and executive function (*refer to* ***Figure 1***). The domain items in Figure 1 are reliable and emphasise 'teachable' and 'actionable' items so that managers and ministries can use ongoing data to reflect, analyse and improve investments, policy and practice. This echoes the aims of many oral reading assessments, such as the Annual Status of Education Report (ASER), Early Grade Reading Assessment (EGRA) and Uwezo.

Figure 1. IDELA domains and items



Source: adapted from Pisani et al, 2015

3. CHALLENGES FACED AND SOLUTIONS OFFERED

The IDELA and the Literacy Boost assessments are not the only measurement tools available and used to assess learning outcomes in an international context. After years of development and practice, however, we do consider these tools to be the most feasibly utilised and the best suited to Save the Children's purposes of intervening to improve systems that support children's development and learning, especially the most deprived. The following sections detail the challenges faced and solutions devised to meet them. Where relevant, we make comparisons to alternative instruments and offer examples of why the IDELA and the Literacy Boost are optimal assessment tools for our goals.

3.1 Ensuring feasibility

Unlike many costly measurement tools on the market, both the IDELA and the Literacy Boost assessment can be administered by non-professionals who train and practice administering the instrument over a seven- to nine-day period during which time they learn to establish a rapport with the children and administer the items in a child-friendly way. Qualifications commonly considered are a background in working with children, patience and clear communication skills. No prior educational work experience is necessary although many country teams call upon retired teachers or education graduate students to fill the ranks in

assessor teams. **Table 3** details the time required to adapt and pilot the instruments, train assessors and assess a child for both the IDELA and the Literacy Boost assessment.

The items required to administer the assessments are equally simple and readily available. The IDELA requires a set of local materials such as beads, beans or bottle caps, a local children's book, a few picture sheets/cards, a laminated puzzle, paper and pencils. The Literacy Boost assessment requires a letter sheet, two sheets of words and a reading passage. Whether the data is collected electronically or on paper, children are familiar with these simple materials and IDELA does not introduce foreign stimuli into the assessment situation.

Findings feed back to the programme team after an average of two months to help with programme planning or improvement. This process is faster if the data is collected electronically, making data entry and cleaning substantially shorter. Save the

TABLE **3**

Feasibility and essential investments of time by assessment

	IDELA	**Literacy Boost**
Instrument adaptation and piloting	4 days	4 days
Training assessors	3-5 days	5 days
Assessment time per child	35 minutes per child/1 hour including caregiver questionnaire	30 minutes (including background interview)

Children's best practice calls for the collection of inter-rater reliability data throughout data collection by having a pair of assessors jointly score two children in the sample from each programme site. We then rate the consistency of their scoring using intraclass correlation coefficients (ICCs) and these average .95 for Literacy Boost across 20 sites and .90 for IDELA across 4 sites.

3.2 Direct child observation

IDELA is a direct child observation tool, not a parent or teacher report. Data from parents and teachers can be an over- or under-estimation of the child's abilities depending on whether the adult has ever asked or noticed the child performing a specific task and the adult's perceptions of the child's performance. The IDELA asks the child to perform a task and records his/her response. The results can certainly be affected by children's willingness to participate but it improves upon the third party reporting used by other assessments, including the Offord Centre for Child Studies' Early Development Instrument (EDI) and UNICEF's Multiple Indicator Cluster Surveys (MICS). All one-on-one oral reading assessments entail direct child observation. This is one of IDELA's strengths in comparison with other school readiness assessments.

In most early grade oral reading assessments, children in Grades 1 to 3 are also asked a series of background questions that supply indicators of socio-economic status, home learning environment, repetition, early childhood care and development (ECCD) participation and chore workload. The Literacy Boost assessment provides more information about the learning environment children have to support them at home than do other assessments (see article by Dowd and Friedlander). This can help shape intervention activities that aim to strengthen the presence, variety and use of books in children's daily lives outside of school. Although these assessments may capture measurement error associated with self-reporting by children, it is cost-effective since it is obtained during a school-based assessment rather than a more expensive household survey to obtain guardian responses. IDELA is also accompanied by a caregiver questionnaire to supply the same information asked of the children in addition to parenting habits. This strategy to capture caregiver data, however, is more likely to succeed because parents accompany their young children to the assessment site or centre to consent to participation and engage directly with assessors themselves.

3.3 Continuous variables to capture continuous skill development

One of the major strengths of IDELA is the continuous scoring system that allows for a more nuanced perspective on learning and development than is possible if items are simply scored as correct or incorrect. For example, a feasible, quality IDELA measure of expressive vocabulary—a precursor skill to vocabulary measures of oral reading assessments—can be taken by asking a child aged 3 to 6 years to name things to eat that can be bought in the market. The number of different items the child names is counted to offer a continuous score across the age range of 3 to 6 years. This sample item, coupled with other language and literacy items in Figure 1 inform our understanding of children's emergent language and literacy skills. The inter-item correlation of the multiple items in the language and literacy domain across 11 countries is .77. The continuous score approach underlies most of the items in IDELA and allows the administration of the tool with a wider age range of children, which helps document where on a specific skill continuum children land. For example, fine motor skills are assessed not by whether or not a child can draw a person when asked but instead by points awarded for each detail drawn (based on complexity)—head, body, arms, legs, hands, feet and facial features.

The Literacy Boost assessment also collects continuous indicators of reading skills, representing a greater depth of information than oral reading assessments that classify the skill level of a child as: knowing letters, reading words or reading passages. For the same investment in time and resources to get to and interact with children, the Literacy Boost assessment uses similar materials—letters, words and passages to read—to collect several continuous indicators of both reading mechanics and comprehension during the interview. This supplies broader variation for equity analyses and offers more flexibility for interpretation and reporting. Indeed, the data from a Literacy Boost assessment can be analysed to provide the same levels of categorisation—reader of letters or words or passages—as other oral reading assessments.

However, the reverse may not necessarily be true. Not all oral reading assessments aim to effectively provide exhaustive details of what children do know to inform interventions and improvements. This relates to another challenge—fitting the assessment to the population and purpose.

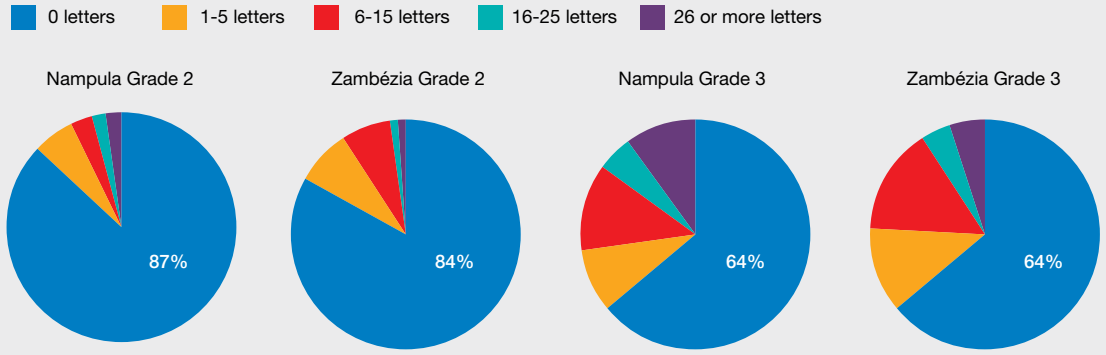
3.4 Fit population and purpose

The Literacy Boost assessment measures key mechanics of reading like the EGRA used by the United States Agency for International Development (USAID) and the World Bank, but with less of a focus on speed and more attention to what is happening at the lower end of the skill distribution. The skills assessed are similar but the tasks around them are shifted in an effort to avoid floor effects. This translates to testing a more basic version of the same skills, which is crucial because Save the Children and our partners often work in marginalised areas where children's skills are at the low end of a national distribution. This practice highlights the importance of selecting the appropriate oral assessment for the purpose and target population of the study (see Recommendation 1 in Chapter 5). The EGRA, often used in national samples, doesn't always capture an intervention sample well and so can be less useful for a programming team.

Consider **Figure 2** from the June 2013 *Aprender a Ler* Mozambique Baseline Report (Raupp et al., 2013) compared alongside the Literacy Boost assessment data collected by Save the Children in Nacala, Mozambique in early 2014 (*see* ***Figure 3***). The EGRA implemented in Nampula and Zambezia for Grades 2 and 3 has an indicator of letter knowledge that asks a child look at a 10 x 10 matrix of letters (letters repeated based on the frequency of appearance in common texts to arrive at 100) and name as many letters as she/he can in a minute. A fairly common assessment for young children in developed countries (Good and Kaminski, 2002), it assumes that children know the alphabet and assesses how fast they can name the letters. The 2013 Baseline Report uses this approach to present the following picture of letter knowledge in these provinces and grades.

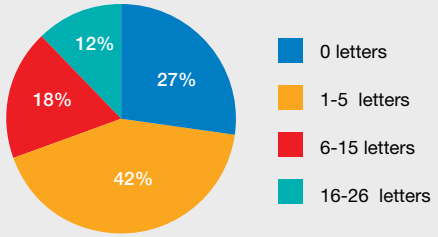
Notice that these are provincial level samples. The very large percentage of students with zero scores suggests that the assessment is too difficult for the students (see Figure 2). As expected, among children in both provinces who are still in Grade 3, they perform better on this task than children in Grade 2. Consider that any value under the total number of letters in the alphabet (26 in English, for example) cannot be interpreted very usefully because the child could know a few common letters, such as A, N, E and D, and find only these repeatedly in the 10 X 10 matrix for a score of 10 or 12 since letters appear more than once. Naming a dozen or fewer letters is unfortunately what almost

Figure 2. Total letters named correctly in one minute by grade and province in Nampula and Zambézia, Mozambique 2013 (n=3,589)



Source: Raupp et al., 2013

Figure 3. Total letters known out of 26 by Grade 3 students in Nacala, Mozambique 2014 (n=702)



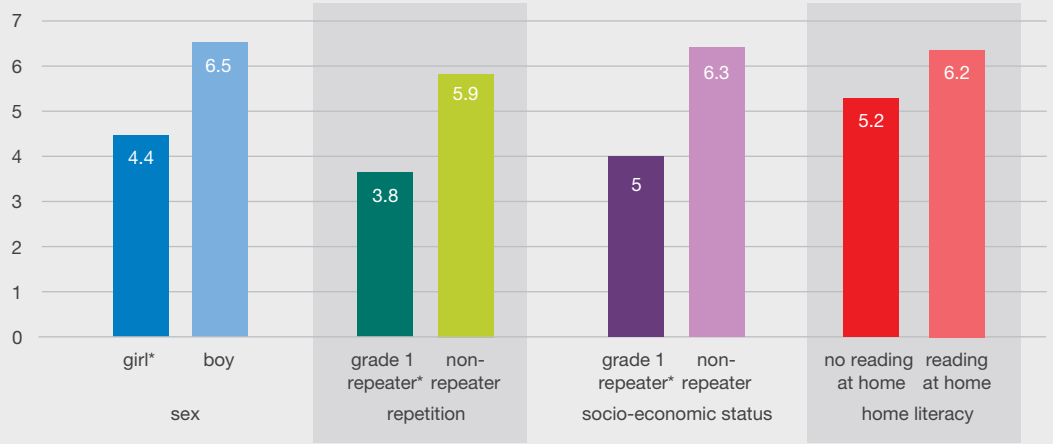
Source: Save the Children, 2014

all children in this sample are capable of doing (the values in the purple portion of the pie chart in Figure 2 may be above 26 but are not labeled). So, this indicator allows us to see a deficit but does not clearly capture what skills there are in this population on which to build. Further, the indicator has little or no variation to help pinpoint inequities in the system, making it difficult to tell whether boys or girls, those in greatest poverty, those in illiterate households, repeaters, or older children face different or greater challenges in achieving basic skills. The question the EGRA letter subtest answers (how many letters a child can name in a minute) is too advanced and while the finding of overall low letter knowledge might be useful for advocacy, it is not useful for intervention.

Because Save the Children uses assessments in these settings to shape improvements for children, we are not as interested in how many of the 100 letters a child can name in a minute as we are in learning about whether or not children know all of the letters, whether there are patterns in which letters pose challenges and/or whether there are patterns in who is struggling to master this foundational skill. The Literacy Boost assessment therefore asks a slightly different question: how many letters of the alphabet does a child know? It then uses a slightly different assessment to capture these data, resulting in a pie chart, such as the one from Mozambique (see Figure 3), that shows variation across the same groups as those established in Figure 2 from the USAID report.

Using this approach, we can also determine which letters are best known (A, I and O) and least known (Y, W and Q). Finally, because this approach offers variation with which to work, we can consider inequities that exist as we begin the intervention. For example, we can see in **Figure 4** that girls and Grade 1 repeaters are central targets for this work

Figure 4. Average number of letters known by group in Nacala, Mozambique 2014 (n=702)



Note: *denotes significantly lower at p<0.001.
Source: adapted from Nivaculo et al., 2014

because their scores at baseline are significantly lower than boys and non-repeaters, respectively.

A non-significant difference exists between those falling in the highest and lowest socio-economic status and those falling in the highest and lowest home literacy environment. These groups will be reassessed at midline to ensure that this gap does not widen as the intervention moves forward.

The Literacy Boost assessment targets existing skills and monitors the progress of an intervention overall and by groups. This model of data capture is more helpful for planning interventions than simply measuring a deficit because it focuses on assets to provide a better idea of the overall scope of a challenge.

3.5 Measuring a range of skills

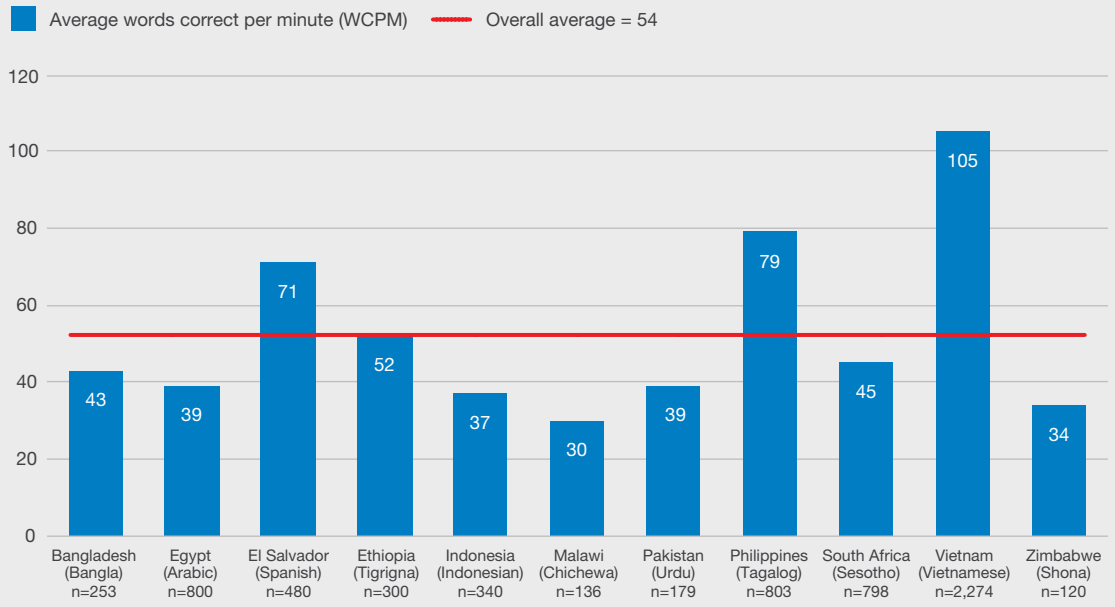
Whether using IDELA or the Literacy Boost assessment, Save the Children teams always collect data on a range of skills. In oral reading, this ranges from the foundational to higher order skills as shown in Table 1. This methodology ensures that the assessment captures skill variation for *all* children. In low-skill contexts, the data show where children fall in terms of more basic reading skills and not just where they are falling short. This was the case at a programme site in Zambia where the baseline sample of 384 children showed only five children could read a grade-level passage. In this setting, variation in letter recognition and reading single words provided critical information for shaping both intervention and advocacy. In contexts of higher overall skills, such as El Salvador, where children were able to correctly identify 78% of letters and 70% of children were able to read a grade-level story independently (reading 35 correct words per minute on average with 64% accuracy and 68% comprehension). Including both foundational and higher order skills in this assessment offered evidence of mastery and indicated a programmatic focus on higher order skills. It also offered variation with which to consider equity and key target groups. For example, it allowed investigation into whether the 30% of children not yet reading independently were struggling with specific skills or were predominantly from a specific group, such as girls, those with lower socio-economic status, repeaters or those with the fewest learning resources at home. These additional data help support programme and advocacy teams when formulating appropriate interventions.

In addition to lessons from individual field sites, data analyses across countries support the practice of assessing a range of reading skills as foundational literacy skills correlate with the higher order skill of reading comprehension. For instance, across 64 datasets we found that although letter knowledge has the weakest correlation with reading comprehension at $r = .31$, the strongest correlation is only $r = .42$ for both reading accuracy and fluency. Thus, if only letter knowledge is obtainable, data indicate that this is still relevant information on children's learning trajectories and it is positively related to comprehension. Further, fluency, accuracy and word reading are highly inter-correlated ($r = .67$), stemming from their common root in children's letter knowledge and decoding ability. Reading comprehension represents the use of these mechanics alongside broader knowledge and vocabulary to make meaning of text and has a relatively lower correlation with them as noted above. Measuring comprehension is therefore essential because it taps into an area that the other measures that focus on the mechanics of reading do not. Simply knowing that kids are fluent or read accurately does not imply they necessarily read with comprehension. For these reasons, Save the Children always recommends that oral reading assessments include a direct measure of reading comprehension. In fact, Save the Children's current best practice asks ten comprehension questions—one summary, six literal, two inferential and one evaluative, in that order.

3.6 Understanding fluency and comprehension

Sometimes, how you measure specific skills rather than which skills are measured can shift our understanding of how skills interrelate. Save the Children administers a reading passage with

Figure 5. Average fluency (WCPM) among readers with comprehension by programme site



Source: Dowd, 2015

comprehension questions but assessors do not stop the child reading when one minute has passed. We do not have any basis for knowing how fast a child 'should' read in many of the languages and cultures in which we assess reading. Unlike English or some European languages, there is not a rich academic literature from which to draw the conclusion that one minute is enough time to complete a text and demonstrate understanding. Further, many contexts feature children learning to read in a second language, complicating the relationship between fluency and comprehension by featuring with novel letters, sounds, vocabulary, grammar, etc. Indeed, **Figure 5** shows that the achievement of reading with comprehension (80% or more questions answered correctly) in programme sites in marginalised communities can correlate to fluency rates both well above and well below the cross-country overall average of 54 words correct per minute (WCPM) (Dowd, 2015).

While it is important to consider automaticity (WCPM) with which a child is reading, many children may read at what some experts tell us are 'slow' rates but with comprehension.

In the Literacy Boost assessment, children are allowed to read past one minute with stop rules so that it remains efficient and not frustrating or tiresome for the child if they are not capable of reading independently. Save the Children also allows the child to look back at the text when they are being asked questions. Therefore, our comprehension subtest is not a test of memory but a test of whether children can make meaning from text. This approach facilitates the collection of a second important indicator—accuracy. This measures the percent of the text that the child read correctly in total, including within and beyond the timed minute. Interestingly, this second untimed measure of accuracy offers a more stable metric. Across the sites in Figure 5, the average accuracy among children who can read their grade-level text with comprehension ranges from 94-99% with an overall average of 96%.

Save the Children contends with the tendency to focus on automaticity measured in WCPM, which is often used to denote 'fluency'. We feel that this focus on speed obscures the goal of helping children read with comprehension. While measuring

comprehension can be difficult and different levels of comprehension exist, Save the Children's interventions are intentionally aimed at enabling children to achieve reading with comprehension and thus so do our measures.

3.7 Data utility

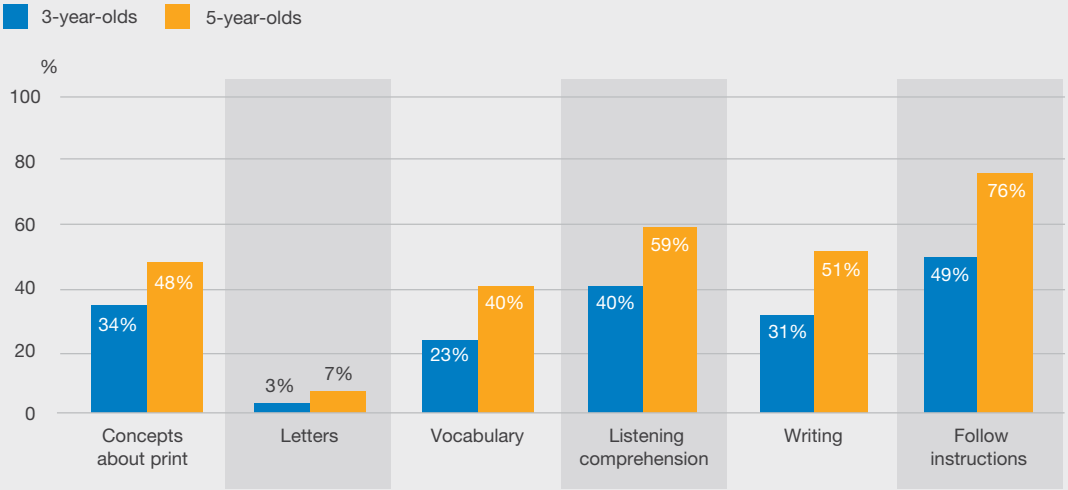
Save the Children uses both oral reading assessments and IDELA to understand differences in developmental levels by age/grade, detect programmatic impact and contribute to global comparisons and dialogue. For example, **Figure 6** shows how IDELA detects variation in emergent language and literacy skills by age in Zambia where children aged five score significantly higher on all subtests than their three-year-old peers on all subtests.

Figure 7 shows that both the IDELA for children ages 3-6 and the Literacy Boost assessment for reading in the early grades of primary school detect the impact of different interventions on children's learning. The IDELA portion of the figure displays the learning gains of five-year-old children with varied levels of early learning support: No Early Childhood Care and Development (ECCD), ECCD and ECCD + Emergent Literary and Math (ELM) intervention. The ELM Intervention is an ECCD package with a specific focus on the skills underlying primary grade reading and math. The difference between IDELA scores at baseline in the No ECCD and both ECCD sites may be due to unobserved differences in the communities that have such centres and/or in the children who are sent to them once they exist. The baselines taken using the Literacy Boost tool on the right portion of the figure are not comparable to the IDELA data among younger children and are more similar to each other as they are from children in schools with and without the intervention. These data display the accuracy gains for children in Grade 3 over the course of a school year with and without the intervention.

The proliferation of primary school reading assessments has raised the issue of basic skills attainment worldwide. Save the Children uses data like those presented in **Figure 8** to support children's learning and contribute to the global dialogue on problems and solutions.

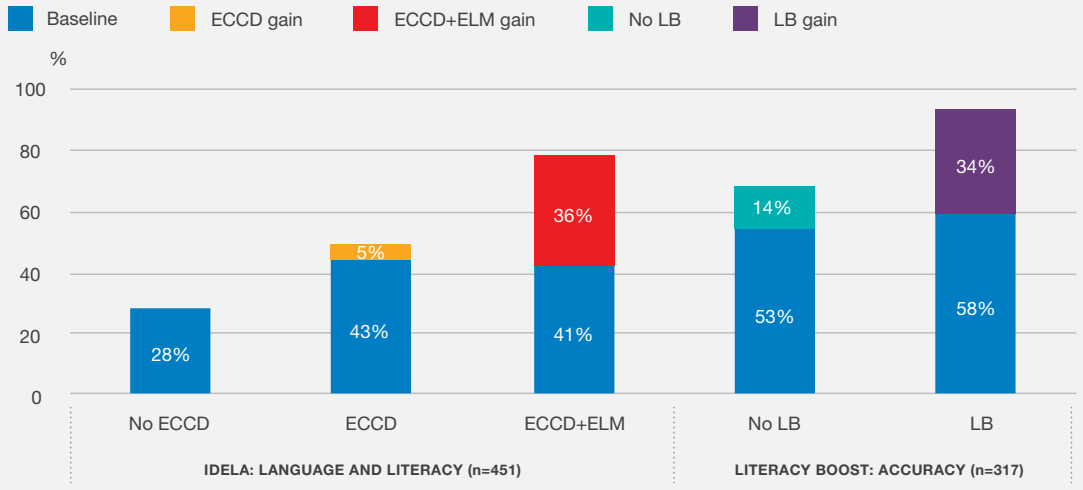
Assessments of reading in early grades have revealed the crisis in education but ECCD assessments highlight the need to support children's learning holistically and at earlier ages. In Zambia, for example, the National EGRA results for Grade

Figure 6. IDELA detects variation by age in Zambia, 2013 (n=273)



Source: Dowd, 2013

Figure 7. IDELA language and literacy and Literacy Boost accuracy average baseline and gains by group in Ethiopia, 2013



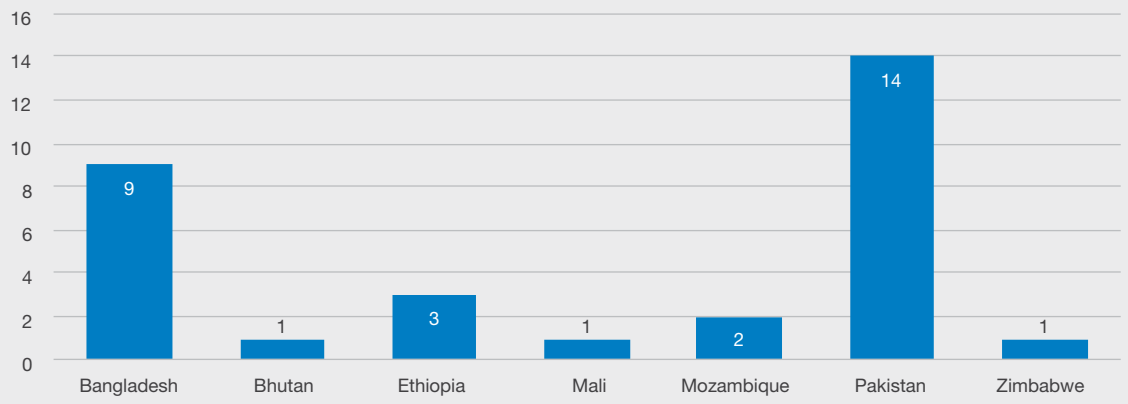
Source: adapted from Amente et al., 2013 and Friedlander et al, 2013

2 and 3 found that over half of pupils tested were below proficiency in oral reading and English literacy. For the oral reading fluency task, pupils were asked to read a short narrative story for one minute. A majority (91%) of Grade 2 pupils and 78% of Grade 3 pupils were unable to read a single word in this passage. This performance is not surprising given what we see in the Figure 7 for school readiness skills of five-year-olds about to enter Grade 1. In Zambia, as in many other countries, the gaps in school readiness are big and it is too late to intervene if assessments are only administered in Grade 2 and 3 to detect these gaps. An instrument like IDELA can shed light on these issues much earlier.

Using IDELA, Save the Children has learned about the foundational skills that children in and out of ECCD programmes have developed prior to primary school. For example, on average, across communities in eight countries, children entering primary school only know five letters of the alphabet, two-thirds of children are unable to complete a simple pattern and about half cannot articulate something that makes them feel happy or sad. These data lend nuance to our understanding of the learning crisis. Specifically, they offer a clearer view of what skills children bring to school to help them learn to read. For instance, Figure 8 depicts basic skills like average letter knowledge across programme sites so that we can weigh children's chances for success in reading a Grade 1 textbook more realistically.

When we look at the data, the learning crisis, becomes more complex than just gaps in language, literacy and reading. Indeed, because IDELA assessments are based on a central framework that is adapted to local language and culture, it shows us that children in each country have different profiles of similar, though not directly comparable skills. For example, five-year-old children in a community in Mali enter primary school knowing one letter of the alphabet on average and possess very limited if any writing skills. In contrast, their peers in a in a programme in a Pakistani community know 14 letters and are more likely to be scribbling and forming early letters. Similarly, on average 45% of children assessed in Ethiopia were able to complete a simple puzzle and provide at least one appropriate response to a conflict with a friend. However, only 2% of children in a community in Bangladesh could correctly complete a puzzle but slightly more than half could provide an appropriate response for

Figure 8. Average number of letters out of 20 identified by five-year-olds by programme site



Source: Save the Children

solving a conflict. Being able to show primary school educators the skills that their new students already possess as well as the skills they need to focus on building in the early grades can better prepare schools, communities and parents to support their children's continued learning most effectively.

3.8 Exploring equity

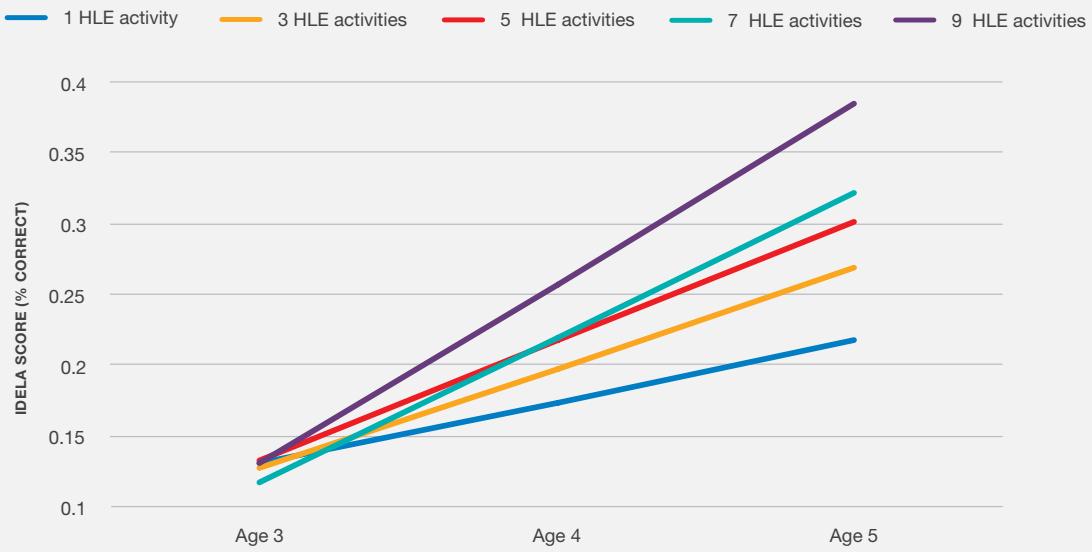
The background data collected from caregivers in IDELA and from students during a Literacy Boost assessment enables equity analyses such as those in Figure 4 above. Taking into consideration the influence factors, such as socio-economic status, gender, repetition, chore workload or home learning environment (HLE) have on early learning enables us to effectively target services to fill specific needs. Knowledge of these factors and their impact on learning also helps fuel advocacy to ensure that all children are learning. For example, multivariate regression analysis results from a 2015 national ECCD study in Bhutan using IDELA and targeting children ages 3 to 5 years show the significantly wider gap between older children who have strong HLEs and those who do not as compared to younger children in similarly diverse HLEs (*see* ***Figure 9***). The differences in IDELA scores (depicted as percent of items answered correctly) by HLE are more extreme in the older the group of children (see Figure 9).

This type of equity analysis at baseline whether using the IDELA or the Literacy Boost assessment informs programme interventions and can ultimately inform both continuous improvement as well as advocacy for effective future investments.

4. CONCLUSION

Save the Children's IDELA and Literacy Boost assessment measure children's early learning and development of reading skills from age 3 years to Grade 3 with feasibility and rigor. These assessments bridge the silos of preprimary and primary classrooms, and present a continuous assessment framework that helps to situate children's learning and development in an asset-based model. The silos around preprimary and primary measurement is somewhat arbitrary. In many parts of the world where Save the Children works, an assessment such as IDELA represents a much more appropriate assessment for children in Grade 1 compared to the more traditional literacy and numeracy primary school assessments. The IDELA and the Literacy Boost instruments represent a fluid continuum of skills. Depending on the context and background of students, a combination of the IDELA and the Literacy Boost assessment may be the most appropriate way to capture a variation of skills and inform programmes to better support these children. It is important to keep in mind that

Figure 9. Relationship of age and HLE with child development in Bhutan controlling for socio-economic status, sex, paternal education and home language, 2015 (n=1,377)



Source: adapted from Save the Children (2015)

children don't necessarily transition from preprimary to primary classrooms. Instead, many enter primary schools without any prior preschool experience. Understanding what children bring to the table at Grade 1 entry can be much more effective in helping to build interventions to support them early on rather than waiting to document concerning trends in Grade 3 and up.

Measuring a range of skills helps ensure that data on learning as well as equity is utilized to strengthen programmes and advocacy. Many other organizations and academic partners work with Save the Children in this effort and we hope to continue to participate in and build this network over time.

Children's skills can be quickly and reliably assessed using the IDELA and the Literacy Boost assessment to inform programmes and policies that promote greater learning and equity. Based on our experience, we recommend the use of these tools to avoid floor effects. This advantage makes them particularly well suited to marginalised populations. Further, we promote the inclusion of a range of continuous indicators spanning from foundational to higher order skills. In early childhood assessments, direct observation is essential while in early grades, indicators of both timed and untimed connected text reading best informs interventions and the evidence base. Finally, measures of comprehension—the goal of learning to read—are non-negotiable.

If you are considering using an assessment to intervene to improve systems that support children's development and learning—especially if you are targeting those struggling the most—then consider the IDELA and the Literacy Boost assessment (access these at www.savethechildren.org). They may well be the early learning and reading assessment instruments that best suit your needs.

REFERENCES

Ali Amente, A., Yenew, A., Borisova, I. and Dowd, A.J. (2013) *Ethiopia Sponsorship-funded Early Childhood Care and Development (ECCD) Program. Emergent Literacy and Math (ELM) Intervention Endline Report*. Washington, DC: Save the Children.

Dowd, A.J. (2015). "Fluency and comprehension: How fast is fast enough?" Paper presented at the CIES Annual Conference, Washington, DC.

Friedlander, E., Hordofa, T, Diyana, F, Hassen, S, Mohammed, O. and Dowd, A.J. (2013). Literacy Boost Dendi, Ethiopia Endline II. Washington, DC: Save the Children.

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S. Chang, S.M. and Grantham-McGregor, S. (2014). "Labor market returns to an early childhood stimulation intervention in Jamaica". *Science*, Vol. 344, No. 6187, pp. 998-1001. (Accessed June 5, 2015).

Good, R. H. and Kaminski, R.A. (eds.) (2002). *Dynamic Indicators of Basic Early Literacy Skills*. 6th edn. Eugene, OR: Institute for the Development of Educational Achievement. http://dibels.uoregon.edu/.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A. and Yavitz, A. (2010). "The rate of return to the HighScope Perry Preschool Program". *Journal of Public Economics*, Vol. 94, No. 1-2, pp 114-128. http://doi.org/10.1016/j.jpubeco.2009.11.001

Lonigan, C. J., Schatschneider, C. and Westberg, L. (2008). "Identification of children's skills and abilities linked to later outcomes in reading, writing, and spelling". *Developing early literacy: Report of the national early literacy panel,* pp. 55-106.

Lonigan, C. J., Burgess, S. R. and Anthony, J. L. (2000). "Development of emergent literacy and early reading skills in preschool children: evidence from a latent-variable longitudinal study". *Developmental psychology*, Vol. 36, No. 5, p. 596.

Learning Metrics Task Force (2013). *Toward Universal Learning: What every child should learn*. Learning Metrics Task Force Report No. 1 of 3. Montreal and Washington D.C.: UNESCO Institute for Statistics and Centre for Universal Education at Brookings. http://www.uis.unesco.org/Education/Documents/lmtf-rpt1-toward-universal-learning-execsum.pdf

Nivaculo,C.C, Sebastiao, M.N, & Pisani, L. (2014) *Literacy Boost Nacala, Mozambique Baseline Report*. Washington, DC: Save the Children.

Pisani, L., Borisova, I. and Dowd, A., (2015). *International Development and Early Learning Assessment Technical Working Paper.* Save the Children. http://resourcecentre.savethechildren.se/library/international-development-and-early-learning-assessment-technical-paper.

Raupp, M., Newman, B. and Revés, L. (2013). *Impact Evaluation for the USAID/Aprender a Ler Project in Mozambique: Baseline Report*. Washington D.C.: United States Agency for International Development. https://www.eddataglobal.org/documents/index.cfm/Final USAID Aprender a Ler Baseline Report 15 june 2013.pdf?fuseaction=throwpub&ID=483

Save the Children (2015). *National ECCD Impact Evaluation Study 2015*. Thimphu, Bhutan: Save the Children.

Scarborough, H. S. (1998). "Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors". Shapiro BK, Accardo PJ, Capute AJ (eds). *Specific reading disability: A view of the spectrum*. York Press: Timonium, MD, pp. 75-119.

Thompson, R. and Nelson, C. (2001). "Development science and the media: Early brain development". *American Psychologist,* Vol. *56*, No. 1, pp. 5-15.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A. and Hecht, S. A. (1997). "Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: a 5-year longitudinal study". *Developmental Psychology*, Vol. *33*, No. 3, p. 468.

Utility of the Early Grade Reading Assessment in Maa to Monitor Basic Reading Skills: A Case Study of Opportunity Schools in Kenya

JOYCE KINYANJUI
Women Educational Researchers of Kenya

ABBREVIATIONS

AET	Africa Education Trust
ECDE	Early Childhood Development and Education
EGRA	Early Grade Reading Assessment
KCPE	Kenya Certificate of Primary Education
L1	First language or native speakers
MOEST	Ministry of Education, Science and Technology
OSP	Opportunity Schools Programme
RTI	Research Triangle Institute
SIL	Summer Institute of Linguistics
SPSS	Statistical Package for the Social Sciences
WERK	Women Educational Researchers of Kenya
WCPM	Words correct per minute

1. INTRODUCTION

The Opportunity Schools Program (OSP) was initiated in Kajiado County Kenya in 2012 with the goal to achieve "Improved learning outcomes for increased enrolment, attendance, progression and completion in Early Childhood Development and Education (ECDE) through Grade 8 by 2016".

The programme seeks to achieve five specific outcomes by 2016:

1. Increase enrolment, attendance and completion by 5%, 60% and 30% respectively by 2016
2. Ensure that 50% of pupils perform literacy and numeracy functions at their required grade level
3. Develop a large-scale, Manyatta based Early Childhood Development and Education (ECDE) model
4. Harness citizen participation to support schooling and learning
5. Institute strong infrastructure to support schooling and learning.

The OSP is implemented by the Women Educational Researchers of Kenya (WERK), a professional association of education researchers whose mission is to contribute to knowledge generation and utilization by linking research to policy and action. To improve literacy, the programme model involves teaching literacy using the mother tongue or native language, which refers to the first language learned by the pupils. Pupils then transfer the literacy skills learned to reading in English and Kiswahili. As the OSP is a dual-language programme, the first language and culture are not replaced by the addition of a second language and culture—this is referred to as additive bilingualism. Due to the rollout of a national literacy strategy *Tusome* ('Let us all read') in 2015 to improve literacy competencies in English and Kiswahili among Grade 1 and 2 pupils, WERK is focused instead on improving the literacy skills of Grades 1, 2 and 3 pupils in the Maa language.

The OSP is implemented in 20 public schools in the Kajiado County that were selected due to their low learning outcomes. According to WERK (2011), on average only 23% of children aged 6-16 years

in Kajiado County could read at Grade 2 level, as compared to 51% nationally. The 20 schools chosen are situated in villages with the worst Uwezo learning outcomes[1] in the Kajiado County. For more information on the Uwezo assessment, please refer to the article by Nakabugo.

This paper provides evidence of the emerging impact of the OSP intervention reading and comprehension skills as assessed using the Maa Early Grade Reading Assessment (EGRA) tools. Since this was an impact evaluation comparing pupils' scores in 2014 and 2015, this article focuses on comparing the assessment results of pupils who were in Grades 1, 2 and 3 in 2014 with results captured among the same cohort of pupils later in Grades 2, 3 and 4 respectively in 2015. Baseline results for the Grade 1 pupils in 2015 have not been included in this paper as the aim is to document the progress of pupils over a one-year period.

The Maa Early Grade Reading Assessment tool can be accessed here.

2. CONTEXTUAL BACKGROUND

2.1 Linguistic context in Kenya

Kenya is a multilingual society with 42 ethnic groups or linguistic communities, each distinctly defined by language. The multilingual nature of Kenyan society is acknowledged in all policy documents as the Constitution of Kenya (2010)—the single most important and fundamental policy provision—advocates non-discrimination on the basis of language. Article 7(3) provides that the state shall (a) promote and protect the diversity of language of the people of Kenya; and (b) promote the development and use of indigenous languages, Kenyan Sign language, Braille and other communication formats and technologies accessible to persons with disabilities. In addition, Article 27(4) provides for the non-discrimination of all on the basis of language. These fundamental rights provide an obligation to the state to develop and promote all the languages. Notably, Article 7(2) lists English and Kiswahili as the official languages in Kenya and goes further to make establishing Kiswahili as the National Language an additional responsibility in Article 7(3).

2.2 Legal and policy framework for language in education policy

The language in education policy has been in force for over three decades, basing its foundation on the recommendations of the Gachathi Report (1976)[2]. This language policy is consistent with the existing legal, constitutional and other education policy provisions in Kenya. The policy borders on all the provisions that guide the national language framework as well as legal, constitutional and policy frameworks within which education is packaged and delivered in Kenya. It identifies mother tongue as the most suitable medium of instruction in the majority of lower primary classes (i.e. grades). Sessional Paper No. 14 (2012) states that the mother tongue shall be used for child care, pre-primary education and education of children in the first grades of primary education (i.e. children 0-8 years of age). When the school community is not homogeneous and lacks a common mother tongue, then the language of the catchment area would be used.

The education policy on using mother tongue as a medium of instruction provides the rationale that:

- Mother tongue acts as a link between home, early childhood development centres and primary school and encourages the child's free expression.
- It is a tool for teaching literacy, numeracy and manipulative skills.

1 Uwezo is a household-based, citizen-led assessment where children are assessed at home and not in school. The catchment areas for the selected schools had among the lowest learning outcomes in Kajiado County.

2 In 1976, a National Committee on Educational Objectives and Policies was set up to evaluate Kenya's education system, define the new set of education goals for the second decade of independence, formulate programmes to stem rural-urban migration, propose plans to promote employment creation and to suggest how to cut the education budget from 15% to 7%. The report popularly known as the Gachathi Report (1976) recommended the use of the mother tongue as the language of instruction from class 1 to class 3 (Grades 1–3).

- A good mother-tongue education programme lays a strong foundation for learning other languages.
- Mother tongue, when used as a medium of instruction, provides children with a sense of belonging, self-confidence and motivates them to participate in all school activities. This provides for a smooth transition experience from home to school.

2.3 Geographical context

Kajiado is derived from the Maasai word *Orkejuado,* which means 'the long river'—a seasonal river that flows west of Kajiado town. Kajiado County lies at the southern edge of the Rift Valley about 80 km from the Kenyan capital of Nairobi and has an area of 21,901 square kilometres. It borders Nakuru, Nairobi and Kiambu to the North; Narok to the West; Machakos and Makueni to the East, Taita Taveta and Tanzania to the South. The main physical features are valleys, plains and occasional volcanic hills.

2.4 Political context

Kajiado County has five constituencies each represented by a Member of Parliament in the National Assembly. These include Kajiado North, Kajiado Central, Kajiado West, Kajiado East and Kajiado South/Loitokitok. The county can also be subdivided administratively into seven subcounties: Kajiado Central, Isinya, Loitokitok, Magadi, Mashuru, Namanga and Ngong. The OSP is implemented in Kajiado Central and Loitokitok subcounties. The entire county has a total population of 807,070 people (49.8% females and 50.2% males) as of 2012 with a population growth rate of 5.5% per year. The county was initially occupied by Maasais (nomadic cattle herders) but other communities have since moved in, including Kikuyu, Kalenjin, Ameru, Kamba, Luhya and Luo. According to the Office of the Controller of Budget (Government of Kenya, 2013), 47% of the inhabitants live below the poverty line compared to the national average of 46%.

2.5 Education context of Kajiado County

School enrolment in Kajiado County stands at 55.8% for boys and 50.0% for girls (Government of Kenya, 2015). There were 461 primary education centres that participated in national exams in 2014 (63.6% public and 36.4% private). In the same year, 14,574 pupils (52.9% boys and 47.1% girls) in Kajiado County took the Kenya Certificate of Primary Education (KCPE) examination. Kajiado County faces the serious challenge of having overage learners—56.4% of pupils registered for KCPE in 2014 were 15 years old and above. There is also the challenge of having underage pupils completing primary education—3.3% of pupils sitting for KCPE are 13 years old and below. According to the Kenyan education system[3], the correct age for completing the primary cycle is at age 14. The primary level completion rate for the county stands at 47.3% for boys and 43.7% for girls against a national average age of 79.6% (Government of Kenya, 2015). The schools follow the national school calendar that runs from January to December of any given year.

2.6 Description of the Maa language

The Maa language or language group has a standard Latin-based orthography. It has several

3 The Kenyan education system comprises two years of pre-primary education, eight years of primary education, and four years of secondary education. The minimum age for entrance into Grade 1 is 6 years.

dialects that are distinct—Maasai from Laikipia, Kajiado, Tanzania and Baringo all known as 'Maa'. It is spoken by approximately by 500,000 Maasai, Samburu, Endorois/Chamus in Kenya and about 500,000 Arusa, Kisonko and Ilparakuyo people in Tanzania (Payne, 2008).

The Maa language consists of the following consonants and digraphs b, ch, d, h, j, k, l, m, n, ng, ng', p, rr, s, t, w, y, sh. It has 15 vowels written using five letters—a, e, i, o, u. For comparison, Swahili and Spanish have only five vowels; while English arguably has 18 vowels (represented in written format as a, e, i, o, u). The Maa vowels are distinguished by the tone: high, low or falling (high-low). A falling tone is where the tone moves quickly from high to low and hence can be perceived as falling. The current Maasai writing system under-represents significant sounds in the language, making it very difficult to learn to read and write. Thus, any new Maasai-language literacy programme is unlikely to be successful unless the writing system issue is addressed first. To mitigate this problem, an orthography review process was launched in July 2012 with the goal of developing an experimental orthography for use in the Maasai literacy instructional materials being developed for the OSP. The experimental orthography in use in the programme includes basic tone patterns written over vowels as high 'á', low 'a' and falling as 'â'. The high and the low tone are taught in Grade 1 and the falling tone in Grade 2. **Table 1** lists the complete set of 15 vowels.

TABLE **1**

Complete set of Maa vowels					
tone	a	e	I	o	u
high	á	é	Í	ó	ú
low	a	e	I	o	u
falling	â	ê	Î	ô	û

Tone is an extremely important feature in the Maa language as the meaning of individual Maa words change depending on tone. The following pairs of words with a similar combination of letters but different tones are used to illustrate this. The word *empírón* means "making fire using traditional methods" while the word *empirón* means "something fat". The word *álé* means "a cowshed" while the word *alé* means "milking".

3. OVERALL APPROACH TO THE MAA EGRA

The Maa Early Grade Reading Assessment (EGRA) used quantitative approaches to sampling, choice of data collection tools and data analysis. A quantitative design was chosen to enable the

comparison of results across years. Samples were drawn to ensure representativeness of all the characteristics within the population of interest in the assessment. As the assessment sought to observe and detect any differences emerging from the interventions, quantitative tools of analysis that included testing statistical significance were used.

Stratified random sampling was used to select the pupils to be assessed. A total of 800 pupils (400 girls and 400 boys with 10 from each of the Grades 1, 2, 3 and 4) from the 20 Opportunity Schools were selected. The pupils were assessed in reading in three languages (English, Kiswahili and Maa) and in numeracy. However, this paper presents the results of the Maa assessment only. The following are some of the characteristics of the Opportunity Schools:

- In 2014, the repetition rate was 6.5%. Repetition in this case refers to repeating a grade within the year. Thus, 6.5% of all pupils enrolled in Grades 1-8 had repeated a grade in that year. The repetition rate was highest at the end of the primary cycle (Grade 8) at 12% while Grade 2 and 3 repetition rates were 9% and 8% respectively. Grade 1 had the lowest repetition rate at 5%, making the average repetition rate for early grades 7%.

- Overall enrolment in Grades 1 to 3 increased by 14% between 2013 and 2014 (*see* ***Table 2***). Due to severe drought in 2014 and 2015, many families migrated in search of water and pasture, which affected enrolment. Enrolment thereby decreased between 2014 and 2015 by 13%.

TABLE **2**

Enrolment in Grades 1-3, 2013-2015

	Enrolment (n)			Percentage change in enrolment	
	2013	**2014**	**2015**	**2013 to 2014**	**2014 to 2015**
Girls	1,302	1,426	1,277	+10%	-10%
Boys	1,335	1,569	1,329	+18%	-15%
TOTAL	**2,637**	**2,995**	**2,606**	**+14%**	**-13%**

- Attendance of pupils enrolled in Grades 1-3 was monitored. During the year 1 (2013) baseline survey, the average pupil attendance rate was 72%. In year 2 (2014), the average daily attendance rate increased to 77%. All data presented was collected in March of that year.

3.1 Adapting an EGRA using open-source assessment tools

The WERK in collaboration with the Africa Educational Trust (AET) and the Summer Institute of Linguistics (SIL) International developed an EGRA in Maa in March 2014. The purpose of the tool was to monitor the progress of the acquisition of reading skills in Maa. The team was also joined by fluent first language or native speakers (L1) who had been involved in developing the Maa grade-level reading materials for the OSP.

The Maa EGRA is an orally administered assessment aimed at measuring the reading skills foundational to later reading (and academic success). The Maa EGRA takes approximately 15 minutes to administer one-on-one with each student and is often combined with a questionnaire to measure a variety of student background variables to later assist in explaining some of the reading outcome findings.

The Maa EGRA is a hybrid between the Research Triangle Institute (RTI) EGRA in English and the Uwezo framework for reading tests in Kiswahili. This shows the possibility of using open-source materials to develop oral assessments. The decision to use a hybrid was informed by the fact that each assessment has unique strengths. Thus, although Uwezo assessment tools are not timed making measurement of automaticity impossible, they have a very clear framework for item development that enables separation into distinct levels of difficulty. Due to this detailed framework, it is easy to develop different tools of the same level of difficulty that can be used over the years. The protocol was adapted from the *EGRA Toolkit* (RTI, 2009). All the subtasks were timed to allow the measurement of automaticity or the quick and accurate recognition of letters, sounds and words without hesitation. The Maa EGRA subtasks are listed in **Table 3**.

TABLE **3**

Maa EGRA subtasks

Subtask	Description	Adapted from EGRA	Adapted from Uwezo	Phase (s) of development
Letter-sound identification	Measures knowledge of letter-sound correspondences. One hundred letters and digraphs of the Maa alphabet are presented in random order in both upper and lower case. Only 27 letters are in upper case while the other 73 letters and digraphs are in lower case. Some of the letters and digraphs are repeated more than once. It is timed to 60 seconds and is discontinued if none of the sounds in the first line (i.e. 10 letters) are produced correctly.	Number of letters	—	Partial alphabetic
Syllable identification	Measures the ability to read individual syllables. Fifty syllables are presented. It is timed to 60 seconds and is discontinued if none of the first 10 syllables are read correctly.	Number of syllables	—	Partial alphabetic
Familiar word reading	Measures the ability to identify individual words from grade-level text. Fifty words are presented. It is timed to 60 seconds and is discontinued if none of the words in the first line (i.e. five words) are read correctly. Words must be familiar and within the pupils vocabulary. They should also be nouns with no plurals, found in the state-recommended textbooks and have two or three syllables.	Number of words	Type of words to be included	Alphabetic
Oral reading fluency	Measures the ability to read a grade-level passage of approximately 60 words. It is scored for accuracy and rate. It is timed to 60 seconds and is discontinued if none of the words in the first line (i.e. about 10 words) are read correctly. The stories are based on Grade 2 texts and all words must be found in the state-recommended books. Number of syllables to be between two and four with one or two words of five syllables.	Length of text	Type of words included in the story	Consolidated-alphabetic
Reading comprehension (without look backs)	Measures the ability to answer questions about the grade-level passage. Question types include explicit and inferential questions, and look backs are not allowed. Each story should have five comprehension questions. The first two are literal, meaning that the answers should be found directly in the text, usually from only one sentence; the next two are textually implicit, meaning that the answers should be found in the text but the pupil would have to draw from two sentences to answer. The last question should be inferential, involving some reasoning. None of the questions should use the pronoun 'I' to avoid confusing the pupils as they may personalise it.	Direct, implicit and inferential questions Number of questions	Direct and inferential questions	Consolidated-alphabetic Automatic
Interview	Gathers information about the child that is related to literacy and language development (e.g. first language, access to print). It is self-reported by the child.	Type of questions to the child	—	Consolidated-alphabetic Any phase of interest

Source: adapted from the EGRA and the framework for development of Uwezo Kiswahili assessment tools

3.2 Data collection

Pupils were assessed in reading in Kiswahili, English and Maa, and numeracy. The assessment took place at the school level in March 2014. Once sampled, pupils who were to participate in the assessment were called out of class by the enumerators one at a time to a quiet place or classroom where they were assessed. This was done in order to minimise disruptions in the class and to ensure that the children being assessed were not withdrawn from class for too long. Two assessors were assigned to each school and each child was assessed by an individual assessor. Interrater reliability was not verified.

3.3 Data entry and analysis

Quantitative data were generated from the field. The data were then coded and organized for verification and completeness. Supervisors verified the data at the material submission stage. Upon verification of the completeness of the data capture, a template that had been prepared, filled and tested on a dummy tool was used to capture the data in Excel for export to SPSS 20 where it was analysed. Data entry was strictly supervised to ensure that no data were left out. Microsoft Excel was chosen due to the quantity of the data received. In addition, the data were straight-forward and thus no complex tool was needed to reliably enter the data. The data were cleaned logically before it was exported to SPSS 20 where it was analysed on the basis of the preconceived output tables. The inferential analyses were conducted in Stata.

4. PRESENTATION OF FINDINGS

4.1 Letter sounds

The first literacy task was to measure pupils' understanding that letters in written words represent sounds in spoken words. This is referred to as the alphabetic principle. For a pupil to acquire this skill, they need to understand the relationship between letters and their sounds referred to as graphophonemic knowledge, and the association between a specific letter and its corresponding sounds—for example, the letter *m* and the sound '*mmm*' as in 'man' is a letter-sound correspondence. The mean number of correct letter sounds identified per minute was 33 correct letter sounds per minute (out of a possible 100 letters presented), with a standard deviation of 22.77).

4.2 Syllables

This subtask measures pupils' ability to recognise and manipulate the individual sounds (phonemes) in spoken words referred to as phonemic awareness. On average, pupils were able to correctly read 25 syllables per minute (out of a possible 50 letters presented), with a standard deviation of 15.10.

4.3 Familiar word reading

This subtask measures the ability to identify individual words from grade-level text. On average, pupils were able to read 23 familiar words per minute (out of a possible 50 familiar words), with a standard deviation of 11.90.

4.4 Reading for fluency

Oral reading fluency measures the ability to read a text correctly, quickly and with expression. However, prosody or the use of appropriate intonation and phrasing when reading or reading with expression was not measured. On average, pupils were able to read 27 words correct per minute (WCPM) (out of a possible 63 words in a text), with a standard deviation of 17.13.

Emergent readers are those who read more than 17 WCPM while fluent readers are those who read at more than 45 WCPM. The programme has opted to use the national Ministry of Education, Science and Technology (MOEST) Kiswahili benchmarks of 45 WCPM for fluent readers and more than 17 WCPM for emergent readers. The decision to use Kiswahili was informed by the similarities in the structure of the language. **Table 4** presents the percentage of fluent and emergent readers measured by the Maa EGRA.

As shown in Table 4, 12% of the pupils were able to read more than 45 WCPM. In Grade 2, 2% of pupils had acquired oral reading fluency while this percentage was 13% for Grade 3 and 22% for Grade 4. Overall, 54% of the pupils could read more than 17 WCPM, meaning that they were either emergent or fluent readers.

TABLE **4**

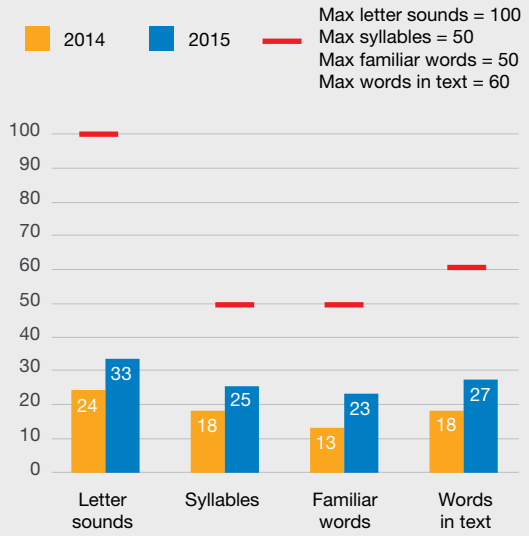
Maa EGRA: Percentage of pupils falling in each WCPM category in 2015

Grade	Below 17 WCPM	17-44 WCPM	45+ WCPM
Grade 2	70%	28%	2%
Grade 3	40%	48%	13%
Grade 4	28%	50%	22%
Overall	46%	42%	12%

Source: WERK, 2015

The study also sought to compare the reading scores of pupils in 2014 to those in 2015 as previously mentioned. **Figure 1** compares the reading scores of pupils in Grades 1-3 in 2014 with those of pupils in Grades 2-4 in 2015 to capture learning progress.

Figure 1. Comparison of the acquisition of reading skills of Grade 1-3 pupils in 2014 to Grade 2-4 pupils in 2015



Source: WERK, 2015

Between 2014 and 2015, there was an increase in scores for reading letter sounds from an average of 24 to 33 correct letter sounds per minute—an increase of 9 correct letter sounds per minute. The reading of syllables increased from an average of 18 to 25 correct syllables per minute—an increase of 7 correct syllables per minute. The reading of familiar words increased from 13 to 23 correct familiar words per minute—an increase of 10 correct syllables per minute. Oral reading fluency increased from 18 to 27 WCPM—an increase of 9 WCPM.

4.5 Comprehension scores

Reading comprehension measures pupils' ability to understand and derive meaning from written language. Children were first instructed to read a passage and then were asked questions based on where in the passage they had stopped reading. Non-readers were excluded from this subtask. On average, 48% of the pupils could not answer a single comprehension question correctly while 14% could only answer one question, 14% could answer two questions, 11% could answer three questions, 7% could answer four questions and only 5% could answer all the comprehension questions correctly.

5. SUCCESSES AND CHALLENGES

5.1 Successes

i. Measuring pupils' learning outcomes

According to LaBerge and Samuels (1974) as quoted by Fuchs et al. (2001), automaticity or the speed and accuracy with which single words are identified is the best predictor of overall reading competence. This skill can only be measured orally. Oral reading fluency therefore becomes the most salient characteristic of skillful reading (Adams, 1990).

ii. Monitoring progress in the acquisition of reading skills

The use of an oral assessment made it possible to monitor the progress in the acquisition of reading skills in Maa by pupils in the programme schools from year to year, across grades, by grade and

gender as well as other variables collected in the accompanying student questionnaire.

iii. Designing a teacher professional development programme

Measuring automaticity by counting the number of correct words read per minute allows the project staff to note the types of decoding errors students make and the kinds of decoding strategies students use to read vocabulary. It also helps determine the pupil's level of reading development. Oral assessments therefore facilitates the designing of a teacher professional development programme to support teachers and enable them to make sound instructional decisions and formulate reading interventions that help pupils learn to read.

iv. Availability of an EGRA in Maa

An EGRA in Maa is now available for anyone interested in administering an oral reading assessment in Maa.

v. Political

Over the last couple of years, the government has called for evidence-based policy decision-making. Sharing data on pupils' reading scores with the government has helped renew the push for the implementation of the Language in Education policy by the government. Other non-programme schools have also shown interest in adopting a similar programme at their schools.

5.2 Challenges

i. Contextual

- Dialects

There are regional language variations including differences in pronunciation, grammar and/or vocabulary. By using the standard orthography, this disadvantaged pupils who spoke a dialect where the words were unfamiliar or had a different meaning.

- Experimental orthography

Maa is a tonal language and based on the tone, different words have different meanings as seen earlier. The current Maasai writing system is difficult to read as the tone is not marked. For certain words, one can only derive its meaning when read in context. Currently, there is a push to revise the Maa orthography. The orthography used is therefore experimental and the programme is in the process of evaluating whether the marking of tone has any effect on acquisition of reading skills.

- Language change and borrowings

Languages change throughout their lifetime. In the process, languages develop new vocabulary and regularly develop new meanings for old words. One major type of language change is borrowing or coining new words for example *engárri* (coined from the Kiswahili word *gari* meaning "car") and *embúku* (coined from the English word "book"). The challenge with these is that the word coined may not be understood by all speakers of the language. Care has to be taken to include words that have been adopted over time and now form part of the vocabulary.

ii. Socio and psycholinguistic issues

- Language of instruction

Despite the fact that Maa is the language used more often and with more proficiency hence the dominant language, the prestige of English as the language of the elite induces many parents to demand that their children learn English and also learn in English. As a result, pupils are taught three languages (English, Kiswahili and mother tongue) at the same time. This has implications in the acquisition of reading skills in all three languages.

iii. Low capacity of government to support a multi-lingual education programme

- Grade-level texts

Maa EGRA is a pioneer project in Kenya and initially, there were no grade-level text books for teaching Maa and as such, the programme had to develop materials while developing the assessment tools at the same time. Thus, the development of the Maa EGRA faced certain resource challenges.

- Lack of a national benchmark on fluency

There is currently no benchmark on reading fluency specific to Maa. There is a need to develop a benchmark specific to Maa instead of using

Kiswahili benchmarks due to similarities in the structure of the two languages.

- Few teachers competent in teaching and assessing Maa

Most of the teachers teaching and assessing Maa were not trained to teach literacy in Maa. Hence, a lot of time is spent training them.

- Lack of clear guidelines for implementation of the language in education policy.

The language policy in Kenya was developed in 1976 and its implementation began then. However, due to lack of clear guidelines on its implementation, many schools are not teaching the mother tongue in lower grades or using it as a medium of instruction.

iv. Political

- Contradictions in government policies

The Language in Education Policy is in conflict with the government policy of posting teachers anywhere in the country based on whether there is a vacancy or not. There are currently non-Maa speakers teaching early grades. This has great implications for the teaching of Maa and the acquisition of skills in Maa. Teachers who do not speak Maa also cannot assess the pupils' Maa reading skills.

6. FUTURE SUGGESTIONS

From a programme perspective, these are some of the suggestions for continued improvements going forward:

1. Currently, there is a need for detailed documentation on the use of Maa teaching and learning materials in order to verify whether the materials contain grade-level texts. This is one way to help ensure that the Maa EGRA is measuring skills at the right level.
2. The WERK is proposing to carry out research on the experimental orthography. A decision needs to be made on whether the new orthography enhances reading in Maa.
3. The WERK proposes to empirically decide on a Maa benchmark for reading fluency based on the evidence available. This will help in accurately measuring and monitoring progress in the acquisition of reading skills.

Based on all the above information, the WERK proposes to review the tools next year to make them more effective.

REFERENCES

Adams, M. J. (1990). *Beginning to Read: Thinking and Learning about Print*. 1st edn. Cambridge MA: The MIT Press.

Fuchs, L. S., Fuchs, D., Hosp, M. K. and Jenkins, J. R. (2001). "Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis". *Scientific Studies of Reading*, Vol. 5, No. 3, pp. 239-256. http://www.specialistedpsy.com/fuchsetalreadfluency.pdf-link.pdf

Kenya Ministry of Education, Science and Technology (2012). *Sessional Paper No 14 of 2012, A Policy Framework For Education And Training: Reforming Education and Training in Kenya.* Nairobi: Government Printer.

Kenya Ministry of Education, Science and Technology (2015). *2014 Basic Education Statistical Booklet.* Nairobi: Government Printer.

Kenya Office of the Controller of Budget (2013). *Kajiado County Budget Implementation Review Report*. Nairobi: Government Printer.

Kenya Judiciary (2010). *The Constitution of Kenya, 2010.* Nairobi: National Council for Law Reporting.

Kenya (1976). *Report of the National Committee on Educational Objectives and Policies (Gachathi Report).* Nairobi: Government Printer.

LaBerge, D. and Samuels, S. (1974). "Toward a theory of automatic information processing in reading". *Cognitive Psychology,* Vol. 6, pp. 293-323.

Payne, D. L. (2008). *The Maasai (Maa) Language: Kenyan Southern Maasai, Samburu*. Oregon: University of Oregon.

Research Triangle Institute International (2009). *Early Grade Reading Assessment toolkit.* USAID Education Data for Decision Making (EdData II). Washington, D.C.: USAID.

Women Educational Researchers of Kenya (WERK) (2011)*. Are our children learning? Annual learning assessment report.* Nairobi: WERK

Learning-by-Doing: The Early Literacy in National Language Programme in The Gambia

PEI-TSENG JENNY HSIEH
University of Oxford

MOMODOU JENG
The Gambia Ministry of Basic and Secondary Education

ABBREVIATIONS

EGRA	Early Grade Reading Assessment
ELINL	Early Literacy in National Language
FIOH	Future in Our Hands
GATE	Gambia Association of Teaching English
GPE	Global Partnership in Education
INSET	In-Service Teacher Training
MOBSE	Ministry of Basic and Secondary Education
NAT	National Assessment Test
NL	National language
SEGRA	Serholt Early Grade Reading Ability

1. INTRODUCTION

The Gambia is a small country in West Africa with a population of 1.8 million. Surrounded by Francophone countries, including Senegal which almost fully embraces the territory, The Gambia is one of the very few Anglophone countries in the region. English and the five major national languages—Jola, Mandinka, Pulaar, Saraxulle and Wollof—are the official languages. Most of the people can speak at least two of the local languages but not necessarily English. The Gambia is ranked 168 out of 187 countries in the United Nations Development Programme's Human Development Index. Nevertheless, the government has shown a strong commitment to education, which is reflected by the high government expenditure on education (20% of total expenditure) and the many innovative education programmes in the country (The World Bank, 2015).

As of 2015, there were 498 government and government-aided Lower Basic Schools (Grades 1-6), and around one third of the children attend pre-school before attending primary schools. The language of instruction beginning in Grade 1 is English but in the majority of government schools, local languages are the dominant languages in classrooms for both teachers and students. Late enrolment (normally due to parents' choice of starting with religious education) and grade repetition rates are significant but not abnormal for the region. Most of the education policies and reforms are managed and implemented by the Ministry of Basic and Secondary Education (MOBSE), and higher education only began in the early 1990s with the establishment of Gambia College as an institution for teacher training.

In recent years, with a relatively stable and steadfast leadership, the MOBSE has successfully introduced a number of nationwide education programmes. The Early Literacy in National Language programme (ELINL) is one of the programmes that have contributed to the most significant changes to the education system in the country. In 2014, an evaluation study was conducted in the third year of the programme's implementation and the outcome provided recommendations to the government for scaling it up. The assessments discussed in this article were part of the evaluation plan to examine the ELINL programme outcome.

2. EARLY LITERACY PROGRAMMES IN THE GAMBIA

There has been a general consensus supported by the result of the annual National Assessment Tests (NATs) and the four cycles of the Early Grade Reading Assessment (EGRA) since 2007 that the achievement level of a large proportion of students is unsatisfactory by government standards. Most alarming is the very low attainment in basic literacy among primary school children as revealed by the EGRA. The 2007 EGRA observed a large floor effect. Using the EGRA tool, half of the assessed Grade 3 children were unable to read a single given English word and 82% of the children assessed were unable to read more than 5 words per minute (Research Triangle Institute International, 2008). Although the most recent EGRA result showed significant improvement in letter recognition, the results for reading comprehension continued to stagnate (MOBSE, 2011).

The result was quite surprising as, unlike a number of countries in the region, the percentage of unqualified teachers was low. In a teacher assessment (MOBSE, 2012) covering about two thirds of the primary teachers in the country, the majority of the teachers (75%) demonstrated that they have at least the subject knowledge required to teach in primary schools—although their pedagogical knowledge and classroom management skills could be further improved.

In response to the low achievement in reading, the government began to trial teaching to read through the Jolly Phonics programme in 2008. The programme is designed to be child-centred and teaches English literacy through learning five key skills for reading and writing (i.e. letter sounds, letter formation, blending, segmenting, tricky words). Jolly Phonics was managed by the Gambia Association of Teaching English (GATE) and was scaled up to national level in 2009. At the same time, Future in Our Hands (FIOH), a Swedish non-governmental organization, introduced an early literacy programme the Serholt Early Grade Reading Ability (SEGRA) to a selected number of schools. The SEGRA approach also provides phonics instruction to teach literacy in English but places more emphasis on extracurricular reading and on involving communities. It also encourages using National Languages (NLs) to aid comprehension.

The ELINL pilot programme was introduced in 2011. The programme aimed to develop phonemic awareness through NLs to provide a foundation for learning to read in English. The rationale for this approach was based on the mounting evidence in applied linguistics and in cognitive psychology that young children can learn to read with better fluency and comprehension when doing so in a language they speak and understand. In addition, the NLs are orthographically transparent and can thus better facilitate the acquisition of decoding and blending than the orthographically opaque English language. Once children have mastered the basic reading mechanisms in their own language, they will be better prepared to read with fluency and comprehension in a second language. The pilot targeted a total of 125 Grade 1 classes in 109 government and government-aided Lower Basic Schools across all six educational regions in the country. There were 25 classes per language for the five main national languages: Jola, Pulaar, Mandinka, Saraxulle and Wolof. With technical support from the Global Partnership in Education (GPE), new teaching and learning materials were developed for the programme in these five languages, including orthography, scripted lessons (in both English and NLs), textbooks (Reader I and II) and flash cards (including letter, word and picture cards).

Pupils in the ELINL classes receive 30-60 minutes of national language lessons per day with the following specifically defined instructional steps:

- Step 1: Revision of letters, syllables, words and sentences (from the previous lessons)
- Step 2: Teaching new letters (letter sound, letter shape, synthesis analysis of key word)
- Step 3: Blending letter sounds to read syllables (blending letters)
- Step 4: Blending syllables to read words (blending syllables)

- Step 5: Reading practice
- Step 6: Writing practice.

In addition to these steps, the programme also highlighted the importance of quick, formative assessments during the lessons. Every child should be given an opportunity to read aloud the new letters, syllables and words taught in a lesson. Simultaneously, in the non-ELINL classes, students continued to receive reading instructions in English through the Jolly Phonics and SEGRA methods.

In order to sensitise the community to the introduction of NLs in teaching and learning, regional consultative meetings and orientation activities were conducted to create awareness and seek approval from the public and communities. MOBSE reported general support and acknowledgement from the public. The MOBSE's In-Service Teacher Training (INSET) Unit together with five assigned national language desk officers coordinated the training and implementation of the programme. A cascade training design was employed to train 25 teacher trainers who coordinated, supervised and conducted subsequent training of other teachers and school-based coaches. Coaches visited schools monthly to provide support to the ELINL classes and employed a structured observation form to gather periodic information for INSET's records. In addition, coaches provide support to head teachers on programme implementation and give immediate feedback to teachers on teaching following lesson observations.

Results of the post-test in 2012 showed some significant gain in reading skills. After approximately six months of literacy instruction, around 40% of the children in the ELINL pilot group could read at least one word in a given English passage while only 14% of the children in the comparison group could do so. Hence, the government made the decision to expand the pilot to include one new Grade 1 class (25-45 pupils) in each of the following academic years in the same pilot schools. An evaluation study was carried out in 2014 to examine the implementation and outcome of the ELINL using an assessment in five language forms that was designed for this purpose (Hsieh, 2014).

3. EVALUATING THE ELINL OUTCOMES

The objectives of the 2014 ELINL evaluation were:

1. To determine if the pilot achieved the intended goals set out for the ELINL programme (for pupils in Grades 1, 2 and 3).
2. To provide evidence to the public and stakeholders of the outcome of the ELINL programme. In particular, to determine whether the programme has been beneficial and whether it has had an impact on pupils' English competencies.
3. To identify key factors that may contribute to the programme's effectiveness.

The success of a literacy programme depends on the programme's characteristics (i.e. teacher preparation and knowledge, curriculum, materials, instruction, assessment methods, etc.) and can be greatly influenced by contextual factors (e.g. student and school characteristics, situational factors, family role). Although it could be interesting to explore a number of aspects in the evaluation of the ELINL, for the scope of this exercise and the needs of the government, the emphasis was placed on quality of programme input.

3.1 Instrumentation

A pupil assessment in five languages, a teacher questionnaire, a school-based coach questionnaire, and a head teacher questionnaire were developed for the study to gather both quantitative and qualitative information on the ELINL programme implementation and outcomes. Many challenges lie in the design of multilingual assessments. The difficulties are not only about translation and adaptation but also the comparability and interpretation of data from the different language forms. In large-scale cross-cultural assessments, comparability can often be questionable even after lengthy and costly preparation. In devising and administering the NL assessments for the ELINL evaluation, limited time and resources were some of the significant factors driving the design and enumeration plan. There were no established

assessment tools in the five national languages—in fact, there was very little written material in these languages if any at all. There is also very limited information regarding the differences in the linguistic complexity of the five NLs at different stages of language development (Goldsmith, 1995; Bird, 2001). The development of the instruments and the content of the test thus relied mainly on the textbooks and scripted lessons designed for the ELINL programme and on expert judgements. MOBSE's national language desk officers played important roles in the design and adaptation of the instruments. Each set of assessments contained a pupil's background questionnaire and sections on 1) letter sound recognition, 2) syllable reading, 3) word reading, 4) connected text reading and 5) reading comprehension. These subtests were included because they are the core components of teaching and learning in the ELINL programme. In addition to the reading tests in national languages, the instruments also included a subtest on connected text reading and reading comprehension in English, which was identical to a subtest in the 2013 Gambian EGRA.

While not assuming complete compatibility across the different forms, orthographic transparency and regularity was similar across the five languages. The common items in letter sound reading, syllable reading, word reading and the section on English connected text and comprehension questions provide some references to the generalisation of outcomes. The use of the EGRA subtests also affords the opportunity for comparison with the national EGRA sample.

3.2 Sampling

The sample for the 2014 ELINL study comprised 2,864 pupils in 91 schools across the six educational regions. The sampling plan for the 2014 study was to select a representative sample of students in each of the NL pilot groups and a group of traceable students in the ELINL programme from the 2012 post-test.

There were some challenges in drawing a comparable sample, specifically in tracing students' progress. A baseline assessment for the ELINL programme was conducted in November 2011 and the sample schools were drawn from the list of schools intended as pilot schools. The sampling criteria was much restricted by the possible logistic arrangement. The first year post-test in June 2012 found only 23-57% of the students surveyed at baseline were traceable due to school administration errors. In addition, due to funding limitations, the requested control schools were conveniently sampled and the requirement for a control group was not taken into much account. These have introduced much difficulty and complexity to the selection of a traceable and comparable sample for the 2014 study. The comparison of the full 2012 and 2014 pupil lists and the selection of 'matched schools' was the attempt to make progress in tracking and to make comparison with the national sample possible.

The estimates were based on the national assessment in 2012/2013 and the ELINL baseline results, with a 95% confidence interval and a 3.5% margin of error. Up-to-date lists of students were obtained by the NL officers and matched with the 2012 post-test student list. Six Madrassa schools on the list were excluded from the sampling due to early school closure for holidays.

In terms of the comparison with the control group, two control schools for each NL were identified from the baseline list to match the selected pilot schools. In pairing up the pilot and matched control schools, the desk officers were asked to take into account the environment and performance (i.e. NAT results, parent perceptions, etc.) of the schools. The aim was to make comparisons only between schools with similar conditions.

In the matched control schools, the sample consisted of ten randomly selected pupils of the same NL as the matched pilot schools in each of the Grades 1, 2 and 3. In addition, the sample also included all Grade 3 pupils assessed at baseline and/or during the 2012 June post-test. **Table 1**

shows the number of candidates assessed in the study.

TABLE **1**

Achieved sample for the 2014 ELINL study

	Jola	Mandinka	Pulaar	Saraxulle	Wollof
Grade 1	186	161	190	149	142
Grade 2	174	161	170	150	134
Grade 3	169	187	178	147	174
Grade 3 comparison group	73	69	72	96	102

Source: Hsieh, 2014

3.3 Outcomes of the evaluation

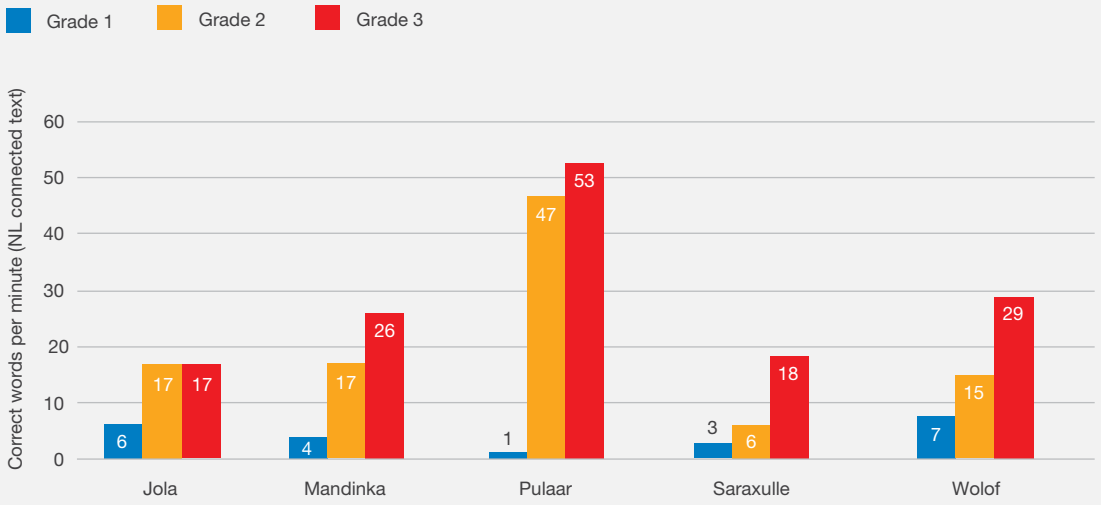
Not assuming full compatibility of the five test forms (for the reasons previously stated) in this section, we present the results for the total sample and by language groups. All results presented in this section and the next are supported by tests of statistical significance set at a probability of 0.05 (p-value ≤ 0.05).

On average, the ELINL children could recognise nearly 80% of the letter sounds (out of 27-33 graphemes, depending on the language) in Grade 1 and more than 90% when they reached Grade 3 in this longitudinal study. They were able to read about 60% of the syllables given (out of 25 frequently used NL syllables) at Grade 1 and 80% of the given terms by the time they were in Grade 3.

Phonemic awareness instruction aids reading comprehension primarily through its influence on word reading. However, for children to understand what they read, they must be able to read words accurately and with fluency. The ELINL pupils were able to accurately read about half of the NL words given (out of 25 commonly used NL words) in Grade 1. They were able to read about two-thirds of the words by the time they were in Grade 2. Nonetheless, the increase in number of words read is only a moderate 4% from Grade 2 to Grade 3.

Figure 1 presents the results of the subtests on NL connected text reading. Although NL reading fluency varies between the groups, generally, there is an increase in fluency from Grade 1 to Grade 3. Pupils in Grade 1 were able to read between 6 and 20 words in the texts of 53-62 words—although they were only able to read with limited fluency at this stage (1-7 words correct per minute).

Figure 1. Number of correct NL words read per minute by NL group and by grade



Source: Hsieh, 2014

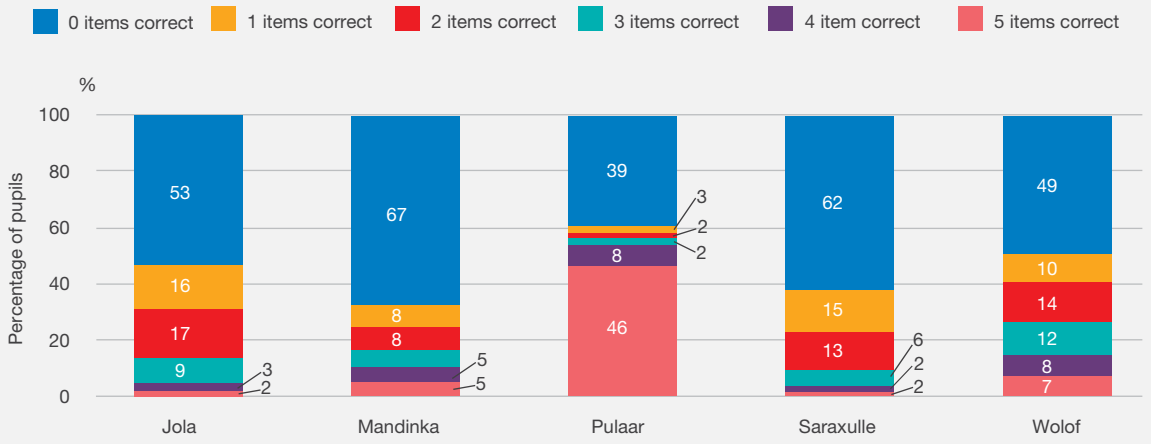
It is worth noting that Pulaar pupils read with unusually low fluency at Grade 1. After verifying the data with the desk officers, we eliminate the possibility of data collection and entry error. More analysis results and qualitative information would need to be obtained to investigate the low Pulaar NL fluency score for the Grade 1 cohort. Grade 2 and Grade 3 Pulaar pupils seem to have more consistent reading fluency between English and NL reading.

The underlying purpose of enhancing reading fluency (or other foundational reading skills) is to achieve comprehension. Although emphasis on pre-reading skills is important for early literacy programmes, the skills to recognise letter sounds and shapes, and blend syllables and words cannot really be considered 'reading'. Ultimately, any reading programme would expect the children to be able to read with comprehension. **Figure 2** and **Figure 3** present the reading comprehension scores achieved by NL group and by grade. The results suggest that children will need further support to read and also to process and understand what they have read. Although pupils in different NL groups were reading on average 17-53 words per minute in Grade 3, more than one-third of those pupils scored zero on reading comprehension.

The result of the English reading and comprehension test was similar to reading in NLs. Even when children were able to read with good fluency in English, comprehension of the text seemed to be limited. Pupils in Grade 3 were reading on average 35 words per minute. Although they could answer on average two out of the five questions related to the text, nearly 40% of the pupils scored zero on reading comprehension. This is likely attributable to the focus on word-level instruction as opposed to comprehension in the classrooms.

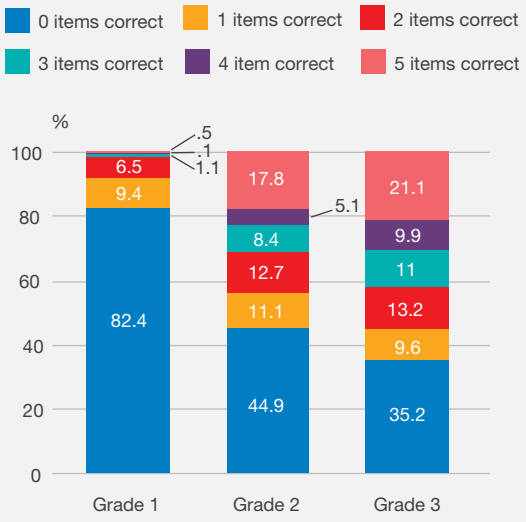
Interestingly, even after accounting for the word complexity between the languages, at Grade 3, children achieved on average a higher level of fluency in English. This might partly be due to more exposure to English written materials in schools. In addition, across all NL groups, there were children reading with good fluency but achieving zero comprehension scores. When reading NL connected texts, there were also students reading with very low fluency and accuracy but who were still able to answer some of the comprehension questions. One possible explanation might be that it is easier for children to infer meaning in a language they already master and use regularly (orally), either through the

Figure 2. ELNIL reading comprehension score by ELNIL group, all grades combined (in %)



Source: Hsieh, 2014.

Figure 3. ELNIL pupils' performance on ELNIL comprehension questions by grade for all languages (in %)



Source: Hsieh, 2014.

limited number of words they were able to read or the questions asked in the NL.

In all the subtests, pupils' performance varied across NL groups. Among them, Pulaar pupils achieved the highest mean scores across all subtests in all grades. In reading English connected text, around 65% of the Pulaar students were able to read one word or more. When we look at reading connected text in the NL, only 18% of them could read one word or more. This might indicate that the higher scores in the subtests achieved by the Pulaar pupils cannot be fully attributed to national language learning but other factors. Observation suggests this might be due to pre-school exposure to phonics teaching or very early encounters with Arabic scripts in Darra (Abu-Rabis and Taha, 2006). Nevertheless, we do not have enough measures of pupils' backgrounds to examine these variables.

4. RESULTS OF THE ELINL PROGRAMME

There is yet little consensus on the standard of reading fluency in these national languages. In examining the outcome of the ELINL programme, we will discuss in this section the results presented in the last section by benchmarking with the ELINL learning goals and by comparing with the comparison group and the national sample.

4.1 Benchmarking with the ELINL learning goals

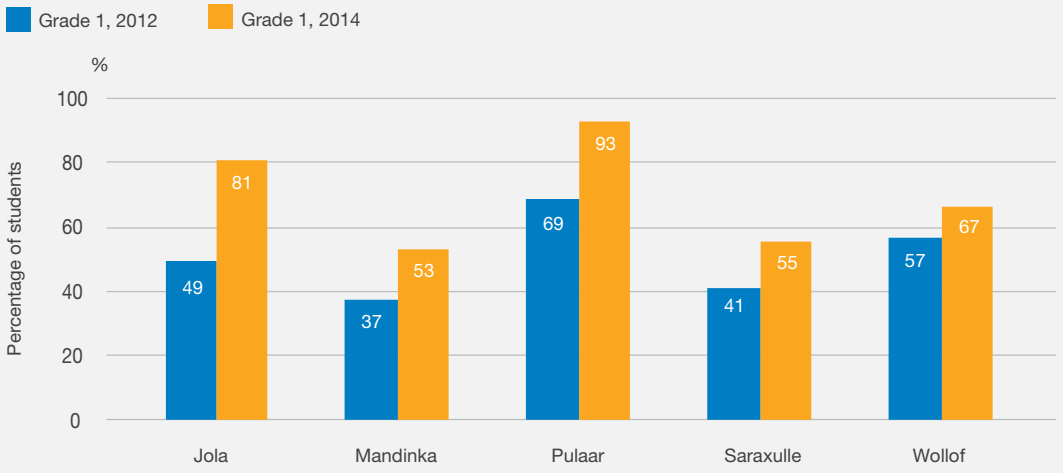
At the planning phase of the ELINL piloting in 2011, the goals set by the GPE for the programme at the completion of 15-20 weeks of scripted lessons were that:

- approximately 85% of children will be able to name 80% of letters
- approximately 85% of children will be able to read at least one word in a connected text in one minute
- all students should read by the end of the trial at different levels.

In addition, it is anticipated that learning to read in their mother tongues would allow the children to acquire basic reading skills and phonemic awareness that can be transferred to reading in English.

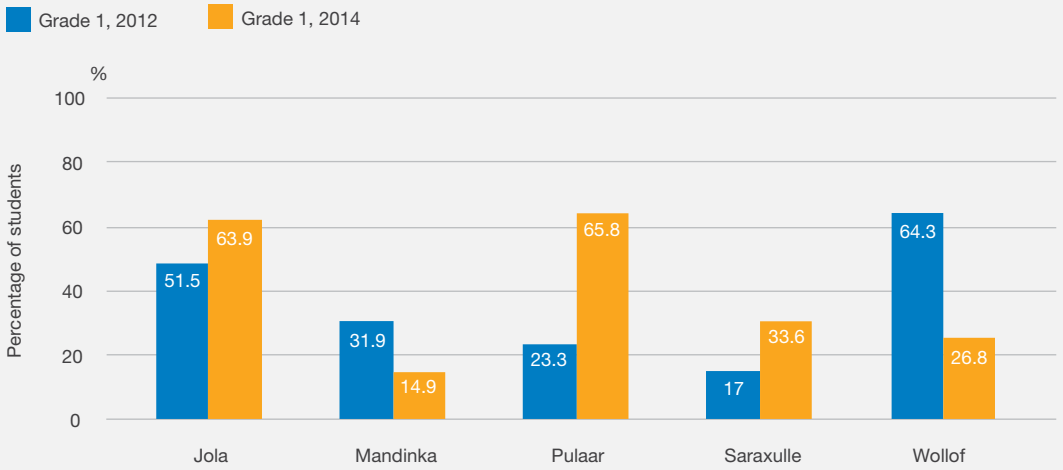
The results presented later in this section of assessments that were conducted at the end of the first year pilot in 2012 suggested that the above goals were not fully achieved. However, more students in the ELINL programme were able to achieve these goals when compared to those in the comparison group.

Figure 4. Grade 1 students recognising at least 80% of the letters, 2012 and 2014 (in %)



Source: Hsieh, 2014

Figure 5. Grade 1 students reading at least 1 word in NL passage, 2012 and 2014 (in %)



Source: Hsieh, 2014

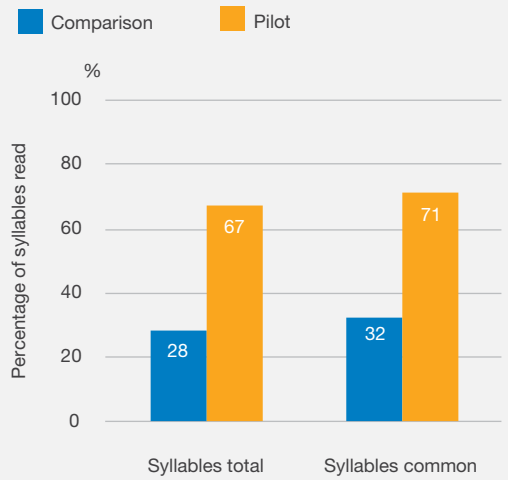
4.2 Comparison of Grade 1 in 2014 and Grade 1 in 2012 (first year pilot)

Figure 4 and **Figure 5** compare the result of the 2012 Grade 1 cohort (in Grade 3 at the time of the study) with that of the 2014 Grade 1 cohort. It seems fair to presume that after three years of the programme trial, those in Grade 1 in 2014 will read better than those in Grade 1 in 2012 and this appears to be the case for most of the NL groups.

4.3 Comparison with control schools and the national sample

At the point of comparison, we examine the results of syllable reading, English connected text reading and reading comprehension. These subtests consist of items that are identical and are considered fair for both the ELINL and non-ELINL pupils (i.e. pupils learning the same skills in other reading programmes). **Figure 6** shows that the ELINL pupils

Figure 6. Pupil performance on syllable reading, by pilot groups (in %)



Note: * scores weighted. ** 'syllables_common' are the ten commonly used syllables across the NL identified by the NL experts.
Source: Hsieh, 2014

Figure 7. Pupil performance on syllable reading, by grade and pilot group (in %)

Source: Hsieh, 2014

from all grades combined were able to read on average more than twice the number of syllables.

We know from the 2011 baseline that hardly any children were able to read at syllable or word level when they began Grade 1. **Figure 7** shows a result that is consistent with the 2012 post-test outcome—after approximately 7 months of exposure to ELINL lessons, the ELINL pupils had already acquired noteworthy gains in syllable reading.

Figure 8 shows the average number of English words read per minute by ELINL pupils and the comparison group pupils. The ELINL pupils read with a greater fluency in English connected text across all grades. The difference was especially notable by the time they reached Grade 3.

Figure 9 shows average reading comprehension scores in percentage for ELINL pupils and the comparison group pupils. The former achieved higher reading comprehension scores across all grades. There were also fewer children who scored zero in the ELINL group.

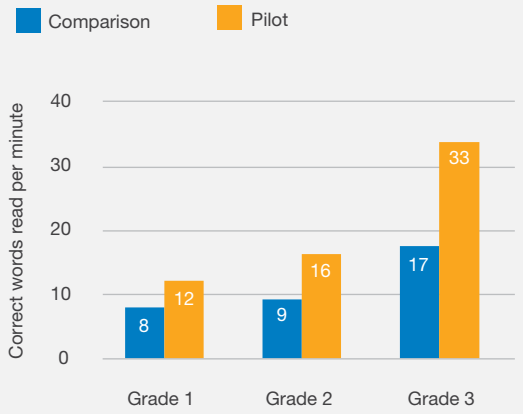
4.5 Comparison with the national sample (the EGRA 2013 results)

The incorporation of the EGRA 2013 connected text and reading comprehension questions allowed comparison with the EGRA national sample. Although not assuming complete comparability due to the necessary difference in sample coverage and possible difference in demographics, the comparison should at least be indicative of the impact of the ELINL programme.

Figure 10 presents the oral reading fluency in English achieved by 1) the total sample of the 2013 EGRA pupils; 2) government and government-aided school pupils in the 2013 EGRA; and 3) the ELINL pupils assessed in 2014 (government and government-aided schools). The result shows that pupils in the ELINL programme read with significantly better fluency when compared to the national sample.

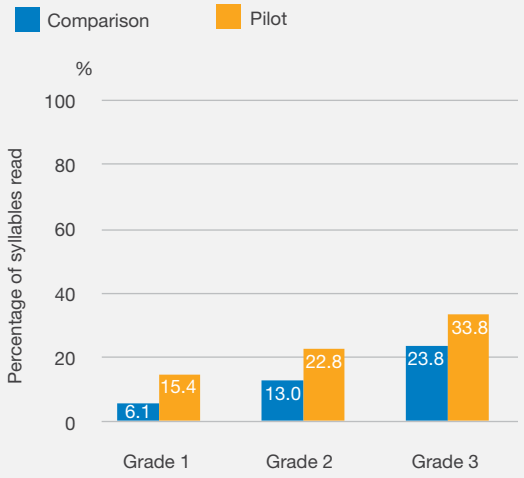
Table 2 presents the score of the English reading comprehension questions (5 points in total) achieved by the EGRA and ELINL pupils. Although the percentages of children who did not get any of the comprehension questions correct were high in

Figure 8. Oral reading fluency in English, by pilot group and grade



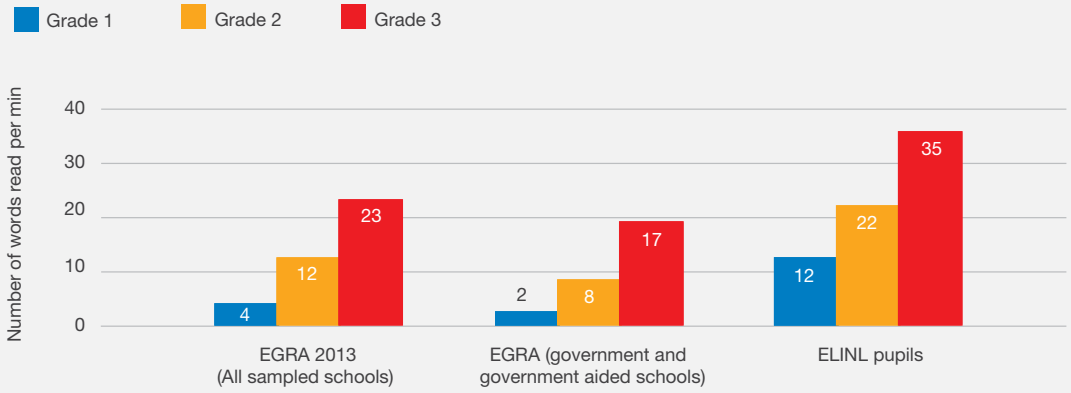
Source: Hsieh, 2014.

Figure 9. English reading comprehension, by pilot group and grade (in %)



Source: Hsieh, 2014.

Figure 10. Pupil performance on English connected text reading, EGRA and ELINL comparison



Source: UNESCO Institute for Statistics

TABLE **2**

Pupil performance in English reading comprehension questions (EGRA and ELINL sample comparison)

Series	Grade 1 EGRA	Grade 1 ELINL	Grade 2 EGRA	Grade 2 ELINL	Grade 3 EGRA	Grade 3 ELINL
0 items correct	88.6%	78.3%	70.5%	52.0%	49.5%	37.8%
1 item correct	8.2%	3.6%	15.1%	13.1%	19.8%	9.4%
2 items correct	1.7%	2.4%	8.0%	9.4%	13.4%	11.3%
3 items correct	0.7%	1.0%	3.0%	4.3%	7.1%	11.6%
4 items correct	0.5%	0.1%	2.3%	4.6%	3.6%	9.1%
5 items correct	0.3%	14.6%	1.1%	16.7%	6.5%	20.8%

Source: Hsieh, 2014

both the EGRA and ELINL groups, there were fewer pupils scoring zero in Grades 1, 2 and 3 in the ELINL sample. It is also worth noting the difference in the percentage of pupils achieving full or almost full comprehension in connected text reading (getting all 5 questions correct) between the two groups.

Qualitative data was captured among a total of 291 teachers, 82 coaches and 85 head teachers who completed the semi-structured questionnaire. More than 15% of the teachers, coaches and head teachers were absent for workshops, confinement leaves, illness or bereavements. There were also a noteworthy number of ELINL teachers and coaches who no longer served in their originally designated schools (due to transfer, sickness or death) but had not been replaced at the time of the survey, posing a threat to the effectiveness of the programme's implementation. The information gathered through the questionnaires pointed to challenges in the ELINL implementation. Exposure time to the programme also varied as a result of different start dates, lesson times and lesson frequency per week, especially for classes in Grade 1. Although there were delays in the distribution of teaching and learning materials and in teacher training, in reality head teachers' knowledge and support of the programme as well as trained teachers' capacity to teach the NL determined students' exposure to the ELINL.

In the questionnaire, teachers were asked to comment on their familiarity and confidence in teaching the NL lessons. They were also asked to provide two summary lesson plans to illustrate how they would organize their NL lessons on a daily basis. The outcomes, which coincide with the desk officers' reports, teacher feedback and general observations, showed that many teachers had trouble mastering the NL orthography while some were still not confident with teaching NL lessons. Teachers were also asked whether they were able to give every child the opportunity to read aloud the new material taught in their lessons. Nearly 90% reported that they did although many teachers also commented that the limited lesson time and class size made it difficult to give every pupil an opportunity to try.

Comparison of results from the coaches' and teachers' questionnaires showed discrepancies between the reported ELINL class implementation. Many of the coaches had trouble responding to the questions regarding lesson time, lesson coverage, teacher capacity and pupil opportunities and performance. The outcome reflects the view of INSET and the head teachers that further training is required to prepare the school-based coaches to support and enhance the teaching and learning of NLs. The result was not too surprising considering the limited training that some of the teachers and coaches had received at the time of the survey.

Interestingly, the point in time when schools started to teach NL lessons in the 2013 academic year (some started in the first month and some only in the second term), the length of the daily lesson and the number of times teachers were trained by INSET appears to have not made an impact on the performance of the NL classes. Difference in teacher and coach professionalism is nonetheless significant between the best and worst performing groups (p-value < 0.001). Teachers in better performing schools were more familiar with the scripted lessons, conducted informal formative assessment more regularly and were more confident in teaching NLs. They were better at describing strategies that can help pupils with reading difficulties and demonstrated the ability to integrate the key instructional steps in the NL lessons. In addition to being present more often, their NL classes were daily and more regular. It is also worth noting that more teachers in the better performing schools were exposed to at least one other phonics programme, presumably giving them more opportunities to acquire a knowledge of phonics.

Coaches in the better performing schools were also more confident with the use of scripted lessons and their capacity to provide useful feedback to NL teachers. These are likely the reason why teachers in the better performing schools received more support from their school-based coaches as their

NL lessons were more frequently observed and they were provided with suggestions to improve teaching.

5. DISCUSSION AND FUTURE PLANS

The results of the ELINL third year evaluation suggest an advantage to adopting the ELINL programme. Students in the pilot group performed better on the reading tasks than those in the comparison group. The ELINL pupils also performed better in English reading and comprehension when compared to the EGRA national sample. Within the ELINL programme, learning outcomes varied among the national language groups and among Year I, II and III cohorts. Some differences may have been due to pre-school exposure to reading and pupils' socio-economic backgrounds while others were most likely attributable to the programme's implementation. The anticipated learning goals set out for the ELINL programme were not fully achieved. Nevertheless, when compared to the 2012 Grade 1 cohort, significantly more children in the 2014 Grade 1 cohort were able to achieve these pre-defined goals.

How to help children to read with comprehension? remains a question that requires immediate attention in The Gambia. While many ELINL pupils demonstrated a good grasp of foundational reading skills (i.e. letter, syllable and word reading), this did not always lead to the processing and understanding of the NLs and English texts. In all NL groups, many pupils read with good fluency but still had difficulty answering the comprehension questions. The four rounds of EGRAs reveal the same tendency in the national sample.

Support for the phonics approach versus the whole reading/language approach has waxed and waned through much of the twentieth century, and as the pendulum swings, each approach to teaching reading has taken its turn to dominate. The increasingly widespread view and practice are that each approach has a different but potentially complementary role to play in the effective teaching of reading. Many maintain the view that "phonics instruction, to be effective in promoting independence in reading, must be embedded in the context of a whole reading programme" (International Reading Association, 1997; Rayner et al., 2002). In The Gambia, much focus has been placed on the phonics approach mainly because of the introduction of the EGRA and the various early literature programmes. While systematic phonics instruction is known to be beneficial to beginning readers, the government could also explore the potential of combining literature-based instruction with phonics, which has been proven by many studies to be more powerful than either method alone (Stanovich and Stanovich, 1995; Gee, 1999). It would also be ideal for the reading programmes to be better integrated with the national curriculum and learning achievement targets.

There were varying degrees of implementation barriers among NL groups and individual schools. Student attrition, regularity of the NL lessons, teacher and coach professionalism and teacher movements require particular attention in the programme management at the school level. While a number of aspects can be further improved to ensure the quality of the programme's implementation, teacher capacity (i.e. professionalism and professional competence) was almost the single most important factor to student learning outcomes. It would be necessary for teachers to master the basic NL orthographies and the scripted lessons. While we are not arguing that these are the only steps required for teaching children to read or that teachers have to follow scripted lessons at all times, these are means to ensure that lessons are purposeful and structured. In the implementation of an early reading programme, it is important for the government to ensure continuous and sustainable training with quality assurance mechanisms in place. At some point, the focus of the training should be shifted from teaching basic orthographies to how to address the learning needs of the student, and to help students read with fluency, accuracy and comprehension.

In introducing NL programmes, there is also the need for continuous sensitisation to the purpose of NL learning. The high student attrition rate in some of the ELINL classes was due to a misunderstanding of NL teaching and learning. The target group for sensitisation should include parents and communities as well as teachers and head teachers who might be unclear about the purpose of the programme.

Based on the findings of the 2014 study, the MOBSE with approval from the general public has developed a work plan and road map to scale up the ELINL programme to cover all public lower basic schools in the country in September 2015. The scaling up has also integrated the other two early literacy programmes in The Gambia through collaborations between the three parties in charge. Student textbooks, scripted lessons and guidebooks for training, assessment and monitoring have been developed for this purpose. By August 2015, around 2,800 primary school teachers had been trained to teach the new integrated reading strategy through a cascade training arrangement. A synchronised timetable for Grades 1-3 has been implemented across the country, allocating double periods of reading in NLs in the first periods of the school timetable each day followed by double periods of reading in English each day. This is meant to give both teachers and pupils the opportunity to reinforce and apply skills learnt in the NL sessions in the English classrooms.

Learning to read in one's mother tongue is very challenging for multilingual countries with diversified local languages. Limited teaching and learning materials (if in existence at all), popular preference for the government-recognised lingua franca and lack of experienced teachers are major obstacles for governments wishing to introduce or maintain the teaching of mother tongues. The government in The Gambia has made an informed decision to introduce the NL programme with a very short preparation time and through a 'learning-by-doing' process. While many aspects in the implementation can be further improved, it has been a remarkable journey that has brought valuable lessons to The Gambia and to countries wishing to adopt similar programmes.

REFERENCES

Abu-Rabia, S. and Taha, H. (2006) "Reading in Arabic Orthography". Malatesha Josh, R. and Aaron, P. G. (eds.), *Handbook of Orthography and Literacy*. Florida: Taylor & Francis.

Bird S. (2001). *Orthography and Identity in Cameroon*. Paper presented at the 96th Annual Meeting of the American Anthropological Association, Washington, November 1997. http://cogprints.org/1446/5/identity.pdf (Accessed June 3, 2015).

Gee, J. P. (1999) "Critical Issues: Reading and The New Literacy Studies: Reframing the National Academy of Sciences Report on Reading". *Journal of Literacy Research*, Vol. 31, No. 3, pp.355-374.

Goldsmith, J. A. (ed) (1995). *The Handbook of Phonological Theory, Blackwell Handbooks in Linguistics*. Oxford: Blackwell.

Hsieh, P. T. J. (2014) *Year III Study of ELINL: Pupil achievement, teacher capacity and programme implementation.* Evaluation report for the government of The Gambia and The World Bank. Washington D.C.: The World Bank.

International Reading Association (1997). *The Role of Phonics in Reading Instructions: A Position Statement of IRA*. Newark, DE, USA: International Reading Association.

The Gambia (2011). *Report on Early Grade Reading Ability Assessment 2011.* The Gambia: Ministry of Basic and Secondary Education.

The Gambia (2012). *Result of baseline teacher assessment for in-service teacher training.* The Gambia: In-service teacher training unit, Ministry of Basic and Secondary Education.

Rayner, K., Foorman, B. Perfetti, C. A., Pesetsky, D. and Seidenberg, M. S. (2002) "How should reading be taught?" *Scientific American*, Vol. 286, pp. 84-91.

Research Triangle Institute International (2008). *The Gambia Early Grade Reading Assessment (EGRA): Results from 1,200 Gambian Primary Students Learning to Read in English*. Report for the World Bank. Research Triangle Park, NC: RTI International.

Stanovich, K. E. and Stanovich, P. J. (1995). "How research might inform the debate about early reading acquisition". *Journal of Research in Reading*, Vol. 18, No. 2, pp. 87-105.

The World Bank (2015). *Data for The Gambia*. http://data.worldbank.org/country/gambia. (Accessed May 23 2015)

Using Literacy Boost to Inform a Global, Household-Based Measure of Children's Reading Skills

MANUEL CARDOSO
UNICEF

AMY JO DOWD
Save the Children

ABBREVIATIONS

ASER	Annual Status of Education Report
DIBELS	Dynamic Indicators of Basic Early Literacy Skill
EGMA	Early Grade Math Assessment
EGRA	Early Grade Reading Assessment
GMR	Global Monitoring Report
GPE	Global Partnership for Education
MICS	Multiple Indicator Cluster Survey
UNICEF	United Nations Children's Fund
RTI	Research Triangle Institute
RWC	Reading with comprehension
SDG	Sustainable Development Goal
UIS	UNESCO Institute for Statistics
WCPM	Words correct per minute

1. INTRODUCTION

The need for a survey of early reading and numeracy skills such as the one proposed here with the Multiple Indicator Cluster Survey (MICS) as its platform, stems from UNICEF's dual focus on learning and equity. It is no longer enough (it probably never was) to focus solely on access to education: we must make sure that children are actually learning and developing the skills necessary for further academic learning. Therefore, development of these skills should be monitored.

Equity is just as much of a priority. Thus, when evaluating the state of learning among a population of children, those who are not currently in school should also be taken into account. As a household survey, the MICS offers a snapshot of learning among all children, including those who are currently attending mainstream schools but also those in other types of schools or non-formal education, those who are not currently attending school, those who have dropped out and those who have never attended any form of educational centre. While admittedly partial and imperfect, this *direct* measurement of reading and numeracy skills (as opposed to a proxy based on educational attainment or other characteristics) represents a huge contribution for both learning and equity.

Across several meetings, members of the Annual Status of Education Report (ASER), Global Partnership for Education (GPE), Research Triangle Institute (RTI), Save the Children, UNESCO Institute for Statistics (UIS), Global Monitoring Report (GMR), World Bank and UNICEF MICS and Education discussed several options for capturing children's reading and numeracy skills during the implementation of the MICS. As this household survey is administered in more than 50 countries across the globe, the inclusion of this module offers an important lens into academic learning and the Sustainable Development Goal for Education (SDG 4) in many settings.

At meetings in December 2014 and June 2015, the group worked through the strengths, weaknesses and challenges of several options for a reading skill assessment that could be administered in a time frame of about two minutes. The resulting reading recommendation for piloting closely resembled

current Save the Children practice. For this reason, UNICEF and Save the Children embarked on a further collaboration to investigate how well a briefer administration of that practice would capture the information that UNICEF hoped to gain. While the group also took up the assessment of basic numeracy[1], this paper focuses on the reading recommendation and further analyses to inform its development for the MICS.

This article explains the methodology for measuring early reading skills among 7-14 year olds in the MICS, including which types of constructs are targeted (oral reading accuracy; reading comprehension: literal and inferential) or not targeted (oral reading fluency; foundational skills, or other types of comprehension), as well as a discussion of the number of comprehension questions for each comprehension skill.

Although the theoretical framework is based on a "simple view of reading" and subsequent reformulations that emphasise fluency, the implications of this theoretical foundation will be balanced against practical considerations. These considerations include the process of data production in a household setting, targeting children as respondents (as opposed to an adult proxy, a first for the MICS) and the expected use of the data, which will be aggregated across different languages within each MICS country (many of which are multilingual) to produce national indicators.

The analytical methods focus on the secondary analysis of existing school-based assessments of reading at Grade 2 and 3. The statistical techniques compare the results of different comprehension measures to each other as well as to the results of a measure of oral reading accuracy.

The data sources are a set of assessments from Save the Children's Literacy Boost programme, which aims to improve children's reading achievement. The data comes from Bangladesh, Burundi, India, Kenya, Lao People's Democratic Republic (PDR), Philippines and Vietnam.

The results show that streamlined versions of the comprehension measure (2-3 questions) yield similar results as longer versions (8-10 questions), while improving parsimony and feasibility. The relationship with accuracy is consistent across different versions of the comprehension measure, whether they are streamlined versions based only on two or three comprehension questions (mostly literal) or longer versions with eight to ten questions of different types. Based on analysis of the data and the time constraints imposed by the MICS, it is recommended that at least two literal comprehension questions are used—instead of just one—alongside an inferential question.

2. A FOCUS ON COMPREHENSION

There is widespread consensus for the notion that comprehension is the ultimate goal of reading. However, there is less agreement on everything else about it, including the importance of different types of comprehension, how to measure them and how to combine them in a meaningful overarching indicator to monitor progress across different grades and ages. A widespread—if not universally accepted—framework for the study of reading comprehension postulates three increasingly complex levels of comprehension: literal, inferential and evaluative (Basaraba et al., 2013). This section, which borrows heavily from Basaraba et al. will describe those three levels and explain why the first two will be included in the MICS's reading tasks, and also why the third one will not be included.

Literal comprehension tasks require a reader to retrieve information from a passage. This is "the focus of the skills and strategies initially introduced to all readers, especially in the primary grades, when they are being taught to read with understanding" (Basaraba et al., 2013). The task has a low level of cognitive demand but a central role in early reading instruction. Therefore, it will be piloted in the MICS as the first comprehension task.

1 A parallel collaboration between the RTI and UNICEF, based on the Early Grade Math Assessment (EGMA), focuses on the numeracy component.

This type of comprehension, however, is not enough. It is the simplest comprehension skill on which more advanced skills depend. Inferential comprehension tasks are also essential and indicative of greater growth. These "require readers to understand relationships that may not be explicitly stated in the passage but are essential for passage understanding, such as the connection between two events in a narrative or understanding a character's motive for a particular action" (Basaraba et al., 2013). At the same time, a text-connecting inference (Baker and Stein, 1978; Williams, 2015) cannot be completed without literal comprehension. An inferential comprehension task in the MICS offers an opportunity to test whether the child is able to go one step beyond the mere retrieval of information stated explicitly to connect facts in the text in order to answer questions. This is why inferential comprehension is included in the MICS as the second comprehension task.

Finally, while "evaluative comprehension tasks require readers to analyse and critically interpret the text based on their prior knowledge and experiences" (Basaraba et al., 2013), such questions pose two key challenges in the specific context of the MICS. First, the task may be more common in school settings, biasing results for children not in school. Second, the variability of acceptable answers to such questions that draw from readers' prior knowledge and experiences among a diverse population of children in terms of both age and grade invites a range of responses so wide that it poses problems for scoring. This is especially complicated in the context of a household-based survey, where most interviewers may not have a background in teaching or assessment. For these reasons, there will be no evaluative comprehension questions in the MICS learning module.

In summary, the MICS will focus on literal and inferential comprehension—the first two of three increasingly complex levels of comprehension—and not include, at least for MICS 6[2], tasks related to evaluative comprehension (the third level). Although we recognise the importance of all these different types of comprehension tasks, the decision to focus on the first two stems mostly from interview time constraints and considerations regarding scoring in the field.

3. DECODING ACCURACY AS ONE OF THE PILLARS OF COMPREHENSION

If comprehension is the universally acknowledged goal of reading, the ability to decode print accurately is widely understood to be a prerequisite and a fairly reliable predictor of comprehension. However, once

2 The sixth round of the MICS will start in late 2016 and will finish in 2018-2019. As in previous rounds, it is expected to include between 50 and 60 countries, mostly in sub-Saharan Africa but also from other regions. http://mics.unicef.org/

again, as with comprehension itself, the consensus ends there. First, there is little agreement on the relative importance of accuracy as compared to other factors, such as oral language skills or fluency, especially across languages. Second, there are several different ways of measuring decoding accuracy and this brief section will explain the practical considerations leading the MICS to focus on oral reading accuracy in connected text.

Three methodological decisions must be made regarding the measurement of decoding accuracy: oral or silent reading, real words or non-words, and isolated or connected text. First, a child can process print **either orally or silently** using a variety of approaches. However, decoding accuracy is much easier to monitor orally, especially in the field setting of the MICS.

Second, decoding skills (sometimes in combination with other skills or knowledge, depending on the situation) allow readers to process text by using their graphophonemic knowledge (Garcia et al., 2013). Either **actual words or non-words** can be used to assess decoding accuracy (ibid). Some scholars propose the use of non-words, arguing that knowledge of letter sound correspondences is difficult to disentangle from word recognition when actual words are used since readers partly rely on their lexical knowledge (i.e. their vocabulary skills) to identify a word as they read it (ibid). Therefore, decoding of non-words can be seen as a more valid indicator of decoding skills as it is not susceptible to lexical knowledge. Conversely, reading non-words is not a common task for children or adults. As a result, if non-words were to be used, more detailed instructions and practice items should be included, which increases interview time. In practice, it is still possible that some respondents would still be confused by the use of these non-words or fail to understand the purpose of the task. Therefore, a decision has been made to focus on the decoding of real words rather than non-words.

Finally, words can be read either **in isolation or as connected text** (e.g. a story) (Grigorenko et al., 2008). Reading words in isolation may provide a more valid measure of a reader's decoding skills than connected text because a reader faced with connected text can rely on grammar, sight words and their prior knowledge of the passage's topic to decode it. This reduces the effectiveness of oral reading accuracy as an indicator of print decoding skills. Using a story as the stimulus for the oral reading accuracy task, however, has an obvious practical advantage: the same story can be used as the stimulus for the comprehension questions as well. Admittedly, this approach also has a disadvantage: the measurement of the decoding skill is not independent from the measurement of the comprehension skills and, as such, this is likely to overestimate the statistical association between these two measures. However, estimating associations between the different skills is not the main purpose of this survey.

4. WHY THE MICS IS NOT MEASURING READING FLUENCY[3]

Oral reading fluency is generally regarded as a good predictor of reading comprehension, especially in early grades (Roehrig et al., 2008; Hussien, 2014). Although it is unclear whether oral or silent reading fluency is the better predictor of comprehension, there are obvious practical difficulties involved in the measurement of silent reading fluency as opposed to its oral counterpart, especially in field situations. There are currently two main approaches to the measurement of oral reading fluency in the field in developing countries. For lack of a better term, we will refer to them as the quantitative and qualitative approaches. This section will describe these two approaches, consider their advantages and disadvantages and ultimately, present the rationale for not including reading fluency as a construct to be measured in the MICS.

The quantitative approach measures fluency as a speed rate, typically in words correct per minute (WCPM). This measure operationalises two of the

3 This section does not intend to present all possible ways of measuring fluency or all arguments for and against doing so as part of this type of endeavor. As the section's title indicates, it merely states the rationale for not measuring fluency in this specific project.

three elements in the commonly accepted definition of fluency: reading accurately, quickly and with expression. This last component—expression—is the one missing here and in many quantitative treatments of fluency as speed. WCPM has been widely used in the US in English and Spanish—for instance, by the Dynamic Indicators of Basic Early Literacy Skill (DIBELS) (Good and Kaminski, 2002) and across the globe by the Early Grade Reading Assessment (EGRA). Its advocates have emphasised that its results do not depend too heavily on the specific text used as stimulus, provided that it is appropriate for the target grade or age. WCPM can also be sensitive to intervention, which may be regarded as an advantage in the context of monitoring and evaluating reading programmes—provided that it actually has an effect on comprehension.

This indicator has been used in a variety of languages, both within the same country or across countries (Abadzi, 2011, 2012; Abadzi et al., 2014; Gove et al., 2011, 2013; Jiménez et al., 2014). However, most studies have generally refrained from drawing direct comparisons across languages because word length and orthographic complexity vary (Piper et al., 2015). This makes WCPM comparisons across languages difficult and poses a serious problem for the use of this indicator in the MICS, where indicators are typically disseminated for each country regardless of language.

Another challenge with the quantitative approach lies in the operational complexities involved in its measurement in the field, especially around timing. This method requires the interviewer to operate a timer or stopwatch while marking the words read incorrectly by the child. Then the interviewer must mark the last word read by the examinee when the allocated time (typically one minute) has elapsed or, if the examinee happens to finish reading the passage before the end of the allocated time, the interviewer must record the exact time elapsed. The reliable administration of this measure routinely requires extensive training as well as supervised practice in the field (see article by Dowd et al.). For this reason, when it comes to oral assessments of reading among children, this timed approach has been confined mostly to school settings using specifically trained assessors.

In household settings, the Literacy Assessment and Monitoring Programme (LAMP) implemented by the UIS has used a combined timed and untimed reading tasks approach for its Reading Components instrument. However, it does not target children but adults (aged 15 years or older)—a fact that may reduce the operational challenges. Moreover, the MICS does not have logistical space for both types of tasks. On the other hand, many household-based assessments of children's reading skills (e.g. the ASER, Beekunko, Uwezo) have so far used a qualitative approach. This approach does not require using a timer or stopwatch since it relies on the interviewers' judgement on whether a child reads fluently or haltingly. Although this method poses fewer operational challenges in the field (i.e. interviewers not having to negotiate use of a timer simultaneously while performing other tasks during the assessment), it requires interviewers to be able to appraise a child's fluency, which may require either extensive training or a background in teaching and/or assessment.

In summary, when it comes to oral reading fluency, both cross-language comparability and operational complexity pose challenges for the MICS. The MICS aims to avoid this challenge, focusing on a measure of accuracy instead of fluency—a tally of words read correctly from which a percentage of all words read correctly can be derived.[4]

5. WHAT ABOUT ORAL LANGUAGE SKILLS?

The MICS also does not measure oral language. However, if the two pillars of the "simple view of reading" (the theoretical framework underlying

4 The debate around oral reading fluency as a construct of interest in oral assessments of early reading goes beyond the challenges it poses to measurement in the field or comparability across languages. For instance, some stakeholders have raised issues related to the policy implications of what might be perceived as a focus on fluency per se as opposed to as a proxy or predictor for comprehension. However, that debate goes beyond the scope and ambitions of this article.

© Dana Schmidt/The William and Flora Hewlett Foundation

the MICS) are oral language and decoding skills, why is the MICS not measuring oral language? A systematic survey of oral language skills would require examinees to demonstrate that they both speak and understand the language. Oral comprehension tasks pose operational challenges because ideally they would require a standardised aural stimulus—in other words, a recording that examinees must hear, understand, retain in their memory and on which they answer questions. One alternative to this direct measurement of oral language skills would be to collect information on the children's language skills or exposure, either from the children themselves or from a proxy respondent (typically, a caregiver). This is the approach chosen by the MICS and it includes questions on home language, means of instruction in the classroom (when applicable) and language preferred by the child for the assessment. These questions are admittedly very imperfect proxies of oral language skills and slightly less imperfect proxies of oral language exposure. However, we make the assumption that children who perform the reading tasks in a language that is both their home language and the means of instruction in their classroom will typically have higher oral language exposure to that language—and probably better oral skills in that language—than children for whom the language of the survey is neither their home language nor means of instruction. That said, the development of the oral language skills instruments that can be used jointly with reading assessments is probably overdue. However, that discussion is beyond the scope of this project.

6. THE PROPOSED INDICATORS

The initial reading recommendation from the expert meetings focused on the collection of the percentage of all words in the passage read correctly out loud as well as asking the child one literal comprehension question that entailed retrieving information from the story's first sentence and one inferential comprehension question to which the answer is present in the text. From these data, UNICEF proposes four indicators:

1. The percentage of children reading with accuracy at a specific threshold (to be determined but possibly at 90-95%)
2. The percentage of children who answer one literal comprehension question correctly
3. The percentage of children who answer one inferential comprehension question correctly
4. An early reading skills indicator that is the percentage of children demonstrating mastery of all three tasks.

This final determination of whether each child is a reader with comprehension is the focus of this investigation.

Save the Children's practice also aims to identify whether a child is a reader with comprehension. Their definition, however, is not limited to a two-minute administration and a child has the opportunity to answer ten questions of the following four types: one summary, six literal (first one from the first part of the story), two inferential and one evaluative. A child is considered a reader with comprehension if she/he answers eight or more of the ten comprehension questions correctly.

Important to consider within the scope of the discussion in this article is the overlap of types of questions as well as the existence of numerous datasets that may be used to determine whether the children that Save the Children determines are readers with comprehension are similarly identified when using the proposed shorter UNICEF MICS approach. Of related interest, from a cross-country comparative perspective is the possibility that readers with comprehension read with similar accuracy, regardless of language.

7. RESEARCH QUESTIONS

- How accurately can the proposed two questions (one literal and one inferential) classify a child who reads with comprehension as compared to (or as benchmarked against) a more complex assessment of comprehension?
- At what rate do readers with comprehension read? Does this vary based on how many questions are used to define readers with comprehension?

8. PROCEDURE

This investigation had the following four steps. First, we defined in the existing Save the Children's ten-question datasets which two questions match the UNICEF MICS's proposed approach. In all of them, the second question was the literal target question as it entailed recall of information from the first sentence of the story. The selection between the two inferential questions in the Save the Children data depended on whether the information was contained in the text. If both questions met this criteria, then one was chosen randomly. The second step was to create a reading with comprehension (RWC) classification indicator based on these two questions (RWC2). Applying the MICS's mastery focus, the indicator RWC2 was scored as one if the student answered both questions correctly. Third, a comparison was made between those readers who were classified as readers with comprehension based on ten questions (RWC10) and based on two questions (RWC2). In addition, the use of only literal and inferential questions (seven of the ten questions) as the yardstick for defining a reader with comprehension was also considered. From these seven questions in Save the Children's datasets, an additional indicator for consideration was created: RWC7. Finally, average accuracy for each group of readers with comprehension was calculated.

9. DATA

Data from the seven countries listed in **Table 1** come from Literacy Boost implementation sites. The children and youth range from ages 5 to 15 years with an overall average age of 8.2 years.

In each instance, these data enabled Save the Children to track learning progress and shift programme investments based on evidence of impact and investigation of equity across key target groups. For this reason, it is important to keep in mind that the population from which these samples are drawn is one in which children have relatively fewer resources than peers in other parts of their country or city.

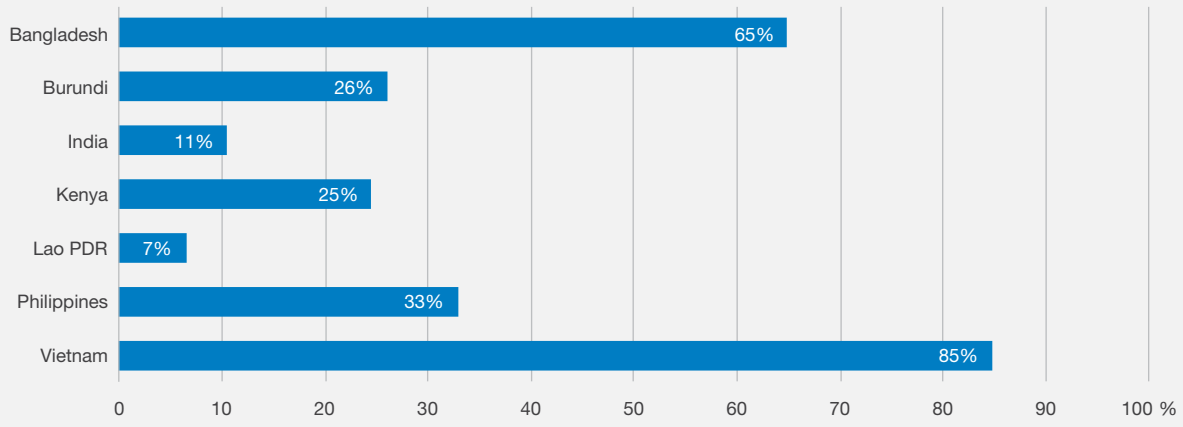
TABLE **1**

Data by country, grade, sample size and age

Country	Grade	n	Age range	Average age	Language
BANGLADESH	3	1,012	7 to 13	8.9	Bangla
BURUNDI	2	549	6 to 15	9.2	Kirundi
INDIA	2	1,159	5 to 15	8.5	Hindi
KENYA	2	1,043	5 to 13	7.7	Kiswahili
LAO PDR	2	714	6 to 15	8	Lao
PHILIPPINES	2	828	6 to 13	8	Filipino
VIETNAM	2	562	6 to 13	7.2	Vietnamese

Source: Save the Children

Figure 1. Percentage of readers in each sample from Table 1



Source: Save the Children

For the purpose of this investigation, it is also important to know that not all students in the sample are readers—defined as those who were able to read the grade-level story out loud. During administration, the child is given a grade-level text and asked to read it aloud to the assessor. If the child struggles with getting started, she/he is encouraged to try. If, however, she/he does not begin reading with some accuracy within half a minute then she/he is thanked for trying her/his best and the assessment ends. **Figure 1** shows the percentage of children who were identified as readers in each sample in Table 1.

10. FINDINGS

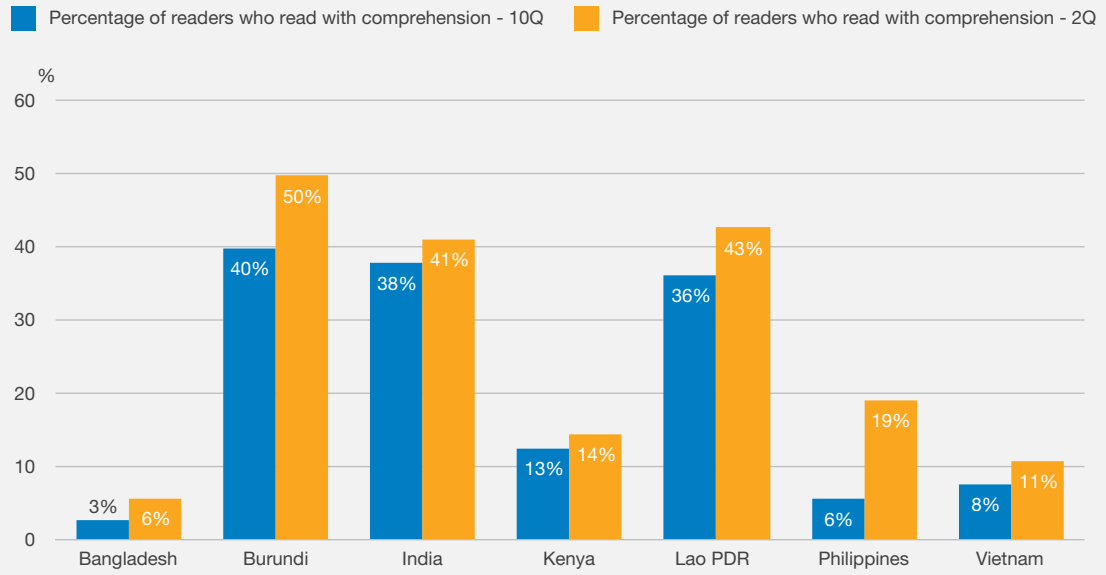
Among these readers, we generated the new RWC2 variable. **Figure 2** presents the comparison of the percentage of readers identified as reading with comprehension using the ten-question approach of Save the Children (in blue columns) to those identified using the proposed MICS two-question approach (in yellow columns).

In the data from each site, there are more students identified as reading with comprehension with the two-question approach than there are by the ten-question approach. From this we surmise that the proposed MICS indicators are more likely to overestimate reading with comprehension as compared to Save the Children's practice.

In terms of agreement, of the children identified as readers with comprehension by Save the Children's ten-question method, 81% are also identified in this way by the proposed two-question approach. However, of the children identified as readers with comprehension by the proposed UNICEF MICS approach, only 56% were also identified as such via Save the Children's ten-question approach. It is clear that Save the Children's approach has a more holistic definition of reading with comprehension. While those who meet this criteria are highly likely to be so identified with the briefer tool proposed by UNICEF, the tool is also likely to include in its group of readers with comprehension those unable to meet the higher bar of fuller comprehension.

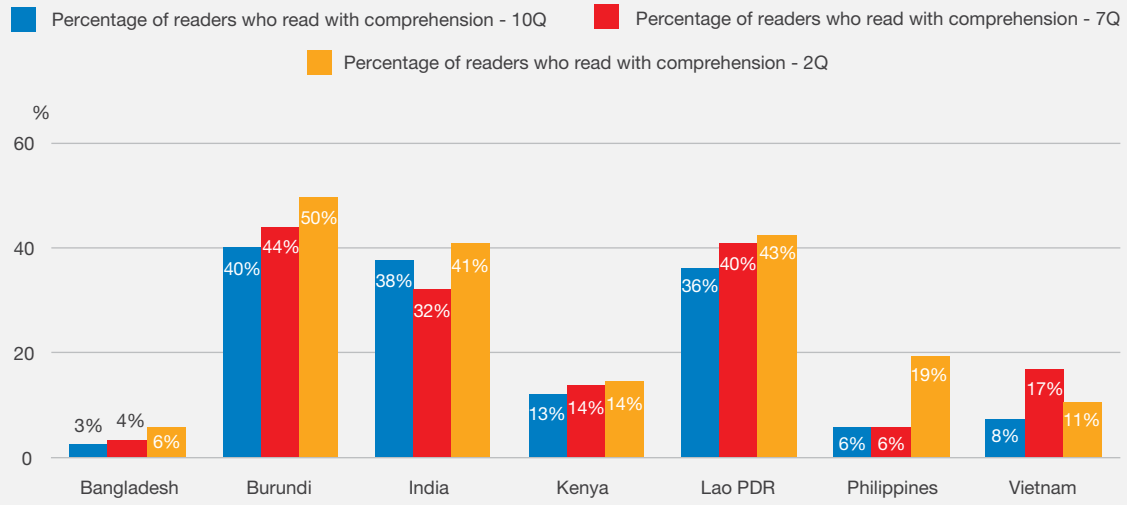
The inclusion of both evaluative and summative questions in Save the Children's reading with comprehension measurement begged a further comparison between the RWC classification based on two questions and the same based on just the seven literal and inferential questions within the Save the Children's assessment. In **Figure 3**, columns to enable this comparison (red) are added to the original set from **Figure 2**.

Figure 2. Readers with comprehension based on 10 and 2 questions, by site



Source: Save the Children

Figure 3. Readers with comprehension based on 10, 7 and 2 questions, by site

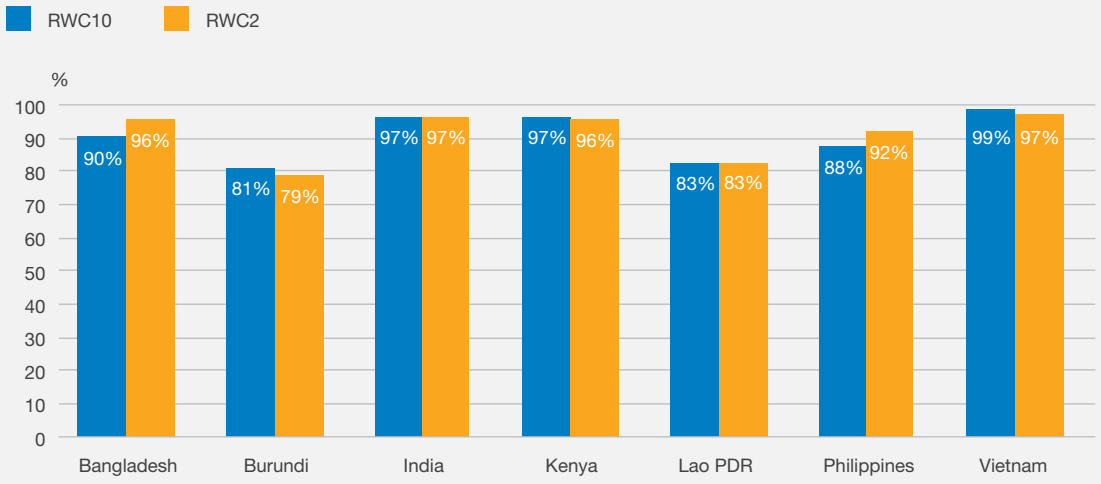


Source: Save the Children

In most sites, there are still more students identified as reading with comprehension via the two-question approach than there are by the seven-question approach. With the exception of Vietnam, overall, the proposed indicator is likely to overestimate reading with comprehension as compared to Save the Children's practice on just literal and inferential questions.

In terms of agreement, of the children identified as readers with comprehension by Save the Children's seven-question method, 78% are also identified in this way by the proposed two-question approach. However, of the children identified as readers with comprehension by the proposed UNICEF MICS approach, only 64% were also identified as such via Save the Children's seven-question approach.

Figure 4. Average accuracy among readers with comprehension, by RWC type and site



Source: Save the Children

The only way to increase the level of agreement is to add additional literal questions to the UNICEF MICS proposed approach. A single literal question added across these sites brought agreement up to 75%.

Finally, considering accuracy in **Figure 4**, the overall average accuracy of readers with comprehension is quite similar within sites regardless of number of questions used to identify them. Across all sites, the average accuracy of readers with comprehension is 91%.

It is clear from the relatively lower columns for Burundi and Lao PDR that there may be some variation in rates across languages and cultures—a possibility to consider when piloting tools for broader use than these small samples.

11. CONCLUSIONS

For most countries with data that can reasonably be compared along these lines, we deem the consistency between the streamlined versions of the comprehension measure (2-3 questions with at least one literal and one inferential) and the more extensive ones (8-10 questions also including summary and evaluative questions) to be acceptable. However, there are important gains in increasing the streamlined version from two to three items (two literal and one inferential).

Children in Grade 2 (or Grade 3 in the case of Bangladesh) who read a grade-appropriate story with comprehension have an average accuracy rate (in reading that same story) of approximately 90%. There is some variation across countries, ranging from approximately 80% in Burundi to approximately 98% in Vietnam. These differences may be due either to inherent characteristics of the orthographies being compared or to differences between the instruments, which strictly speaking were not designed to be comparable across languages. It could also be a combination of both factors. Further exploration of this issue is recommended.

Within a country and language, however, the consistency in the relationship between accuracy and the two measures of comprehension is more than acceptable for our purposes. In fact, six percentage points (in Bangladesh) is the biggest difference in the average accuracy rate of readers with comprehension as defined by the two different criteria. This means that, at least in Grade 2, the relationship between accuracy and comprehension is fairly consistent across several measures of

comprehension. This makes benchmarking in accuracy easier.

As for the issues considered in the introduction regarding the need to measure fluency, different types of comprehension or oral language skills directly (as opposed to through the use of imperfect proxies), it is not our intention to draw any general conclusions about these. The decisions made in these regards are partly based on the constraints imposed on the MICS operation and are not meant to put into question the approaches used by other practitioners in this field.

12. NEXT STEPS

The most immediate next step involves developing general guidelines and specific instruments in at least two languages and testing this approach in the field in two countries. The module will be tested initially on its own in order to determine its feasibility and to improve on the details of its implementation. This pilot will use instruments adapted from the EGRA (for which we thank the RTI and USAID).

The following step would entail conducting studies of concurrent validity of the assessment in the field with the awareness that although this approach has been inspired by the likes of the ASER, the EGRA and Literacy Boost, its purpose is not identical to theirs.

Finally, following the field tests and concurrent validity studies, a rollout in a number of countries involved in the MICS 2016 will require setting up a quality assurance mechanism to ensure that the stories and questions developed by the implementing countries are in accordance to the general guidelines and will generate the type of information that will enable the production of comparable indicators at the country level.

REFERENCES

Abadzi, H. (2011). *Reading fluency measurements in EFA FTI partner countries: Outcomes and improvement prospects*. Global Partnership for Education. GPE World Paper Series on Learning No. 1. Washington D.C.: The World Bank. http://documents.worldbank.org/curated/en/2011/09/18042914/reading-fluency-measurements-efa-fti-partner-countries-outcomes-improvement-prospects

Abdazi, H. (2012). *Developing Cross-Language Metrics for Reading Fluency Measurement: Some Issues and Options*. Global Partnership for Education. GPE World Paper Series on Learning No. 6. Washington D.C.: The World Bank. http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2013/07/26/000356161_20130726155230/Rendered/PDF/797740WP0wpm0e0Box0379789B00PUBLIC0.pdf

Abadzi, H. and Martelli, M. (2014). *Efficient Reading for Arab Students: Implications from Neurocognitive Research*. Paper presented at the World Summit of Innovation in Education (WISE), November 5, 2014, Doha, Qatar.

Baker, L., and Stein, N.L. (1978). *The Development of Prose Comprehension Skills*. Center for the Study of Reading Technical Report No. 102. University of Illinois at Urbana-Champaign: The National Institute of Education.

Basaraba, D., Yovanoff, P., Alonzo, J. and Tindal, G. (2013). "Examining the structure of reading comprehension: do literal, inferential, and evaluative comprehension truly exist?" *Reading and Writing*, Vol. 26, No. 3,pp. 349-379.

García, J. R. and Cain, K. (2013). "Decoding and Reading Comprehension A Meta-Analysis to Identify Which Reader and Assessment Characteristics Influence the Strength of the Relationship in English". *Review of Educational Research*, Vol. 84, No. 1, pp. 74-111.

Good, R. H. and Kaminski, R. A. (2002). *DIBELS Oral Reading Fluency Passages for First through Third Grade.* Technical Report No. 10. Eugene, OR: University of Oregon.

Gove, A., Habib, S., Piper, B. and Ralaingita, W. (2013). "Classroom-up Policy Change: early reading and math assessments at work". *Research in Comparative and International Education*, Vol. 8, No. 3, pp. 373-386.

Gove, A. and Cvelich, P. (2011). *Early Reading: Igniting Education for All. A report by the Early Grade Learning Community of Practice*. Research Triangle Park, NC: Research Triangle Institute. http://www.uis.unesco.org/Education/Documents/early-reading-report_gove_cvelich.pdf

Grigorenko, E. L. and Naples, A. J. (eds) (2008). *Single-word reading: behavioral and biological perspectives*. New York: Taylor and Francis.

Hussien, A. M. (2014). "The indicating factors of oral reading fluency of monolingual and bilingual children in Egypt". *International Education Studies*, Vol. 7, No. 2, p. 75.

Jiménez, J. E., Gove, A., Crouch, L. and Rodríguez, C. (2014). "Internal structure and standardized scores of the Spanish adaptation of the EGRA (Early Grade Reading Assessment) for early reading assessment". *Psicothema*, Vol. 26, No. 4,pp. 531-537.

Piper, B., Schroeder, L. and Trudell, B. (2015). "Oral reading fluency and comprehension in Kenya: reading acquisition in a multilingual environment". *Journal of Research in Reading*, Vol. 00, No. 00, pp. 1-20.

Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F. and Torgesen, J. K. (2008). "Accuracy of the DIBELS oral reading fluency measure for predicting third grade reading comprehension outcomes". *Journal of School Psychology*, Vol. 46, No. 3, pp. 343-366.

Williams, J. C. (2015). "The New Salford Sentence Reading Test (2012) and the Diagnostic Reading Analysis (2008) assess 'inference'– but what forms of inference do they test?" *English in Education*, Vol. 49, No. 1, pp. 25-40.

A Longitudinal Study of Literacy Development in the Early Years of School

MARION MEIERS and JULIETTE MENDELOVITS
Australian Council for Educational Research

ABBREVIATIONS

ACER	Australian Council for Educational Research
IRT	Item Response Theory
LLANS	Longitudinal Literacy and Numeracy Study

1. INTRODUCTION

A longitudinal study of children's literacy and numeracy development through the primary school years was conducted in Australia from 1999 to 2005. The study, called the Longitudinal Literacy and Numeracy Study (LLANS) was undertaken by the Australian Council for Educational Research (ACER). The LLANS was designed to identify patterns of growth in literacy and numeracy achievement. A random sample of Australian students was followed across seven years of primary schooling. The key research question was "what is the nature of literacy and numeracy development amongst Australian school children?"

The literacy component of the LLANS investigated children's development in reading and writing. This article focuses mainly on the reading component of the study in the first three years, which was assessed orally in one-on-one interviews.

1.1 Why a longitudinal study?

Longitudinal studies collect data from a cohort of individuals on multiple occasions over an extended period of time. These studies are challenging to conceptualise, administer and sustain, and require an ongoing commitment of resources. A longitudinal study designed to investigate development in an area of learning makes it possible to study progress over time at the individual level. This is in contrast to the more common cross-sectional studies that collect data from many individuals at one point in time. A longitudinal study can identify patterns of development as well as processes of skills and knowledge formation. The LLANS was conceived to trace the processes of development in the key areas of literacy and numeracy.

2. PLANNING FOR THE LLANS

2.1 The literacy construct

An extensive literature review of recent national and international studies of literacy in the early years of schooling was performed in the planning stages of the study (1998). The literature review revealed the prevalence of a view of literacy as broadly defined. The influential report from the committee established by the US National Academy of Science to investigate the prevention of reading difficulties in young children indicated that adequate reading instruction requires that children use reading to obtain meaning from print; have frequent and intensive opportunities to read; be exposed to frequent, regular spelling-sound relationships; learn about the nature of the alphabetic writing system; and understand the structure of spoken words (Snow et al., 1998).

TABLE **1**

Critical aspects of literacy assessment in the first three years of the LLANS study

1st year of school March	1st year of school November	2nd year of school March	2nd year of school November	3rd year of school May
Reading print in the environment (using photos) (MT)	Comprehension of a picture story book read aloud (MT)	Letters and sounds (PA)	Comprehension of a picture story book read aloud	Comprehension of simple text (MT)
Phonemic awareness (PA)	Writing (WR)	Independent reading of simple text (RF)	Sounding out words (PA)	Spelling (WR)
Book orientation (CP)	Word recognition (PA)	Spelling (WR)	Phonemic awareness (PA)	Segmenting and sounding out words (PA)
Comprehension of a picture story book read aloud (MT)	Independent reading of simple text (RF)	Rhyming words (PA)	Spelling (WR)	Comprehension of a picture story book read aloud (MT)
Concepts of print (CP)	Conventions of print (CP)	Comprehension of a picture story book read aloud (MT)	Independent reading of simple text (RF)	Writing (WR)
		Writing (WR)	Writing (WR)	

Note: codes for the critical aspects are making meaning from text (MT), reading fluency (RF), concepts about print (CP), phonemic awareness (PA) and writing (WR).

For the purposes of studying growth in literacy learning, the following critical aspects of literacy were identified as the focus for data collection (codes are included after each aspect for reference later in this article):

- Making meaning from text (MT)
- Reading fluency (RF)
- Concepts about print (CP)
- Phonemic awareness (PA)
- Writing (WR).

Data relating to the first aspect, making meaning from text, is discussed in this article within the context of the broader definition of literacy.

2.2 Developing the assessment instrument

The assessment tasks were developed through a rigorous process of collaborative work by the ACER item writers and were trial tested in schools. Criteria for the development of the assessment tasks included the following key principles:

- The tasks should be research based, assessing critical aspects of literacy
- The tasks should be built around contexts likely to be familiar to students in the early years of school
- In the first three years of the study, the tasks would be administered during a one-on-one interview, if possible by the student's own teacher
- The tasks should be easy for teachers to administer and should be supported with clear and explicit marking and recording guides
- The tasks should be designed to be administered in a reasonable time, taking into account the attention span of students in their early years of school as well as teachers' workloads.

The task types developed for the first five surveys are shown in **Table 1**.

2.3 Assessment design

The design of the assessment comprised two data collections in the first two years, and one in each subsequent year. This schedule took into account the evidence from many studies that development in literacy (and numeracy) is most rapid at the earliest stages. Two collections in each of the first and second years would allow for more detailed tracking of these developments. **Table 2** outlines the data gathering schedule, including the month in the school year.

TABLE **2**

LLANS data gathering schedule

Month	1st year of school 1999	2nd year of school 2000	3rd year of school 2001	4th year of school 2002	5th year of school 2003	6th year of school 2004	7th year of school 2005
March	Survey 1	Survey 3					
May			Survey 5	Survey 6	Survey 7	Survey 8	Survey 9
November	Survey 2	Survey 4					

Note: this is a southern hemisphere school year, starting in February and finishing in December.

2.4 Method of data analysis

The data generated in the study were intended to help reveal the typical progression in literacy learning and enable the identification of changes over time in what students know and can do.

The method of analysis chosen needed to allow for the calibration of all literacy items from all of the surveys on the same scale, in order of increasing difficulty, thus showing how literacy skills and knowledge developed in sequence. In addition, the analysis should be capable of showing the literacy achievement of students on the same scale, thus indicating what skills, knowledge and understanding students demonstrated at each data collection point. Such a scale would make it possible, for example, to track the progress of an individual student over time or to compare the relative achievement levels of particular cohorts of students at each stage of schooling. Rasch analysis, a subset of item response theory (IRT) which maps item difficulty and student achievement on a single scale, was the ideal methodology for these research purposes.

3. IMPLEMENTING THE STUDY

3.1 Selecting and retaining the sample

One thousand children drawn from a random Australia-wide sample of 100 schools (selected in proportion to the population size of each state and territory) formed the original cohort for the LLANS. The sample was drawn from children who commenced school in 1999. One hundred schools were selected from the ACER Australian sampling frame as a random sample. With the approval of the relevant education authorities, the principals of these schools were invited to participate in the study. If a principal was unable to commit the school to participation in a seven-year study, a replacement school from the sample drawn from the ACER national sampling frame was approached (Meiers, 2006).

At the beginning of the 1999 school year, 10 students were randomly selected from class lists provided to ACER by the 100 schools, and approval was obtained from the parents of these children for their participation in the study. This created a total sample of 1,000 students for the study.

This sample was designed to be large enough to allow for possible attrition over the seven years of the study. Throughout these years, close relationships were maintained between the schools and the ACER research team. Schools were kept informed about the progress of the study and the assessment schedules, and full details were sent to all schools each year—this helped ensure that assessment continued to stay on track despite any changes in school administration. Reports detailing the achievements of participating students were sent to schools at intervals throughout the study.

During the course of the study, many students transferred to other schools. Where possible, principals of the transferring students let the research team know the school to which the student was moving. This information was not always available and the total cohort did become smaller each year as the study progressed. On the other hand, the movement of students meant that the number of participating schools increased. Almost all students in the sample (980) completed the literacy assessments in their first year at school.

In 2002, when the students were in Year 3, there were 154 schools and 559 students in the study. In the final year of the study there were 160 schools participating and 413 students who completed the assessment process.

3.2 Collecting literacy data

The assessment tasks used hands-on activities and authentic texts (e.g. high quality children's picture storybooks). Using information about the difficulty of the tasks that had been ascertained from trial testing, the tasks were arranged in a series of assessment forms with the easiest tasks assigned to the first administrations and successively more difficult tasks to later administrations. Common items were used in two adjacent forms to allow linking of all the literacy tasks together in analysis through common item equating.

All materials required for the assessments were provided to schools to ensure that the tasks would be administered in a standardised manner across schools. These materials included simple reading books and full colour picture storybooks. Each participating school was sent a package containing all resource materials and assessment booklets. All costs for delivery and return of the materials and for the materials themselves were covered by the LLANS project.

The assessment booklets included all instructions for administrating the tasks and marking guides for each question with space for recording five students' responses in each assessment booklet. A sufficient number of booklets were sent for recording responses to accommodate all the students selected in the school's random sample.

The full colour picture storybooks and simple books for independent reading were included in the package together with any worksheets. Where it was known that the random sample included students from more than one class in a school, resources were sent for each class.

The package also provided clear written instructions for the paid return of completed assessment booklets to ACER for data entry. A date for the final return of materials was provided.

The classroom teacher administered the assessments to the children in her class during one-on-one interviews that took approximately half an hour. In order to achieve as much commonality in administration as possible, detailed guidance on the administration of the interviews was provided with the assessment materials each year.

3.3 Collecting background information

A questionnaire seeking key information about the students was sent to all schools in 1999. Seven hundred and sixty responses were received. The missing data limited the scope of the analyses that could be carried out. Further student background information was collected in the final year of the study, 2005, from a large proportion of the students who remained in the study. The student background information collected in 1999 indicated that among the sample of 760 returns, 53% were female and 47% were male. Four per cent were of Aboriginal and Torres Strait Islander background. Nine per cent spoke a language other than English at home.

4. ANALYSING THE DATA AND CONSTRUCTING THE LITERACY SCALE

Data collected in the first five LLANS assessments provided information needed for the calibration of the LLANS items and the LLANS instruments. Data collected from the children's responses to the assessment items were used to create achievement scales for literacy and numeracy, respectively. A Rasch IRT model was used for this purpose (Rasch, 1980; Masters, 1982). At the end of the first three years of the study, the items used in the five surveys were mapped onto a single literacy scale and the students' achievement on each occasion was also mapped onto the same scale. As mentioned above, the equating of the tasks was made possible by embedding common items in the assessment instruments.

4.1 What were the findings of the study?

The data collection and analysis of the first five surveys allowed for a mapping of the process of development in literacy learning. A described literacy scale was constructed from the data collected in the study, covering the full range of proficiency of the students in the sample and showing typical progression through stages of literacy learning in the first three years of school.

Figure 1 shows the map of reading and writing development during the first three years of school based on the LLANS study.

The Literacy Progress Map in Figure 1 is a scale of developing literacy achievement. The skill descriptions on the left side of the figure have been selected from the five surveys conducted in the first three years of school. Each description refers to a single item. The skills at the top of the list

Figure 1. LLANS literacy progress map: The first three years of school

Descriptors of skills assessed

Spells correctly some familiar and unfamiliar words eg whiskers.(WR)

Uses and controls a variety of common punctuation in own writing. (WR)
Interprets meaning of a passage of a narrative read aloud. (MT)
Writes a well connected piece showing a recognisable structure eg narrative, recount. (WR)
Writes simple sentences joined with simple conjunctions eg like, but, then. (WR)
Spells some common words with irregular patterns eg basket. (WR)
Identifies key events in a story after listening to a picture story book. (MT)
Pronounces correctly words that require blending of at least 3 syllables. (PA)

Uses a full stop and capital letter appropriately when writing a sentence. (WR)
Explains character's actions in a simple reading book read independently. (MT)
Reads simple reading book (repetitive structure, varied content) with word for word accuracy. (RF)
Explains the overall message of a simple fable. (MT)

Infers information from obvious clues in a simple reading book. (MT)
Explains explicitly stated ideas in a simple reading book. (MT)
Reads simple common words correctly from labels on chart. (RF)
Matches text and meaning accurately in a simple reading book. (MT)

Locates relevant information after listening to an information text read aloud. (MT)
Makes a direct link to meaning of text after viewing illustration in a picture story book. (MT)

Writes one or more simple sentences in response to a task. (WR)
Manipulates sounds in words eg swaps c in camp with l to make lamp. (PA)
Retells key aspects after listening to a picture story book. (MT)
Identifies a capital letter correctly. (PA)

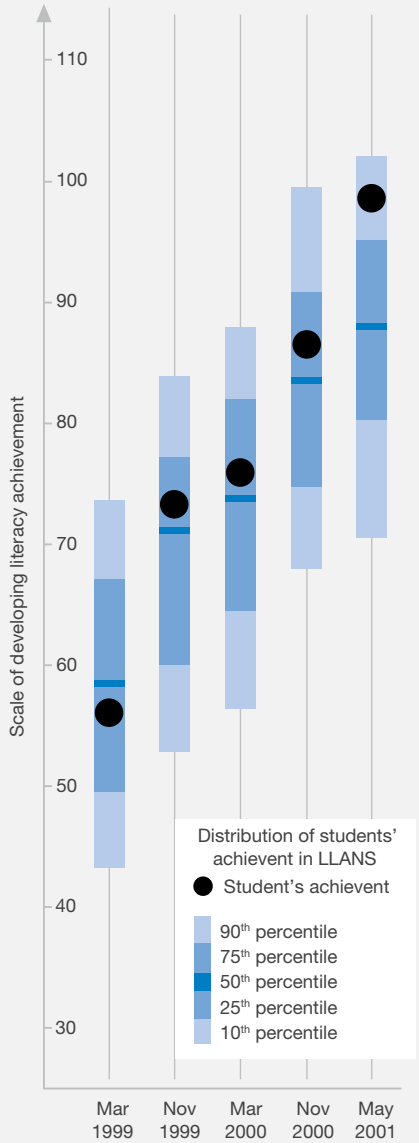
Predicts plausible story for a simple reading book after looking at cover. (MT)
Gives a literal interpretation of illustration from a picture story book. (MT)
Writes about a picture using combination of scribbles and some letters. (CP)

Reads correctly one or two words from the title of a simple reading book. (RF)
Identifies letters correctly in a given word from a simple reading book. (PA)

Identifies words with same first sound from list of three. (PA)
Identifies a word. (CP)

Identifies main character in a simple reading book. (MT)

Describes some main events shown in an illustration after listening to a picture story book. (MT)



Note: codes for the critical aspects of the descriptors are making meaning from text (MT), reading fluency (RF), concepts about print (CP), phonemic awareness (PA) and writing (WR).
Source:Meiers et al., 2006

were those that most students found difficult while those at the bottom of the list were those that most students found easy. Only a very small selection of skill descriptions from the five surveys could be included in Figure 1. The location of these skill descriptions is described in detail in **Table 3**, which indicates skills that were difficult or easy for students at the time of the particular survey.

The five shaded bands on the right side of Figure 1 show the distribution of performance of students participating in the study, with the highest achievement at the top of the scale. Figure 1 shows that there is an upward trend in the median scores for students across the five surveys. Overall, the progress map shows a clear pattern of growth in literacy achievement in the first three years of school.

A number of features of this map are noteworthy. First, the map shows the interweaving of skills across the difficulty range, with phonemic awareness, fluency and concepts about print developing alongside making meaning from text, both through listening and through independent reading. This indicates that learning to decode should not be regarded as a necessary precursor to making meaning from text but rather, that the two elements of decoding and meaning-making are concurrent and complementary.

Second, it can be seen that there was a wide distribution of literacy achievement at school entry (captured in the first survey in the bar furthest to the left) and this continued through the first three years at school. Moreover, it can be seen that there was very large overlap in achievement across the five data collections. This key finding indicates the complexity of the task of providing appropriate learning opportunities for all students.

Finally, the single student whose progression is mapped across the first five surveys in Figure 1 illustrates how unpredictable an individual's learning

TABLE **3**

LLANS findings indicating developing reading comprehension in the first three years of school

	Teacher prompt	Skill description	Achievement level
Survey 1 (school entry)	Now, tell me the story I read you. Where does the story begin? Where do I go next?	Retell a narrative in a picture storybook, including some key events. Locate the front of a picture storybook and understand the directional sequence of text.	Above 90th percentile (difficult). In 10th percentile (easy).
Survey 2 (end of first year at school)	Now, I'd like you to read the story to me.	Read all of a simple storybook with words for word accuracy, read "would" as a sight word and write a recognizable sentence.	In the 75th percentile (moderately difficult).
Survey 3 (start of second year at school)	What is this story about? The boy is asleep. How do you know he is dreaming? (response to a question about a specific illustration).	Identify key events after listening to a picture book. Students whose achievement fell in the 10th percentile were "likely to be able to give a literal interpretation of an illustration in a picture storybook."	Above 90th percentile (difficult). In 10th percentile (easy).
Survey 4 (end of second year at school)	This word says "plentiful." What do you think "plentiful" means? What does this page tell you about why mice chew things?	Able to use context to provide meaning for unfamiliar words in an informational text. Likely to be able to locate specific information in a simple informative reading book read independently.	Around 75th percentile (relatively difficult). In 10th percentile (easy).
Survey 5 (third year at school)	Why did the rich farmer give Mo Chin a bag of gold at the end of the story?	Likely to be able to explain a character's actions in a simple reading book read independently.	Around 50th percentile (moderate).

trajectory can be. This child began, in the first survey, a little below average in his literacy achievement. By the time of the second survey, his literacy was a little above average, and this position continued in the third survey. In the fourth and fifth surveys (especially in the fifth), his literacy achievement was well above average. This individual's progression is not typical—the majority of children whose literacy development begins at a comparatively low level, remain in a relatively low position in relation to their peers although their achievement improves. However, the individual trajectory shown indicates the existence of different patterns of development. Such variations are a reminder of the individuality of literacy development—a fact that teachers must deal with. It is important to note that the typical development of reading knowledge and skills charted in the progress map provides a broad guide to reading curriculum and pedagogy.

The achievement levels that were presented in Figure 1 are by percentiles and indicate items that most students found difficult and those that most students found easy. Table 3 presents some of the findings from the five surveys conducted in the first three years. The selection includes both easy and difficult items.

5. OUTCOMES OF THE STUDY

At the classroom level, the process of assessing students in one-on-one sessions in itself provided teachers conducting the assessments with important and useful insights into the progress and achievement of their students. Further, the reported individual students' achievement provided teachers with a sound basis for planning future teaching strategies to meet the needs of all students.

The model of individual student interviews developed for the early years component of the LLANS has since been used to assess literacy development in several research projects and large-scale student assessment programmes. A large-scale national study of teaching effectiveness in the early years described the LLANS as Australia's benchmark of early literacy procedures (Louden et al., 2005). A number of Australian state education systems have commissioned programmes using the individual literacy assessments and the strategy of one-on-one teacher interviews. The assessment activities and resource materials in these large-scale projects have been print-based and hands-on, although in some cases the teachers have been able to record the student data electronically, making it possible for student achievement to be reported to participating schools almost immediately.

Most recently, the study has informed the development of a computer-based reading/literacy assessment for 6-year-olds in Australia's Northern Territory, which has a large indigenous population. Reaching beyond Australia, the study has provided the construct for pilots of tablet-based reading assessments for the early years in Afghanistan and Lesotho.

5.1 Notes on planning and implementing a longitudinal study

The LLANS was successful in its aim to develop a progress map of early literacy development and in providing empirical evidence on the different rates of growth in literacy of individual children. The achievement of the study can be attributed to:

- Careful planning
  - Framing the research questions to be investigated to inform the planning and design of the study.
  - Having a clear concept of the nature of the construct (in this case literacy), derived from a literature review to shape the nature of the assessment tasks.
  - Using an assessment design that allows linking of the data from one assessment to the next, thus affording a continuous map of literacy development.

- Drawing a sample of sufficient size
  - Applying a scientifically drawn random sample of students large enough to tolerate substantial

attrition given the number of years over which the survey is conducted.
  - Building relationships with schools and principals in order to encourage them to keep the survey researchers informed and to maintain as many as possible of the sampled children in the survey over the years.

- Ensuring that the data are collected in a consistent way
  - Providing all survey materials to the schools.
  - Giving clear administration guidelines to teachers so there is confidence that the assessment is being delivered in a similar way to all the children.

- Using appropriate tools for analysis
  - Using analysis tools (such as a Rasch model or other IRT models) that allow the calibration of all tasks and all participating children on a single scale so that progress can be tracked over time.

A further important feature of a successful longitudinal study is to access funding and resources that will cover the duration of the study.

Finally, of great importance to the success of a longitudinal study is maintaining continuity in the research team so that the wisdom and experience gained can be carried from the early stages of planning through to data collection, analysis and the reporting stages.

REFERENCES

Louden, W., Rohl, M., Barratt Pugh, C., Brown, C., Cairney, T., Elderfield, J., House, H., Meiers, M., Rivalland, J. and Rowe, K. (2005). "In Teachers' Hands: Effective literacy practices in the early years of schooling". *Australian Journal of Language and Literacy,* Vol. 28, No. 3.

Masters, G.N. (1982). "A Rasch model for partial credit scoring". *Psychometrika,* Vol. 60, pp. 523-547.

Meiers, M., Khoo, S.T., Rowe, K., Stephanou, A., Anderson, P. and Nolan, K. (2006). ACER Research Monograph 61: *Growth in Literacy and Numeracy in the First Three Years of School*. Camberwell, Australia: Australian Council for Educational Research. http://research.acer.edu.au/acer_monographs/1/

Rasch, G (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: MESA Press (original work published 1960).

Snow, C.E., Burns, S.M. and Griffin, P. (eds) (1998). *Preventing reading difficulties in young children*. Washington DC: National Academy Press.

Assessing Young Children: Problems and Solutions

CHRISTINE MERRELL, PETER TYMMS
Durham University

ABBREVIATIONS

app	Application
iPIPS	International Performance Indicators in Primary School
OECD	Organisation for Economic Co-operation and Development
PIPS	Performance Indicators in Primary Schools

1. INTRODUCTION

Assessing young children around the age of starting school or earlier presents considerable difficulties and more arise if comparisons are to be made across different cultures and contexts. This article begins by expanding on some of those issues. We then proceed to suggest solutions to the apparently formidable difficulties. These solutions are based on the experience with the Performance Indicators in Primary Schools (PIPS) assessment—an assessment for use with children at the start of school. The authors have gained over 20 years of experience developing, adapting and successfully using the PIPS assessment in several countries. The PIPS Baseline Assessment was originally designed for formative use within schools (see Tymms, 1999; Tymms and Albone, 2002 for examples). It has subsequently been expanded to an international project for the study of children starting school and the progress that they make during their first school year in different parts of the world (www.ipips.org).

2. CHALLENGES OF DESIGNING RELIABLE AND VALID ASSESSMENTS OF YOUNG CHILDREN'S COGNITIVE DEVELOPMENT

There are different approaches to assessing cognitive development, including posing questions, either in written format or verbally, asking children to perform a practical activity (possibly an open-ended investigation) and observing their responses or the way that they work and interact within an educational setting. Each of these methods is associated with a range of issues, some of which are common to all and others that are more assessment-specific. For example, we know from practice and research that:

- Many young children cannot read when they start school (Merrell and Tymms, 2007) and therefore traditional group assessments with administration instructions that require a certain level of reading ability are not feasible.
- Group assessments, such as a pencil-and-paper test, require an element of self-management and young children tend not to have the capacity to cope in such situations.
- Young children generally have a limited concentration span (Sperlich et al., 2015) and, consequently, the length of time that an assessment should take needs to be correspondingly short. This becomes an issue particularly if the method of assessment requires

a child to focus on answering questions or completing a directed activity.

- Young children have limited short-term memory capacity. Whilst an adult can be expected to hold seven novel pieces of information plus or minus two for a short time, a young child might only be able to concurrently hold two or three pieces of information (Demetriou et al., 2015). This means that obtaining reliable information from complex questions is not feasible.

- There are significant differences in the developmental levels within an age cohort. For instance, among children who are aged 4 years old in affluent countries, some will be reading fluently, doing complex sums and have an extensive vocabulary while others have not realised that text on a page is a code that carries meaning let alone possess the ability to identify letters (Tymms et al., 2014; Wildy and Styles, 2008a, 2008b; Merrell and Tymms, 2007). The latter group are probably unable to perform simple counting, don't recognise any digits and possess a vocabulary that may be extremely limited.

- If an assessment of children's cognitive ability is made entirely on the basis of observations, it is possible that they may fail to display their full potential. For example, a child may have an advanced understanding of mathematical concepts but if the activities in the setting do not challenge them to display this understanding, it will be missed. A further issue with observations is bias against individuals and groups (Harlen 2004, 2005; Sonuga-Barke et al., 1993; Wilmut, 2005).

- In addition to the challenges outlined above, consideration needs to be given to what should be assessed—the official curriculum of the country, variables that predict later success/difficulties or skills that are most malleable at the age of assessment? (Thompson and Nelson, 2001).

3. CHALLENGES OF DESIGNING ASSESSMENTS OF NON-COGNITIVE DEVELOPMENT IN YOUNG CHILDREN

The term 'non-cognitive skills' describes a collection of attributes and traits that represent the ways in which we think, our feelings, emotions and behaviour (Borghans et al., 2008). Non-cognitive skills continue to develop throughout our lives (Bloom, 1964). They include critical thinking and problem-solving skills, persistence, creativity and self-control. A recent report by the Organisation for Economic Co-operation and Development (OECD, 2015) emphasises the importance of non-cognitive skills for positive outcomes in life, using the big five personality dimensions that are widely recognised in psychology (openness, conscientiousness, extraversion, agreeableness and neuroticism) as an organizing framework. There are many descriptions of the big five traits, for example: Openness has been defined by Costa and McRae (1992) as the degree to which an individual is open to fantasies, aesthetics, feelings, new ideas and experiences. Conscientiousness was defined by Trapman et al. (2007) as the degree of dependability, organizational ability and degree to which an individual persists to achieve a goal. Extraversion is defined by Trapman et al. (2007) as the quantity and intensity of interpersonal interaction. Agreeableness is associated with being flexible in behaviour, cooperative and tolerant (Trapman et al., 2007). Neuroticism is described by adjectives such as anxious, touchy, nervous and unstable (Costa and McRae, 1992).

It cannot be assumed that young children possess a conceptual understanding of these attributes and traits or the ability to evaluate their own behaviours and actions in an objective way through an assessment. Their vocabulary is emerging and unlikely to be sufficiently sophisticated to be able to understand what is being asked and to be able to respond appropriately. Indeed, Soto et al. (2011) suggested that self-report questionnaires are only appropriate for children aged 10 years and over.

On the basis of research such as that of Soto et al. (2011), we must rely on adults who know the children to assess these non-cognitive areas on the basis of their knowledge built up through observations and interactions. But relying on adults to conduct these assessments has its challenges. Large classes can mean that teachers may not know each child well enough to provide a proper assessment. Most parents know their children well but some have low levels of literacy, making written surveys unreliable while interviews either in person or by phone are expensive to conduct. A further complication is that assessors interpret questionnaire statements in different ways.

4. CHALLENGES OF INTERNATIONAL COMPARISONS

Additional challenges are faced when making international comparisons of cognitive development and non-cognitive skills. Different interpretations of constructs and items arise across cultures as well as within cultures. For example, in a study by Merrell et al. (2013), teachers in Australia, England and Scotland were markedly different in their ratings of children's inattentive, hyperactive and impulsive behaviour. All international studies face issues with adaptations from one language to another given the subtle nuances conveyed through language. In reading assessments, there is the added challenge of working with different writing systems that could involve a major distinction between the alphabetic writing systems and the logographic writing systems used in some parts of Asia.

Ratings of non-cognitive skills are likely to be influenced by prevailing norms of behaviour and by individual perceptions (e.g. Merrell et al., 2013) and by variation in the standards applied by the assessors (Hosterman, 2009; Duckworth and Yeager, 2015).

For an international comparative study of children in their first school year, the varying ages at which they start school throughout the world needs to be taken into consideration because development in the early years varies greatly (Tymms et al., 2014); one year at this stage can be a quarter of a child's life. In England, for example, the mean age at the start of school is 4.5 years and children may start school just after their fourth birthday. By contrast, children in Russia commonly start school at the age of seven (Kardanova et al., 2014). Family and cultural expectations will influence development as will a country's policy on early development. Can international comparisons of children's development and progress in their first year of school be valid? Can it yield useful information for policy and practice as well as increase our understanding of child development in general? In the next parts of this article, we identify ways to move forward in the face of the challenges arising from the issues discussed thus far.

5. ASSESSMENT PURPOSE: USING RESULTS FOR FORMATIVE, RESEARCH OR ACCOUNTABILITY

In an ideal world, we would want an assessment that provides information that is useful for the teacher, for national statistics and for research purposes. While assessment information is certainly useful for research and national statistics, when assessments become means of public accountability, they lose their formative purpose. In the case of the PIPS, the assessments are primarily intended to provide formative information for teachers. In order to ensure large scale use of the PIPS without harmful impacts, we create agreements that limit its use to formative purposes. We typically achieve this by creating an agreement on how the information is to be used that emphasises confidentiality. Results from the PIPS assessment are fed back to schools via a secure website where each participating school can see only their own results. The reports are a combination of charts and tables with both norm referenced and raw scores. The teacher can use the norm referenced scores to compare the development of their pupils with a representative sample. The raw scores provide detailed information about which questions each child answered correctly and incorrectly, revealing strengths and areas for development.

We do allow for the data to be used for research that aims to inform wider practice and policy but not for accountability. We formally agree with all stakeholders that the data is confidential—pupils, teachers and schools will not be identified when analysed and published for research purposes. We go so far as to say, in some areas, that if a school publishes their own data, they will be dropped from the project. This apparently aggressive stance provides schools with a reason not to disclose their results and yield to the pressures exerted from the public, journalists or higher authorities.

6. CONTENT OF THE ASSESSMENT

If the initial information from an assessment is to guide the teacher—as has been our intention when developing the PIPS Baseline Assessment—the assessment needs to include content that can provide comprehensive, reliable and valid information on what children know and the skills they possess on the road to becoming literate and numerate (Tymms et al., 2009). Depending on the purpose of the assessment, the results from the assessment should indicate each child's zone of proximal development so that the teacher can plan tailored learning experiences for specific children. If a sample of the class is assessed, the results will give the teacher an idea of its general ability level and its variation, which has some use but is more limited than information on all individuals. There are compromises to be made between collecting detailed information on all children in a large class and the time that it takes to accomplish this.

Information from an assessment of children's cognitive ability, such as the PIPS assessment, can be used as a predictor of later success or difficulties, and it can also be interpreted as an outcome measure for the time prior to assessment. It could be used as an outcome measure with which to evaluate early childhood development policy.

We have not focused on developing content specific to a country's official curriculum as many children at the start of school will not have followed an official curriculum yet. Curricula in early years tend to focus on developing general skills, such as personal and social development, basic language and precursors to reading, numeracy and motor skills rather than specific areas of learning, such as a specific period in history.

7. DESIGNING ASSESSMENTS FOR USE WITH CHILDREN DURING THEIR FIRST YEAR OF SCHOOL

Foremost, it should be noted that due to the stage of development among children in their first year of school, any assessment of a young child's cognitive development that is conducted before they start school or during their first school year must be conducted on a one-to-one basis with the assessor and the child if high-quality information is to be obtained. For the same reasons, the assessment must be completed within 15 to 30 minutes. Beyond this time, the validity of the data collected will drop as the children tire and their concentration ebbs. The administration costs will also rise if trained assessors are used to collect data for a research project.

The assessment must be robust so that it can produce reliable and valid results independently of the administrator, otherwise it is subject to the potential bias of that administrator (see Harlen, 2004, 2005 for examples of how assessments can be prone to bias). This is important if the results from assessments conducted by different assessors in different settings are to be meaningfully compared. Specialist knowledge and training should not be a necessary pre-requisite to obtaining high-quality information otherwise this limits the use of the assessment. The content of the assessment must be appropriate for children of a wide range of abilities within an early year's cohort. If the format and content of the assessment is to be appropriate for a wide range of ability and yet be administered within the time limit suggested earlier, then the only way to achieve this is to use an adaptive approach. If a child does not answer easy items correctly, they are not presented with more difficult ones but if a child is moving rapidly through the assessment and answering questions correctly, she/he is rapidly moved forward to more difficult

content. This approach not only addresses the range of abilities but it also decreases the time spent on the assessment. Furthermore, questions that are too difficult are not administered, reducing the probability of demoralising children with lower ability.

Can we envision a one-to-one adaptive assessment that can be carried out with the assessor using pencil and paper? Our own experience indicates that the assessors do not always follow the rules. An alternative method would be to use laptop computers but laptops can be expensive and security can sometimes be an issue when working in areas such as the favelas in Brazil or the townships of South Africa. We need an intelligent device that is inexpensive and widely used, and on which the assessment itself can be easily deployed. We found that a smartphone or tablet alongside a booklet provides the answer. The child and the assessor look at the booklet together. An application (app) is accessed by the assessor through the smartphone or tablet. The app selects items and the assessor records the child's responses electronically. The app contains rules that govern which questions are presented to the child on the basis of their answers. This relieves the assessor from having to follow adaptive rules and deciding which question to present to the child, which means that they can focus more on the child. We have used this approach successfully in the Western Cape of South Africa and in Russia. The assessment is divided into sections such as vocabulary, concepts about print, letter recognition and so on. Within each section, questions are organized in order of increasing difficulty in a series of sequences with stopping rules. Children start with easy items and if they answer questions correctly, they are presented with more difficult ones until they make a certain number of mistakes. The assessment then moves on to another sequence. This may be a more advanced section—for example, letter recognition if a child has demonstrated competence in concepts about print or a simple section of a different area of development such as counting. We have found that this approach can generate high quality data in just 15 to 20 minutes about children's early language, reading and mathematics development while allowing the precocious few to demonstrate the full extent of their ability.

For assessments of non-cognitive development, we believe that we need to work with teachers rather than parents because their professional judgements are based on a wide experience of children of a similar age. For large classes, we suggest sampling because of the daunting workload that an assessment of every child in a class of 50 pupils would imply. However, the purpose of the assessment should guide the decisions on parameters. If the purpose is to inform policy, sampling children within classes and schools will provide a picture of children's development. If it is to inform classroom instruction, a sample will provide teachers with a profile of their class' abilities. If individual children with particular needs are to be identified, all children in the class should be assessed.

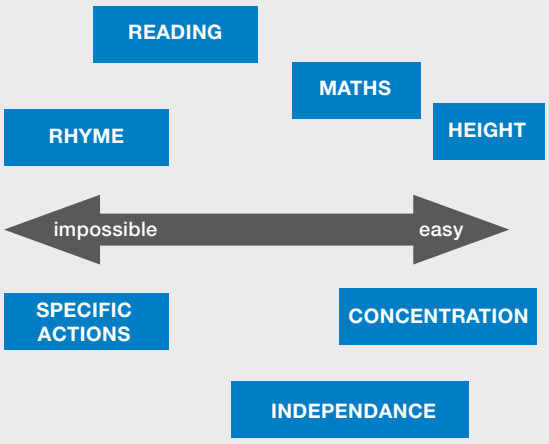
We have recently been developing a method to make comparable observations of children's behaviours from different contexts and cultures. It is debated if this can be achieved but we believe that it can be using short video clips of children (suitably anonymised) exhibiting different levels of a particular behaviour, such as attention, in a range of contexts. Teachers are asked to make their own ratings of the children in the clips and then to rate the children in their class. By analysing each teacher's ratings against the clips, we have evidence that we can establish the reliability, validity and severity of their scores and even evaluate if the construct being addressed is meaningful in different cultures.

8. COMPARISONS ACROSS COUNTRIES

Based on our experience, we found that comparisons across countries of some variables are virtually impossible while other variables lend themselves more easily to comparisons. The chart in **Figure 1** details this hypothesis, suggested by the authors.

Based on our research using the PIPS assessment, we found that some areas of development and

Figure 1. Comparison hypothesis



skills can be easily adapted across languages and cultures with few problems (e.g. simple arithmetic). Other areas cannot be adapted and compared so easily, such as the ability to spot rhyming words. In this case, words that rhyme in one language would not necessarily rhyme when translated into a different language. Alternative words would need to be used and this would change the level of difficulty of the items. It may be possible to devise an assessment of nonsense sounds and words which rhyme and would be unfamiliar to children in different countries but it is questionable whether valid data could be collected using this approach. There are certain behaviours that are particular to certain cultures, such as the use of head movements (for example, in the Indian sub-continent a head movement which means "yes" is seen in the west as meaning "no"). However, there are some behaviours and aspects of personality that can be compared—for example, conscientiousness, curiosity or the ability to empathise. Others may be linked to cultural norms. For example, it may be acceptable for a child to question a request made of them by an adult in one culture, indeed valued as a mark of curiosity or independence, but the same behaviour would be considered unacceptable in another and therefore not displayed. Caution is needed when interpreting the behaviour of children from different cultures.

In summary, some areas of development and skills can be compared across all cultures, others can be compared across some cultures and some are unique to particular situations. All of these ideas are being put into practice within the International Performance Indicators in Primary School (iPIPS) project. Currently, data are being collected and analysed from South Africa, Russia, China, Brazil, England and Scotland.

CONCLUSION

We believe that despite the daunting challenges faced at the outset, with the PIPS and iPIPS we have developed an assessment system that works in different cultures and contexts. We are able to collect reliable and valid data in a short time with young children at varying developmental stages, which is useful to schools and at the same time can provide analyses for policymakers. Very similar assessments can be used with suitable adaptations across cultures and costs can be kept down so that work can be carried out effectively and efficiently.

REFERENCES

Bloom, B.S. (1964). *Stability and Change in Human Characteristics*. New York: John Wiley & Sons.

Borghans, L., Duckworth, A.L., Heckman J.J. and Bas ter Weel. (2008). "The Economics and Psychology of Personality Traits". *Journal of Human Resources,* Vol. 43, No. 4, pp. 972-1059.

Costa, P. T. Jr., and McRae, R. R. (1992). Revised *NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, Florida: Psychological Assessment Resources, Inc.

Duckworth, A.L. and Yeager, D.S. (2015). "Measurement matters: Assessing attributes other than cognitive ability". *Educational Researcher,* Vol. 44, pp. 237-251.

Demetriou, A., Spanoudis, G. and Shayer, M. (2015). "Mapping Mind-Brain Development". Farisco M. and Evers K. (eds.), *Neurotechnology and direct brain communication*. London: Routledge.

Kardanova E., Ivanova A., Merrell C., Hawker D. and Tymms P. (2014). *The role of the iPIPS assessment in providing high-quality value-added information on school and system effectiveness within and between countries*. Basic Research Program Working Papers. Moscow Higher School of Economics.

Harlen, W. (2004). "A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes". *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Harlen, W. (2005). "Trusting teachers' judgement: research evidence of reliability and validity of teachers' assessment used for summative purposes". *Research Papers in Education*, Vol. 20, pp. 245-270.

Hosterman, S.J. (2009). Halo Effects and Accuracy in Teacher Ratings of ADHD Symptoms: Influence of Ethnicity and Developmental Level. Ph.D. thesis, Lehigh University, USA.

Merrell, C. and Tymms, P. (2007). "What Children Know And Can Do When They Start School And How This Varies Between Countries". *Journal of Early Childhood Research*, Vol. 5, No. 2, pp.115-134.

Merrell, C., Styles, I., Jones, P., Tymms, P. and Wildy H. (2013). "Cross-country Comparisons of Inattentive, Hyperactive and Impulsive Behaviour in School-Based Samples of Young Children". *International Research in Early Childhood Education,*Vol. 4, No. 1, pp 1-17.

OECD (2015). *Skills for Social Progress: The Power of Social and Emotional Skills*. OECD Skills Studies. Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264226159-en

Sonuga-Barke, E. J. S., Minocha, K., Taylor, E.A. and Sandberg, S. (1993). "Inter-ethnic bias in teachers' ratings of childhood hyperactivity". *Journal of Developmental Psychology*, Vol. 11, pp. 187-200.

Soto, C.J., John, O.P., Gosling S.D. and Potter J. (2011). "Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample". *Journal of Personality and Social Psychology*, Vol. 100, No. 2, pp. 330-348.

Sperlich, A., Schad, D. J., & Laubrock, J. (2015). "When preview information starts to matter: Development of the perceptual span in German beginning readers". *Journal of Cognitive Psychology*, *27*(5), 511-530.

Thompson, R. A. and Nelson, C. A. (2001). "Developmental science and the media: Early brain development". *American Psychologist*, Vol. 56, No. 1, p. 5.

Trapmann, S., Hell, B., Hirn, J.W., Schuler, H. (2007) "Meta-Analysis of the Relationship Between the Big Five and Academic Success at University". *Journal of Psychology*, 215, 132—151.

Tymms, P. (1999). *Baseline Assessment and Monitoring in Primary Schools: Achievements, Attitudes and Value-added Indicators*. London: David Fulton Publishers.

Tymms, P., Jones, P., Albone, S. and Henderson, B. (2009). "The first seven years at school". *Educational Assessment and Evaluation Accountability*, Vol. 21, pp. 67-80.

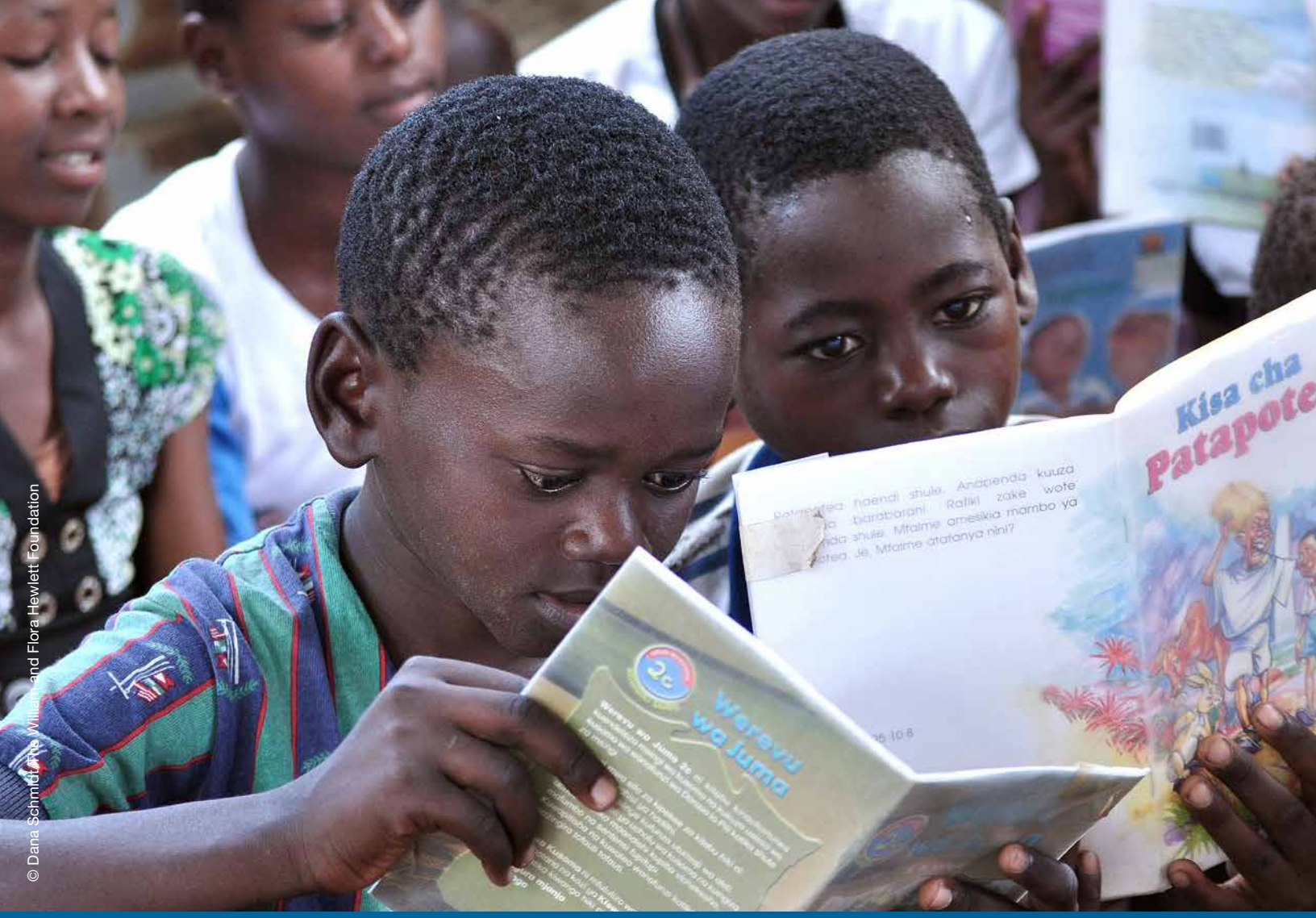
Tymms, P. and S. Albone (2002). "Performance Indicators in Primary Schools". A.J. Visscher and R. Coe. (eds.), *School Improvement Through Performance Feedback*. Lisse/Abingdon/Exton PA/ Tokyo: Swetz & Zeitlinger, pp. 191-218.

Tymms, P., Merrell, C., Hawker, D. and Nicholson, F. (2014). *Performance Indicators in Primary Schools: A comparison of performance on entry to school and the progress made in the first year in England and four other jurisdictions.* Department for Education: London. https://www.gov.uk/government/publications/performance-indicators-in-primary-schools

Wildy, H., & Styles, I. (2008). Measuring what students entering school know and can do: PIPS Australia 2006-2007. *Australian Journal of Early Childhood*, Vol. 33, No. 4, pp. 43-52.

Wildy, H., and Styles, I. (2008b). "What Australian students entering primary school know and can do". *Journal of Australian Research in Early Childhood Education*, Vol. 15, No. 2, pp. 75-85.

Wilmut, J. (2005). *Experiences of summative teacher assessment in the UK.* Qualifications and Curriculum Authority. London: QCA.



© Dana Schmidt/The William and Flora Hewlett Foundation

Chapter 3
Translating Reading Assessments into Practice

The articles in this chapter discuss strategies to optimise the impact and utility of household-based and school-based assessments. They describe different ways to collaborate with individuals, school systems and government structures to improve learning.

Assessing Children in the Household: Experiences from Five Citizen-Led Assessments

JOHN KABUTHA MUGO, IZEL JEPCHIRCHIR KIPRUTO AND LYDIA NAKHONE NAKHONE
Twaweza East Africa

SAVITRI BOBDE
ASER Centre, Pratham

ABBREVIATIONS

ASER Annual Status of Education Report
NAS National Achievement Survey

1. INTRODUCTION

Around a decade and a half ago, an article published by the *Guardian* titled "Children taught at home learn more",[1] evoked energetic debate on the value of the home environment in promoting learning. Indeed, the home environment is acknowledged as critical to literacy development and is indisputably a significant factor in the educational success of children. Moreover, it plays an important role in its many facets of parental and sibling support, social and language development, and esteem building in younger children among others. A recent review has established that while the school is central to learning, it may be also responsible for creating a cultural disconnect between young children and teaching methods that are abstract and removed from their everyday experience (Nag et al., 2014). The review also identifies the home environment as a critical enabler of the development of early literacy. In assessing learning, the household has been credited for offering cultural variety where one is able to connect a child to the home circumstances that confront them day-by-day (Wagner, 2011).

The evolution of the citizen-led assessments that started with the Annual Status of Education Report (ASER) in India[2] in 2005 and has since spread to eight other countries, emerged from the recognition that existing assessments of learning had failed to recognise the home as the cradle of learning for young children, especially in the acquisition of basic literacy and numeracy skills (see articles by Banerji as well as by Aslam et al.). Critical characteristics of citizen-led assessments include being conducted in the home rather than at school and that they combine learning measurement approaches (from the education field) with citizen-monitoring approaches (from the transparency and accountability field) to engage ordinary citizens in the assessment of children's learning (Results for Development, 2015). The assessments are conducted orally, one-on-one with the child, use simple tools that are easy to administer and are conducted on an unprecedented scale.[3]

This article provides a reflection on the opportunities and challenges linked to assessing children at the household level by sharing experiences from five assessments: the ASER-India, the ASER-Pakistan, Uwezo (Kenya, United Republic of Tanzania and Uganda), Beekunko (Mali) and Jàngandoo (Senegal). This article opens with a description of a study conducted to fill the knowledge gaps related to the question of assessing at the household level.

1 See http://www.theguardian.com/uk/2000/aug/13/education.educationnews1

2 The ASER is an annual assessment of learning started by Pratham, and covers almost all rural districts in all states of India with a report launched every year since 2006.

3 The citizen-led assessments are conducted in almost all districts in the participating countries and jointly assess over one million children every year. Read more at www.palnetwork.org

Key methodological considerations in assessing children at the household level are then presented followed by a discussion of the opportunities linked to assessing children at the household level. The challenges experienced in assessing children are also discussed followed by a presentation of the strategies employed by the various assessments to mitigate against these challenges. The article closes with a discussion on the critical perspectives and new horizons to explore in strengthening assessments at the household level.

1.1 How did we collect the data?

Considering the dearth of literature on how data is collected, we resolved to begin a conversation with the people who have implemented household assessments in the last seven years (since 2009) in various countries. A total of 86 participants responded to our questions, combining key informant interviews, a focus group discussion and questionnaires as summarised in **Table 1**. The participants are listed and acknowledged in **Appendix I**.

1.2 Methodological considerations

Citizen-led assessments adhere to similar methodological principles as those used in other household-based assessments. Five principles were used across the samples.

i. Random sample of households

For each assessment, a sample of enumeration areas or villages is randomly selected using the official national sampling frame, as provided by the national bureau of statistics in each country. From each village, a random sample of households is produced either from a full list of all the households in the selected village or enumeration area (e.g. Uwezo and ASER-Pakistan urban areas), or according to a pre-defined sampling rule (e.g. ASER-India and ASER-Pakistan rural areas). The methodology for sampling households varies across the assessment implementation areas—either maps from the national bureau of statistics are used or a drawing of the map of the sampled village is produced on site with the help of the village leader and others. Simple random sampling produces a sample containing both households with and without children. In each of the sampled households, ALL children within the target age range are listed and assessed regardless of their schooling status or level. Please refer to the section "Overview of oral reading assessments" in this ebook for more details on the target population in citizen-led assessments.

ii. Administration of oral assessments

The assessment questions are administered orally and one-on-one with each child. In assessing literacy, the child responds orally to a set of sub-tasks presented by the assessor (see article by Nakabugo), including reading comprehension questions.[4] In assessing numeracy, children are given a combination of oral and written tasks to solve. The oral questions are administered in the child's language of choice. The assessment is administered in conversational tone to ease tension with the child and make them as comfortable as possible while performing the assessment tasks.

In addition, the head of the household responds orally to questions on various household indicators that are later used in the analysis stage to run

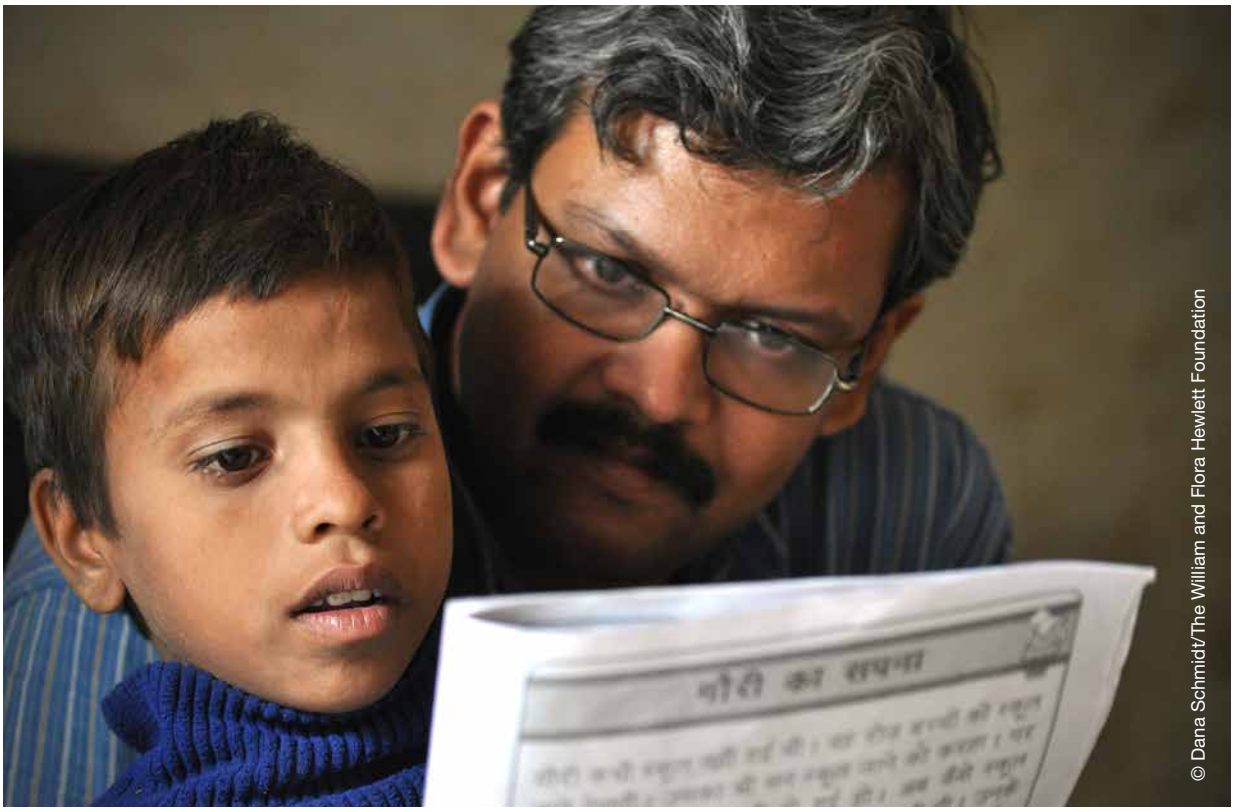
4 ASER-India and ASER-Pakistan have not assessed comprehension in all years.

TABLE **1**

Summary of key informants by type and assessment

	ASER-India	ASER-Pakistan	Bεεkunko	Jàngandoo	Uwezo	Total
Key informant interviews	1	1	2	1	7	12
Focus group discussion	0	0	0	0	9	9
Questionnaire	34	0	0	0	31	65
Total	35	1	2	1	47	86

Source: Key informant interviews conducted by the authors



regressions and confirm general hypotheses on household factors that influence learning outcomes. Finally, volunteers visit the public schools that the majority of children attend in that sampled enumeration area (private schools are also visited in ASER-Pakistan and Uwezo assessments). In schools, teachers respond orally to questions on school-related factors or inputs that support learning. The volunteers begin with the school survey then proceed to the household assessment.

iii. Local volunteers assess the children

In all of the countries, volunteers are trained on how to conduct assessments in literacy and numeracy. It is common for a pair of volunteers to conduct the assessment in each of the sampled villages (e.g. the ASER-India, the ASER-Pakistan and Uwezo), although this is not the case for all the citizen-led assessments (e.g. Beekunko). Furthermore, Uwezo also emphasises gender parity among the pair of volunteer assessors (when feasible). In the same way, for some assessments, volunteers who come from and live in the sampled villages (or at least nearby) are preferred while in other areas, assessments utilise teacher trainees who do not necessarily come from the sampled villages—as long as they are proficient in the language spoken in the sampled area. As documented by Results for Development (2015), this participation of local people is designed to broaden the audience that usually consumes assessment data (policymakers, pedagogues, education authorities) to include a wider range of people—all of whom have a stake in the educational outcomes of the country's children.

iv. Call-back to include all children in sampling households

In all five assessments (across seven countries), strict instruction is given to the volunteers to return to the households to assess children who were not present during the time of visit. This practice is employed to reduce non-response in the sample of children. To increase the possibility of getting all the children in the household, the assessments are conducted over the weekends and in the evenings when a weekday extension is necessary.

2. BENEFITS OF ASSESSING CHILDREN AT THE HOUSEHOLD LEVEL

2.1 Overview

Assessing children at the household level has many advantages—for the children, the partners, the volunteers and other implicated stakeholders. Each of the key informants were asked to rate the benefits of assessing children at the household level. The greatest perceived benefit among the Uwezo key informants is the privilege of talking to parents and giving them instant feedback. The greatest benefit among the key informants from the ASER-India is related to the opportunities created for volunteers and citizens to engage and develop their skills. See **Figure 1** for details.

i) Benefit 1: reaching ALL children, including the most excluded

The purpose of learning assessments is to establish the learning competences of children. By assessing children at the household level, we can increase the heterogeneity of the sample—it includes children that are enrolled in school and those who are not; those who attend formal and non-formal schools; those in public or private schools (and even those that are schooled at home); and those who attend school regularly as well as those who don't. As Aslam (2014) contends, if we are focusing on ensuring that every child has foundational skills, the best place to reach ALL children is their homes. **Box 1** provides a comparison to school-based assessments in India.

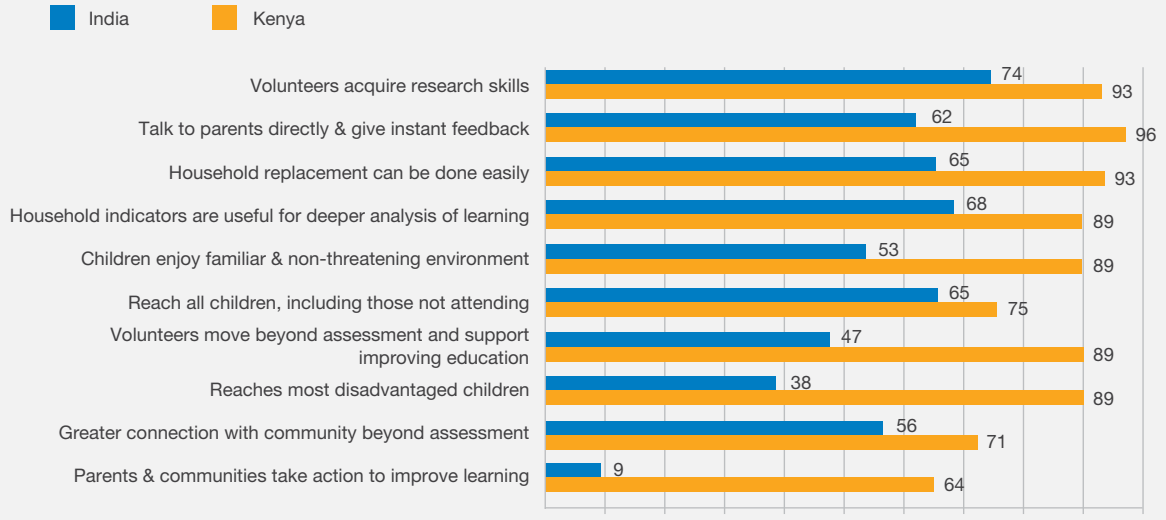
ii) Benefit 2: children relax from the tensions linked with school

When children are assessed outside under a tree or inside the house where they rest and play, they are able to differentiate the assessments from school examinations. Younger children tend to be shy and the proximity of the parent, sibling or friend reassures them. The informal environment is often supportive since there is always greater excitement after the child has read or completed a numeracy task and awaits their friend to also take it.

iii) Benefit 3: engaging with parents and communities to discuss learning

When we assess children at home, we encounter parents face-to-face. Many school-going children have parents who never attended school or they

Figure 1. Benefits of assessing children at home assessments by percentage of respondents who agree with the statements



Source: based on the data collected through a questionnaire (28 respondents from Kenya and 34 respondents from India) specifically developed for this article

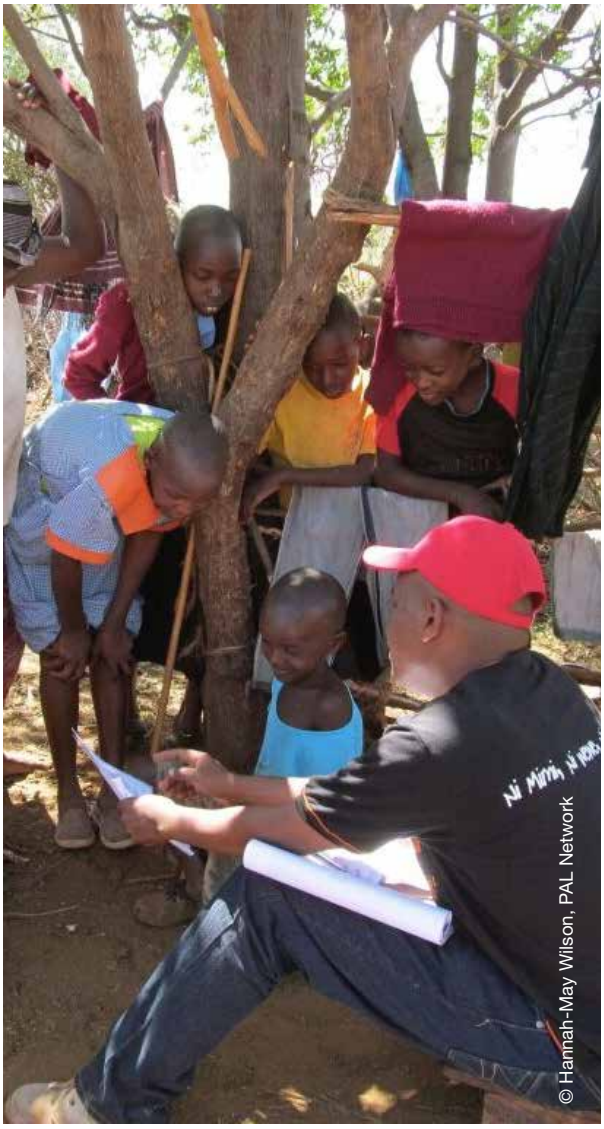
Box 1: Standardised school-based assessments in India

Standardised school-based assessments are done in two ways.

1. Public examinations that are conducted at exit points from the system: at the end of primary education (Grade 5), at the end of lower-secondary education (Grade 10) and at the end of upper-secondary education (Grade 12). They are intended to certify outcomes for all children in schools and are affiliated to particular examining boards. These results are not representative of all children since different boards have different assessments.

2. The National Achievement Survey (NAS) is conducted every three years for children in Grades 3, 5 and 8 in government schools. This low-stakes and sample-based assessment generates estimates of learning at state and national levels in a limited number of domains. Since the NAS is limited to students enrolled in government schools, it does not provide any information on the learning levels of children studying in other types of schools, such as private or non-formal schools. Furthermore, being grade-level tests, they are not useful in a context where most children are several grade-levels behind.

All public examinations and national learning assessments are traditional paper-pencil tests. They hide the fact that many children who are enrolled in school are unable to read fluently (Suman Bhattacharjea, ASER-India, from key informant interviews conducted in May 2015, see Appendix I).


© Hannah-May Wilson, PAL Network

have to travel long distances between schools and households. This presents a difficult gap to bridge. The conversations that emerge during the assessment are a powerful first step towards engaging parents in supporting their children's learning.

After assessing children in the household who fall within the assessment age, parents get instant feedback on what the children can or cannot do. This ritual spurs interest among some parents, especially those who have never connected with the reality of what their children can do and the levels of their ability. Usually, the assumption of every parent is that because they go school, children must be learning.

Different from the school's end term report card, which at times never gets into the parents' hands or do not make sense to less literate parents, the oral conversation with parents is always a refreshing encounter. For many parents, listening to their child reading is a unique experience.

iv) Benefit 4: accessing household indicators that enrich analysis

Rather than inquiring about children's backgrounds from school personnel, encountering children in the household connects us to the realities and circumstances of learning at the household level

(*see example in* ***Box 2***). Collecting household information in schools proves especially difficult when the target group is young children who cannot answer many questions about their families' socio-economic status.

Box 2: Volunteer excerpt interview from Uwezo

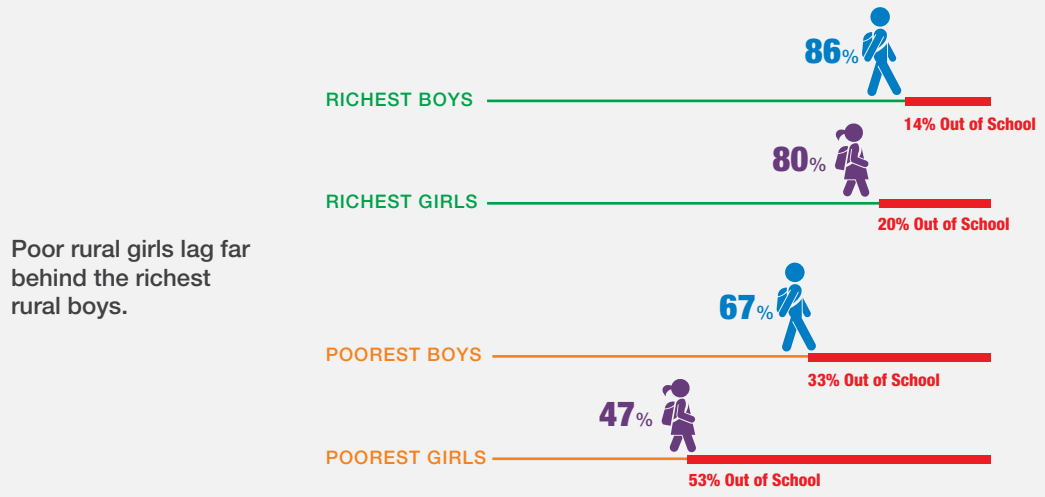
"Last week, I was in a household in Kangemi (a slum in Nairobi) assessing a child. The mother was selling liquor and men kept coming in and out, some sitting on her bed. I agreed with the boy that we should walk and find a quieter place outside the house. I was badly struck by the challenge the boy was confronting in his learning" — partner volunteer, Uwezo in Nairobi

The citizen-led assessments include a context questionnaire that collects information on the household. The data generated can produce indicators on a wide range of issues—including out-of-school children, socio-economic status, attendance and enrolment in school. These indicators are then used to enhance understanding of the socio-economic factors underlying learning outcomes. **Figure 2** provides an example of an analysis from the ASER-Pakistan on the inequality in school enrolment among boys and girls in rural areas, highlighting the large gender gaps even in rural communities.

v) Benefit 5: volunteers find greater purpose in developing their skills and engaging meaningfully with their communities

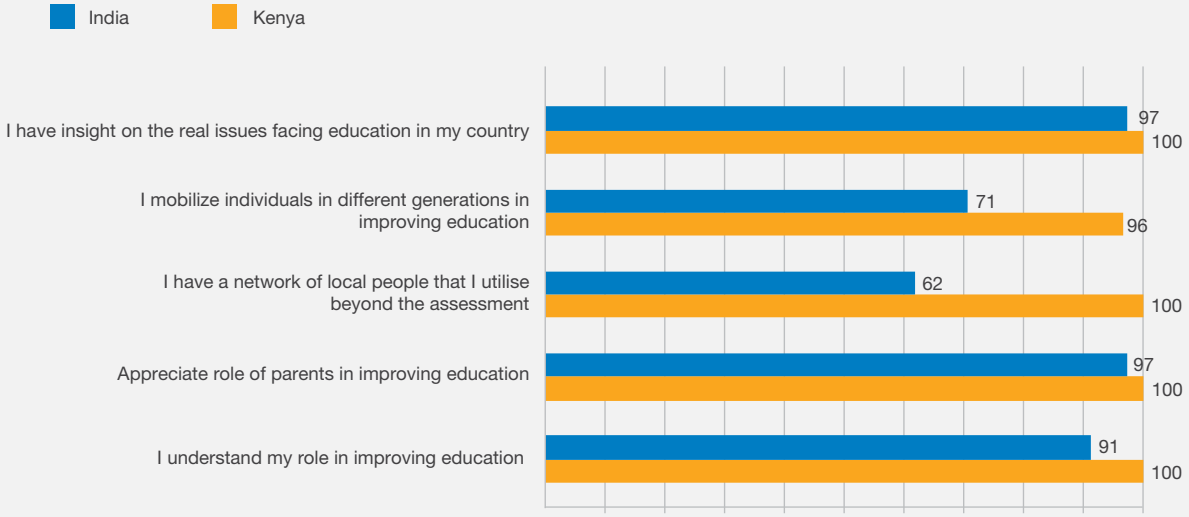
The use of simple tools to assess children in the community has led us to discover ordinary people with extraordinary potential to improve the learning outcomes of children. Over the years, we are glad to have been at the centre of connecting passions. The volunteers and assessors have enjoyed five advantages in varied measures (*see* ***Figure 3***). First, through the training and hands-on assessment of children and holding conversations with local community members, the volunteers have grown into a tremendous group of data collectors and researchers and their skills have led to not only self-fulfilment but in their utility in other community-based studies and interventions. Second, the assessors report gaining insight into the issues that affect education in their own communities and to encounter the realities of their own neighbourhoods. Third, they understand their role in improving education and have discovered opportunities for their own agency. Fourth,

Figure 2. Example of analyses from the ASER-Pakistan, rural areas



Source: adapted from Aslam (2014), using data from ASER Pakistan

Figure 3. Benefits of household assessments to the volunteers, by percentage of respondents who agree with the statements



Source: based on data collected through a questionnaire (28 respondents in Kenya and 34 respondents in India) specifically developed for this article

Box 3: Excerpt from a volunteer assessor for the Uwezo assessment

"Assessing children with Uwezo over the last five years has given me such sense of greater purpose and fulfilment. I feel so good when I assess kids. Sometimes when I fail to assess the non-sampled ones, they run after me calling out: *mwalimu, hata mimi naomba kusomeshwa* (teacher, I beg that you also teach me). I pause, get out the materials and assess them, without recording, and we all walk away happy" — partner, Uwezo.

especially for those who come from the sampled villages, volunteering builds up networks of parents, neighbours, teachers and local leaders with varied opportunities for community engagement. Last, being in the household, assessors can access the intergenerational realities in understanding learning—sometimes stretching conversations to four generations. Grandparents, parents, children and even grandchildren engage as a community with the learning of children in their family. In addition, volunteers are either current or future parents, so we hope that these experiences will instil a sense of responsibility in ensuring a proper education for their own children. From an informal comparison between India and Kenya,[5] these benefits seem greater when the volunteers come from the sampled communities. See Figure 3 for a list of benefits weighted by percentage of volunteers and **Box 3** for an excerpt from an Uwezo volunteer assessor.

3. CHALLENGES OF AND POSSIBLE SOLUTIONS TO ASSESSING CHILDREN IN THE HOUSEHOLD

3.1 Overview

The quantitative study between the ASER-India and Uwezo partners established greater unity in identifying challenges faced when administering assessments at the household level. The top challenge encountered was from children being unable to read because someone was watching, usually parents. Linked to this, many parents were shocked when their children could not complete tasks and it is not uncommon to find parents who threatened their children. **Box 4** provides some

5 In Kenya, volunteers largely come from the sampled villages while this is not the case for India.

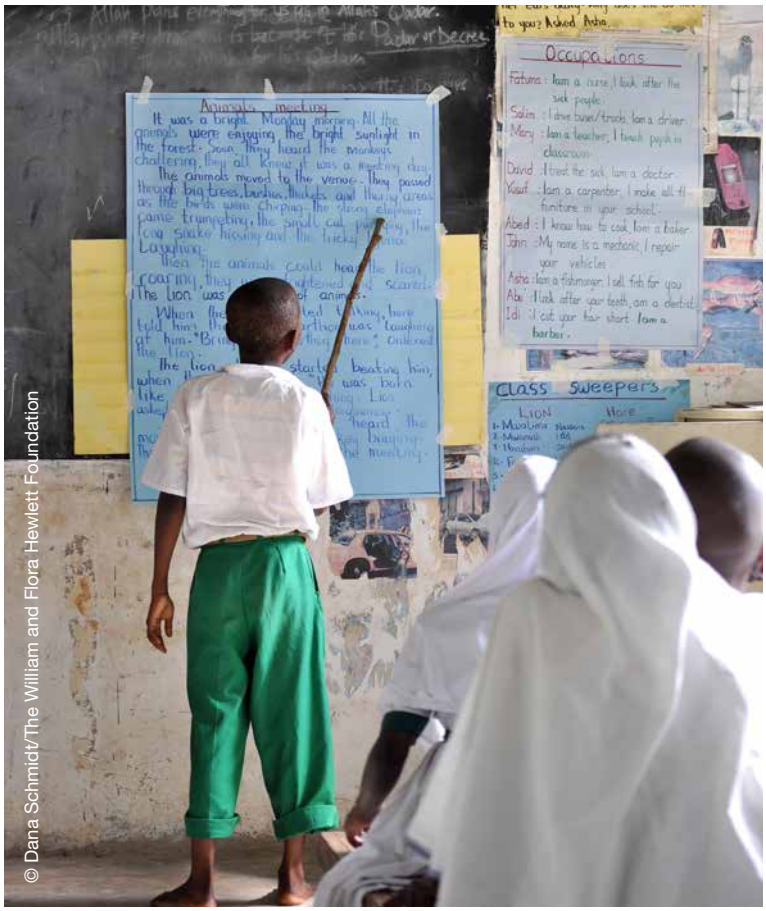
Box 4: Excerpts from key informant interviews

"Rarely have there been cases where the child is forced to interrupt the test to cook, perform a task or respond to a request by the parent, which would extend the test time"
—Respondent, Jàngandoo Senegal

"The availability of parents and often their negative reactions after finding the low level of learning of their children is a challenge to us"
—Respondent, Beekunko Mali

"Some parents out rightly confront the child: 'You mean you cannot do this?' The level of tension is higher among children who cannot read. Children who can read want to show it off and rarely face problems"
—Respondent, Uwezo Kenya

"We do not entirely miss out on teachers, because many teachers are also parents"
—Respondent, Uwezo Kenya



© Dana Schmidt/The William and Flora Hewlett Foundation

examples of these common challenges. Another challenge was related to access. In the arid parts of some countries, the villages can be very vast and one must brave the long, tiresome walks to reach the households randomly selected. A third challenge, found mostly in East Africa, is that many children are not found at home over the weekends because they are attending tuition classes and volunteers have to return to the household multiple times to complete the assessment. In Senegal, many children move away for the holidays in winter making it difficult to reach all the children in the sample. The least cited challenges in the ASER-India and Uwezo included children being busy with household chores and lacking concentration for the assessment, volunteers being tempted to fake data to avoid the long walks through the village, household heads not allowing their children to be assessed and finally the burden of seeking consent multiple times (rather than only once at the school). **Figure 4** summarises the extent to which the ASER-India and Uwezo assessors experienced the challenges listed.

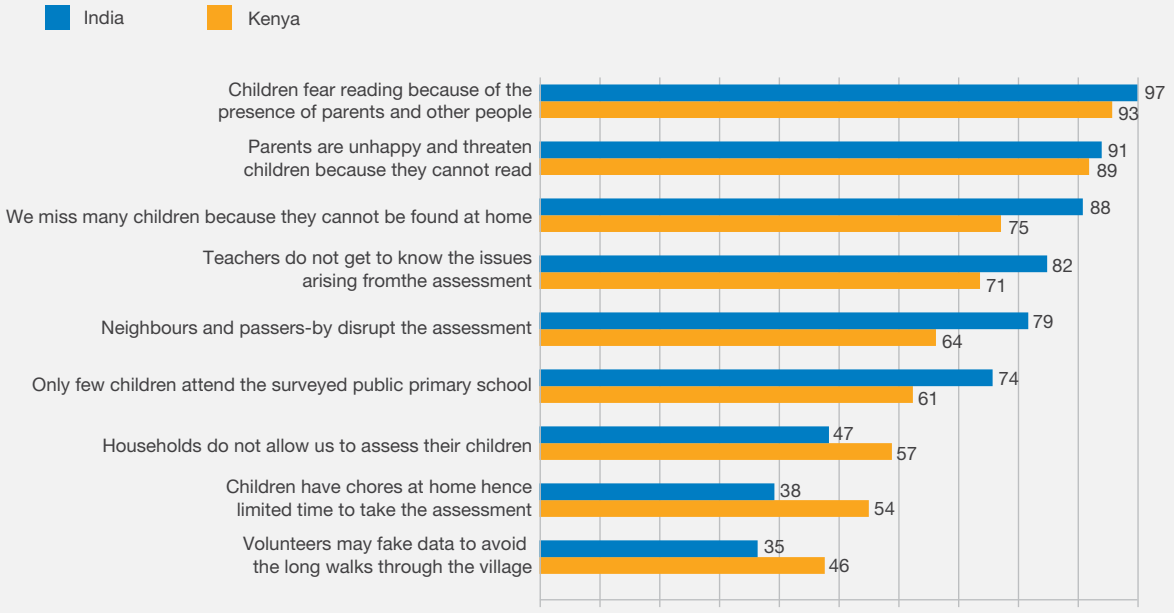
3.2 Practical approaches to address challenges

Citizen-led assessment volunteers encounter common challenges in the process of conducting the assessment. **Table 2** highlights some of these common challenges and the types of actions taken collectively to address them. See additional perspectives from Jàngandoo in **Box 5**.

4. FINAL THOUGHTS

Over the years, we have made incremental improvements to address challenges and seize emerging opportunities. Much improvement has been made to the processes of tool development and implementing the survey in order to tighten standardisation and quality control. If we were to start over again with significantly more resources,

Figure 4. Challenges encountered by the ASER-India and Uwezo volunteer assessors, by percentage of respondents who agree with the statements



Source: based on data collected through a questionnaire (28 respondents from Kenya and 34 respondents from India) specifically developed for this article.

we would probably try to build in more activities to generate impact at the village level. It becomes apparent that the desired connection to local communities will not happen in the 20 minutes used to conduct the household survey, but through sustained momentum for change following the assessment. Some of the recent improvements to the assessment process include the following initiatives:

- Uwezo has been testing a few ideas on engaging communicators at the local level in 2015, and we look forward to the lessons we will learn.

- The ASER-Pakistan has taken considerable steps to shed light on issues of access and learning with respect to children with disabilities (Singal and Sabates, 2016). Citizen-led assessments have started exploring possible partnerships with expert institutions to improve access to and assessment of children with disabilities, which continues to be a challenge among the various initiatives.

Jàngandoo has introduced technology (i.e. tablets) in the assessment process. The ASER-India also piloted the use of technology in 2014, and Uwezo in 2015. Certainly, the costs and risks are high, considering the scale of the assessments, but we must keep trying until we arrive at an affordable option. We conclude that weighing the benefits against the detriments, household assessments offer a unique solution to the Sustainable Development Goal for Education which calls for the provision of inclusive and equitable quality education as well as the promotion of lifelong learning opportunities for all. Foreseeably, despite all the challenges, the household is the only place where all children in developing economies can be included and the only way to tell the whole story.

Despite the many advantages that accrue from household-based assessments, the complementarity with school-based assessments cannot be overstated. School-based assessments provide teachers with opportunity to directly engage with the assessment process and results while reflecting on strategies for enhancing learning among the children that they teach.

TABLE **2**

Possible solutions to common challenges

Challenge	Actions taken when faced with this challenge
Parents are unhappy and threaten children because they cannot read.	▪ We assure parents that with time children improve, if they get the required support. This conversation often turns into an important one in which parents voice their views about their children, schools and learning. ▪ We work mostly in pairs so in such cases, one person can engage the unhappy parent in discussion away from the child while the other assesses the child. Note, this only happens in isolated cases.
Children fear reading because the presence of parents and other people (neighbours and passers-by) disrupts the assessment.	▪ We politely request people to give us time with the child alone and we tell them that we will explain the results after we finish. ▪ We divide responsibilities such that one of us converses with the crowd while the child is being assessed in a quieter place. ▪ We ask to take the child away from the crowd for testing by asking for permission from relatives. ▪ We train our volunteers to seek out these solutions.
Missing many children because they cannot be found at home during the time of visit.	▪ We make call backs later in the day or the following day. ▪ During our visits to the school, we announce that we would be visiting the homes so that children are aware and can wait for us. ▪ We use village leaders to inform households of our arrival prior to the visits.
Households do not allow us to assess their children.	▪ We take time to introduce the assessment and our work, and seek consent. ▪ We use village leaders to inform the people prior to our visits, and if possible also walk with us during the assessment. ▪ Some volunteers are well known to the people, which facilitates access to households.
Children are busy with chores at home and have limited time to take the assessment.	▪ We ask if we can come back during their free time. In cases of many siblings, we make agreements and test them in turns in ways that are convenient for them.
Volunteers may fake data to avoid the long walks through the village.	▪ We conduct monitoring (spot checks) and post-assessment recheck in large samples to confirm that visits were conducted and if not, send them or someone else back to correct the data. ▪ We randomly call heads of households to confirm that they were visited before we clear the volunteer.[6]
Teachers do not get to know the issues arising from the assessment.	▪ We share results during education days and visit some schools. ▪ The government schools in the sampled villages are visited and the purpose of the survey and the previous year's findings are discussed with the teachers.
Only few children attend the surveyed public primary school.	▪ Most of the children in rural areas are in public schools. We are clear in that our target is not the school but children. For example, in Pakistan, we also visited private schools.

Box 5: Solutions to reduce the effect of challenges—perspectives from Jàngandoo, Senegal

- Assess in a quiet place. Under the Jàngandoo programme, it is clearly written in the volunteer agreement signed by all staff on the ground that the child must be isolated in a quiet place during the assessment to make them feel more at ease and not afraid by the presence of parents or all other possible disturbances that can interfere with their performance.

- Capture data using tablets. We developed the 2014 software (using tablets to collect data) which was designed to ensure that all children in the household are assessed before the data is submitted.

- Engage teachers in the assessment process. Teachers are appointed to evaluate the assessment in households. Through their experiences, they can discover the weaknesses children have in the different learning processes and competencies assessed by the test. This leads them to question the usual teaching methods and education governance.

6 See example of ASER-India's recheck framework on http://img.asercentre.org/docs/Aser%20survey/Ensuring%20data%20quality/qualitycontrolframework.pdf

APPENDIX I: LIST OF QUESTIONNAIRE RESPONDENTS

We wish to acknowledge the following persons who supported the writing of this article in their various capacities.

Category	Country	Participants
Key informants interview	India	Suman Battacherjea
	Kenya	Joyce Kinyanjui, Daniel Wesonga, Amos Kaburu
	Mali	Massaman Sinaba
	Pakistan	Baela Raza-Jamil
	Senegal	Pr. Abdou Salam Fall
	Uganda	Mary Goretti Nakabugo
Focus group discussion	Kenya	Charles Kado, Walter Kwena, Francis Njuguna, Grace Mwathe, Rosemary Njeri, Beatrice Kiminza, John Kariuki, Zacharia Kabiru
Questionnaires	India	Bhalchandra, Devyani Malgaonkar, Dharmendra Kumar, Ketan, Zerah Z C Rongong, Anuradha Agrawala, Shikha, Bhumi Bhatt,Hashini Silva, Heli pathak, Chandrika Prasad maurya, Neelam Dipakkumar Kanjani,Mangi Lal RJ, Tajudin, Vina, Manisha, Bhavya Karnataka, Kamalakshmma, V.S Varalakshmi, Dr. Denita Ushaprabha, Manjunatha B G, Suraj Das, Rashmani , Shailendra Kumar Sen, Omprakash Kolare, Rajesh Bandewar, Vinod Gurve, R.Chokku, M. Jayasakthi, Lakshmi Kandan, A.L.E. Terrance, Thangavel N
	Kenya	Anthony Odhiambo, Alois Leariwala, Kibet Kipsang, Larry Wambua, Chrispinus Emusugut, Geoffrey Ngetich, Joseph Halkano, Gideon Koske, Polycarp Waswa, Sospeter Gitonga, Eunice Lubale, Lucky Mwaka, Noah Amrono , Mike Njeru, Erick Kipyator, Lucy Mutono, Maryvine Nyanchoka, Achuchi Jane Okello, Ibrahim Hassan, Shem Ongori, Rashid O. Miruka, Peter Chem, Chris Kung'a, Sam Mukundi, Paul kepkemboi, Mary Chepkemoi, Stephen Kamau and Mohamed Golicha
	Tanzania	Gerald Samwel Ng'ong'a, Fortunata Manyeresa
Translation		UNESCO Institute for Statistics

REFERENCES

Aslam, M. (2014). *Citizens reshaping education through household-based learning accountability initiatives: Evidence from Pakistan and beyond*. Presentation at a Technical Meeting on 21-23 July 2014 on Understanding What Works in Oral Assessments of Early Reading. http://www.uis.unesco.org/StatisticalCapacityBuilding/Workshop%20Documents/Education%20workshop%20dox/Montreal%202014/08.Household%20based%20learning%20accountability%20initiatives%20in%20Pakistan_EN.pdf

Nag S., Chiat S., Torgerson C. and Snowling M.J. (2014). *Literacy, Foundation Learning and Assessment in Developing Countries: Final Report*. Education Rigorous Literature Review. London: Department for International Development.

Plaut, D. and Jamieson Eberhardt, M. (2015). *Bringing Learning to Light: The Role of Citizen-Led Assessments in Shifting the Education Agenda*. Washington, D.C.: Results for Development Institute. http://r4d.org/sites/resultsfordevelopment.org/files/resources/Bringing%20Learning%20to%20Light_English.pdf

Singal, N. and Sabates, R. (2016). *Access and learning are equally important for children with disabilities*. Global Partnership for Education blog. http://www.globalpartnership.org/blog/access-and-learning-are-equally-important-children-disabilities (Accessed February, 2016).

Wagner, D.A. (2011). *Smaller, Quicker, Cheaper: Improving Learning Assessments to Developing Countries*. Paris: UNESCO-IIEP. http://unesdoc.unesco.org/images/0021/002136/213663e.pdf

Assessment in Schools

EMILY KOCHETKOVA AND MARGARET M. DUBECK[1]
RTI International

ABBREVIATIONS

EGRA Early Grade Reading Assessment
MOE Ministry of Education
IRR Inter-rater reliability

1. INTRODUCTION

This article provides an overview of the process of administering a school-based Early Grade Reading Assessment (EGRA) and covers many of the considerations an implementer will need to bear in mind when planning and conducting such a survey. Readers are encouraged to download the revised edition of the *EGRA Toolkit*[2] for more in-depth information on each of the topics covered here as well as for information not covered, such as descriptions of and research undergirding individual EGRA subtasks and drawing of the school sample (RTI International, 2015).

2. ADVANTAGES OF SCHOOL-BASED ASSESSMENT

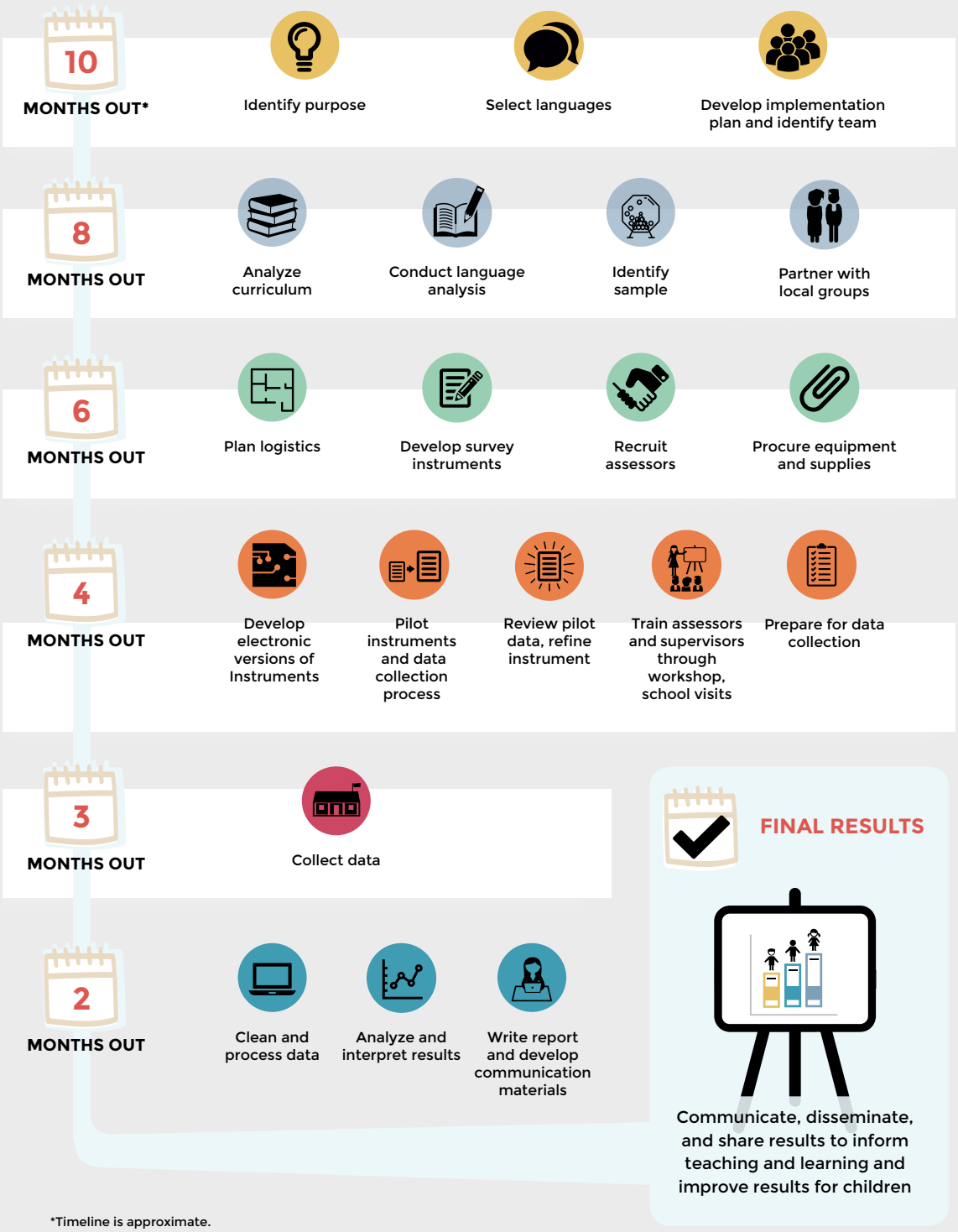
There are several ways to gather information on children's learning. For example, depending on the research question(s) posed by the stakeholders sponsoring the survey, demographic household-based surveys can include a few education indicators about the children living at the selected residence. Children in the home may be interviewed or given a few brief tasks to assess their skill level with regard to literacy or numeracy. In addition to the descriptive statistics that this assessment can provide, researchers can derive variables to look for correlations between the children's abilities and other indicators in the survey, such as parental involvement in schoolwork, eating habits or expectations for labor, to help explain to some degree the learning outcomes data. Likewise, community-based surveys can measure similar indicators while collecting information on how a local community supports education.

However, these broad types of surveys may not be designed to collect detailed data on learning outcomes. The EGRA, in contrast, requires that a trained assessor sit with the individual—randomly selected pupils for 10 to 20 minutes each—to administer the series of subtasks contained within the assessment. Another disadvantage of many household and community-based surveys is that aside from anecdotal information gleaned from interviews with participants about schools that the children in the family or community attend, they often do not gather specific information on the school context, such as enrolment data, infrastructure or teaching and learning materials. Although contextual observations and interviews are not part of the EGRA given that its administration takes place in schools, researchers can take advantage of the opportunity to conduct interviews

1 The authors would like to acknowledge and thank our colleagues Jason Boyte, Amber Gove and Erin Newton for their respective contributions to this article in design, technical review and editing.

2 Go to https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&id=929

Figure 1. The Early Grade Reading Assessment Timeline



*Timeline is approximate.

with school staff, such as teachers and head teachers. Researchers can also take inventory of the resources available at the school and even conduct classroom lesson observations if time and funding allow. Any and all additional data collected have the potential to shed light on factors that influence reading performance as reflected in the EGRA results (provided that proper protocols for linking data sets are in place).

3. LOGISTICS

i) Timeline (see Figure 1)

The EGRA can be administered at any time during the school year. The timing should be driven by the implementers' research question. If the implementers are interested in learning about pupils' reading skills at a certain grade level, data collection should be planned for near the end of the school year. For example, an assessment of Grade 2 reading skills should take place near the end of Grade 2 to ensure that the study captures close to all possible learning gains made in that grade. When contemplating the overall timeline of activities, setting the dates for data collection (usually spanning at least two weeks)[3] should be a first priority and activities leading up to this can be scheduled backward from the start of fieldwork. The dates established for data collection should take into account public and religious holidays or other special events, such as national elections, as well as weather patterns that could impact a school's normal operations or accessibility.

ii) Materials

The primary materials required for conducting an EGRA in schools include electronic tablets or other handheld devices (and associated accessories) for digital data collection as well as paper and pens or pencils. Even when the implementer is planning to collect data electronically, the EGRA requires paper for training activities, for the pupil stimulus books and as backup versions of all instruments and protocols to be carried by data collection teams in the event of a malfunctioning electronic device.

iii) Transportation and lodging

EGRA implementers will need to make provisions for the data collection teams to travel from school to school. This may involve hiring cars or motorbikes or simply providing a transportation allowance to teams to enable them to make their own transport arrangements. Teams will need to spend the night nearby before data collection begins at a given school to give them ample time to arrive at each school before it opens in the morning on the day of the visit.

iv) Permissions

The implementer will need to secure documentation from the sampled schools' governing bodies (i.e. Ministry of Education or other non-governmental organization), stating that the requirements for carrying out the study (e.g. approval by the institutional review boards for conducting research with human subjects) have been met and that permission is granted to conduct the EGRA. Data collection teams should carry a copy of this documentation with them to share with schools.

4. RECRUITING AND SELECTING ASSESSORS

4.1 Government employees vs. private sector

The quality of the assessment data collected is ultimately in the hands of those who collect it. These assessors or enumerators are responsible for the sometimes grueling work of visiting school after school, day after day, often over difficult terrain and long distances. There are a variety of criteria for the study leaders to consider when selecting assessors to conduct an EGRA. The criteria are described in the following section but a first consideration is whether to draw from the public or private sector.

i) Private sector assessors

Assessors from the private sector may be university students or other individuals familiar with research or survey practices. Students planning to enter the

3 Data collection should not surpass four weeks (two or three is ideal) since early reading skills change quickly at this stage—a child's skills at day 1 of a data collection effort would be quite different at day 40.

field of education can make for reliable assessors as they are typically young and energetic, have a vested interest in learning how well pupils in their country are reading, are likely to be comfortable with technology in the case of electronic data collection, and tend to have the required flexibility to attend training workshops and conduct days or weeks of fieldwork. Private assessors must be paid a per diem allowance as well as a fee for their labor.

ii) Public sector assessors

These assessors are usually drawn from the ranks of the Ministry of Education (MOE). They may be teachers, teacher trainers or education officers in some capacity. While some of the characteristics of private sector assessors may be harder to find among government workers, there are several advantages as well. One advantage has to do with buy-in or acceptance of the assessment results. The more that MOE officials are able to witness firsthand the performance of children being assessed, the more likely they are to believe and support the validity of the assessment results—even if they may not function particularly well as assessors. Ultimately, MOE involvement tends to decrease the possibility of refuting the findings. Secondly, if the EGRA is being conducted with an eye toward eventual adoption by the government as a routine assessment, it is best for assessors to come from among government staff. Public sector assessors typically are paid only a per diem allowance for the data collection as the work takes place during their normal working hours and can be seen as an extension of their duties within the MOE.

When assessors are drawn from the ranks of government staff, to avoid tension during fieldwork, it would be wise for the study leaders to consider hierarchy or seniority when composing teams ahead of data collection. Any supervisor, field monitor or team leader roles should be filled by assessors with more seniority. Additionally, if a cascade training approach is planned (i.e. training a small cohort of local individuals who will then lead the training of all assessors), these 'master trainers' should not be considerably more junior than the assessors they will be training.

Often government staff have a background in the teaching profession. While trained teachers such as these have the advantage of being experienced and comfortable with children, they often find it difficult to step out of the role of 'teacher' and into the role of 'researcher' during an assessment, especially when a pupil they are assessing is struggling or giving incorrect responses. Trained teachers tend to want to take the opportunity to instruct, correct or otherwise guide the pupil toward the correct response, which undermines the validity of the data collected. Thus, when assessors with a teaching background have been chosen for the training, it is important to continually emphasize that during data collection, they are researchers and not teachers and that they are simply there to observe and record not instruct.

4.2 Qualifications

i) Educational background

Assessors should have completed a post-secondary degree of some kind. Ideally, they will have previous experience or at least exposure to research and/or survey methodology—but this is not required. University students can make reliable assessors as can trained teachers. Assessors should be comfortable interacting with children.

ii) Language considerations

Assessors must be able to speak and read fluently in the language of the assessment they will be administering to pupils. It is also helpful if they speak the lingua franca or the national languages that may be used in part of the training. Locating and recruiting speakers of the language can be relatively simple but it should not be assumed that a native speaker is also a proficient and confident reader of that language, especially in cases of languages that are spoken in only one region of the country or that have a fairly new orthography. The ability to read confidently and with feeling is necessary so that assessors follow the assessment protocol 'script' exactly (thus ensuring that all pupils have the same experience) and so that what they say to the pupil sounds natural and is easily understood.

4.3 Training

A workshop to train assessors for data collection should be planned to span five to eight days, depending on the number of instruments that will be administered (refer to the 2015 *EGRA Toolkit*). Training will involve an overview of the purpose of the assessment, the general plan for administration (time frame, location, team composition, etc.), thorough explanation of the instruments, ample time for practice, visits to schools, and inter-rater reliability (IRR) testing (a discussion of IRR is included in section 6.5).

i) Venue considerations

Training of assessors should ideally be residential. A hotel with conference facilities that includes a large space for plenary sessions and smaller rooms for group work will lessen training and practice time lost to daily commutes. If electronic data collection is planned, the venue must have strong WiFi and internet capabilities. Additionally, the venue should be near one or two local schools that use the language of the assessment so that assessors can make a planned visit to practice.

ii) Language experts

Opinions may vary among the assessors as to the vocabulary used for some of the assessment tasks as well as the correct pronunciation of words and letter sounds. Having a language expert present at the training can settle any disputes or confusion about what assessors should consider correct and incorrect during the assessment. The language expert should have formal training in linguistics for the language in question. In addition to settling any discussions that may arise about the language, the language expert can spend time during the training workshop to drill and practice the assessment subtasks with assessors to help tune their ears for hearing correct and incorrect responses.

iii) School practice

The agenda for the training workshop should include two school visits for practice purposes. If the schools are nearby, typically this can take place during the first half of the day and the latter half can be used to debrief and conduct more training or practice sessions at the venue. The opportunity to practice in schools with real children (as opposed to another assessor acting out the role of the child during practice at the training venue) is

crucial to help assessors gain confidence and hone their skills.

The first practice school visit should take place after assessors have been introduced to all of the instruments. They do not need to have very much prior practice before the first school visit. The purpose of the first school visit is simply to expose the assessors to the context in which they will ultimately be working during data collection so they get a feel for what it is like to interact with pupils, gain a general sense of the timing and organization of the work, and most importantly, use the instruments. The first school visit should take place on the second or third day of training. Often the first school visit is an eye-opening experience for novice assessors. They typically return to the training venue with a much clearer understanding of their responsibilities as assessors and are ready to put in the serious time and effort needed to master the assessment protocol.

The second school visit should take place two or three days after the first and usually falls on the second to last day of the training workshop. Between the first and second school visits, assessors will have spent many hours practicing and fine-tuning their technique. At this point, they will be more familiar with the instrument content as well as any electronic devices and software to be employed and they will have a greater idea of what to expect. Thus, the second school visit is often a gratifying and confidence-building experience because assessors can see their own improvement. While the first school visit is focused mostly on using the instruments, the second school visit should include everything assessors will be required to do during data collection. This includes meeting and introducing themselves to the head teacher, arranging the assessment space, sampling the pupils, administering the assessment, and returning the assessment space to its original condition. Trainers might also set a goal for each assessor to assess a certain number of children (at least five) to encourage a focus on efficiency in addition to correct technique.

iv) Assessor performance monitoring

During assessor training, it is important to formally test that assessors are consistently scoring the same observation in the same way. This is particularly important in the case of the EGRA. Since the assessment is oral rather than written, assessors are not afforded the opportunity to revisit the child's response at a later time to be sure that they have captured it accurately. Above all, the desired consistency in scoring depends on effective training.

Assuming that assessors have been effectively trained, the purpose of monitoring assessor performance is threefold:

1. Determining priorities for training
2. Selecting assessors
3. Reporting on the preparedness of the assessors to stakeholders.

See the revised *EGRA Toolkit* (referenced earlier) for more details on conducting assessor performance tests during assessor training. In summary, the assessors being trained are arranged in a large room in which a mock assessment of a child is given (using either two adult trainers, a video or a combination of the two). Assessors evaluate the responses of the 'child' and record them on their device as if they were the one assessing the child. The child's responses are based on a script created in advance (known as the Gold Standard) so that the errors made are deliberate and there is little or no ambiguity about whether the responses are correct or incorrect. The evaluation of the child by the assessors is captured and used to calculate their assessor performance test score(s). A score of 90% is recommended as the 'passing' threshold. Any assessors who are not able to reach the pre-determined passing threshold by the last scheduled assessor performance test should not be sent into the field to collect data. If there is concern over having sufficient numbers of assessors to complete the work according to schedule, additional assessors can be placed on a reserve list and called on if the need arises—but these additional assessors should receive additional targeted training

in the specific areas that the testing results show are in need of improvement.

Note that these instructions were written assuming that the assessor performance testing is conducted in a context in which the data are collected by means of electronic data-capture software using a tablet. It is important to note that this testing can also be conducted in a paper-and-pencil format, which leads to some additional steps in the processing of the data, including the capture of the assessors' responses into a database. See more on measuring assessor performance during fieldwork in the section on data collection (section 6).

5. PILOT TESTING

The pilot testing of the instruments can take place before or after assessor training. There are advantages and disadvantages to both approaches and the decision often comes down to logistics and context.

If no experienced assessors are available (from a prior administration of the assessment), it may be best to schedule the pilot test to take place immediately after the assessor training workshop ends. Typically, pilot testing will take only one or two days to complete if all trained assessors are dispatched. An advantage of this approach is that the pilot test, in addition to generating important data on the instruments themselves, also provides valuable insight into the performance of the assessors. Those analysing the pilot data can look for indications that assessors are making certain common mistakes, such as rushing the child or allowing more than the allotted time to perform certain tasks.

A disadvantage of pilot testing after assessor training is completed is that the instruments used during assessor training are not yet finalized because they have not been pilot tested. In many cases, earlier less-formal pretesting of the instruments will contribute to their fine-tuning so that the formal pilot test typically does not give rise to major instrument revisions. Still, in this scenario, assessors should be informed that the instruments they are practicing with during training may contain some slight changes during later data collection. The implementer should thoroughly communicate any changes that take place after the pilot test to all assessors before they go into the field.

When pilot testing takes place immediately after assessor training, it is recommended that a period of at least two weeks elapse between the pilot test and full data collection to allow for analysis of pilot test data, instrument revisions, printing, updating of electronic data collection interfaces and distribution of materials to assessment teams.

In other cases, it is preferable to conduct pilot testing prior to assessor training. In contexts where an EGRA has taken place previously in the recent past (no more than two years prior) and hence trained assessors are available, a brief refresher training over one or two days can be sufficient to prepare for the pilot test. An advantage of this approach is that the instruments can be finalized (based on data analysis from the pilot test) before assessor training begins. Similar to the recommendation above, it is prudent to allow for at least two weeks between pilot testing and assessor training so that all materials can be prepared not only for training but also for data collection. In this scenario, data collection can begin as soon as possible after training ends.

6. DATA COLLECTION

6.1 Notifying schools

At least a month before data collection begins, the implementer should contact the sampled schools to confirm their location and verify that they are still operating; that they meet the definition of the population of interest (e.g. if only public schools are being assessed, that the sampled school is not a private school); that they have pupils enrolled in the grade to be assessed (usually ten pupils per grade, per school are sampled); that school will be in session during the dates scheduled for data collection; and that the language of instruction at

the school matches that of the assessment. The implementer will replace any school in the sample that does not meet all of the criteria for selection with a predetermined replacement school (see section 6.3), which should also be contacted for verification.

6.2 Weather/terrain considerations

The implementer should take into account the time of year that data collection will take place and any associated weather conditions that could impact fieldwork. For example, in some countries, the end of the school year may correspond with the rainy season, worsening road conditions, which could potentially impact school attendance or operation in certain areas.

6.3 Replacement schools

For every school in the sample, the implementer's survey statistician should plan for at least one—or ideally, two—associated replacement schools in the event that the originally sampled school cannot be visited. Replacement schools should be selected based on their similarity to the originally sampled school, such as location, type (public or private), enrollment, etc. Reasons for replacing a school might include a school being closed or a school having fewer than half of the number of pupils needed for assessment (e.g. if ten pupils in Grade 2 are to be assessed but only four pupils are present at the school on the day of the visit). Sampled schools that are located in difficult-to-reach areas should not be replaced simply for convenience although in some cases, a sampled school will be totally unreachable to assessment teams due to weather or road conditions and will have to be replaced. If a school needs to be replaced due to conflict in the area, it is likely that the associated replacement schools in the sample will be impacted by the same conflict. In this case, a new school sample will have to be redrawn. The implementer should carefully document and justify all replacements.

6.4 School visit protocol

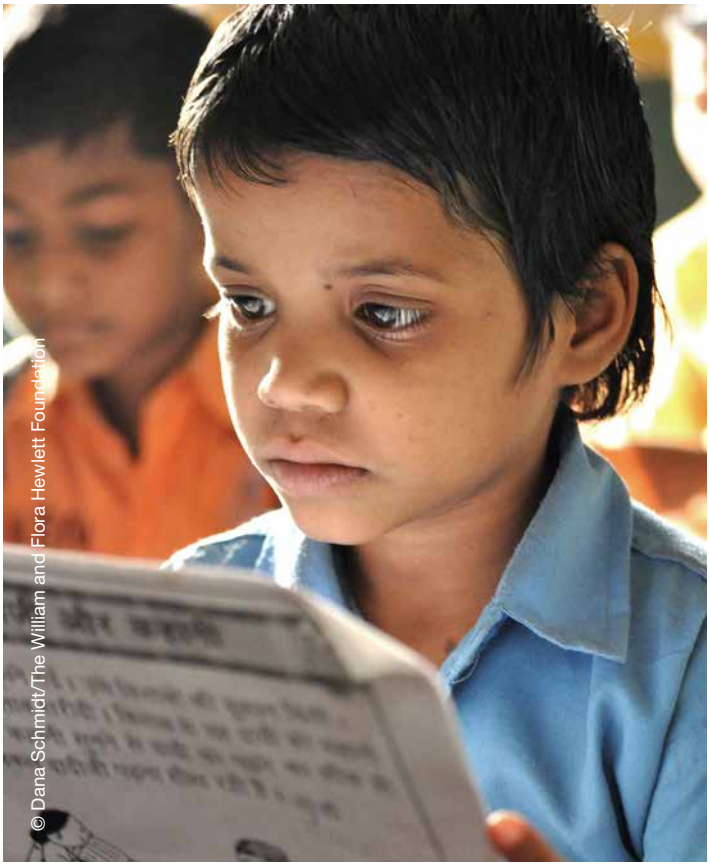
i) Pre-visit logistics

Once data collection begins, each team of assessors (typically two individuals if only one grade is being assessed) will administer the EGRA at one school per day. In some contexts, a day for traveling between schools may be required. Whenever possible, the team should lodge in the vicinity of the school the night before so they can arrive at the school as soon as it opens for the day (if not a few minutes earlier). The team should have all materials ready and organized for the day. This includes making sure that any electronic data collection devices have enough battery power to last through the entire school visit.

ii) Introductions and setup

The first order of business upon arrival at the school is to greet the head teacher or principal. Although school administrators will be expecting them as earlier contact should have been made (to verify the sample assumptions), teams should have a copy of an official letter that explains the study and indicates the entity commissioning the work—typically the MOE. When meeting with the head teacher, the team will need to explain the pupil sampling protocol they will follow as well as what they will need from the school in order to administer the assessments.

After meeting with the head teacher and before selecting the pupils, the team should set up the assessment space. Ideally, there will be an empty classroom at the school where assessors can conduct the EGRA in opposite corners. A table or desk and two chairs or benches will be needed per assessor so that he or she can sit opposite the pupil. In some contexts, there is no such empty room available or the room is adjacent to a noisy classroom, making it difficult for the assessor and pupil to hear one another. Carrying tables and chairs to a shady place a short distance from the school is also an option. Whatever the location, it is important that no one else be present to observe the assessment, such as the teacher, head teacher, or other pupils, as this could intimidate or distract the pupil and impact performance.



iii) Pupil sampling

The EGRA requires a sample of pupils only from the grade of interest at each school. Usually, ten pupils are selected at random. In many countries, the gender ratio in the school population is fairly balanced such that it is not necessary to sample boys and girls separately. In such cases, even though more boys will be randomly sampled in some schools while more girls are sampled in others, when all of the data are collected, the gender ratio in the sample will correspond to that of the overall population. If the difference in the ratio of gender representation for a population is greater than 45/55, it is a good idea to stratify by gender. In this case, an assessor will create two lines of pupils, one for girls and one for boys, and use the established protocol to sample from each line.

To sample the pupils, an assessor should gather together all of the pupils in the grade who are in attendance at the school on the day of the visit and ask them to form a line. No particular ordering is needed. In some contexts, the assessment team administers a pupil interview questionnaire to each pupil after finishing the EGRA. Typically, the questionnaire includes items in which the assessor can note what reading materials (textbooks, exercise books) the pupil has. If this is the case, it is important that all pupils bring their materials with them to the sampling line.

The assessor will count the total number of pupils and divide the number by ten (the number of pupils to be assessed). This will provide the sampling interval, which will be used to select the pupils out of the line. For example, if there were 38 pupils in the line, the sampling interval would be 4 (round up or down to the nearest whole number). Thus, the assessor would remove from the line the 4th, 8th, 12th, 16th, 20th, 24th, 28th, 32nd, 36th and (going back to the beginning) 2nd child. The assessor should then use the sampling interval to select two extra children who will serve as replacements in the event that a selected pupil refuses to assent to the assessment. Once the pupils are selected, the assessor should write down their names on a sheet of paper that will be destroyed by the assessor after the school visit is finished.[4] The first pupils to be assessed can be identified and the remaining selected pupils can be sent back to their classroom from which they will be fetched individually by name when the time comes. More information about pupil sampling can be found in the 2015 *EGRA Toolkit*.

The EGRA Tooklit, Second Edition can be accessed here

iv) Assessment steps

During the assessment, it is critical that the assessors follow the EGRA protocol exactly as they were trained. This includes establishing good rapport with the pupil before beginning, using the scripted instructions, leading the pupil through examples prior to each subtask and keeping up with the pupil, especially on the timed tasks. There should be no clutter on the desk or table aside from what is needed for the assessment (typically only the

4 Names are collected to reduce the chance that selected pupils are replaced by the teacher, who may misunderstand the nature of the assessment and want to ensure that the 'best' pupils are chosen. However, if one person on the team is able to stay with the pupils while they wait to be assessed, their names do not need to be collected.

pupil stimulus book). If tablet devices are used for electronic data collection, the assessor can rest the tablet on the table but must ensure that the screen is not visible to the pupil.

After each assessment concludes, the assessor should thank the pupil and hand her or him a small gift (a pencil, eraser or sharpener is a sufficiently relevant and inexpensive gift) then fetch the next selected pupil. In the event that a selected pupil sits down but then does not want to take the assessment, the assessor should provide a gift and then fetch one of the replacement pupils who were selected.

After all assessments have been conducted, the team should return the room to its original state, thank the teacher and head teacher and ensure that all information has been collected and is accounted for. If collecting data electronically and if technologically possible, the team should upload the data from the devices to a central server before leaving the school. If this is not possible for connectivity reasons, they should do so as soon as they are able. This will reduce the chance of losing data due to a lost, stolen or damaged device.

6.5 Quality control monitoring during fieldwork

Another advantage of electronic data collection is that researchers can monitor the quality of the data in real time. In most cases, teams are able to upload the data from each school on the day of the visit. Analysts can be assigned to perform a series of quality control checks on the data on a regular basis, checking for missing or irregular-looking data and even monitoring assessor performance. For example, it is possible to look at EGRA data and determine whether an assessor has been rushing the child through the timed tasks or allowing more time than is stipulated in the protocol. All of these checks allow for communications to be made to the assessors so that they can correct errors along the way.

Finally, it is possible and advisable to measure IRR during fieldwork by having the assessors on a team pair up to assess one of the selected pupils together each day. One assessor will take the lead in interacting with the child and the other will sit silently and simply mark responses. Once the data for the day are uploaded, analysts who are monitoring the incoming data can check that the two assessors are marking responses from the same child in the same way and if there are major discrepancies, the assessors can be contacted.

7. CONCLUSION

Schools are microcosms where education policies are put into practice. School-based surveys and the EGRA in particular can yield valuable data on pupil learning outcomes, including what pre-reading and reading skills children have mastered and where they are struggling. Other contextual instruments often administered alongside the EGRA can yield insights into the possible reasons that children are doing as well or poorly as the survey reveals. With thorough preparation and training, data collection teams can be equipped for the inevitable challenges of fieldwork and maintain a high standard for data quality. The data collected can inform policy and practice decisions, providing an evidence base for needed changes in instructional approach, resource allocation or other facets of the education landscape.

REFERENCES

RTI International. (2015). Early Grade Reading Assessment toolkit, Second Edition, prepared for USAID under the Education Data for Decision Making (EdData II) project, Research Triangle Park, NC: RTI. https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&id=929

Conducting an Early Grade Reading Assessment in a Complex Conflict Environment: Is it Worth it?

KARYN BEATTIE
Concern Worldwide

JENNY HOBBS
Concern Worldwide and University College Dublin

ABBREVIATIONS

CEC	Community Education Committees
EGRA	Early Grade Reading Assessment
GCE	Global Campaign for Education
GPI	Gender parity index
IRR	Inter-rater reliability
NGO	Non-government organizations
RTI	Research Triangle Institute
SD	Standard deviation
UNFPA	United Nations Population Fund
WCPM	Words correct per minute

1. INTRODUCTION

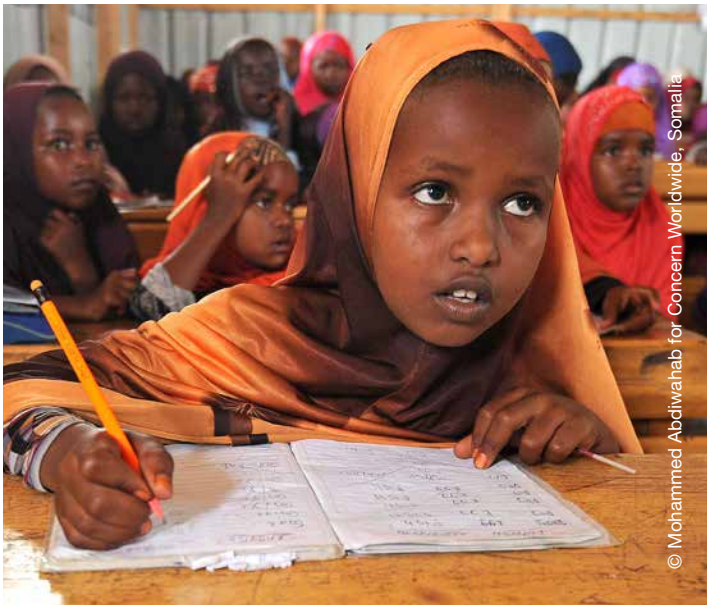
This article documents the experiences of Concern Worldwide in South-Central Somalia. An Early Grade Reading Assessment (EGRA) was conducted in 2013 in collaboration with the Ministry of Education in schools directly supported by Concern in Mogadishu. The assessment had the dual purpose to provide baseline data on which to measure change from a literacy intervention and to draw attention to early grade literacy levels in Somalia. This was the first EGRA administered in this complex, insecure and disaster-prone country. Findings of the assessment are documented here along with the challenges and opportunities that arose. The value of investing in literacy assessments in the context of a country affected by conflict is discussed and recommendations are provided for practitioners, academics and governments considering future EGRAs within similar contexts.

2. BACKGROUND

To most people, Somalia invokes thoughts of a country where violent conflict has been the norm for more than two decades. It is the home of the film-worthy pirates and a region that experiences cyclical famine, producing images of under-nourished children with which the world has become so familiar. In reality, Somalia's people live in a complex, protracted emergency context in which they are forced to rely on humanitarian assistance and a fledgling government overwhelmed by the challenges. Within this environment, a sense of resilience and adaptation can be seen, especially in classrooms where families have succeeded in securing access to education for their children in spite of chronic conflict and poverty.

Since the collapse of the Siad Barre government in 1991, Somalia experienced on-going conflict which destroyed infrastructure and the institutions that should have provided basic services. In August 2012, the Federal Government of Somalia was established and began the difficult task of addressing the competing priorities created by years of instability.

United Nations Population Fund (UNFPA) statistics from 2014 estimate the population to be a little over 12.3 million people. Of these, approximately 13.4% live in and around the capital Mogadishu. The south and central parts of Somalia, which include Mogadishu, are significantly worse off than the self-declared state of Somaliland in the north and semi-


© Mohammed Abdiwahab for Concern Worldwide, Somalia

autonomous region of Puntland in the north-east where institutions are at least functional. Overall, approximately 73% of Somalis live in poverty and access to basic services, such as water, sanitation, health care and education is extremely limited, particularly in rural areas. For example, only 27% of the population have access to a good water source and just 22% have access to sanitation facilities.

For all of these reasons, life is difficult for children (and adults) in Somalia. Neighboring countries have seen rapid growth in access to primary education, largely due to the expansion of public education services and the introduction of free and universal primary education following commitments made by governments under the Education for All movement. In comparison, Somalia's education ministry supports only 13 schools, all within the capital city of Mogadishu. Most schools are run by non-government organizations (NGOs), such as Concern, UN agencies or private institutions and individuals so there are multiple curricula and no standard exams. This also means that fee-paying institutions are the main service providers for education at all levels, including primary education. Obviously, this presents a serious challenge for poor households, such as those who have been internally displaced or those from marginalised clans who typically do not have the kinship networks to support them. These households struggle to afford basic necessities and education costs may simply be too much for them. In this scenario, girls are particularly affected since boys are usually prioritised for schooling in a household with limited means.

As a result, Somalia has one of the lowest enrolment ratios in primary education in the world. The Global Campaign for Education (GCE) report names Somalia as the world's most difficult place to go to school, and states that Somalia is one of four countries where more than 70% of the population is illiterate.

Despite these challenges, education has traditionally been highly valued in Somalia. Following independence in 1960, a system of formal education provision expanded very quickly with pre-1991 governments investing in the construction of hundreds of schools, training tens of thousands of teachers, and most significantly, investing in national literacy strategies. The Somali language was recognised as a national asset and a Latin script was developed to facilitate widespread cultural and functional use of Somali literacy. The Somali language was widely used within all levels of education, including in the universities that flourished during the 1980s (Abdi, 1998).

"And, as Somalia became an independent republic on 1 July 1960, mass education was promoted as the country's best available venue for socio-economic advancement. As a sign of the times, Abdillahi Qarshe, a prominent Somali singer/composer, buoyantly sang this popular nationalist song:

Aqoon la'anni waa iftiin la'aane
waa aqal iyo ilays la'aane
Ogaada, ogaada, dugsiyada ogaada
O aada, o aada
Walaalayaal o aada".

(Lack of knowledge is lack of enlightenment
Homelessness and no light
Be aware, be aware of schools
And go to schools, go to schools
brothers and sisters, go to schools)
(Afrax, 1994).

The centrality of poetry to Somali culture and history may be one of the reasons the language enjoys such prestige and widespread use in Somalia and surrounding countries. Somali is classified within the Cushitic branch of the Afro-asiatic language family and is the best documented of the Cushitic languages (Lewis, 1998; Lecarme and Maury 1987). Pitch is phonemic and changes of pitch are used for grammatical (rather than lexical) purposes (Saeed, 1999). Somali is an agglutinative language, generally following a subject-object-verb structure.

Evidence of written Somali has been found dating back to the late 19th century and many written scripts have been used, including Arabic script and Wadaad writing (Ministry of Information and National Guidance, 1974). In 1972, the Somali Latin alphabet was adopted as the official script and is now most widely used. All letters of the English alphabet are used except p, v and z. There are five basic vowels, which provide twenty pure vowel sounds (front and back variation, short and long versions), and three tones (high, low and falling). There are no special characters except for the use of the apostrophe for the glottal stop. Three consonant digraphs are used: DH, KH and SH. Tone is not marked, and front and back vowels are not distinguished.

Government schools and many schools run by NGOs apply Somali as the medium of instruction. There are many private institutions often referred to as 'umbrella schools' due to their governance structures (several schools under one private service provider) that use other languages including Arabic and English. The EGRA in this article relates only to Somali-medium primary schools.

3. TEACHING AND LEARNING NEEDS

Concern Worldwide has been working in the education sector in Somalia since 1997, supporting the poorest and most marginalised children. Programmes currently operate in Mogadishu and Lower Shebelle—areas vulnerable to recurrent conflict and mass displacement in addition to famine and poverty. In keeping with the organizational mandate, the programme has identified a target group of girls and boys living in extreme poverty, unable to pay school fees and unlikely to access education without NGO support. It is important to consider the fluid nature of the conflict and displacement in Somalia when planning for this context—children may arrive in Mogadishu and stay for short-term protection then return to remote villages when conflict subsides. Children may be living in temporary camps or unsafe neighbourhoods and they may be living without the protection of their parents or staying with extended family. Flexible, responsive education services are needed to meet the social, emotional and academic needs of these children and to ensure their protection during their school life.

Accurate statistics on educational enrolment, attendance and retention are difficult to find in Somalia, particularly in the most fragile parts of the country. A household survey conducted in slum areas targeted by Concern's schools in 2012 found that only 8.6% of children were enrolled in school (Concern Worldwide, 2013). More boys are enrolled in education than girls largely due to cultural norms and poverty-related barriers to girls' education. Clearly, gender parity is a major issue but the situation for *both* boys and girls is stark. The age range of students attending Concern-supported schools (primary level) varies considerably—children may need to start or progress in school much later due to poverty, displacement, attendance in Koranic schools prior to entry to primary school or absenteeism during periods of crisis. Demand for entry into alternative basic education (accelerated) classes is high but a high acceptance of older learners is also a positive feature of education at all levels.

School-level statistics from this EGRA are not representative of the wider population. Concern-supported schools prioritise gender parity and have a gender parity index (GPI) of 0.91 with continued attempts to increase the participation of girls. At the time of data collection for the EGRA (2013), the schools had a mean attendance rate of 83% across the year and a grade retention rate of 73% across primary school. Attendance and retention

can remain high in Mogadishu even following unrest and mass displacement due to the high demand for spaces. Schools in Lower Shebelle may be closed for long periods of time due to conflict or data collection may not be possible due to insecurity so these statistics are more representative of the population attending school in urban slums.

Concern recognizes the role of governments as duty-bearers to the right to education for all and, as such, works in partnership with the Ministry of Education while building capacities. Somalia presents unique challenges to this principle as there is essentially no effective national-level public education system to support. Within this context, the programme must maintain a delicate balance to ensure that the poorest children can access education without setting up parallel education structures that may further weaken government ownership. To counter this, schools are owned and run by Community Education Committees (CEC) under the regulation of the Ministry of Education where possible. CEC members are typically parents and caregivers in addition to community leaders. They are supported to lead all aspects of school management, including teacher recruitment and payment, resource provision and supervision. Resources and training are provided directly to the CEC and school staff on child protection, pedagogy and school management. This has effectively allowed almost continuous support to schools despite recurrent conflict when NGO access is extremely limited. This system of school governance also empowers parents and community leaders to identify contextually and culturally appropriate solutions to issues, such as corporal punishment, gender equity and conflict mitigation.

Concern adheres to a holistic approach to education programming so other services have been integrated into the education programme, including the provision of nutrition, water and sanitation, hygiene promotion and a school health initiative. The main objective, however, is to increase children's access to good quality education.

4. THE EARLY GRADE READING ASSESSMENT

In 2012, the education team in Mogadishu met with Ministry of Education staff and partners to plan the next phase of programming. The need for stronger data on student learning was identified, particularly on early grade literacy skills. It was decided that an EGRA would be jointly conducted by Concern, partners and the Ministry of Education to serve two purposes. Firstly, it would provide baseline data against which to measure change from a literacy intervention being rolled out in Concern-supported schools in 2013. Secondly, the data would identify learning gaps among students, providing an evidence-base for the need to strengthen early grade literacy pedagogy across all schools.

The first EGRA was conducted in 2013 in Grades 2, 3 and 4 in Concern-supported schools in Mogadishu. At the time of the assessment, six primary schools in Mogadishu and 19 community-based schools in the region of Lower Shebelle in southern Somalia were supported by Concern Worldwide. It was intended that all 25 schools would be included in the EGRA but this was not possible due to security constraints in Lower Shebelle. Security restrictions limited the movement of NGO staff and support was provided to teachers remotely by phone. In addition, at the time, an armed opposition group had issued warnings to all citizens against using mobile data on phones or tablets. This made it impossible to transport tablets, which are used for data collection, into the Lower Shebelle area that was controlled by an armed opposition group to the Federal Government at the time. The team decided to limit the assessment to five (of the six) Concern-supported schools in four districts in Mogadishu. The sixth school had recently opened and was still enrolling children so it was not included.

The assessors were selected from Concern staff, teachers and Ministry of Education staff. In total, 16 people were trained and the best 10 were used to conduct the assessment. The training took two weeks: The first week covered phonics training and

Box 1: Tangerine software platform

Tangerine is a versatile software that runs on devices using the Android operating system. Tangerine was developed by the RTI specifically to allow reading assessments to be captured digitally. In the case of the EGRA in Somalia, Concern used the Tangerine platform to create a digital version of the Somali EGRA and to store the results. Concern used Samsung tablets with seven-inch screens for the assessment. Concern's experience using Tangerine in other countries proved it to be more efficient than paper-based tests. Data can be reviewed daily to spot any problems and to address them. Also, there is no need for lengthy data inputting, and data analysis is less time consuming based on the way Tangerine reports the data set.

Source: UN Economic and Social Council Statistical Commission. December 17, 2015.

the second, the use of the tablets and Tangerine (*see **Box 1***). The phonics training was particularly important since phonics has not been taught in Somalia since the late 1960s which means that most of the assessors had not learned to read using phonics. The training was crucial to ensure that the assessors developed some consistency in their own pronunciation. An essential tool was the use of an inter-rater reliability (IRR) test, to ensure high degrees of agreement among assessors. The use of Tangerine and the tablets was covered in the second week. One observation from this section of the training is that although the participants learn to use the tablets fairly quickly, it takes much longer for them to master conducting an EGRA using the tablet.

It is important to note that the results of the assessment could not be generalised to other schools in Mogadishu for two reasons. First, because only schools supported by Concern were included and secondly, there is considerable variance between the different schools.

Random sampling was used to identify participants in Grades 2, 3 and 4. Even though these grades had an average GPI of 0.91 (indicating that there are more boys than girls enrolled in these grades), it was decided that the sample should have equal numbers of girls and boys to allow analysis of results disaggregated by sex. This was particularly important due to the gendered barriers to education in Somalia. Typically, EGRAs are conducted at the end of an academic year of Grades 1, 2 and/or 3. However, due to a number of delays, this EGRA was conducted at the start of the academic year and, therefore, the sample included students from Grades 2, 3 and 4.

Based on the total number of students in these grades, the sample required was 400 students—34 additional students were added to compensate for records that may have needed to be removed due to errors in administration.

Random selection was conducted using the attendance registers on the day of the assessment. The EGRA tool allows each student the opportunity to opt out of the assessment, however, all students gave their consent. The sampling strategy also took into account the schools' double shifts. The breakdown of students sampled from each grade is provided in **Table 1** by grade and shift.

TABLE **1**

Number of students assessed by grade and shift

	Shift 1	Shift 2	Total	Percentage of total number of students in the 5 Concern schools
Grade 2	105	109	214	12%
Grade 3	66	44	110	9%
Grade 4	67	43	110	13%

Source: Concern Worldwide, 2013

5. ASSESSMENT DESIGN

The EGRA was designed to assess children's reading skills. When administered as a baseline, results can inform policy and programme planning and can also provide a reference point against which changes can be measured for future assessments.

The EGRA version used had been developed by the Research Triangle Institute (RTI) International and the Ethiopian Ministry of Education with funding from USAID for use in Ethiopia with a Somali-speaking population. With their approval, the instrument was adapted for use in Mogadishu and rendered into Tangerine (see Box 1). The Ethiopian EGRA included six subtasks—letter sound identification, familiar and unfamiliar word reading, reading passage fluency, reading comprehension and listening comprehension. This was analysed by a team of educationalists in Somalia to refine it for the purposes of this assessment.

Three key changes were made to the instrument:

1. A thorough review of the reading passage and questions was conducted to ensure alignment with the dialect of Somali used in Mogadishu and to adapt to the context where needed. This resulted in minor word changes, spelling changes and some sentence restructuring based on word changes.

2. Two subtasks were added. An additional reading passage was designed that was aligned to expected reading skills of children in Grade 5 with an accompanying comprehension sub-task. This was added to avoid the ceiling effect of students who read the initial passage with ease, which was predicted by educationalists in Mogadishu due to the variance in educational standards across schools. This would provide a more accurate data set on children who may be in multi-age classes or who may have attended different schools due to displacement and therefore be above the level expected at that grade. It was anecdotally known that some proficient readers were in lower grade classes but there was no evidence on how proficient they were or how many children fell into this category. To avoid asking children to read the second passage if they struggled with the first, which would cause stress for the child, skip logic was written into the testing programme. This meant that children who could not correctly answer three comprehension questions from the first passage would not be asked to read the second passage at all—the test would automatically skip to the end. This was just one of the advantages of using Tangerine.

3. To minimise the time of individual testing and to compensate for the new subtasks, the number of subtasks overall were reduced. The revised instrument included six subtests: letter sound identification, invented word reading, oral passage reading (levels 1 and 2) and reading comprehension (levels 1 and 2) (*see* ***Table 2***).

For all subtasks a skip logic is in-built to ensure that children who cannot correctly read a specified number of consecutive items (ten letters or seven words) can stop the task and move to the next.

TABLE **2**

EGRA Instrument: subtests

Instrument subtask	Skill demonstrated by students' ability to:
1. Letter-sound fluency	Say the sound of each letter fluently. Children were presented with commonly occurring letters in a timed test. This was scored as the number of letter sounds said correctly per minute.
2. Invented word oral reading fluency	Process words that could exist in a given language but do not. These are invented words and hence unfamiliar to children. The objective of using non-words is to assess the child's ability to decode words fluently and efficiently. This subtest is measured by counting the number of invented words read correctly per minute.
3. Connected-text oral reading fluency	Read a grade-appropriate text. This is a timed test measuring words read correctly per minute.
4. Reading comprehension in connected text	Answer several comprehension questions based on the passage the pupil read in sub-task three. Five questions are asked and the score is provided as the percentage correct. Questions were factual based on information provided in the text with no inference required.

This is a standard feature of all EGRA tools to avoid undue stress for children who cannot complete a task.

6. ASSESSMENT FINDINGS

6.1 Letter-sound identification

This sub-task assesses children's ability to link sounds to letter symbols and should generally be mastered in the first year of formal school in Somalia. Provided with 60 seconds, we would expect children in Grades 2 and 3 to progress quickly through most or all of the 100 letters in the test with some completing the task with time remaining on the clock. Children's scores in this subtask are low although some progress is seen as children move through each grade (*see* ***Figure 1***). Mean scores for children in Grade 4 were 26 correctly identified letter sounds per minute (SD=9.28). This is higher than scores at the Grade 2 level (mean=17, SD=12.9, $p < 0.001$) but still very low and letter-sound identification does not appear to be a skill children have learned with automaticity after three years in school. Considering that children were not explicitly taught phonics in these schools, the relatively low scores were expected by assessors and were used to inform a phonics-based teacher training course.

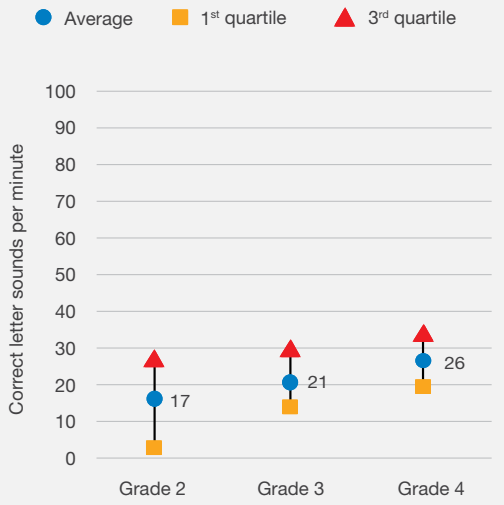
In Grade 2, one out of five students (or 20%) was unable to identify any of the letters correctly. However, in Grades 3 and 4, this percentage drops to 8% and 3% respectively.

6.2 Invented word reading

In the second subtest, children were asked to read invented words and to decode the correct pronunciation based on their knowledge of phonics. Somali is based on a consonant-vowel-consonant structure so this activity provides an opportunity for children to apply their knowledge of letter sounds and blend them together to form words. Invented words are used to avoid testing children's sight vocabulary, restricting the assessment to active blending.

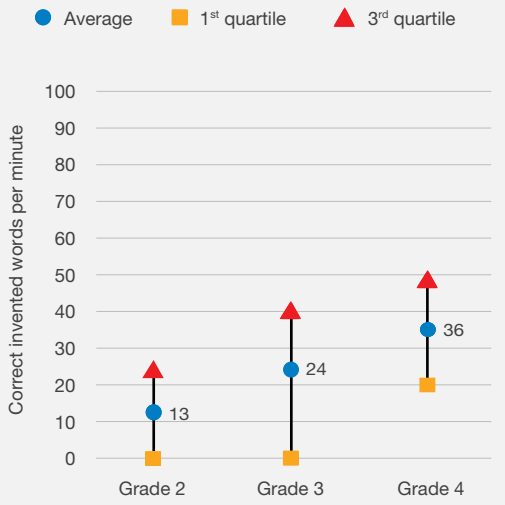
This subtask showed a wide range of competencies within Grades 2 and 3 (*see* ***Figure 2***). For these grades, children at the 25th percentile scored zero as they were unable to read any words presented.

Figure 1. Average scores for the letter sound identification subtest per minute



Note: Figure 1 shows the average correct letter-sound identification scores by grade with the average for the first and third quartile also shown (n=434).
Source: Concern Worldwide, 2013

Figure 2. Average scores for the invented words subtest per minute



Note: Figure 2 shows the average number of correctly processed words by grade with the average for the first and third quartile also shown (n=434).
Source: Concern Worldwide, 2013

Within the same grades, scores at the 75th percentile were 25 and 41 words correct per minute (WCPM) ($p < 0.001$), respectively. Almost half (46%) of children tested in Grade 2 were unable to read any words in the list and following another year in school, one third (31%) were still unable to read any words. Although scores are higher in Grade 4 ($p=0.009$), 13% of children were unable to read any words ($p=0.001$). Scores at the 25th percentile rise to 21 correct words in Grade 4 ($p < 0.001$). Mean scores increase significantly for each consecutive grade, from 13 WCPM in Grade 2 to 24 WCPM in Grade 3 ($p < 0.001$), then to 36 WCPM in Grade 4 ($p < 0.001$). Although this is not sufficiently high in a child's fourth year in school, it does indicate that children's skills in blending are gradually building each year with practice.

6.3 Oral reading fluency

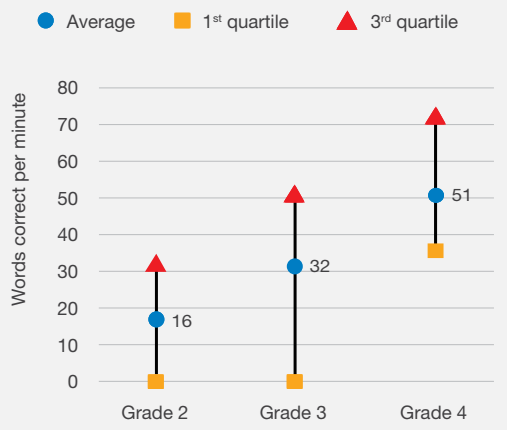
The third and fourth subtests were oral reading passages with related comprehension questions. These are considered to be the most important subtasks in the assessment as the overall target is for all children to be able to engage with real text, read with speed and understand the information within the text. Children's reading speeds are calculated using a timing mechanism built into the survey. At the time of the assessment, in collaboration with education experts within the Ministry of Education and partner organizations, a target was set for Somalia of 60-65 WCPM for children in Grade 3. This target has not been verified to date through more extensive comparative research but is accepted as an interim target until further guidance is available.

All children attempted the first reading passage leveled for students in Grade 2, and consisting of 64 words. Almost half of students (47%) in Grade 2 could not identify a single word. This reduced to one quarter (25%) of children in Grade 3 and 11% of children in Grade 4.

The average number of WCPM for the first passage is shown in **Figure 3**. Mean scores for this task mask wide variance in scores for each grade. The mean score for Grade 2 students (16 WCPM, SD 20) needs to be considered against the median score of just 2 WCPM. This demonstrates that the vast majority of children are unable to read the text although a few children can read quite well with a maximum score of 78 WCPM, bringing the average score upwards. It is therefore important to look at scores at the first and third quartile to better understand the spread in ability throughout each grade.

Scores at the 25th percentile for both Grades 2 and 3 are zero (0 WCPM), which is contrasted by a score of 35 WCPM at the 25th percentile in Grade 4 ($p=0.002$). Steady growth can be seen in scores at the 75th percentile—from 33 WCPM in Grade 2 to 52 WCPM in Grade 3 ($p < 0.001$), rising to 73 WCPM in Grade 4 ($p < 0.001$). This shows that while there remains a group of children who cannot read any words at all grades, this group becomes smaller as the grades progress and the majority of students make some incremental progress each year. This is a positive finding—the education system (within this small number of schools) has the capacity to meet the needs of most learners but does so at a delayed pace and children are not meeting targets at the appropriate grade.

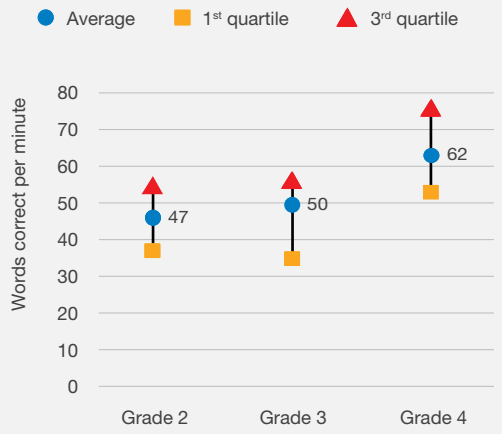
Figure 3. Average WCPM for the first oral reading passage subtest



Note: Figure 3 shows the average WCPM for the first oral reading passage by grade with the average for the first and third quartile also shown (n=434).
Source: Concern Worldwide, 2013

One in five students (22%) progressed to the second reading passage, which was more challenging as it consisted of 143 words. Of the 93 students tested, the vast majority were in Grades 3 and 4. As would be expected, the students who qualified to participate in this subtask were proficient readers (i.e. they had successfully read the first passage and responded to at least three comprehension questions correctly). Mean scores for this test were 47, 50 and 62 WCPM respectively for Grades 2, 3 and 4 (*see* ***Figure 4***). This reinforced the perception that there are a small number of proficient readers in each grade as predicted by the team designing the assessment. Documenting evidence that these children are in the minority has provided ways for the education programme to challenge teaching practices that might 'teach to the top'—leveling teaching to align with the high performing children without sufficient differentiation in teaching strategies to engage struggling and emerging readers.

Figure 4. Average WCPM for the second oral reading passage

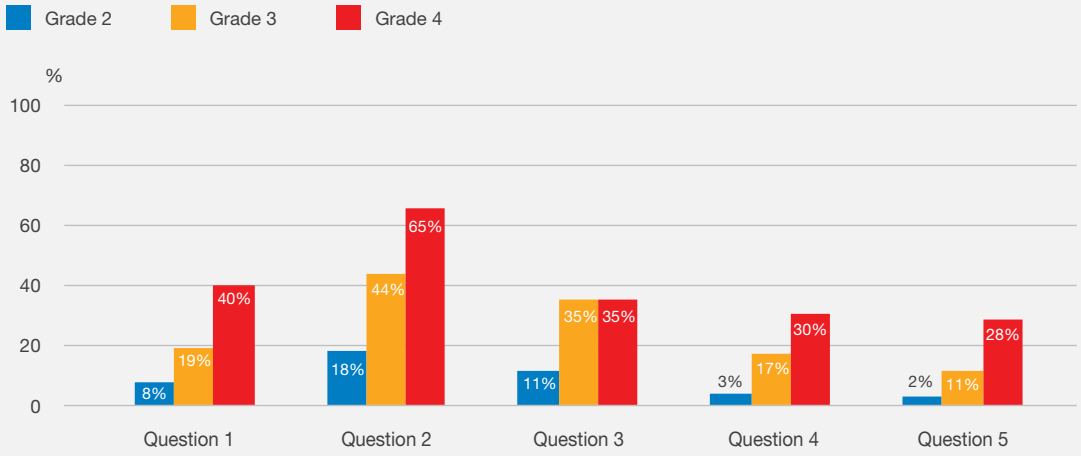


Note: Figure 4 shows the average WCPM for the second oral reading passage by grade with the average for the first and third quartile also shown (n=434).
Source: Concern Worldwide, 2013

6.4 Reading comprehension

For both oral reading passages, there were corresponding comprehension questions. This is an essential part of the reading assessment as it signifies that children can generate meaning from a given text. **Figure 5** shows the percentage of children per grade who were able to correctly answer the questions corresponding to the first oral passage. It shows that across all grades, there were children able to correctly answer all five questions

Figure 5: Percentage of children who correctly answered the comprehension questions for the first oral reading passage*



Note: * Between Grade 2 and Grade 3, difference in percentage who answered: Q1: significant (p=0.003); Q2: significant (p < 0.001); Q3: significant (p < 0.001); Q4: significant (p < 0.001); Q5: significant (p=0.001). Between Grade 3 and Grade 4, difference in % who answered: Q1: significant (p < 0.001); Q2: significant (p=0.001); Q3: not significant; Q4: significant (p=0.026); Q5: significant (p=0.001).
Source: Concern Worldwide, 2013

but that the majority of children in all grades could not respond correctly to any questions.

Conclusions to be drawn from this subtask are limited by the high numbers of children unable to read the passage. The results do not reflect the language comprehension of students—a child cannot be assessed on their comprehension if they have not read the story themselves and it is not read to them (the previous subtest is reading comprehension). For this reason, the utility of data from this subtask is limited, particularly for Grades 2 and 3.

For the questions relating to the second reading passage, no child was asked more than three out of a possible six questions before the exercise was discontinued.

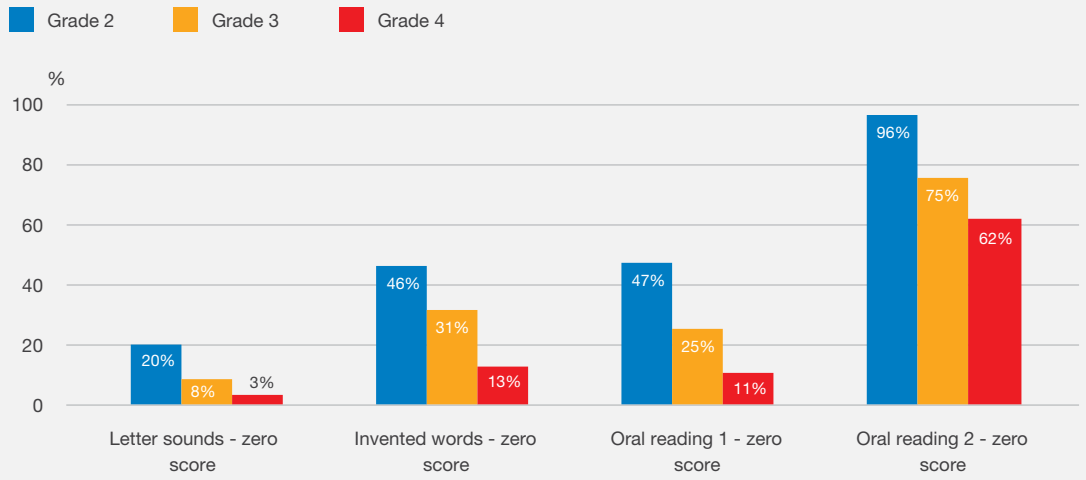
7. CONCLUSION

Overall, assessment scores for children in Mogadishu were low although not as low as might be expected given the challenges. While there are children unable to succeed in basic literacy tasks, such as letter-sound identification or reading short passages, this number decreases through each grade. This is illustrated in **Figure 6**. As expected, fewer children struggled with the identification of single letters than reading full words or texts.

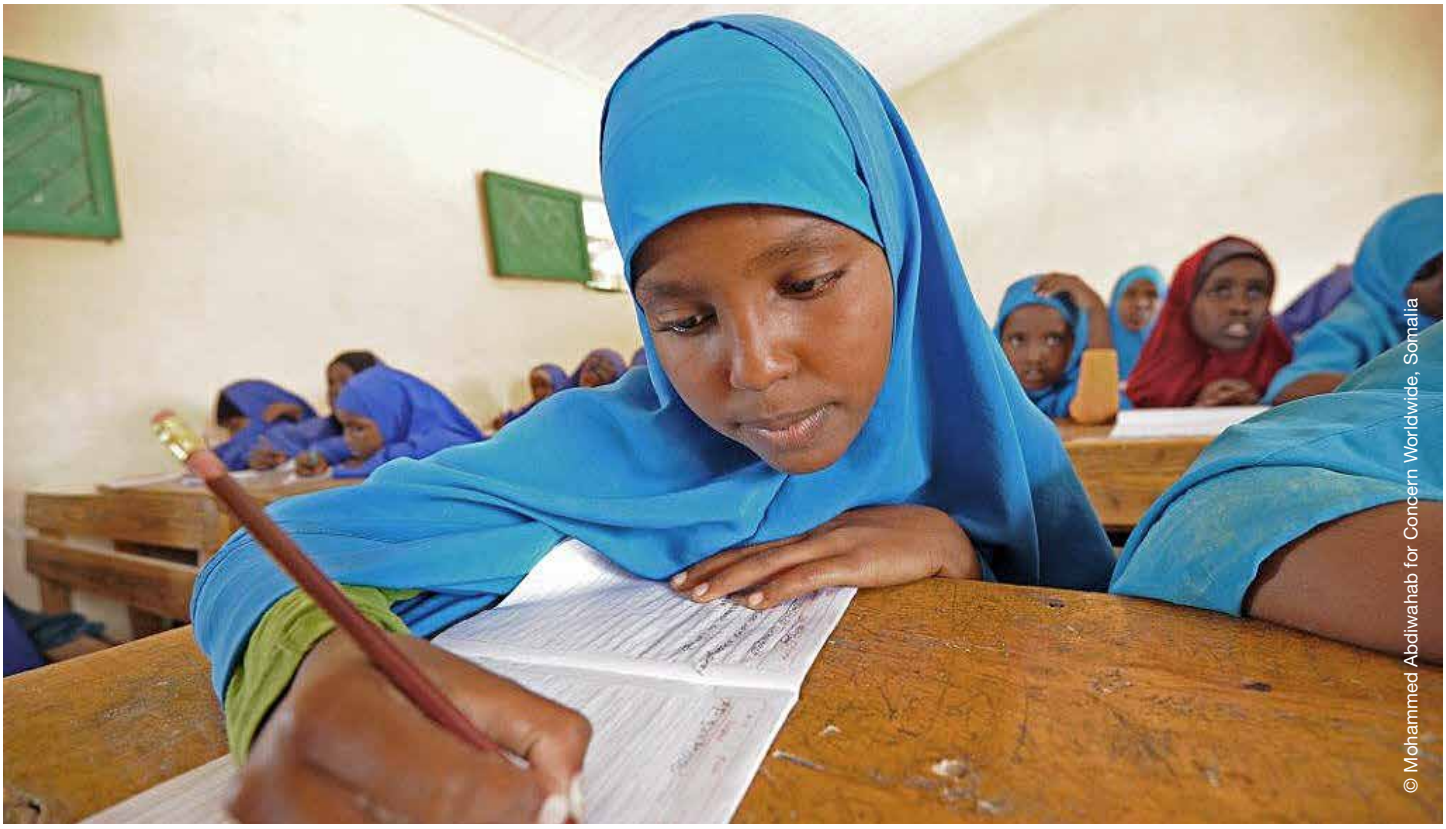
The assessment identified foundational skills for reading that children were not acquiring in Somalia —letter sound knowledge, skills for blending sounds to decode unfamiliar words and oral reading fluency. By designing and administering Concern's EGRA in Somalia, the team was able to work directly with teachers and partners to plan appropriate solutions to address learning gaps. This provided new opportunities for policy dialogue, in-classroom support, planning for materials provision for early grades and generally recognizing early grade literacy needs. In a context of such turbulence, it is essential to provide as many opportunities to gather evidence on learning, facilitate informed discussions and identify mechanisms that might work in this context.

Concern's EGRA has become an annual process since 2013. It is used to test literacy interventions that can inform Ministry of Education practices (and partners) in other schools as well as curricular development. A phonics-based teacher training course is currently being piloted in target schools along with the introduction of new reading materials aligned to meet emergent readers needs and the

Figure 6. Percentage of children scoring zero per subtest by grade



Source: Concern Worldwide, 2013



introduction of in-classroom coaching for early grade teachers. Should security allow, an EGRA is planned for schools in Lower Shebelle in the next phase of the programme.

7.1 Successes and challenges

Conducting the EGRA was particularly challenging for a number of reasons:

- Since Concern was the first organization to conduct an EGRA in Somalia, there were no opportunities to build on tools already existing or to draw from existing expertise in the country.

- Although access was possible in the five schools in Mogadishu, this did not extend to international staff, meaning that the staff member responsible for managing the assessment was not able to accompany the assessors. However, the use of digital devices to capture data contributed to reliable data collection since data could be reviewed at the end of each session (morning and evening) and feedback given directly to the assessors. This would not have been possible with a paper-based assessment.

- Training the assessors to conduct the EGRA required more than simply training them in the specifics of the EGRA because most of them had never been taught Somali phonics—it was removed from the school system in the late 1970s. Concern had to find an experienced Somali phonics teacher, which proved extremely difficult and caused a three-month delay in conducting the assessment.

- The low internet speeds in Mogadishu presented challenges when downloading the data and in particular, when an amendment needed to be made to the instrument. In order to download the amended application, one of the Concern staff had to go to a hotel in the evening to gain access to a better Internet connection. As innocuous as this may seem, it had implications for this staff member as he was in a city where the security was volatile and movements after dark were not recommended.

- The RTI team provided invaluable support throughout the process but because of the time difference between Somalia and the United

States, there were unavoidable delays in resolving problems.

These constraints were not easy to overcome and within such a volatile context, there was no guarantee that it would be possible. However, the outcome was quite significant. Apart from gathering robust data on which to develop a literacy intervention, the process had a big impact on the Concern team, the teachers and even the Ministry of Education staff who participated. The process of listening to young children's reading skills, analysing and discussing the results, brought about a dramatic change in focus for Concern staff and partners. Through this process, the team was quickly convinced of the need to prioritise phonics and increase support to early grade literacy instruction and they became strong advocates for foundational literacy prioritisation. This change was not unique to Concern's team in Somalia—the same transformation and re-focusing was seen in all eight countries in which Concern conducted EGRAs over a two-year period.

8. RECOMMENDATIONS

- Learning to use Tangerine and the tablets can be quick but using the tablet to conduct an EGRA with a primary school child takes time. It is essential to build in time for the assessors to practice with children either at a school or by bringing children to the workshop. The assessors were trained over a two-week period.

- It was extremely difficult to find a Somali language phonics expert to train the assessors. In contexts where this may also be a problem, it is advisable to start the search as soon as possible.

- The assessors who were mostly teachers were initially resistant to conducting an EGRA, arguing that the children had not been taught phonics. It took time to convince them that it was a worthwhile exercise and would act as a baseline from which a treatment could be planned. In some cases, the assessors were clearly nervous that results would reflect badly on them. Taking the time to address these concerns is crucial.

- Perhaps the greatest tool that can be used in training the assessors is the IRR test. An IRR test should be included before the end of the training as it clearly shows the importance of uniformity in scoring and the impact on the robustness of the data.

- In settings like Somalia, we have the tools but not the access to conduct assessments. For example, the EGRA was only conducted in Mogadishu. In future, Concern would like to expand assessments to the schools outside of the city where we have worked for many years but where access for assessors is limited by insecurity. This is similar to our programmes in Afghanistan, Niger and places like Sierra Leone that were temporarily deemed Ebola-affected areas. In these settings, using an EGRA as a baseline is risky because access is unpredictable and therefore subsequent assessments may not be possible.

- Our aim is to develop tools that could be used more easily in these insecure contexts. By necessity, they may be not as rigorous and we may have to compromise on sample sizes and on supervision but we feel that until we are administering assessments in these contexts, we are missing some of the poorest and most vulnerable children. We need to know where the gaps are and provide commensurate support[1].

- Education in emergencies cannot just be about the provision of inputs and hardware because as we have seen in Somalia, conflict can go on for decades. In these difficult and complex environments, measuring learning outcomes is problematic. There seems to be a common misconception that the EGRA can be used to measure a child's progress but this is not the point of an EGRA. The EGRA is a tool to assess

1 We have varied the tools used in different contexts. For example, where governments have existing tools and they are satisfied with those, we will use those tools instead.

the efficacy of the education system in teaching literacy and should never serve as a high-stakes test of individual children. This needs to be made clearer. If teachers misunderstand the purpose of an EGRA, they are likely to misinterpret the figures or even misrepresent them.

- In emergency contexts where expertise is limited, we need clear and simple guidelines; regular training in administering EGRA and using Tangerine; tools that can be used; and a database of experts who can help. Concern teams benefited greatly from technical support provided by the RTI—much of this provided without cost. However, as the number of organizations conducting EGRAs grows, there is a need for a more robust help-desk function where organizations can pay for sporadic technical support.

REFERENCES

Abdi, A.A. (1998). "Education in Somalia: History, destruction, and calls for reconstruction". *Comparative Education*, Vol. 34, No. 3, p. 327.

Afrax, M.D. (1994). "The mirror of culture: Somali dissolution seen through oral culture". A.I. Samatar (ed.), *The somali challenge: From catastrophe to renewal?* Boulder, CO: Lynne Rienner Publishers.

Concern Worldwide (2013). *Irish Aid Programme Funding Baseline Report.* https://www.concern.net/sites/www.concern.net/files/media/page/concern_iapf_2013_rfs.pdf

Lewis, I. (1998). *Peoples of the Horn of Africa: Somali, Afar and Saho*. Trenton, NJ, USA: Red Sea Press.

Lecarme, J and Maury, C. (1987). "A software tool for research in linguistics and lexicography: Application to Somali". *Computers and Translation*, Vol. 2, No. 1, pp. 21-36.

Saeed, John (1999). *Somali*. Amsterdam: John Benjamins Publishing Company.

Ministry of Information and National Guidance (1974). *The Writing of the Somali Language*. Somalia: Ministry of Information and National Guidance.

Administering an EGRA in a Post- and an On-going Conflict Afghanistan: Challenges and Opportunities

HOMAYOON SHIRZAD AND AINE MAGEE
Concern Worldwide

ABBREVIATIONS

BEACON	Basic Education for Afghanistan Consortium
EFA	Education for All
EGRA	Early Grade Reading Assessment
IRC	International Rescue Committee
RESP	Rural Education Support Programme
WCPM	Words correct per minute

1. COUNTRY CONTEXT

In Afghanistan, about 3.5 million school-age children are out of school of whom around 75% are girls (Ministry of Education, 2013). Decades of prolonged war and violence in Afghanistan has taken a substantial toll on many services in the country and the public education system is no exception. Although great strides have been made since 2001 to improve the education system, access to quality and safe education remains a challenge. According to the Ministry of Education statistics in 2008, the net enrolment rate for primary school children was 52% (42% girls and 60% boys). In 2010 in the Badakhshan district where Concern works, it was estimated that 32% of boys and only 13% of girls completed primary school (United Nations Assistance Mission Afghanistan, 2010). The result is that adults in Afghanistan receive on average 3.1 years of education throughout their entire lives.

In Afghanistan, conflict and insecurity are key obstacles to universal access to education. Hundreds of schools have been destroyed or remain closed following prolonged conflict. In particular, the rural areas of Afghanistan still lack basic physical school infrastructure despite the prioritization by the international community for the reconstruction of schools since 2001. Natural disasters, such as flooding, heavy rain, avalanches and earthquakes have also taken their toll on existing schools. In 2010, the Ministry of Education reported that nearly half of the existing schools do not even have buildings and many existing buildings are too damaged or unsafe to use (Ministry of Education, 2010).

Poverty also plays a role in blocking access to education, especially for girls as forced or early marriage is seen as a way to alleviate economic pressures through additional income in the form of a bride price. In most families, children are an integral part of the household livelihood strategy either by providing support through income generation activities (farm work or work done within the home, such as carpet weaving), seeking formal employment, begging or other means (Jackson, 2011).

Another huge issue in relation to access to education for girls are cultural beliefs and local traditions that are biased against the education of girls. In Afghanistan, the ratio of girls to boys in school (gender parity index) remained at 0.7 between 2011 and 2013 (Ministry of Education, 2010; World Bank, 2013). This lack of access to education for girls is reflected in literacy statistics where the estimated national adult literacy rate

(aged 15 years and above) for males is 50% whereas it is only 18% for women. In rural areas, the situation is even bleaker—an estimated 90% of women and 63% of men cannot read, write or compute (Ministry of Education, 2012).

The terrain in Afghanistan also presents a challenge in terms of access. Villages in mountainous regions are very remote and isolated and often do not have a school. Children have to walk long distances over difficult terrain to get to the nearest school and there are valid concerns for their safety as they face the risk of landslides, flooding, kidnapping or attack by wild animals (Kiff, 2012).

In functioning schools, there are concerns about the quality of education with recurring problems of illiteracy among children who have attended school for many years. Teachers in remote schools have few qualifications and the pupil-teacher ratio for Grades 1-3 is 180 children per teacher. This unmanageable number of students per teacher is evidence that the country still lacks enough qualified and motivated teachers to deliver a quality education (Ministry of Education, 2014).

Another recurring problem is the actual time spent by students in the classroom. In Afghanistan, schools are often closed for extended periods due to insecurity, conflicts and natural disasters thus limiting the contact time between teacher and student and consequently, reducing the opportunity to learn. The attendance rate for boys is 64% while it is much lower for girls at 48% (UNICEF, 2008). This shows that even for the 40% of girls lucky enough to be enrolled in schools, they miss out on half of the education curriculum due to low rates of attendance.

Although international education goals have not yet been achieved, the Islamic Republic of Afghanistan has a strong commitment to Education for All (EFA) and the government has endorsed sector policies and strategies to move towards providing all children and adults with relevant quality education (Ministry of Education, 2010). To deal with the issue of access to education in remote locations, the Ministry of Education, with the support of international organizations and donors, has initiated the Community Based Education system. This system includes community-based schools in remote locations and an accelerated learning programme for out-of-school youth. Concern is working to support the Community Based Education system through its innovative Rural Education Support Programme (RESP).

2. CONCERN WORLDWIDE EDUCATION PROGRAMME OVERVIEW

Concern Worldwide has been supporting education in Afghanistan since 2001, mainly through school renovations. However, the RESP which started in 2012 changed the focus of the programme to specifically support access to education for marginalized children. This is in line with Concern Worldwide's strategic goal to increase education provision for the poorest groups of society with a specific focus on female participation. The RESP seeks to improve access to education for boys and girls through the establishment of a quality community-based education system that aims to improve learning outcomes in four districts in the Takhar and Badakshan provinces.

Currently, Concern is supporting six government hub schools and has established 22 community-based schools with the enrolment of more than 500 children (52% of whom are girls). These community-based schools are located in remote villages so that the problem of distance, accessibility and safety are simultaneously addressed for local children. Schools are equipped with latrines and safe water supplies. As an interim strategy to support the expansion of the Ministry of Education's reach in rural areas, teachers' salaries are paid and textbook, stationary and basic furniture are provided by Concern. Staff and students also have access to psycho-social support activities; training on children's' rights, HIV and AIDs; and disaster management. The main focus of the programme, however, is on improving children's learning in these schools, specifically by increasing teachers' capacity to teach literacy in early grades through training and on-going coaching and support.

TABLE **1**

EGRA subtasks and corresponding skills assessed

Instrument subtask	Procedure to assess skill level
1. Letter-name fluency	Students are provided with 11 rows of 10 random letters to be read from right to left, from top to bottom (direction of Dari). The individual student is instructed to read aloud as many letters of the alphabet as they can within one minute. Students are scored on how many letters were correctly identified by name within one minute.
2. Letter-sound fluency	Students are provided with 11 rows of 10 random letters to be read from right to left, from top to bottom (direction of Dari). The individual student is instructed to say the sounds of each letter, in order. Students are scored on how many letter sounds they correctly identified within one minute.
3.Invented word oral reading fluency	These are made-up words and hence unfamiliar to children. The objective of using non-words is to assess the child's ability to decode words fluently and efficiently. This subtask is measured by counting the number of invented words read correctly per minute.
4.Connected-text oral reading fluency	Students are asked to read a simple story. This is a timed test measuring connected-text words read correctly per minute.
5.Reading comprehension in connected text	Provide correct responses to five comprehension questions based on the story read in subtask 4. Assessors ask each question orally and students are required to respond orally. The score is provided as the percentage of correct responses out of five questions.

Source: Concern Worldwide (2014)

3. EARLY GRADE READING ASSESSMENT IN AFGHANISTAN

A focus on literacy as a key foundation skill necessary for all future learning is a vital component of the RESP. It was thus essential to ensure that the support provided to teachers on teaching literacy responded to the needs of children in this regard. To assess learning needs and ultimately to enable measurement of learning outcomes in the community and hub schools, the Concern team invested in an extensive assessment of reading levels in these target schools. Results of this Early Grade Reading Assessment (EGRA) were used to guide the design and delivery of the teacher training component of the programme and to support teachers when addressing the fundamental components of literacy in their classes.

The EGRA was conducted in early September 2014 among 323 children (180 boys and 143 girls) in Grades 1, 2 and 3 in both community based (58 boys and 60 girls in Grade 1) and government hub schools (81 boys and 17 girls in Grade 2 and 41 boys and 66 girls in Grade 3). The test was composed of five subtests, all in the Dari language as outlined in **Table 1**.

The results of the assessment were quite alarming. Average scores for the letter naming test were 19, 30 and 29 letters correct per minute in Grades 1, 2 and 3 respectively with no significant difference between boys and girls in any grade. Unfortunately, there were still a large proportion of children who scored zero on this simple subtest—30% in Grade 1, 27% in Grade 2 and 19% in Grade 3.

The letter-sound identification subtask yielded even worse results. Among Grade 1 students, 60% could not identify a single letter sound. Likewise, 41% of Grade 2 students and 50% of Grade 3 students scored zero on this subtest. Again, there are no significant differences between boys and girls. Research indicates that the knowledge of letter sounds is a key learning skill for children to decode letters into words so this finding is particularly worrying.

As a result of low awareness of letter sounds, decoding of invented words was weak as well among students across all grades. Grade 1 children decoded an average of 1.8 words correct per minute (WCPM). There was no significant difference between children's decoding ability in Grade 2 versus Grade 3. Grade 2 students could read an

average of 7 WCPM while children in Grade 3 decoded 5.7 WCPM. Nine out of ten children in Grade 1 scored zero on this subtest. Two thirds of children in both Grade 1 and Grade 2 scored zero on this subtest.

The key subtask in the EGRA is the connected text oral reading. This is a measure of reading fluency—the skill vital to comprehension and ultimately learning in other subject areas. Almost all children in Grade 1 scored zero in the oral reading fluency test. This might be expected since Grade 1 students had only completed three months of schooling at the time of the test. However, it was very concerning that in Grade 2, four out of five children scored zero on this test. Likewise, children in Grade 3 did not fare much better as seven out of every ten children in Grade 3 scored zero on the oral reading fluency test. There was no significant difference between boys and girls in any grade. Unsurprisingly, average scores (measured in WCPM) were very low. The average scores on the oral reading fluency test is zero in Grade 1, 6 WCPM in Grade 2 and 9 WCPM in Grade 3. These scores are far below the government standard of 45-60 WCPM by Grade 3.

Another key learning from the EGRA data was that there was no significant difference (at 95% confidence level) between the scores of children in Grade 2 versus those in Grade 3 in any test. This shows no progression in literacy despite an extra year of schooling. The result also underscores the deficiencies in the education system where students progress through grades without actually developing new literacy skills.

4. CHALLENGES TO CONDUCTING AN EGRA IN CONTEXTS LIKE AFGHANISTAN

While the importance and usefulness of the EGRA in programme design and focus is undisputed, conducting the EGRA in the Badakshan and Takhar provinces of Afghanistan was not an easy task for the Concern team.

Firstly, Concern Worldwide's Afghanistan team themselves had no experience with conducting an EGRA and found it difficult to bring learning from other Concern contexts due to the different script used in the Dari language. The technical and specialized nature of the test resulted in difficulties in sourcing a partner with the expertise to build internal staff capacity. In fact, it took over a year to find a partner to support the process. The International Rescue Committee (IRC) within the Basic Education for Afghanistan Consortium (BEACON) finally fulfilled this role by supporting Concern in the training of assessors, training on the EGRA survey tool, training the data cleaning officers and training Concern key education staff on the process.

As the target schools are located in remote areas, it was difficult to recruit staff to undertake the assessments. The assessors recruited had to undergo extensive training in phonetics and phonics as they themselves had not learned to read in this way.

Having completed the assessment itself, it was then difficult to source external support to clean and analyse the data and produce the EGRA results in a timely manner. This problem was confounded by the fact that the EGRA was conducted as a paper and pencil test, which added another layer of complexity by creating the need to source data clerks to diligently and accurately enter the data into an appropriate format for analysis.

There were also many logistical issues. The team administering the test experienced delays and threats to safety due to the presence of the Taliban in the target communities. Moreover, the timing of the EGRA coincided with a period of heavy rain and flooding. Bad weather combined with remote destinations resulted in transportation issues.

Insecurity also negatively impacted the attendance of children in school, resulting in low numbers available on the day of the test. Some children who were present on the day of the test were not used to interacting with strangers and undertaking such testing. The children were very shy and some were

unwilling to interact with assessors—some children started crying before the test was conducted. The team responded with compassion and worked with teachers to reassure students but this level of tension for students is likely to have led to underperformance and should be considered in the interpretation of the results.

Finally, financing such an assessment in these remote locations of Afghanistan has proved very costly[1]. Transportation costs as well as assessors' salaries increase greatly when there are unforeseen delays or interruptions to the assessment plan. This was another challenge the team had to surmount in order to successfully complete the EGRA.

5. OPPORTUNITIES

Despite the challenges, the EGRA assessment provided a valuable opportunity to identify key issues and gaps contributing to literacy problems in Afghanistan. The timing of the first EGRA conducted by Concern was crucial. Having just set up the community based schools, the results of each subtest of the EGRA informed specific components of the teacher training modules. Key issues were identified, including lack of phonics and weak phonological awareness among children as well as more practical issues, such as ensuring children had adequate time in class with a teacher in order to learn. The following are the mains strategies Concern are using to address these issues:

Improving the class environment

To assist students in their literacy development, they need to have a classroom with adequate space for active learning, where teaching and learning materials can be safely stored and where posters reinforcing literacy concepts can be displayed. Concern is supporting classroom provision and rehabilitation of classrooms as spaces conducive to learning.

Improving literacy instruction

What the teacher does with the class during the literacy lesson time is extremely important. Most primary classes are held for 2.5 hours per day over six days. There is not a lot of time to cover all the subjects required in this schedule.

Concern is advocating that teaching time and methods used in early grades should focus on the skills necessary for learning. Foundational literacy skills are prioritized and embedded in classroom activities throughout the day.

Teacher training has now prioritised literacy, which has led to a change of the focus in classes. A curriculum has been developed for a daily 60-minute lesson that focuses specifically on literacy. This is part of the usual timetabled Dari lesson but it has an increased emphasis on phonological awareness and developing phonics which were highlighted as skills that were particularly weak in the EGRA. Concern staff support the teachers in the delivery of this lesson through regular and systematic in-class observation and in-depth feedback.

Ensuring adequate time on task

Teachers and students need to be present for sufficient periods of time for instruction to occur. Learning will happen when students attend class regularly and the teacher reciprocates with punctuality and attendance. Concern is supporting School Management Committees to monitor both student and teacher attendance. The Committees have been trained on a simple monitoring tool to hold teachers accountable for their presence at school and to provide support to teachers when contextual challenges make school attendance difficult (such as ensuring they have a safe place to stay within the community). Attendance records of students are completed daily by teachers and a Child Protection Officer has been appointed to follow up on cases of prolonged absences.

1 US$3,323 for trainings, practice and tool adaptation. US$3,085 for administration in the field to cover food, accommodation, salaries, transport etc. and US$712 for data entry and data cleaning.

Provision of reading materials

Once a student has learned the skill of reading, they need to continually practice that skill in order to become proficient and fluent readers. Access to reading materials is thus essential. The more remote the communities, the more likely it is that families will be illiterate and that fewer reading materials will be available. Concern provides reading material to the most vulnerable children to ensure that they have the opportunity to practice skills they learn in school.

Engaging parents

Parents play a crucial role in their child's education. Concern's international research has shown that children who receive support from fathers and mothers make significantly better progress than their peers (Concern Worldwide, 2014). Thus, Concern staff work with parents to ensure that they are aware of their children's progress in education and to encourage them to support learning outside of the school environment, particularly through reading at home or looking at books together.

Monitoring progress

A robust literacy teaching programme must also include an assessment component. Through regular assessment, teachers can keep track of children who may require additional support. Concern is investigating assessment tools to use in Grade 1 to support teachers in this continuous assessment. Concern is also working with other NGOs and the Department and Ministry of Education to design a reading catch-up for children who are not making sufficient progress in literacy classes. Further, the EGRA will be administered at various stages of the RESP to monitor increases in the quality of literacy classes, which should be reflected in increasing EGRA scores in all subtests.

Engaging at the nation level

Having high quality research on early grade literacy through Concern's 2014 EGRA results provides Concern with evidence to use in advocacy efforts at the national level. In particular, Concern is engaging with the Ministry of Education to persuade them to allocate qualified teachers to Grades 1-3 in primary schools. EGRA results that show no progress in the literacy of children in these grades is a key advocacy tool in this regard.

6. THE VALUE OF THE EGRA

For many years, the international community has supported education in complex contexts, mainly through building and rehabilitating infrastructure as well as providing school supplies. Deviating from this familiar approach requires a lot of internal capacity building, time investment, commitment and energy. The challenges of working in a context such as Afghanistan are complex and include overlapping barriers to learning, such as poverty, conflict, gender inequality and food insecurity. To attempt to implement a totally new approach to education programming in such a country is a mammoth task. Yet, the Concern team invested in this new approach due to continued commitment to the belief that every child has a right to a quality education.

Conducting an EGRA was the first difficult step in a long process to addressing the long neglected quality issue in education in Afghanistan. The execution of the assessment itself was fraught with difficulties due to transport, weather and security problems that compounded the existing technical complexity challenges already associated with the administration of the assessment. Yet, the team prevailed to provide in-depth research that highlights the literacy problems in Afghan schools.

Already the EGRA results have been used to guide teacher training and module development and children are benefiting from new methods of literacy instruction. However, this is just the first step. The EGRA results can be used to highlight the education standards in Afghanistan at national and international fora, influencing changes in funding, curriculum and donor support. The baseline EGRA results can be used as a bench mark against which to measure the progress in the programme—not

to measure an individual child's progress but to continuously monitor the new system of instruction introduced and whether it is imparting key literacy skills to children in early grades.

REFERENCES

Concern Worldwide (2014). *Lost for Words: An Analysis of Early Grade Reading Assessments in the Most Vulnerable Communities in Five of the Worlds' Poorest Countries from 2012-2014*. Concern Worldwide. https://www.concern.net/sites/default/files/media/resource/g2569_lost_for_words_report_final_2.pdf

Ministry of Education, Islamic Republic of Afghanistan (2013). Afghanistan country paper: Learning for All Ministerial Meetings. http://planipolis.iiep.unesco.org/upload/Afghanistan/Afghanistan_UNGA_Learning_for_All_2013.pdf

Jackson, A. (2011). *High Stakes—Girls' Education in Afghanistan.* A Joint NGO Briefing Paper. Oxfam. https://www.oxfam.org/sites/www.oxfam.org/files/afghanistan-girls-education-022411.pdf

Kiff, E. (2012). *Contextual Analysis for Concern Worldwide in North East Afghanistan in the Provinces of Badakhshan and Takhar*. Concern Worldwide.

Ministry of Education, Islamic Republic of Afghanistan (2010). *National Education Strategic Plan for Afghanistan (2010-2014)*. Ministry of Education, Department of Planning and Evaluation, p. 3. http://planipolis.iiep.unesco.org/upload/Afghanistan/Afghanistan_NESP_II_2010-2014_draft.pdf

Ministry of Education, Islamic Republic of Afghanistan (2012). *Afghanistan National Literacy Action Plan (2012-2015)*. Kabul: Ministry of Education.

UNICEF. Afghanistan statistics 2008. http://www.unicef.org/infobycountry/afghanistan_statistics.html

United Nations Assistance Mission Afghanistan (2010). "The Education for all Edition". *Afghan Update*, Vol. Summer, No.23, p.7

World Bank data (2013). http://data.worldbank.org/indicator/SE.ENR.PRIM.FM.ZS

Evaluating Reading Skills in the Household: Insights from the Jàngandoo Barometer

DIÉRY BA, MEISSA BÈYE, SAME BOUSSO, ABDOU AZIZ MBODJ, BINTA AW SALL, DIADJI NIANG
Laboratoire de Recherche sur les Transformations Économiques et Sociales (LARTES), Jàngandoo

ABBREVIATIONS

EFA Education for All

LARTES *Laboratoire de Recherche sur les Transformations Économiques et Sociales* (Research Laboratory on Economic and Social Transformations)

PALME *Partenariat pour l'Amélioration de la Lecture et des Mathématiques à l'École*

PASEC *Programme d'Analyse des Systèmes Educatifs de la CONFEMEN* (Analysis Programme of the CONFEMEN Educational Systems)

SNERS *Système National d'Evaluation du Rendement Scolaire* (National Evaluation System of Educational Achievement)

1. INTRODUCTION

Assessment is essential to the education process as it helps measure learners' performance, the effectiveness of implementation strategies and the relevance of defined policies. Assessment happens at different stages of the teaching and learning process—before, during and after any kind of educational activity is undertaken. Aside from learning assessments conducted in educational institutions, there are other forms of large-scale assessments outside the school environment. The results of these assessments are of interest to international organizations as well as state and local authorities. When done effectively, assessments can provide a necessary diagnosis to guide educational policies and ensure their effectiveness in covering certain domains of learning.

Inspired by the ASER experience in India (see article by Banerji) and other countries undertaking similar initiatives (see article by Aslam et al.), the Jàngandoo Barometer, initiated by the *Laboratoire de Recherche sur les Transformations Économiques et Sociales* (LARTES), is an independent citizen-led assessment targeting children aged 6 to 14 years. It is conducted in all 14 regions of Senegal. In each selected household, all children within the target age group are assessed in reading, mathematics and general knowledge (personal development, knowledge of the social and ecological environment and openness to the world).[1]

In the Wolof language, *Jàngandoo* means 'learning together'. The Jàngandoo Barometer is designed to measure the status and quality of learning of Senegalese children. It uses a standard benchmark called the 'median level', which corresponds to the basic competencies that students in Senegal are expected to acquire by the end of Grade 3. One of the goals of the assessment is to underscore the issue of education quality as a key concern for authorities, parents and education partners and to inform the implementation of changes to the education system. The Jàngandoo results

1 Visit the *Catalogue of Learning Assessments* for more information on the Jàngandoo assessment: http://www.uis.unesco.org/nada/en/index.php/catalogue/173

are shared with families, communities, education authorities and other stakeholders.

In this article, we examine key principles related to reading assessments conducted in households within the context of the Jàngandoo Barometer. We seek to answer the following questions: what are the skills assessed in reading? What is the purpose of this assessment in educational terms? What are the approaches used? What are the issues associated with and the contribution of this type of assessment?

The importance of reading

Reading is an indispensable tool and a fundamental skill for further learning. It is the action of recognising, forming mentally or sounding out graphemes, phonemes or combinations of these and attaching a meaning to them.

Reading is evaluated at all levels of basic education, regardless of the type of education provided. While most early grade reading assessments focus on phonological awareness and fluency, attention to reading comprehension is less frequent—despite its importance. In the Jàngandoo Barometer, we believe that reading assessments should measure decoding skills, fluency and comprehension based on the education levels of learners. It is equally important, however, to measure both children's level of knowledge of the mechanisms required for reading and the understanding what one is reading. Reading cannot be reduced to either, or to just deciphering the code or the construction of meaning. Rather, reading is the dynamic implementation of several processes, ranging from micro-processes to metacognitive processes. Giasson (1990) distinguishes five of these processes:

- micro-processes used to understand the information in a sentence
- macro-processes oriented towards global understanding of the main text and using the text structure
- integration processes that serve to make links between clauses or sentences
- construction processes that allow the reader to go beyond the text (i.e. mental imagery, reasoning, etc.)
- metacognitive processes that are used to guide understanding.

Measuring reading skills is one of the main components of the Jàngandoo assessment and the remainder of this article focuses on the measurement of reading skills only.

2. ASSESSING READING COMPETENCIES

This subsection describes the two Jàngandoo Barometer reading tests: the median-level test administered to all children in the sampled household aged 6 to 14 years of age and the complementary test administered to those who perform well on the median-level test.

2.1 The median-level test

A combination of tests are used to measure reading skills using a scale that ranges from simple to more complex levels. First, foundational reading skills such as phonological awareness are assessed, then reading fluency, and finally reading comprehension—all skills that should be acquired by Grade 3. The competencies assessed are categorised in the following levels:

- At Level 1 (phonological awareness), the knowledge assessed is the child's ability to identify and read sounds and syllables. Phonological awareness and the alphabetic principles are assessed through a global acquisition of the letters of the alphabet. These capabilities are the first step in the process of reading skills acquisition.
- At Level 2 (reading familiar words individually or in connected text), the skill assessed is the child's progress in automatic decoding. Reading is not

mere word recognition and the child must be able to read the words in connected text in order to progress as an independent reader.

- At Level 3 (reading comprehension), the skill assessed is the ability to read connected text fluently and fluidly, and to answer a few comprehension questions about the text.

2.2 The complementary test

The results of the Jàngandoo Barometer (Fall et al., 2014) reported that 28% of the 26,014 children assessed achieved the minimum expected equivalent to reading proficiency at a Grade 3 level. Analysing the data revealed that there are children aged 6 to 14 years that have competencies that are superior to the Barometer's median level of performance—yet, the assessment does not provide much information on the different levels or the optimum level of performance achieved beyond that level. In other words, the test is unable to discriminate among the children at the upper-end of the performance scale. Therefore, we sought to answer questions, such as what is the actual performance level of these children who are performing above the threshold? What is the distribution of performance levels beyond the median level?

An additional complementary test was therefore developed to determine the performance threshold and measure the level of attainment by focusing on the actual performance of children beyond the median level. The complementary test is based on a higher rating scale derived from a set of complementary and gradually more difficult reading tasks that a child should be able to perform by age 14 years and certainly by age 16 years. In 2015, the complementary test was administered to all children who completed all components of the median-level test successfully. Beyond providing data on the optimum performance level, the complementary test will provide elements that will further ground the analysis of the performance of children as they progress in their learning. This information will be a capital gain from the results of the median-level assessment of the Barometer as it allows a more thorough analysis of the quality of learning as opposed to the results produced by the median-level test only.

i) Developing the complementary test

A gradual test was developed to achieve this measurement. A test of reading fluency must gradually be more complicated, which can be achieved by increasing the length of the text and by varying the types of text used. The same is true for the questions used to capture the understanding of the text. It should also be ensured that the complementary test respects the taxonomic levels of performance.

The complementary test was therefore structured in four steps of increasing difficulty to correspond with the reading competencies to be acquired at the following levels:

- Level 4 corresponding to Grade 6 (primary education)
- Level 5 corresponding to the second year of secondary education
- Level 6 corresponding to the third year of secondary education
- Level 7 corresponding to the fourth year of secondary education

In addition, the skills assessed in the complementary test refer to the ability to read different types of text fluently and expressively; to exhibit an understanding of various texts (both documentary or literary texts); to understand a statement or an instruction; to identify the main idea of a text one reads or hears; and to be cognisant of one's reading.

ii) Using the results from the complementary test

The results of the complementary test are integrated into the detailed analysis plan. Children's performances are analysed progressively by age and level of performance. Initially, descriptive statistics will be used for analysis as the correlations with

existing variables may be insufficient to explain the performance of children due to the size of the subsample (18% of the children who successfully completed the median-level test). Also, a qualitative analysis could be made in a specific and detailed study that could be conducted later on this sample of children to identify patterns and determinants of performance.

3. CAPTURING COMPETENCIES IN THE CONTEXT OF THE LEARNING CRISIS

Learning to read is a long process that includes several stages acquired through primary education—yet at each step, the targeted competency is the same: constructing meaning. Reading is enriched and tasks associated with reading competencies become gradually more demanding as children proceed through primary education. More specifically, in the early years of primary education, children should be able to read a text aloud and answer simple/literal questions on the text. After two years in school, they should demonstrate oral reading fluency and be able to answer more complex questions. Towards the end of primary education, children should be able to read a text silently, answer questions based on the text as well as make judgements or inferences about the text. In all three 'stages of reading', understanding is the underlying competency that is being sought.

The competency measured in Levels 4 to 7 (described in section 2 on the complementary test) is reading comprehension. The types of questions used to assess reading (in increasing order of difficulty) are:

- Literal questions that require the reader to relate to or extract elements explicitly mentioned in the text.
- Inferential questions that require the reader to provide a response that is not formulated as such in the text. In essence, the reader is asked to make inferences or connections between pieces of information.
- Critical judgement questions that may give rise to different answers by different readers as they are based on the relationship between the text and the experiences of every respondent (Senegal National Ministry of Education, 2013).

These questions all seek to capture children's understanding of text, which is crucial from the beginning to the end of the reading process. According to Senegal's former Minister of Education, Ndoye Mamadou (2015):

"To better understand the issues related to the assessments, such as those initiated by Jàngandoo, they should be placed in the context of the learning crisis. According to the Education for All (EFA) Monitoring Report 2013-2014, out of 650 million school-age children, more than one third, meaning 250 million do not master the basics of reading and counting. This learning crisis is manifested differently in the regions and countries of the world .... Within each country there are significant disparities between rich children and poor children, children from urban areas and those in rural areas, boys and girls. Faced with such a crisis, the issues are about:

- Precise measurement of the magnitude of the learning crisis and its different dimensions in each country in order to promote awareness and consequent mobilisation around the challenges, demands and emergencies;
- Identification of the factors fueling the crisis and their causes that should guide thinking and action for a way out;
- The quality of policy dialogue to widely share lessons learned and make quality improvements to policies and strategies with the participation of all stakeholders at different levels of the education system;
- The implementation in the field of learning where quality of education, cultures and practices are decisively taking place".

4. EDUCATIONAL ASSESSMENT IN SENEGAL

There are many ongoing learning assessments in Senegal aimed at measuring the learning competencies of children and youth—these include national learning assessments, cross-national initiatives and public examinations. Typically, national learning assessments and many of the cross-national initiatives such as the *Programme d'Analyse des Systèmes Educatifs de la CONFEMEN* (PASEC) are curriculum-based assessments. Therefore, they are intended to assess school curriculum and gauge the learning levels of students in school at specific grades.

Table 1 lists the school-based assessments in Senegal that are intended to measure student learning outcomes at given grades. The assessments are developed to reflect curricular competencies at these grades.

The Jàngandoo Barometer is grounded on the principles of other large-scale cross-national initiatives as well as the Senegalese national learning assessments. However, in comparison, the Jàngandoo Barometer does not relate exclusively to school acquisitions in terms of measuring what is taught but rather assessing what every child in Grade 3 of their elementary education should have mastered in reading, mathematics and general knowledge. The idea behind the Jàngandoo

TABLE **1**

Educational assessments in Senegal

Educational Assessment	Grades assessed	Latest year of administration
Système National d'Evaluation du Rendement Scolaire (SNERS)	Grades 2 and 4	2014
Partenariat pour l'amélioration de la lecture et des mathématiques à l'école (PALME)	Grades 2, 4 and 6	2014
PASEC	Grades 2 and 6	2014

Note: This list excludes public examinations that are intended to certify competencies and are mandatory for progression to the next educational level.

assessment strategy is to go beyond targeting only in-school children by developing standards that are equally applicable to all children in school as well as those who have dropped out, unschooled children, those attending non-formal and informal types of education as well French and Arabic schools. As such, the Barometer targets all children in Senegalese households with no discrimination based on social or linguistic differences.

5. PRINCIPLES OF THE JÀNGANDOO BAROMETER

The Jàngandoo Barometer uses standardised tests, just as other large-scale learning assessments conducted in Senegal. The assessment is not curriculum-based nor does it evaluate alternative educational programmes but rather it produces a median level exit profile that corresponds approximately to the learning acquisitions associated with the end of the third year of schooling (Grade 3). It does not target a particular type of education and is not administered in classrooms. Instead, the assessment is administered at the household level through an inclusive and participatory approach. This is also true for the complementary test that was introduced to gather more in-depth information on the learning achievements of children with higher levels of performance. The assessments are conducted by trained facilitators. Since 2014, data are captured using tablets with software developed for this purpose.

The development of the Jàngandoo Barometer assessment instruments are based on the following principles:

1. Equity

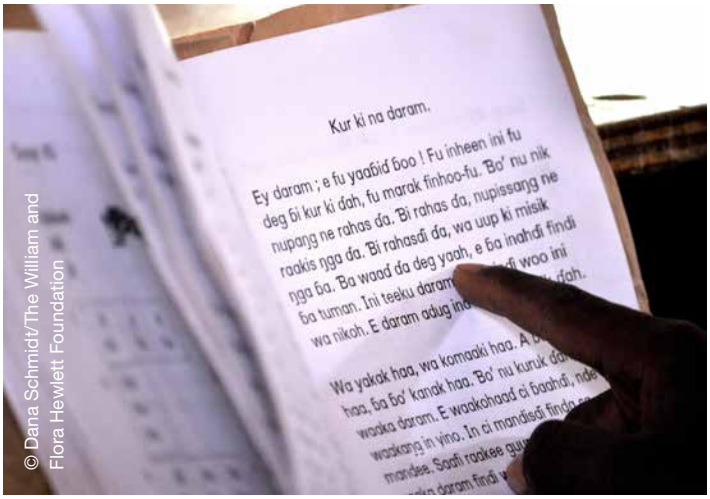
- The items at each level and in each area of assessment belong to the same category of situations (i.e. level of requirements, equivalence), including those for girls and boys of the same age group.
- The testing conditions were adapted as best as possible to fit all socio-economic and psycho-pedagogical characteristics found across the spectrum of children being assessed.
- Special attention is given to ensure the elimination of biases related to certain stereotypes (i.e. gender, physical and sensory disability, ethnicity, religion, social background, area of residence, etc.). Biases in the test (level of content, stimuli and concepts used) are controlled by a central team and then by an external evaluation conducted by an education task force.

2. Reflection of the socio-cultural universe

- The items in the assessment refer to children's experiences, practices and environment.
- The exercises are adapted to the socio-cultural realities of all communities (i.e. linguistic, ethnic, religious, etc.).
- The cultural environments and educational contexts of assessed children are taken into account so as not to refer to only one type of educational provision.

3. Compliance with the pedagogy of success

- Gaining trust is key so from the very first contact, parents and children are reassured, marking a shift from the learning environment approach and the types of student/teacher or Koranic master/disciple relationships.
- The availability of a series of tests with the same level of requirements in Arabic and French gives the child the possibility to choose the test series and the language in which they wish to be assessed in each area of the assessment.
- The test items progressively increase in difficulty in each of the domains. Items are ranked from simple to complex. The objective of this selection is to encourage the child to find solutions and progress in dealing with new challenges (Fall, 2015).



© Dana Schmidt/The William and Flora Hewlett Foundation

4. Comparability

- Comparability is the process by which the equivalence of assessment instruments is established in terms of required skills, content, parameters and characteristics of the targets to be evaluated. Ensuring comparability in the assessment design provides all children with the same opportunity to demonstrate their learning, contributing to the production of fair and reliable results.

- Comparability is taken into account at all stages of the assessment—from development to implementation by ensuring that:
  - The tests are similar from one language to another (i.e. they respect the same principles, use the same level of language and have the same levels of difficulty). The evaluation criteria are the same for both languages.[2]
  - The items at each level for each area of the assessment respectively belongs to the same universe of reference, regardless of language.

6. HOUSEHOLD-BASED READING ASSESSMENTS: ROLES AND BENEFITS

Whether diagnostic, predictive or certificate-driven, assessment is an instrument of standards.

It is this perspective that draws the interest of demographers and education specialists. The entry point for assessments on a demographic basis is through human communities (i.e. households, socio-professional, socio-cultural groups, etc.), which touches several domains of study and investigation.

Household-based assessments are inclusive and cover a broad spectrum of social concerns. First, they can provide information on non-school factors that influence the quality of learning. They can also give us an idea of the situation of those excluded from the education system (e.g. school dropouts) as well as the quality of other types of education used by children, such as community schools and Arabic educational institutes like the daara in Senegal. This form of assessment can reveal the impact of the literate environment on the skills of learners. In fact, more and more education specialists are interested in the effect of the literate environment on the maintenance and improvement in school performance (see article by Dowd and Friedlander).

Assessments conducted at schools target a more homogeneous category (i.e. grades, levels, steps, cycles). The domains assessed refer to programme and taxonomic levels, and are meant to measure the internal and external efficiency of education systems. They focus on learners, teachers and programmes, and enable interventions to target schools and learning by increasing effectiveness or efficiency thereby adding more meaning to the action of educating students.

The two types of assessments (school- and household-based) complement each other. Population-based assessments can serve education in several ways: (i) they broaden the operating field and arguments; (ii) they can support the determinants of environment, and (iii) identify schools and learning in all their diversity. Thus, assessment becomes a culture in populations that will eventually be in a position to objectively appraise the product that is delivered to them by education and training providers.

2 In 2016, Item Response Theory will be used to produce the performance scale. The team will use the ConQuest software developed by the Australian Council for Educational Research as it provides a comprehensive and flexible range of item response models to analysts, allowing them to examine the properties of performance assessments, traditional assessments and rating scales.

The Jàngandoo Barometer is an articulation of these two modes of assessment—it enters through the household using the demographic divisions, targeting census divisions and offers a large critical mass of information. The assessment does not depart from the type of assessment in schools either as it refers partly to school learning by the standard of assessments used.

7. GOING FROM ASSESSMENT TO INTERVENTION

In 2014, the assessment reported that only 19% of children (all ages included) successfully completed the overall median-level test (reading, mathematics and general knowledge). In reading, the success rate was 28%. These results reveal a rather worrying picture of the performances of Senegalese children and show that they face major difficulties with reading fluency and comprehension.

Presenting the performances of children to families has incited them to start questioning the causes of the current situation and linking the poor performance of children to dysfunctions in the educational system in general. Overall, the levels of children's performances are often lower than parents' expectations. The debate initiated with parents and the education community during the sharing of results highlighted several factors that contribute to poor pupil performance. These factors include teaching/learning methods not being adapted to children; the shortage of textbooks; a foreign language chosen as the language of instruction; low parent and community involvement in managing places of learning; and shortcomings in the training of teachers.

To remedy this problem, it is necessary for the education community to mobilise, innovate and make changes in the quality of learning being delivered to children. To raise the quality of education provision, simple, realistic strategies that are suited to all environments are needed.

It has emerged from discussions with communities that the Barometer should go beyond the publication of yearly results. This need has motivated the Jàngandoo programme to take action and move from assessment to intervention. In the quest for solutions and building on its experience to improve the quality of learning, the programme now offers coaching to citizens through the implementation of remediation strategies. A careful review of the results was made and a group of experts (teaching specialists, sociologists, cultural players, etc.) has been convened to develop a guide to instill irreversible reading skills. Once this guide is developed and tested, the goal is to move from experimentation to scaling up and ultimately induce a profound change in education quality.

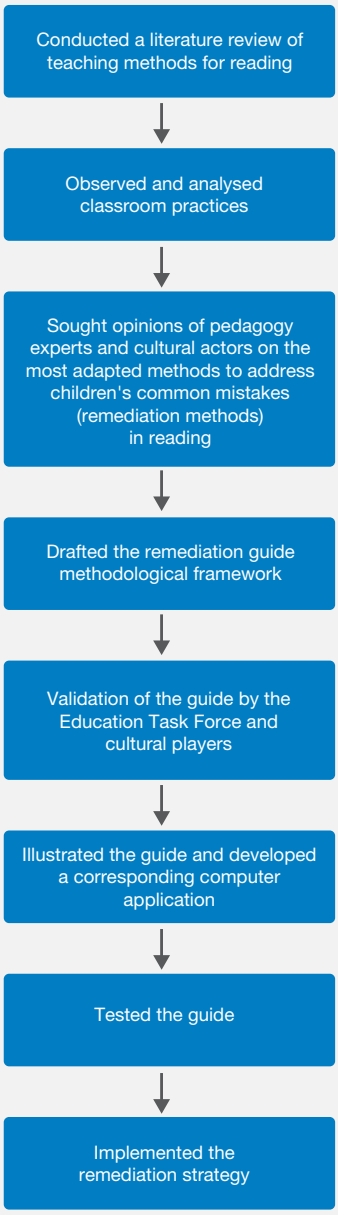
7.1 Developing the remediation guide

An essential step in the process of developing the remediation guide was the identification of the mistakes made by children during the test. This was the trigger for the remediation process. Indeed, identifying the reading difficulties of children provides knowledge of the real needs of these children and insights to developing the right solutions for the problems identified. The challenges identified include difficulties associated with the confusion of sounds, substitution of letters or sounds, addition or deletion of syllables in a word, reversal of the order of letters in a word, bad pronunciation of sounds and an inability to provide timely information on a text. **Figure 1** illustrates the steps that were taken to develop the remediation strategy.

7.2 The remediation strategy

The Jàngandoo Barometer assessment is community- and citizen-driven. The assessment happens at the community level and is implemented by actors from the community. The Jàngandoo programme provides information to parents and must negotiate with them for their children to participate in the remediation process. A 'remediator' is not the child's teacher, but someone in the community who is capable of implementing the strategy with the child. These remediators are recruited based on certain criteria, agreed upon

Figure 1. Steps undertaken in the implementation of the remediation strategy



and defined with the Jàngandoo partners. They are trained and are assigned to implement the strategy in a specific community. The remediation programme is implemented at the level and in a location chosen by the household (in the household or outside). The remediation programme is therefore implemented through a personalised approach and based on the needs of each target and each local authority. Children are placed in small groups of five and are guided by a pair of remediators (one for reading and one for mathematics). A post remediation assessment is conducted at the end of the process to measure each participating child's performance.

In conclusion, we can say that corrective actions that are integrated in the educational process led the children to overcome the difficulties that disrupt their progress in learning, especially in reading. Children who receive educational support through remedial courses gradually regain confidence by successfully overcoming the difficulties that prevented them from advancing in their learning. It is therefore possible with increased awareness and involvement by families and local communities to take a promising initiative and make it effective at the community level to meet the challenge of improving the quality of learning.

8. READING WITH COMPREHENSION AND LEARNING FOR ALL

Assessing reading is one of the goals of the Jàngandoo Barometer. The Barometer aims to assess the quality of learning by measuring different skills and abilities in reading—from phonological awareness to reading fluency and comprehension. The Jàngandoo assessment is administered at the household level and takes into account guiding principles, such as fairness, respect for the cultural universe and the pedagogy of success. The presentation of the results to all education stakeholders promotes the awareness of what can be called a learning crisis that impacts the quality of education.

Reading is a combination of decoding and construction of meaning. The assessment process therefore focuses on all components of reading by giving them each the weight they deserve in the test. However, the Jàngandoo experience reveals that reading comprehension is the most difficult skill to assess. It is therefore important that reading comprehension become the central focus of the remediation strategy. This is why the teaching tools,

such as the remediation guide, have been designed to enable children to analyse a text and make the link between text content and the questions asked. A causal relationship can be established between the reading proficiency level and the implementation of a remediation activity as there is evidence that it positively influences performance.

The Jàngandoo assessment complements other forms of academic evaluations but focuses on communities, grassroots actors and policymakers. This citizen-led assessment empowers the community to diagnose and participate in the research and implementation of solutions to children's learning. This increased community awareness on education helps bring together the different actors involved in improving the quality of education. Moreover, a key determinant of the quality of education happens to be the establishment of a political dialogue on this issue with authorities at the central and local level. Change is inevitable when there is synergistic interaction between all stakeholders striving for quality learning. The Jàngandoo programme is built on the belief that every child has learning potential and quality education means achieving successful learning for all children without exception.

REFERENCES

Fall, A.S., Ba, D., Bousso, S., Cisse, R. and Mbodj, A.A. (2014). *Réflexions relatives à l'approche Jàngandoo*. *Note pédagogique*. Dakar, Senegal: Universite Cheick Anta Diop de Dakar, Institut Fondamental D'Afrique Noire Cheikh Anta Diop et Laboratoire de Recherche sur les Transformations Économique et Sociales. http://lartes-ifan.org/pdf/Note%20pu00E9dagogique.pdf

Fall, A.S. (2015). "Jàngandoo, un baromètre citoyen en construction". *le Monde de l'Education*, No. 016, pp. 3-12.

Giasson, J. (1990). *La compréhension en lecture*. Montreal: Editions De Boeck.

LARTES (2015). "Jàngandoo en bref". Presentation. http://lartes-ifan.org/pdf/Pr%C3%A9sentation%20Jangandoo%20dernier%20jour%20formation.pdf (Accessed February 2016).

Mamadou, N. (2015). "Jàngandoo, une nouveauté dans le paysage de l'évaluation". *le Monde de l'Education*, No. 016, pp. 2-12.

UNESCO Institute for Statistics Catalogue of Learning Assessments. http://www.uis.unesco.org/nada/en/index.php/catalogue/learning_assessments. (Accessed March 2016).

Annual Status of Education Report (ASER) Assessment in India: Fast, Rigorous and Frugal

RUKMINI BANERJI
ASER Centre, Pratham India

ABBREVIATIONS

ASER	Annual Status of Education Report
DIET	District Institute of Educational Training
SSA	Sarva Shiksha Abhiyan
SMS	Short message service

1. INTRODUCTION

Since 2005, the Annual Status of Education Report (ASER) assessment has been conducted across rural India. It engages citizens in evaluating and understanding basic learning outcomes of a representative sample of children across the country. While enrolment rates are above 90%, the real challenge has been to track if children are learning. The ASER assessment exemplifies a nationwide initiative to answer this question and to shift the focus from inputs and outlays to outcomes (incidentally, the word *aser* in many Indian languages means 'impact').

Facilitated by Pratham, the ASER assessment is conducted by a local organization or institution in every rural district in the country. Pratham is an Indian non-government organization working to ensure that every child is in school and learning well.[1] Since 2000, Pratham programmes in low income urban and rural communities have found that enrolment levels are rising but a large proportion of in-school children need immediate help with acquiring foundational skills. Without being able to read fluently and without basic arithmetic skills, children cannot move ahead in the education system. Around 2002-2003, Pratham developed a simple tool to understand children's reading levels. The tool fed directly into instructional practice as children were grouped for instruction based on their level in the reading assessment. The tool was also helpful in explaining to parents where their children were and where they needed to be. It was also useful for tracking children's progress over time. This tool later became known as the ASER assessment tool.

The ASER assessment is a household survey that uses simple tools one-on-one with children to assess reading and arithmetic levels. The tools are both easy to administer and to understand. Using standard sampling methods, over 600,000 children in approximately 16,000 villages and over 565 rural districts are assessed each year. The ASER exercise has been conducted for ten years and has demonstrated that it is a reliable approach to generating data annually for a large and representative sample of children at a relatively low cost. Over these years, the ASER has been the largest and only annual source of information on children's learning in India. Although data on access to and enrolment in school was widely available even ten years ago, information on children's basic reading and arithmetic was not. In fact, for many

1 Pratham runs a variety of programmes in 21 states around India and works directly with communities and schools as well as with governments to work towards these goals. In 2014-2015, Pratham reached approximately a million children through direct interventions with schools and communities. Working in partnership with governments, an additional five million and more children were impacted (see www.pratham.org for more details)

years, policy and practice in India were focused on inputs rather than outcomes. The ASER exercise has contributed significantly to shifting this focus and issues related to children's learning outcomes have moved to the center of all discussions on education in India.

A decade ago in India, the culture of measurement at least in the education sector could be characterised broadly in the following four ways. First, evidence on outcomes was not widely used to formulate plans despite the rhetoric of outcome-based planning. In fact, evidence tended to be used on inputs and expenditures and not on outcomes. Further, current data was usually not available—you had to make-do with information from some years ago. Second, the capacity to conduct measurement at the state/district levels by government departments or by civil society was low. Third, thanks to a dearth in the availability of outcome data, citizens were unable to hold the government accountable for quality public services or indeed plan for meaningful action. Fourth, research was seen as the exclusive domain of experts, academics and universities. All of these features of the education eco-system reinforced the need for a separate autonomous initiative that focused on generating useful data on important outcomes; building capacity to conduct measurement; enabling a wide spectrum of people to participate; and creating an environment where policy, planning and action could be based on evidence.

The ASER survey has several distinctive features. From inception, one of the main objectives was to focus on children's learning. Year after year, the ASER has succeeded in bringing out the issue of what children learn on to the national stage. Evidence collected and facts observed on a national scale are available in the public domain annually. There is a lag of only a few months from when data is collected to when the data is made public.

Prior to the ASER assessment, there was hardly any attention paid to foundational skills that need to be built in the early school years if a child is to have a reasonable chance of completing elementary

schooling in a meaningful way. Reading is one of these essential capabilities without which future progress is impossible. Without learning to read, a child cannot move forward in the school system. Interestingly, ten years ago there were hardly any efforts to measure reading in the primary school years in most developing countries. Much of the 'testing' conducted to assess student achievement used pen-and-paper tests, which inherently assume that children can read. The ASER effort has shown that even after five years of schooling, less than half of all school-going children in India can read simple text. Thus, the assessment of reading is crucially important if the foundations of children's ability to learn are to be built well. There is only one way to assess reading when we are unsure whether children can read at all—that is to ask children to read. Hence, oral assessments needed to be devised that could be used on scale to understand how to help children learn better.

Every year, the ASER assessment completes the entire cycle—the enormous task from design to data collection to dissemination in 100 days.[2] From the collection of the first household survey to the analysis and compilation of data, all tasks are done in the *same* school year. This means that

2 The entire ASER survey cycle is designed to be completed in 100 days. The speed is needed so that data for the current school year can be available in the same year.

action based on the ASER findings can be taken immediately.

This paper provides a quick glimpse into the internal workings of the ASER assessment—a brief overview of the processes that go into the ASER effort—both in terms of implementation on scale and of completing the entire cycle of tasks at a fast pace.

The details of how sampling decisions were made are not included in this paper. The ASER each year carries a sampling note. Similarly, details of tool creation are also not included here. The ASER assessment uses tools in 20 Indian languages—these are the languages that are used as a medium of instruction in government schools around the country. Notes on the development of tools is available on the ASER Centre website (see www.asercentre.org).

2. PACE: THE NEED FOR SPEED

Across much of India, the school year begins in April. After a few weeks in their new grade, children go off for their summer vacations. Classes resume again from June to July and continue till March of the next calendar year.[3] In many states, the government finalises the enrolment numbers by the end of August or early September and by October or November, children are in the middle of the school year. Thus, the period between September and November is a good time for doing any kind of measurement especially for administering assessments of learning for that year. This is the period when the ASER assessment is in the field each year.

Every year, the ASER is released in the middle of January. From start to finish, the entire exercise each year takes about 100 days. The timing of the ASER is determined by a number of factors—the primary one being that the ASER report should be available before the plans or allocations for elementary education for the following academic year are made. The union budget is presented to the Indian Parliament at the end February each year and the annual work plans for elementary education are usually finalised in March. At ASER, we believe that plans for the next year should be based on current information and data. Usually, current data on enrolment and other inputs are available, but until recently, the ASER was the only source of data for learning available for the current school year. This is another major reason for the need for speed in generating estimates for basic learning in India.

The 100-day time period for the ASER assessment implies very tight timelines that need to be adhered to with immense discipline. In a context where delays are common, the clockwork nature and predictability of the ASER has been an important feature of the entire initiative. So far, the ASER has achieved its targeted timelines year after year for a continuous period of ten years. This discipline also sets the precedence for such work and proves that rigorous deadlines can be met if planning and execution are controlled tightly. Further, it is worth highlighting that the ASER assessment is not only fast and rigorous but also frugal. The entire cost from start to finish each year is well below US$1.5 million. Considering that close to 650,000 children are reached every year, the cost per child is less than US$2.50, which is very low when compared to other assessments of student achievement internationally.

3. PARTNERS, PEOPLE AND PARTICIPATION

One of the unique features of the ASER assessment is the decentralised and localised nature of implementation and dissemination. Local ownership and participation is an important element of the architecture of the ASER assessment and is crucial

3 State governments in India set their own school calendars. The summer vacation period varies from state to state. In Bihar, the holiday period is short—barely three weeks in June. In Uttar Pradesh, they are longer, going from mid-May to the beginning of July. In Maharashtra, vacation spans all of the month of May and schools open in June. The Right to Education Act specifies that elementary schools should have at least 225 working days in a school year. Only one or two states in India still operate on a January-December school year. Also, mountain regions of Himachal Pradesh as well as Jammu and Kashmir have longer winter vacations than the rest of the country.

to building awareness, improving accountability and initiating action towards improving elementary education. From its inception, a key component of the ASER process has been to involve local organizations and institutions.

For the ASER assessment, the district represents the 'local' unit. India has more than 600 districts—575 of which are largely rural.[4] A rural district in India may have anywhere between 1,000 to 3,000 (or more) government elementary schools. The ASER assessment is intended to be aligned with planning and implementation as it is envisaged in the elementary education framework of India. Sarva Shiksha Abhiyan (SSA) is a programme for universal elementary education and is the government's flagship programme for providing quality education at the elementary level. SSA falls under the jurisdiction of the Department of School Education and Literacy in the Ministry of Human Resource Development. SSA requires detailed work plans to be made in each district every year. Plan documents include reviews of progress made in the previous year, planned activities for the coming year and budgets to operationalise these plans. The SSA guidelines state that the district's annual work plans should incorporate habitation level planning, participatory processes, community mobilisation and participation as well as collaboration of government and local non-government groups in the process of developing plans.

Every year in each district of India, there is a local organization or institution that carries out the ASER assessment. These local groups include self-help groups, women's organizations, youth groups, well-known non-government organizations, local community based organizations, district colleges and universities. In recent years, teacher training institutions at the district level have participated in large numbers in the ASER effort. In most districts in India, the state government has set up teacher training institutes for pre-service and in-service teacher training called District Institute of Educational Training (DIET). In the ASER 2014, more than 260 DIETs participated in the ASER assessment as partners. Many of these institutions felt that participation in a systematic, structured and well supported national effort like the ASER assessment was beneficial for the training of future teachers who would be exposed to children and families in the community, to be able to discuss issues of learning and also learn how to collect and use evidence. More than half of the participating DIETs have requested that the ASER Centre (the unit of Pratham that leads the ASER effort) have a year-long engagement for capacity building. The resources for this large-scale citizens' participation come from individuals as well as institutional donors.[5]

To ensure comparability and consistency of tools and results over time, the ASER sample design and the assessment tools are centrally designed. However, the actual work is conducted in each rural district is by the ASER partner for that year. The local partner is involved in data collection and also in dissemination of the results. Ideally, the local group (organization or institution) that collects the information for the ASER can potentially be a member of the core group that facilitates the planning process and development of the annual work plan for elementary education in the district.

3.1 What kinds of people are needed for the ASER effort?

There are a variety of tasks that need to be conducted in the 100-day ASER exercise. Here is a brief overview of how many people participate at which level and perform what kinds of tasks (***Table 1*** *provides a summary*):

- Village: At the village level, there is usually a team of two people who conduct the survey. In the ASER, the assessment of children's learning and the collection of schooling information as well as other background data is done in the household.

4 The ASER assessment is only done in rural districts. Census village lists are used for sampling. Urban areas are not covered in the ASER due to the unavailability of community lists that can be used for sampling.

5 Each year, the ASER report lists names of partners and also those of donors and other supporters.

TABLE **1**

Structure and roles of different teams in the ASER process

Team member type/level	Role in the ASER assessment	Number of team members
ASER national team	This team manages all aspects of the ASER assessment across the country. This includes changes in basic design, piloting tools, preparing training materials, conducting the national trainings, tracking progress of the assessment, leading the quality control measures, data analysis and report production.	15-20 team members
ASER state team: core lead teams for each state (these are full-time team members)	For each state, there is a core team that leads the work for that state. This includes planning, recruitment of additional team members (as needed), preparation and training of master trainers, coordination of the roll out of the assessment, all aspects of ensuring quality control, managing the data entry process and participating in state level as well as national level rechecks.	75-100 team members for the entire ASER assessment for India
Master trainers for each district in each state (these team members come on board for the duration of the survey)	At least two master trainers manage all aspects of the ASER assessment in a district (assessment of 30 villages). They conduct trainings, ensure data quality (monitoring, recheck) and manage all other assessment logistics. They are responsible for all assessment related activities in the district until the survey booklets are sent to the data entry centre. On average, more than 560 rural districts are reached each year.	A total of approximately 1,100 master trainers in most years for the ASER assessment in India
Volunteers for village survey	Two ASER volunteers/assessors carry out the assessment in the village assigned to them. A total of 30 villages are assessed in each district.	Approximately 25,000 volunteers participate in the ASER assessment in India each year

Twenty households are randomly selected (following a process in which the survey team is trained) and all children in the age group of 3 to 16 years in each of the selected households are surveyed. Basic information about enrolment is collected on each of them. Enrolment information about pre-school and school is noted as well as the type of school. Children who are five years old and above are assessed one-on-one on basic reading and arithmetic tasks. The survey team also visits one government school in the sampled village to conduct observations and collect data on the school attendance of students and teachers as well as basic data on school infrastructure.

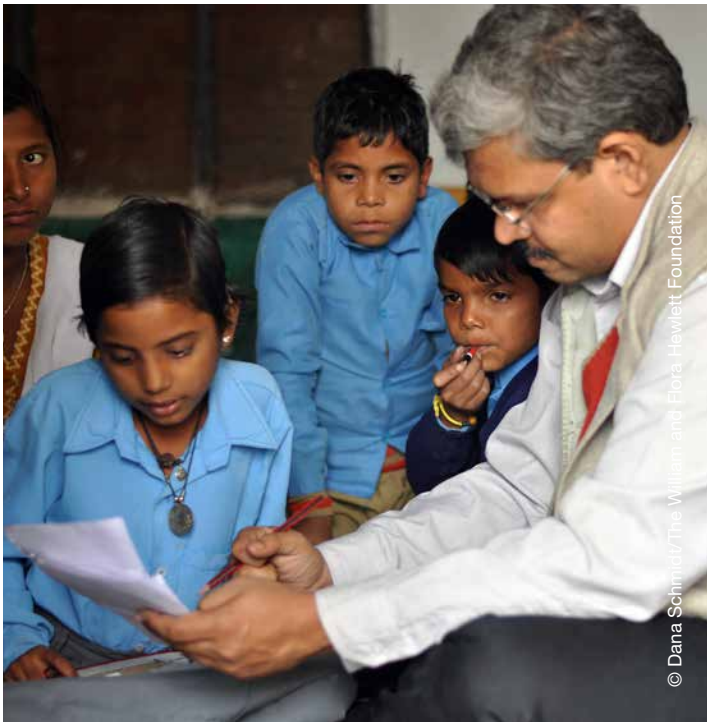
- District: For each rural district, 30 villages are randomly selected from census village lists using the probability proportional to size sampling method. There are typically two models in which the ASER survey is implemented in a district. The first model is when there are 60 ASER volunteers/surveyors. In this model, a pair of surveyors goes to a village and hence 30 villages are covered in one weekend. Since the assessment is conducted in the home, it is crucial to maximise the probability that children can be found at home. Hence, the ASER assessment is carried out over the weekend when schools are closed. The other model is implemented when there are 30 volunteers. In this model, each pair of surveyors covers two villages over two weekends. For any combination that has less than 60 volunteers, the second approach is followed where some pairs do more than one village. There are always at least two master trainers for each district. These trainers conduct the three-day training then accompany the teams into the field on the days of the survey—the accompanying of teams into the field on the days of the survey is referred to as 'monitoring'. After the village survey is completed, the trainers visit a set of villages to verify that the assessment was done and that the data were collected. This activity is referred to as 're-checking'.

- State: There is a great deal of variation in states across India in terms of the number of rural districts in the state. For example, India's biggest state, Uttar Pradesh has more than 70 districts whereas Goa has only two. For each state, the number of master trainers is decided based on the size of the state and the manner in which the ASER assessment is to be rolled out in that year.

If a state decides to conduct the assessment in all districts simultaneously, then the number of master trainers is usually a little more than double the number of districts (so that there are a few backup trainers). If the decision is to conduct the ASER assessment in a state over three phases, then each pair of master trainers may end up training and monitoring teams in at least one district in each phase. Master trainers are recruited, prepared and trained accordingly. A small core team at the state level (full-time team members) manage all aspects of the ASER assessment in their state.

ASER state team members are called ASER Associates or ASER Regional Team members. They are 'fellows' with the ASER Centre for a period of two to three years. In addition to leading all ASER related activities in their state, they also participate in a course run by the ASER Centre on assessment, survey, evaluation, research and communication. This course had earlier received certification from Indira Gandhi National Open University and now has certification from the Tata Institute of Social Sciences—a premier educational institution in India.

In the early years of the ASER assessment, many of the master trainers were deputed to the ASER assessment from Pratham for a period of one to three months. In recent years, master trainers are taken on-board for the period for which they are associated with the ASER assessment. In some cases, master trainers are members of the partner organizations that conduct the ASER assessment in a given district. In other cases, master trainers are individuals who join the ASER effort for a period of one to two months. There are also situations in which one or two institutions at the state level partner with the ASER assessment not as a source of volunteers but to provide master trainers for the ASER assessment for that year. If there is an institutional relationship of this type (usually through a college or university), master trainers may get credit in their educational institutions for participation in the ASER assessment as master trainers. Sometimes based on their good



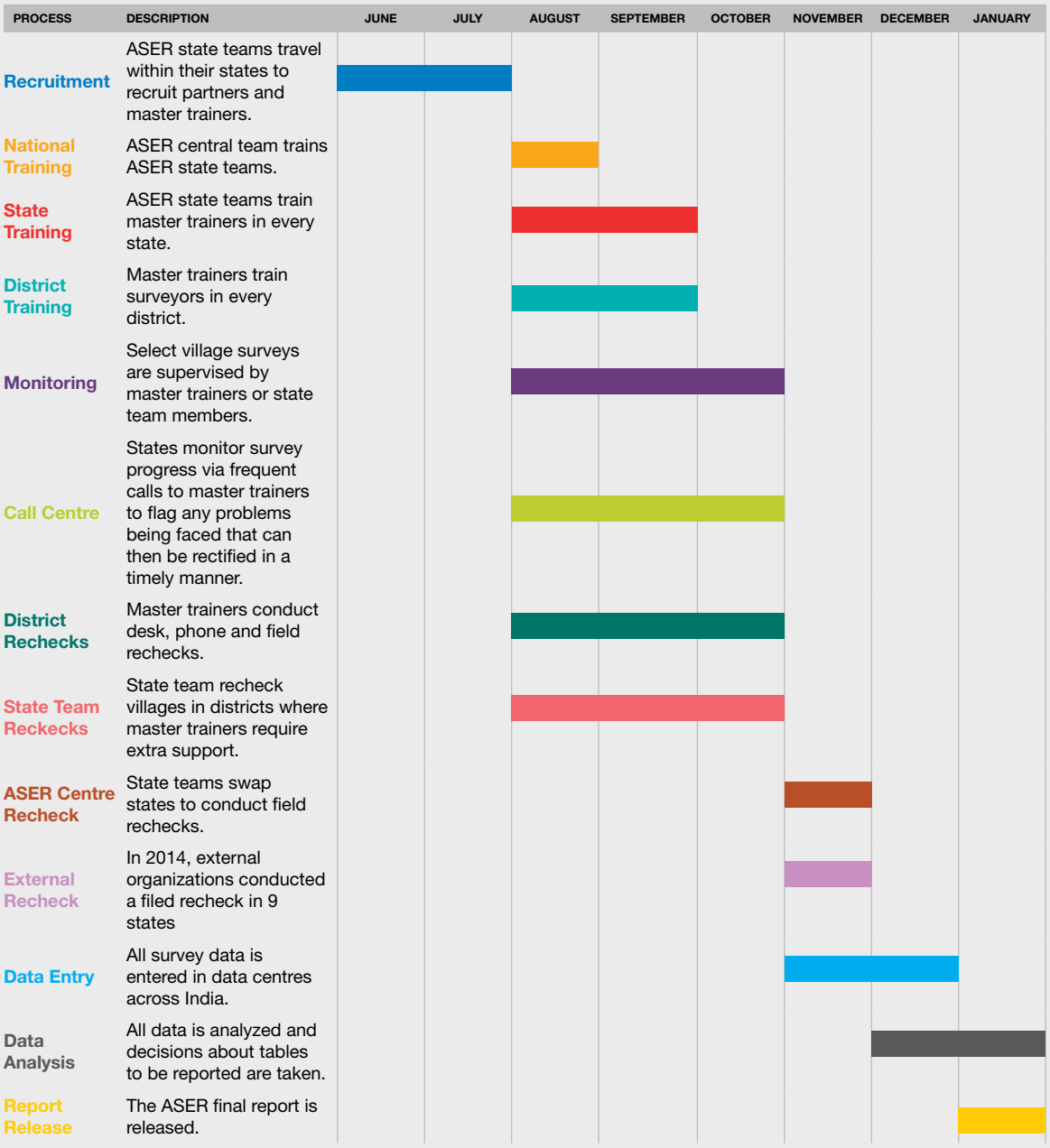
performance in the ASER effort, master trainers may be recruited to become full-time ASER team members for the next ASER. There are also cases where ASER master trainers return each year in the ASER season to participate in the process.

4. PLANNING AND PREPARATION

Early each year, well before the 100-day period begins, a day-by-day schedule is planned at each level (national, state and district) taking into account festivals, examination schedules as well as other possible constraints. The timetable also takes into account events such as elections or censuses as well. The calendar is put in place by working backwards from the day of the release of the report in January.[6] There is also buffer built into each state's schedule for unanticipated delays and difficulties. The importance of meticulous planning cannot be overemphasised. The ability to visualise each and every step clearly is critical if tight timelines are to be met. Lessons learned from each year's experience are ploughed into the next year's planning. The flow of activities for ASER 2014 is summarised in **Figure 1**.

6 The report release date each year from 2006 onwards has been on a day between 12 and 18 January.

Figure 1. ASER assessment timeline and activities



Source: ASER Centre, India

5. PROCESSES AND QUALITY CONTROL

Preparing people to carry out this massive assessment exercise is perhaps the biggest challenge for the ASER assessment. Remember, it is not only the scale that is daunting; the conviction that widespread participation of citizens from all walks of life is an essential and desirable feature of the effort has to also come alive on the ground. Add to this the requirement that the entire process be done frugally and rigorously using available resources.

Due to the assessment's enormous coverage, maintaining quality control at every stage of the

process is crucial. At the same time, balance has to be maintained between cost and scale to deliver the highest quality possible. Many of the quality control measures and mechanisms have evolved over time at every level. Every year, processes were fine-tuned and streamlined based on lessons learned the previous year.

5.1 Training

Training is one of the most important processes that help to equip the ASER volunteers/surveyors with the skills necessary for surveying a village and assessing children. Typically, the ASER assessment follows a three-tier training structure. The national ASER workshop is followed by a state-level training in every state. Both of these are residential trainings and can last anywhere from five to seven days with several field days during the training period. This is followed by district-level training where the ASER volunteers are trained to conduct the ASER survey.

There are a set of key principles that are maintained through the ASER assessment training process at every level:

- There must be at least two trainers (preferably three) for every training session.
- No training can be conducted without the trainees having all the training/instructional material in hand. All materials must be printed in time and be available at the training session for use.
- Training at every level must include field practice. Each volunteer has to be observed by his or her master trainer surveying and assessing children in a village. The field practice session is followed by detailed discussions and clarifications about the process.

i) National workshop

During this workshop, the ASER state assessment teams are oriented on the ASER assessment processes and materials. The workshop is also used to plan for state-level trainings and partner selection. Each ASER state assessment team comprises anywhere between two and five fulltime staff, depending on the size and complexity of the state. The national workshop is the final point each year after which no changes can be made in any process. The instruction manuals are finalised at the workshop. The only changes that take place after this point are related to translation of the procedural documents.

The sessions in the national workshop are of three types: 'doing' the ASER assessment, training for the ASER assessment, and practicing recheck processes and planning. On one day, an actual and complete 'dress rehearsal' of the entire district ASER assessment process is conducted. This entails surveying 30 villages and 20 households in each village. This is done to ensure that everything that has been planned can actually be carried out in the time that is available to a survey team. Similar practice sessions are done with the recheck process.

Since the national workshop is the place where lead teams for every state are prepared, mock trainings are an important part of the proceedings. Team members take turns to conduct specific training sessions and they are graded by other team members. All of these activities are also carried out in the state-level trainings.

ii) State-level training workshops

These workshops prepare master trainers who will then take charge of rolling out the ASER survey in their districts. Master trainers are usually drawn from the district's local partners and Pratham team members. Over 1,000 master trainers are trained and equipped in the state-level workshops. Usually, state-level trainings are organized to run over five to six days and have four main components:

- Classroom sessions: To orient and ensure that all the participants know the content of the ASER assessment process thoroughly. Presentations and case studies are used to help state teams carry out these sessions. Training films are used to highlight specific examples. Case studies are also used to demonstrate different scenarios.
- Field practice sessions: During the workshop, participants and trainers go to nearby villages

and actually conduct the ASER assessment in 30 villages. This takes a full day and is treated as a complete 'dress rehearsal' of the actual process in a district.

- Mock training: These sessions are intended to prepare and improve the training capabilities of district level master trainers. The focus of these sessions is not only on ensuring that complete content is being delivered but also to build the skills and methods of conducting effective training.

- Quiz: A quiz is administered towards the end of each state-level training and immediate feedback is provided to participants. This helps to ensure that all participants have understood the ASER assessment process and to identify participants who may not have obtained the minimal understanding required to conduct the ASER assessment.

Performance in mock trainings, field visits and quiz results are analysed to ensure that master trainers are well prepared. These processes also help to identify weak master trainers who are then either dropped or provided with additional support during district trainings. Master trainers also receive financial training as they are responsible for distributing payment for the district-level training and disbursing small stipends to the volunteers and accounting for the receipts. All of these expenditures have to be accounted for in a clear and transparent manner.

iii) District-level training workshops

The district-level trainings for preparing and equipping ASER volunteers/surveyors are generally held for three days. Like state level trainings, the key elements of district trainings include classroom sessions, field practice sessions and a quiz. Typically, volunteers who do not achieve a satisfactory score on the quiz are either dropped or paired with strong volunteers to carry out the survey.

Due to the scale of the survey and the large number of participants at every level of training, ensuring the quality of these trainings is crucial. The two most important aspects for quality control are:

- Ensuring standardisation and completeness of information cascading from one level to the other. This is achieved through:
  - Providing comprehensive training schedules
  - Using session wise training schedules with checklists
  - Ensuring that all master trainers have training materials (videos, posters and other supporting materials)
  - Creating regional language manuals
  - Having a central team presence at all state-level trainings and close coordination with state team members during the district-level trainings.

- Ensuring that participants are prepared for their role in the survey. This means ensuring that participants have a holistic understanding of the survey processes and their role in it. This is achieved by involving participants in:
  - Classroom sessions
  - Question-answer and clarification sessions
  - Field visits
  - Quizzes
  - Phone support (to answer questions after training is over).

In all district trainings, records are maintained for each of the ASER assessment volunteers. These records contain attendance data for every person for each day of training and quiz scores for all volunteers. The data in this sheet is extensively used to guide volunteer selection for the ASER assessment.
The ASER assessment training system provides a strong foundation for effective implementation of the survey. However, the trainings are viewed not only as a means to collect quality data but also as an opportunity to educate and orient 25,000 people across the country on the importance of measurement. And, in this process ask a simple yet extremely significant question: can our children read and do basic arithmetic? Please refer to the section entitled 'About the Survey' in each year's ASER report for detailed descriptions of the process. Also, see the section entitled 'ASER Survey' on the homepage of the ASER website (www.asercentre.org) for notes and documents on the ASER survey process.

iv) Monitoring

Monitoring refers to quality checks carried out when the survey is being conducted in the field. Monitoring is done at two levels. One level comprises monitoring of the surveyors by the master trainers and the other is the monitoring of master trainers by the state ASER assessment team. Monitoring of the assessors by the master trainers is done through field and phone monitoring. The master trainers visit villages during the survey. Often the villages are selected if a survey team seems to be weak or under-confident or if the village is very remote.

v) Call centre

The monitoring of master trainers by the state team is done by means of a call centre, which is set up in each state.[7] The call centre is usually manned by one or two people who are designated to make and receive phone calls from the field. The call centre system was introduced during the ASER 2011 and has a two-fold objective. The first is to track progress of the assessment at each stage (almost daily) and to be able to make timely decisions and take immediate action. This enables the ASER state assessment teams to efficiently monitor districts in which they are not physically present and then travel to them as required. This 'live monitoring' is also useful for preventing any problems during the survey that may become insurmountable if not attended to in time. The second is to increase the accountability of master trainers. Making regular calls to master trainers in each district also helps them feel supported through the entire survey process.

vi) Recheck processes

There are two elements to the recheck process at the district level:

- Desk recheck

Master trainers do a detailed 'desk check' that entails checking the completed survey booklets that have been handed in so as to identify incomplete or problematic data and then verifying this with the assessors. A checklist helps master trainers carry out a systematic desk recheck. Here is an actual description of what is done:

> Master trainers fill in a compilation sheet for their district that shows children's learning levels for all surveyed villages. Since data for the entire district is summarised in a single format, it can be used to analyse simple trends and identify inconsistent data in specific villages. For example, if the format shows that the total number of children assessed in a village is much higher or lower than other surveyed villages in the same district, this could be an indication of the quality of the data gathering in the village. This compilation sheet is reviewed by master trainers usually after an assessment weekend is completed so that they can identify any problems as soon as possible before the next assessments scheduled for the following weekend.

- Field recheck

After the desk and phone recheck, 'problematic' villages are selected for field recheck by the master trainer. After discussing these problematic villages with the state team, master trainers each recheck three villages per week and together recheck at least 12 of the 30 surveyed villages in a district. In the ASER 2014, 63% of all surveyed villages were either monitored or rechecked.

One of the important features of the monitoring process for recent ASER assessments has been the immediate availability of summary information for each monitored and rechecked village using cell phones and short message service (SMS). The use of SMS started during the ASER 2012. The data that is collected is uploaded on a common online portal, which enables the ASER assessment teams—both at the central and state levels—to receive information on a daily basis about these villages.

Under the guidance of the central core team, state team members often conduct a state-level recheck process to verify the quality of data that has been collected. In November, a national exercise of this kind is also carried out by the national ASER assessment team members who visit other states as assigned. Thorough process audits have also been

7 In the ASER 2014, Manipur, Mizoram, Nagaland, Tripura, Arunachal Pradesh, Meghalaya, Sikkim and West Bengal, did not have a call centre.

conducted in different years. Each year's ASER final report carries details of the training, monitoring and recheck processes under the section entitled 'About the survey.' See www.asercentre.org for all ASER final reports from 2005-2014.

6. PROGRESS AND TRACKING

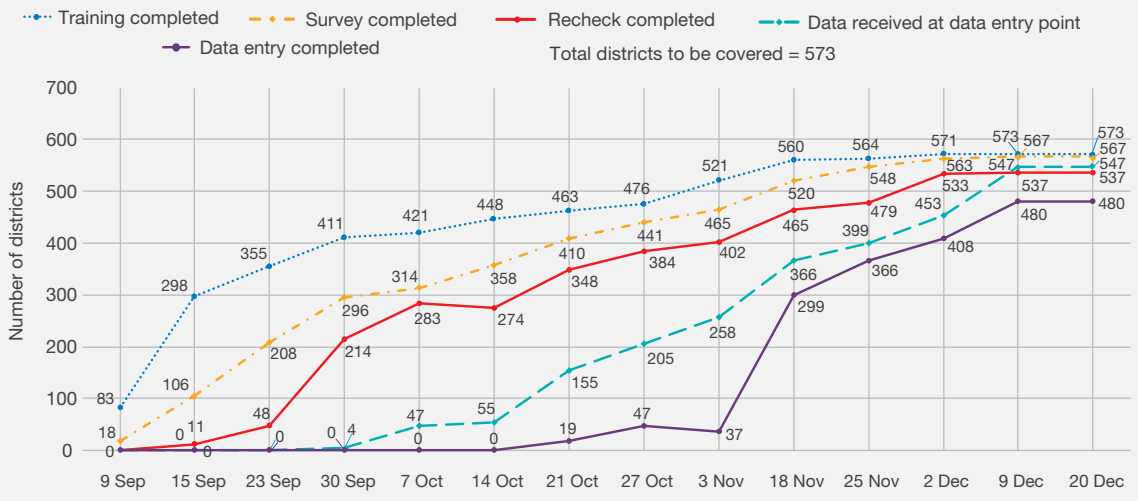
The entire data process is tracked very closely. Between the call centres and data entry centres, close watch is kept on a daily basis on the processes that have been completed and the adherence to scheduled timelines. The pace of activities, especially from the survey and data collection to data entry is tracked closely. **Figure 2** outlines the details.

To begin preparation for data entry, first the software is designed. The software is then tested several times using dummy data to make sure that it is entirely functional. As the software is being developed for the assessment, states select a data entry centre that is conveniently located where they will send the hard copy survey booklets. Each state then finalises their data entry location, keeping factors such as number of data entry operators, number of districts in the state and the cost of data entry in mind. Data entry operators are trained either on-site or by telephone on how to enter the assessment data. The finalised software is then sent to these data entry locations.

A few years ago, an exercise was conducted to analyse the quality of data entry across different data entry centres in the ASER assessment that year. The best quality data entry centre was found to be in the Rajasthan centre. On further investigation, we found that this centre was brand new. It was set up to provide computer training and livelihoods for rural women. Data entry for the ASER assessment for that year was the first job that this data entry centre obtained. Many of the women who were entering the data were also learning how to use computers and had never been outside their village. The organization called Source for Change is now an established data service centre in the region (see www.sourceforchange.in for more details).

In 2014, there were 12 data entry centres across India. Data entry is done in multiple centres to speed up the process and also to cut down on time and cost allocated for the transportation of assessment booklets. Since the ASER assessment is conducted in 20 languages, several of the states using regional languages have data entry done locally so that the assessment booklets can be read by the data entry operators.

Figure 2. Roll out of the ASER assessment process in 2014



Source: ASER Centre, India, 2014

© Dana Schmidt/The William and Flora Hewlett Foundation

During the assessment, master trainers are instructed to submit the assessment/survey booklets directly to data entry centres or to their respective state teams as soon as the survey is completed. Data centres track the number of districts in which the assessment has been completed to plan for how much data entry needs to be done at a time. Sometimes, assessments take longer than expected and so by tracking assessment progress, data centres can plan appropriately for data entry.

The ASER assessment maintains strict confidentiality norms—no identifying information (for individual, household or village) is available in the public domain. In fact, village lists are also strictly guarded at all times.

Once data entry begins, checking for mandatory information that needs to be filled such as school name, household number, gender and age is done. Data validation is also done where some data entry fields can only have specific values. Certain manual checks are also put in place. For example, every fifth household (four households from each village) is cross checked. If five or more mistakes are found in this checking, then all households in the village are rechecked. The ASER state assessment teams also visit the data entry centres to do random data entry cross checks. Compiled data is then sent to state teams for further verification, if required.

The data that has been entered is uploaded into a central server after rechecking. Analysis is done exclusively by the central team in Delhi.

7. PUBLICATION, REPORTING AND DISSEMINATION

In the month of December, as the data cleaning and data analysis is being conducted in Delhi, an in-house production team also moves into place and the process of putting the report together begins. For example, all photographs in the published report are selected from hundreds of photographs taken by volunteers and ASER team members from across the country. The time between the arrival and analysis of data is short and the time available for report production is even shorter. The process of layout and formatting begins in the third week of December and the report goes to the printer soon after the New Year. A team of about 15-20 people work round the clock for the last few weeks in December and early January to produce the data and the report on time.

The structure, content and format of the basic report at state and national level is kept simple so that data can be easily understood by a wide cross-section of people. Each state report is usually six pages with four pages for reporting data collected at the household level with a page each to present enrolment figures, reading data, arithmetic information and other information. Two pages are allocated for reporting data based on the school observation. These basic set of tables are translated into Hindi and into other regional languages so that they can be disseminated easily.

The national ASER final report also has a series of short essays commenting on trends and data. These are written by in-house team members as well as invited guests. As a matter of policy, there is no commentary or interpretation provided on the actual data—only explanations for how to read tables or graphs.

At the time of the release, a series of slides/presentations/notes are also generated for each state. This helps in communicating the key findings for the year to a diverse and wide audience. The media in India uses these documents extensively. The state report cards are also printed in a two- or four-page format for large scale distribution at different levels.

8. CONCLUSION

Even in 2005, 'schooling for all' was well understood by policymakers, planners, practitioners and parents in India. Enrolment levels were at an all-time high—well above 90% even at that time. But the big, and at the time, new question that needed to be asked was: are children learning? The ASER initiative was designed to shift the focus from access and provision to 'learning for all' and to bring children's learning to the centre of all discussions and debates on education in India. It was assumed that one of the important ways to achieve wider awareness about the issue of learning would be through the participation of a broad-based cross-section of people around the country. Widespread involvement of local citizens in conducting the assessment in each district in India was therefore crucial to the fundamental architecture of the ASER assessment.

Large scale participation, however, has important implications for key aspects of the ASER assessment's design and implementation:

- Simplicity of the assessment tool and administration protocol

Widespread participation of citizens in 600 districts meant that it was important to plan on massive scale for training and implementation. Therefore, the process needed to be relatively straightforward in terms of actual testing of children (process and time for each child and each subject) as well as the time taken to complete a sampled village. The assessment tools and administration protocol have been designed keeping in mind that the ASER assessment is a household survey that will be carried out on a huge scale.

- Volunteer model

Large-scale participation has important cost implications. More than 25,000 volunteers participate in the ASER assessment each year. They are trained, mentored and monitored by over 1,000 master trainers. ASER volunteers reach 600,000 to 700,000 children annually in 15,000 to 16,000 villages. ASER volunteers are remunerated only for travel and costs they incur. Hence, the ASER assessment is truly a citizen-led initiative with thousands of people volunteering their time.

The fact that there was going to be large scale citizen participation, however, could not jeopardise the rigour or the reliability of the data. Methods and mechanisms had to be designed to keep data quality issues at the forefront of the design and implementation. Accompanying these challenges were two other critical considerations—need for speed and the necessity of keeping the cost of the effort low. All of this was being attempted in a context where there was no history of measuring outcomes and where the culture or capacity for educational assessments was not strong.

Ten years of administering the ASER assessment have generated many lessons. The foremost learning is that even with limited resources, a robust, rigorous and reliable exercise can be carried out at a fast

time-bound pace by the citizens of a country. This has been done not once or twice but year after year for a decade. The ASER assessment has created a new benchmark for many elements of evidence-based work not just in a country like India but also for a global context. The ASER assessment has challenged assumptions about who can do research and how such data is to be gathered. The entire credit for the success of the ASER assessment goes to the thousands of individuals and hundreds of institutions who have participated and given their commitment, energy and time unstintingly year after year to this massive effort. In addition, none of this would have been possible without the cooperation and collaboration of millions of Indian children and families.

We believe that "when ordinary people are empowered with knowledge, they can bring about extraordinary change" (extract from the ASER Centre mission statement). Measurement, which is critical to generating knowledge, has been an exclusive domain of experts. We believe that measurement needs to be rigorous but easy to understand and to act upon. When ordinary people learn to measure what affects their lives, they can communicate with each other across villages, states, nations and continents to identify and understand their problems, take steps to resolve them and change the world for the better.

REFERENCES

Banerji, R. (2013). "The Birth of ASER". *Learning Curve,* Issue XX, Section C, p.p. 85-87. http://img.asercentre.org/docs/Publications/External%20publications/banerji_p85_birthofaser_learningcurvexxaug2013.pdf

Banerji, R. (2013). "From Schooling to Learning: ASER's Journey in India". M. Barber and S. Rizvi (eds.), *Asking More: The Path to Efficacy*. London: Pearson.

Banerji, R., Bhattacharjea, S. and Wadhwa, W. (2013). "Annual Status of Education Report". *Research in Comparative and International Education*, Vol 8, No.3, p.p. 387-396. http://img.asercentre.org/docs/Publications/External%20publications/aser_rcie_fullversion.pdf

Banerji, R. and Bobde, S. (2013). "Evolution of the ASER English Tool". Berry V. (ed.), *English Impact Report: Investigating English Language Learning Outcomes in Primary School in Rural India*. London: British Council. http://www.britishcouncil.in/sites/britishcouncil.in2/files/english_impact_report_2013.pdf

Banerji, R. and Chavan, M. (2013). "The Bottom Up Push for Quality Education in India". H. Malone (ed.), *Leading Educational Change Global Issues, Challenges, and Lessons on Whole-System Reform*. New York: Teachers College Press.

Eberhardt, M. Plaut, D and Hill, T. (2015) *Bringing Learning to Light: The Role of Citizen-led Assessments in Shifting the Education Agenda.* Washington, DC: Results for Development. http://r4d.org/knowledge-center/bringing-learning-light-role-citizen-led-assessments-shifting-education-agenda-0

Ramaswami, B. and Wadhwa, W. (2010). *Survey Design and Precision of ASER Estimates of ASER*. ASER Centre Working Paper. New Delhi: ASER Centre. http://img.asercentre.org/docs/Aser%20survey/Technical%20Papers/precisionofaserestimates_ramaswami_wadhwa.pdf

Vagh, S.B. (2009). *Validating the ASER Testing Tools: Comparisons with Reading Fluency Measures and the Read India Measures*. ASER Centre Working Paper. New Delhi: ASER Centre. http://img.asercentre.org/docs/Aser%20survey/Tools%20validating_the_aser_testing_tools__oct_2012__2.pdf

To review the ASER Survey key documents, go to www.asercentre.org.

© Dana Schmidt/The William and Flora Hewlett Foundation

Chapter 4
Using Assessment Data: Interpretation and Accountability

The articles in this chapter present case studies about the use of oral reading assessments data - from interpreting and communicating the assessment results with stakeholders to designing interventions for improving reading skills. The success and challenges of using assessments to measure outcomes are discussed.

Is Simple, Quick and Cost-Effective Also Valid? Evaluating the ASER Hindi Reading Assessment in India

SHAHER BANU VAGH
ASER Centre, Pratham India

ABBREVIATIONS

ASER	Annual Status of Education Report
EGRA	Early Grade Reading Assessment
DIBELS	Dynamic Indicators of Basic Early Literacy Skills
J-PAL	Abdul Latif Jameel Poverty Action Lab
ORF	Oral Reading Fluency
RI	Read India

1. INTRODUCTION

The Annual Status of Education Report (ASER), conducted since 2005, is a citizen-led, household-based nationwide survey designed to provide basic yet critical information about foundational reading and basic numeracy skills of Indian children between the ages of 5 and 16 years. The ASER surveys approximately 600,000 children every year. A decade of surveys have highlighted the pervasive challenges that children encounter in acquiring basic reading skills and have been successful in shifting the education debates towards learning outcomes and not merely issues of access and infrastructure (for more information on the ASER assessment, see article by Banerji).

The ASER reading assessment is orally and individually administered in order to enable the assessment of children who are beginning to read or struggling readers who would have difficulties taking a pen-and-paper assessment in a group format. The assessment is administered in 20 Indian languages and in English. It is administered in an adaptive format and the testing time required is about five minutes (see www.asercentre.org for the complete set of instructions and the testing tools).

The ASER assessments have several advantages—they are simple, quick, cost-effective, easy to administer and the results are easy to understand. All of these are desirable features as it makes an annual survey of the scale and scope of the ASER feasible (Wagner, 2003). Two additional advantages of the ASER survey are that: (a) the processing of vast amounts of data is done fast so as to enable the release and dissemination of results in a timely manner and (b) the reporting of results (for reading and arithmetic) is done in an easy to understand format so that findings are comprehensible for all stakeholders—parents, practitioners, planners, policymakers or the common citizenry. Discussions are common at the village level and it is easy to include parents in the debate as the ASER assessment reading tasks and reporting format of the results are easy to engage with and understand. For instance, telling a parent that their child reads single *akshara* but cannot read words provides an understanding of their child's reading level that is easy to grasp.

Despite the relevance of the ASER survey and the many advantages of the assessment format, a pertinent question that needs addressing is how robust are the findings based on such a basic, short and quick assessment of reading? In other words, does the ASER reading assessment provide valid information on children's early reading ability? To

provide supporting evidence for the validity of the ASER reading assessments, this paper presents the content validity of the ASER reading assessment and examines children's concurrent performance on the ASER Hindi reading assessment compared to other assessments of reading. Data are drawn from a large-scale randomised evaluation of Pratham's[1] reading and arithmetic intervention programme by the Abdul Latif Jameel Poverty Action Lab (J-PAL) for which the ASER reading assessment was used along with a battery of other reading and arithmetic assessments. Specific and detailed comparisons are made between the ASER reading assessment and another widely prevalent model of assessing early reading ability, the Early Grade Reading Assessment (EGRA), which had been adapted for use in Hindi (also see www.eddataglobal.org) (Dubeck and Gove, 2015; Gove and Wetterberg, 2011). In addition, the ASER is compared to the Read India Literacy assessment, a pen-and-paper assessment of children's basic and advanced Hindi literacy skills.

2. VALIDITY IN THE CONTEXT OF THE ASER READING ASSESSMENT

Validity indicates whether a test assesses what it purports to assess and is an evaluation of a test's inference and not of the test per se. A test can be put to different uses, such as examining average school performance or making diagnostic decisions for individual students. Each of these uses or inferences "has its own degree of validity, [and] one can never reach the simple conclusion that a particular test 'is valid'" (Cronbach, 1971, p.447). Several forms of evidence are collected to evaluate the validity of the inferences that are based on the test results. As such, validity is an accumulating body of evidence (AERA et al., 1985). One form of evidence is content validity, which indicates the extent to which the content is a representative sample of the domain of interest and whether it assesses the desired or targeted skills and abilities. Another form of evidence, a common empirical investigation termed "concurrent validity" involves comparing performance on the assessment of interest with performance on a comparable assessment that serves as a criterion measure. The criterion assessment is typically another assessment of known psychometric properties that assesses similar abilities or constructs. In this study, as noted above, we are comparing the performance on the ASER reading assessment with children's performance on the EGRA and the Read India Literacy assessment. Strong and positive associations between the three assessments will contribute to evidence that serves as one part of building a full validity argument for the ASER reading assessment.

2.1 Content of the ASER reading assessment

The ASER reading assessment is designed to capture children's early reading skills in 20 Indian languages. For the Indian orthographies, the basic orthographic unit is the '*akshara*', which represents sounds at the syllable level with its constituent parts encoding phonemic information. The *akshara* can vary from simple to complex depending on the extent of the phonemic information encoded. The primary forms of vowels and consonants with an inherent vowel that is unmarked comprise the set of simple *akshara* and contrast with the complex *akshara* that comprise ligatures that are consonants with vowel markers or consonant clusters with a marked or unmarked vowel. Given that the ASER reading assessment is designed to assess early and basic reading skills, its subtasks assess children's knowledge of the simple *akshara*, ability to accurately decode simple words and ability to fluently read a Grade 1 and Grade 2 level passage. The selection of the subtasks is based on the premise that the acquisition of symbol-sound mapping and the ability to decode symbol strings are among the early set of skills that contribute to reading in the Indian alphasyllabaries akin to the alphabetic orthographies (Nag, 2007; Nag and Snowling, 2011). Although the pace of acquisition of the *akshara* tends to be extended given the extensive set of orthographic units that children

1 Pratham is an Indian non-government organization working on large scale to ensure that every child is in school and learning well. Pratham runs a variety of programmes in 21 states around India and works directly with communities and schools as well as with governments to work towards these goals. See www.pratham.org for more details.

have to master, the acquisition of simple *akshara* is expected to be complete by Grades 1 and 2 (Nag, 2007; Sircar and Nag, 2013) and is therefore appropriate for use on an early grade reading assessment.

The content of the ASER reading assessments—i.e. the selection of words, length of sentences and vocabulary—is aligned to Grade 1 and 2 level state-mandated textbooks (curriculum). Analysis of the early grade textbooks highlighted some common expectations, which are that children can a) read simple sentences by the end of Grade 1 and b) read words in connected texts of 8-10 sentences (at a minimum) by the end of Grade 2. Moreover, even though simple and complex *akshara* are introduced in the early years, the primary focus in this formative period is on the learning of simple *akshara*. As noted above, mastery of the simple *akshara* is expected to be complete by Grades 1 and 2 and is therefore appropriate for use on an early grade reading assessment. Hence, the ASER reading assessment includes simple *akshara* drawn from the set of primary forms of vowels and consonants on the *akshara* subtask, two- and three-syllable simple words that do not include any complex *akshara* (i.e. consonant clusters) on the word decoding task and controlled vocabulary in the passages where a minimal number of high frequency complex *akshara* (conjoint consonants) are included (see ASER 2014 for the detailed set of guidelines). The ASER reading assessment then essentially represents an assessment of a baseline or basic set of reading skills from among the early reading competencies. For this reason, it tends to be referred to as a 'floor' test of reading ability. In addition to the standard components listed above, comprehension questions have been administered in two rounds of the survey in 2006 and 2007 (Pratham 2006; 2007). The scoring of the ASER assessment is based on pre-defined performance criteria for each subtask (*see* ***Table 1***), thus it is a criterion-referenced assessment. The ASER assessment reports performance on an ordinal scale indexing children's levels of reading at 'beginner', '*akshara'*, 'word', 'Grade 1 level passage' or 'Grade 2 level passage'.

2.2 The Early Grade Reading Assessment Hindi adaptation, an ideal criterion measure

The EGRA, based on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), is a widely

TABLE **1**

The ASER, EGRA and the Read India Literacy assessments: test content, administration, scoring and score reporting specifics

	ASER	EGRA	Read India Literacy Assessment
Test inference	An assessment of early reading ability	An assessment of early reading ability	An assessment of early and advanced literacy ability
Assessment format	Individually and orally administered; untimed assessment	Individually and orally administered; timed assessment	Individually administered in a pen-and-paper format
Testing materials	Assessment sheets and scoring sheets	Assessment sheets, scoring sheets and stop watch	Assessment sheets
Grades	Common assessment for Grades 1-5	Common assessment for Grades 1-5	Separate assessments with overlapping content for Grade 1-2 and Grades 3-5
Reading sub-tasks and test length	a. Simple akshara (5 items) b. Simple words (one and two syllable words, 5 items) c. Grade 1 level passage d. Grade 2 level passage e. Comprehension questions	a. *Akshara* comprising the Hindi alphasyllabary (52 items) b. Consonant-vowel pairings (*barakhadi*, 52 items) c. Simple words (one and two syllable words, 52 items) d. Pseduowords (52 items) e. Grade 1 level passage f. Grade 2 level passage g. Comprehension questions	a. *Akshara* knowledge b. Reading vocabulary c. Word and sentence construction d. Knowledge of syntax e. Sentence and passage comprehension
Test administration time	5 minutes	10 minutes	10-20 minutes
Administration	Adaptive format: testing begins with the Grade 1 level passage subtask and if child is judged to read it fluently with 3 or fewer mistakes then the Grade 2 level passage is administered, if not, then the word reading task is administered	Non-adaptive format: testing begins with the akshara reading subtask. If the child fails to identify any words correctly within the span of one minute on the 'simple word' reading task then the assessment is discontinued	
Reporting metric	Reading levels reported on an ordinal scale where 1 = beginner, 2 = akshara level, 3 = word level, 4 = Grade 1 reading level and 5 = Grade 2 reading level.	Fluency, that is the number of subunits of text read correctly in the span of one minute reported for each subtask. A composite score of all subtests was created by averaging as the associations between all subtasks was high (with Spearman's rank correlation coefficient ranging from .81 to .94)	Total test score
Scoring criteria	Beginner level—could not read 4 out of 5 simple *akshara* *Akshara* level—can read 4 out of 5 simple *akshara* but not words Word level—can read 4 out of 5 simple words but cannot read words in connected text fluently Grade 1 level passage—can read the passage fluently with 3 or fewer mistakes but cannot read the Grade 2 level passage Grade 2 level passage—can read the passage fluently with 3 or fewer mistakes	None	None
Reliability estimates	Cohen's kappa estimate based on decision consistency across repeated administrations is .76. Cohen's kappa estimate for inter-rater reliability is .64 and the weighted kappa estimate for inter-rater reliability is .82 (Vagh, 2009)	The median coefficient alpha estimates averaged across 5 test samples is .93. Test-retest reliability coefficients for the subtests of the EGRA ranged from .83 to .98	Internal consistency was estimated in two ways. The first approach used a simple count of the total number of items and in the second approach, each question category on the assessment was treated as an individual item thus providing a more conservative estimate. The coefficient alpha estimates for Grades 1-2 based on item and category counts is .93 and .86 respectively and for Grades 3-5 is .93 and .88
Assessment timeline	a. Pilot study 1 b. Pilot study 2 c. Bihar baseline evaluation d. Uttarakhand baseline evaluation e. Bihar midline evaluation	a. Pilot study 1 b. Pilot study 2 c. Bihar baseline evaluation d. Uttarakhand baseline evaluation e. Bihar midline evaluation	a. Pilot study 1 b. Pilot study 2 c. Bihar baseline evaluation

used assessment of reading that has been adapted for use in more than 65 countries and in over 100 languages (Good and Kaminski, 2002). The EGRA is composed of a battery of subtasks that assess print knowledge, phonological awareness and orthographic knowledge, and draws on research in alphabetic languages, notably English (e.g. Snow, Burns and Griffin, 1998). Administration time may vary from 10-20 minutes depending on the number of subtasks administered. The Hindi adaptation of the EGRA comprised tasks that assess children's efficient and accurate decoding of singleton *akshara* (consonants with the inherent vowel that comprise the Hindi alphasyllabary referred to as the '*varnamala*'); *akshara* that comprise the pairing of the consonant 'k' with the secondary forms of all the vowels (referred to as the '*barakhadi*'); simple and familiar words; pseudowords; and words in connected Grade 1 and Grade 2 level text as well as associated comprehension questions. The inclusion of these tasks is based on the premise that the efficient and accurate decoding of singleton *akshara*, *akshara* combinations, words in list form and words in connected text are important and robust correlates of early reading ability and comprehension. Automaticity of these lower-level skills ensures that limited cognitive resources, such as attention and memory, can be freed and allocated to the higher-level skills of meaning-making (LaBerge and Samuels, 1974; Perfetti, 1977, 1985). Although questions remain about the strength of the association between oral reading fluency and comprehension in the *akshara* languages (e.g. Hindi) relative to English, the importance of mastering singleton *akshara* and combinations of *akshara* is an important precursor to skilled reading in the *akshara* languages (Nag, 2007; Nag and Snowling, 2011; Sircar and Nag, 2013; Vagh and Biancarosa, 2012).

Similar to the ASER assessment, the EGRA is an assessment of early reading ability that is orally and individually administered. See Table 1 for a detailed comparison between the ASER and the EGRA. Although the ASER and the EGRA comprise similar subtasks, the EGRA is a lengthier assessment comprising substantially more items at the *akshara* and word reading levels and is a timed assessment of 'fluency' with examiners using a stop watch to ascertain time taken to read the sub-units of texts (i.e. singleton *akshara*, consonant-vowel pairing, words in list form and words in connected text). In addition, the ASER assessment reports results in terms of reading levels, while the EGRA reports fluency rates for each of the reading units. As a consequence while the ASER assessment captures variation in terms of reading levels, the EGRA captures variation between and within reading levels.

The similarities and differences between the ASER and the EGRA thus make the EGRA an ideal candidate for a comparison measure to evaluate the validity of the ASER reading assessment as it allows us to ascertain: a) the degree of association between two assessments of early reading ability which are similar in content but differ in test length, administration format and manner of reporting results and b) the appropriateness of the criterion on the basis of which children are categorised at different reading levels on the ASER assessment by examining fluency rates for each reading level. Overall, the comparison between the ASER assessment and the EGRA allows us to evaluate the agreement in reading ability across tests that are administered independently and with some differences yet comparable in content and designed to assess the same abilities or skills.

3. RESEARCH DESIGN AND METHODOLOGY

3.1 Sample

As part of an evaluation of a reading and math intervention programme, the ASER reading assessment was administered along with a battery of assessments of basic and advanced reading and math ability. The study involved several rounds of data collection: (1) An initial pilot study (Pilot 1) with 256 children from Grades 1-5 in the state of Bihar, (2) a second pilot study (Pilot 2) conducted with 412 children from Grades 1-5 in Bihar, (3) a baseline evaluation conducted in two districts in the state of Bihar (n = 8,866) and of Uttarakhand

(n = 7,237) with children from Grades 1-8, and (4) a midline evaluation in Bihar conducted with 5,612 children from Grades 1-8. The pilot studies were conducted to evaluate the psychometric properties[2] of the assessments. The baseline and midline evaluations were conducted to monitor Pratham's Read India programme. Results from the four rounds are presented separately as they were collected at different time points in the academic year. Most importantly, the results provide replication across different samples and from two districts of two Indian states[3].

3.2 Assessments

All children were assessed individually in their household by a pair of trained examiners. The reading assessments administered were the (a) ASER, (b) EGRA and (c) the Read India (RI) Literacy assessment. The first two assessments were orally administered while the RI Literacy assessment was a pen-and-paper assessment. See Table 1 for details of all three assessments.

3.3 Analysis plan

To evaluate concurrent validity, the degree of association between the ASER reading assessment and the EGRA and the RI Literacy assessment were estimated on the basis of Spearman rho correlation coefficients[4] (see scoring metric in Table 1 for unit of analysis). The expectation is that the ASER reading assessment will be strongly correlated with both assessments but its correlation with the EGRA will be relatively higher than its correlation with the RI Literacy assessment as both the ASER assessment and EGRA have more in common in terms of inference on early reading ability and comprise similar subtasks. In addition, the association of the ASER assessment with the RI Literacy assessment will help determine whether a basic assessment of early literacy correlates with a broader range of literacy skills administered in the traditional pen-and-paper format, as is the expectation.

Given the similarities and differences between the ASER and the EGRA as noted above, additional analyses were conducted to understand (a) how children at the different ASER assessment reading levels performed on the EGRA and (b) whether children who read three or fewer *akshara* or words on the EGRA classified at the 'beginner' or '*akshara*' level on the ASER assessment. These additional explorations for the ASER and EGRA will help evaluate the appropriateness of the ASER's criterion measures and in turn, provide additional evidence for how a simple and short assessment, such as the ASER compares with a longer assessment of identical skills. These analyses are reported for the Bihar baseline and midline and Uttarakhand baseline evaluation samples for simplicity as the results for the pilot studies are in keeping with the findings of these larger samples.

4. RESULTS

The Spearman's rho coefficients presented in **Tables 2-5** indicate that the ASER reading assessment is strongly correlated with the EGRA with correlation coefficients ranging from .76 to .94. The coefficients were estimated separately for Grades 1-2 and Grades 3-5 or 3-8 as the RI Literacy assessment is a separate assessment with overlapping content for these grades (see Table 1 for additional details). Note that the magnitude of the coefficients are even higher, ranging from .90 to .94 when estimated for the full sample rather than separately for Grades 1-2 and Grades 3-5 or 3-8. As expected, the attenuated variation and floor effects when splitting the sample tend to mitigate the magnitude of the associations, especially for Grades 1-2. Additionally, the ASER reading assessment is also moderately to strongly

2 In the second pilot study, the issue of fatigue was also examined. No detrimental effects were noted from administering all assessments in one session versus in two sessions (J-PAL, Pratham and ASER, 2009).

3 Some minor changes were made to the administration format for the Fluency Battery and a few poorly performing items on the Read India Literacy assessment were changed based on the pilot study results. However, the format and content of the ASER reading assessment remained unchanged for all rounds of data collection.

4 A correlation is a statistical measure that indicates whether and how strongly pairs of variables (e.g. scores obtained from two tests) are related. Since the ASER test score is an ordinal or ranked variable, the Spearman rho or Spearman rank correlation coefficients were estimated.

correlated with the RI Literacy assessment with coefficients ranging from .65 to .87.

TABLE **2**

Spearman's rho coefficients for the ASER, EGRA and Read India (RI) Literacy assessments based on Pilot 1

	ASER	EGRA	RI Literacy
ASER	—	.92***	.87***
EGRA	.86***	—	.88***
RI Literacy	.81***	.89***	—

**p < .05; **p < .01; ***p < .001*
Note: the coefficients for Grades 1-2 (n = 96) are above the diagonal and for Grades 3-5 (n = 94) are below the diagonal. The validity coefficients for the ASER assessment and the EGRA for the full sample (i.e. Grades 1-5) is .91.

TABLE **3**

Spearman's rho coefficients for the ASER, EGRA and the Read India (RI) Literacy assessments based on Pilot 2.

	ASER	EGRA	RI Literacy
ASER	—	.82***	.82***
EGRA	.85***	—	.88***
RI Literacy	.77***	.83***	—

*p < .05; **p < .01; ***p < .001
Note: The coefficients for Grades 1-2 (n=171) are above the diagonal and for Grades 3-5 (n=220) are below the diagonal. The validity coefficients for the ASER assessment and the EGRA for the full sample (i.e. Grades 1-5) is .90.

TABLE **4**

Spearman's rho coefficients for the ASER, EGRA and the Read India (RI) Literacy assessments based on the Bihar baseline evaluation

	ASER	EGRA	RI Literacy
ASER	—	.76***	.65***
EGRA	.82***	—	.68***
RI Literacy	.76***	.81***	--

*p < .05; **p < .01; ***p < .001
Note: the coefficients for Grades 1-2 (n = 3,818) is above the diagonal and for Grades 3-8 (n = 3,035) is below the diagonal. The validity coefficients for the ASER assessment and the EGRA for the full sample (i.e. Grades 1-8) is .91.

TABLE **5**

Spearman's rho coefficients for the ASER and EGRA assessments based on the Grades 1-8 Uttarakhand baseline (n = 7,179) and Bihar midline (n = 5,612) evaluation

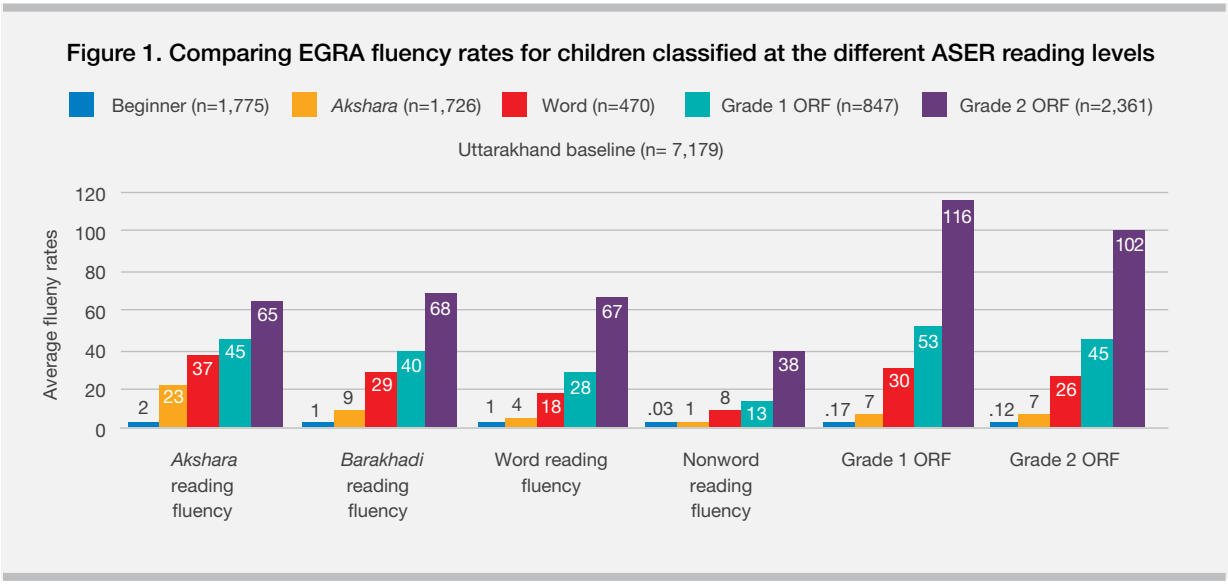
	ASER	EGRA
ASER	—	.94***
EGRA	.94***	—

*p < .05; **p < .01; ***p < .001

The EGRA reading fluency rates for children at the different ASER reading levels illustrated in **Figure 1** indicate that reading fluency rates increase with the increasing ASER reading levels. In other words, children categorised at the 'Grade 2 reading level' read the simple *akshara*, the *barakhadi*, words, nonwords, and words in Grade 1 and Grade 2 connected text of the EGRA with greater fluency then children classified at any of the lower levels of reading on the ASER reading test. The increasing fluency rates with higher ASER reading levels are reflected in the strong Spearman's rho coefficients between the ASER assessment and the EGRA.

Given that the ASER reading levels are mutually exclusive categories, children classified at the '*akshara* level' meet the criteria for the '*akshara* level' but not the 'word level' (i.e. correctly names four out five simple *akshara* but cannot read four out of five words), and so on. The expectation then is that children at the 'beginner level' should not be able to read four or more *akshara* on the *akshara* fluency subtask and children classified at the '*akshara* level' should not be able to read four or more words on the word fluency subtask of the EGRA and so on. Average performances illustrated in Figure 1 substantiate this claim. For instance, averaging across the three samples, children classified at the 'beginner level' demonstrate *akshara* fluency rates of 2 *akshara* (SD = 3.8), children classified at the '*akshara* level' demonstrate word fluency rates of 3 words (SD = 5.8), children classified at the 'word level' demonstrate Grade 1 level oral fluency rates of 25 words (SD = 19.6), children classified at the 'Grade 1 reading level' demonstrate Grade 2 level oral fluency rates of 44 words (SD = 25.1).

The ASER assessment *akshara* and word reading subtests are extremely short tests that comprise only five items. As a result, it is possible that children can be misclassified due to item sampling error. The EGRA, on the other hand, gives children the opportunity to read 52 *akshara* and 52 words on the *akshara* and word reading subtasks, respectively. Hence, to evaluate the efficacy of the ASER assessment, the percentage of children who identified between 0-3 *akshara*/words and those who identified four or more *akshara*/words on the simple *akshara* and word reading subtasks of the EGRA was calculated. This enabled comparing the decision consistency between the ASER assessment and the EGRA on the *akshara* and word reading subtasks.



The results presented in **Table 6** indicate that of the children classified at the 'beginner level', 82% of the children in Uttarakhand, 94% of the children in the Bihar baseline study and 95% of the children in the Bihar midline study could not correctly identify four or more *akshara* on the simple *akshara* reading fluency subtest. Of the children classified at the '*akshara* level', 96% of the children in Uttarakhand, 80% of the children in the Bihar baseline study and 85% of the children in the Bihar midline study could in fact correctly identify four or more *akshara*. Of the children classified at the 'word level', 98% of the children in Uttarakhand, 87% of the children in the Bihar baseline study and 96% of the children in the Bihar midline study did correctly read four or more words correctly. This is a high level of consistency across the two tests.

Further examination of the misclassifications or decision inconsistencies between the ASER assessment and the EGRA (*see* ***Table 6***) indicate that although the children categorized at the 'nothing level' read four or more *akshara* correctly on the EGRA, they demonstrated low rates of fluency in comparison to their counterparts who were categorised at the '*akshara* level'. Similarly, children categorised at the '*akshara* level' who read four or more words correctly on the EGRA, demonstrated low rates of fluency in comparison to their counterparts who were categorized at the 'word level'. In the absence of normative data for fluency rates, evaluation of fluency rates are based on relative comparisons.

5. DISCUSSION

The ASER reading assessments are simple, quick and easy to administer and the findings from the series of studies reported in the present paper

TABLE **6**

Comparison of average fluency rates of children

	***Akshara* level**		**Word level**	
	Consistent M (SD)	**Inconsistent M (SD)**	**Consistent M (SD)**	**Inconsistent M (SD)**
Uttarakhand baseline	23.5 (11.9)	8.76 (6.2)	18.3 (11.3)	9.06 (6.2)
Bihar baseline	19.2 (12.7)	13.75 (13.9)	20.9 (13.6)	12.9 (11.9)
Bihar endline	14.7 (10.39)	7.7 (7.02)	13.5 (8.4)	7.2 (4.1)

Note: Comparison of average fluency rates of children whose ASER assessment reading level is consistent with their EGRA performance and average fluency rates of children whose reading level is inconsistent with their EGRA performance. M = median; SD = standard deviation.

indicate that a brief, basic and quick assessment like the ASER assessment can provide a valid snapshot of children's early reading ability.

Much effort has been invested to ensure that the content of the ASER reading assessment maps onto Grades 1 and 2 curricula standards. Compelling evidence for the concurrent validity of the ASER reading assessment is demonstrated by the strong associations of the ASER with the concurrently administered assessments of literacy—i.e. the EGRA and the RI Literacy assessment. The ASER and EGRA assessments are both assessments of early reading skills administered individually and orally albeit different in administration format and in the length of some of the subtasks. Given the similarities in content and inference but differences in administration format, strong and positive associations between the two assessments provide favorable evidence for their concurrent validity. Moreover, the strong and positive associations of both these assessments with the more detailed assessment of literacy in the traditional pen-and-paper format that assesses a broader range of skills, including vocabulary knowledge, spelling ability, knowledge of syntax and writing comprehension corroborates the importance of the early skills tapped by the ASER assessment and the EGRA for literacy acquisition.

Further substantiating concurrent validity, comparisons of the decision consistency between the ASER assessment and the EGRA indicate that there is a high level of consistency across the two tests at the 'beginner', '*akshara*', and 'word' levels for the purposes of presenting aggregate district and state level estimates. Although there were a small percentage of inconsistencies with children at the 'beginner level' correctly reading four or more *akshara* on the EGRA and with children at the '*akshara* level' correctly reading four or more words on the EGRA, the respective fluency rates of these suspect misclassifications were clustered at the lower end of the fluency continuum. Moreover, given that the ASER reading levels are mutually exclusive categories, it implies that children categorised at the '*akshara* level' do not demonstrate competency at the 'word level' or any other higher level. As a result, the fluency rates of children at the '*akshara* level' are bound to be lower than the fluency rates of children who are classified at the 'word' or higher level. This expectation is supported by the data and is in keeping with the viewpoint that fluency in reading words in connected text requires fluency at the levels of smaller units, such as simple and complex *akshara* (Foulin, 2005; Wolf and Katzir-Cohen, 2001). Consequently, an important instructional implication of this finding is that children categorised at the '*akshara* level' are demonstrating 'minimal' mastery as opposed to 'complete' mastery of *akshara* knowledge and need to further improve their *akshara* knowledge if they are to successfully decode words in list form or connected text. Similarly, children classified at the 'word level' are demonstrating 'minimal' mastery of their decoding knowledge and need to further improve their decoding skills in order to fluently read and comprehend words in connected text.

6. CONCLUSION

Oral assessments of reading to monitor learning outcomes in developing countries are increasingly becoming the norm. As orthography and context-specific adaptations of the ASER assessment are developed, accumulating evidence for the validity of these assessments will help provide a more complete picture of the utility of such easy to implement assessments.

REFERENCES

Abdul Latif Jameel Poverty Action Lab (J-PAL), Pratham and 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*. Unpublished manuscript: J-PAL, Chennai, India.

American Educational Research Association (AERA), American Psychological Association (APA) and National Council on Measurement in Education (NCME) (1985). *Standards for educational and psychological testing*. Washington DC: AERA.

ASER Centre (2014). ASER assessment and survey framework at http://img.asercentre.org/docs/Bottom%20Panel/Key%20Docs/aserassessmentframeworkdocument.pdf

Cronbach, L. J. (1971). "Test validation". R. L. Thorndike (ed.), *Educational measurement*. 2nd edn. Washington, DC: American Council on Education, pp. 443-507.

Dubeck, M.M. and Gove, A. (2015). "The Early Grade Reading Assessment (EGRA): Its theoretical foundation, purpose, and limitations". *International Journal of Educational Development*, Vol. 40,p.p. 315-322.

Foulin, J.N. (2005). "Why is letter-name knowledge such a good predictor of learning to read?" *Reading and Writing*, Vol. 18, 129-155.

Good, R.H. and Kaminski, R.A. (eds.) (2002). *Dynamic indicators of basic early literacy skills*. 6th edn, Eugene, OR: Institute for the Development of Education Achievement. http://dibels.uoregon.edu/

Gove, A. and Wetterberg, A. (2011). "The Early Grade Reading Assessment: an introduction". A. Gove and A. Wetterberg (eds.). *The Early Grade Reading Assessment: applications and interventions to improve basic literacy*. Research Triangle Park, NC: RTI Press, p.p.1-37.

LaBerge, F. and Samuels, S.J. (1974). "Toward a theory of automatic information processing in reading". *Cognitive Psychology*, Vol. 6, p.p. 293-323.

Nag, S. (2007). "Early reading in Kannada: the pace of acquisition of orthographic knowledge and phonemic awareness". *Journal of Research in Reading,* Vol. 30, No. 1,pp. 7-22.

Nag, S. and Snowling, M. J. (2011). "Cognitive profiles of poor readers of Kannada". *Reading and Writing: An Interdisciplinary Journal*, Vol. 24, No. 6,pp. 657-676.

Perfetti, C.A. (1977). "Literacy comprehension and fast decoding: Some psycholinguistic prerequisites for skilled reading comprehension". Guthrie, J.T. (ed.), *Cognition, curriculum, and comprehension.* Neward, DE: International Reading Association. pp. 20-41

Perfetti, C.A. (1985). *Reading ability.* London: Oxford.

Pratham (2006). *Annual Status of Education Report* Delhi: ASER Centre. http://asercentre.org

Pratham (2007). *Annual Status of Education Report* Delhi: ASER Centre. http://asercentre.org

Pratham (2008). *Annual Status of Education Report* Delhi: ASER Centre. http://asercentre.org

Sircar, S. and Nag, S. (2013). "Akshara-syllable mappings in Bengali: A language-specific skill for reading". Winskell H. and Padakanayya P. (eds.),

South and South-East Asian Psycholinguistics, Cambridge, UK: Cambridge University Press, pp. 202-211

Snow, C.E., Burns, M.S. and Griffin, P. (eds.) (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Vagh, S.B. (2009). *Evaluating the reliability and validity of the ASER testing tools*. New Delhi: ASER Centre. http://www.asercentre.org/sampling/precision/reliability/validity/p/180.html

Vagh, S.B. and Biancarosa, G. (2012). Early literacy in Hindi: the role of oral reading fluency. Poster presentation at the International Association for the Study of Child Language, Montreal, Canada.

Wagner, D.A. (2003). "Smaller, quicker, cheaper: alternative strategies for literacy assessment in the UN Literacy Decade". *International Journal of Educational Research*, Vol. 39, pp. 293-309.

Wolf, M. and Katzir-Cohen, T. (2001). "Reading fluency and its intervention". *Scientific Studies of Reading*, Vol. 5, No. 3, pp. 211-239.

USAID Lifelong Learning Project: The Linguistic Profile Assessment[1]

LESLIE ROSALES DE VÉLIZ, ANA LUCÍA MORALES SIERRA, CRISTINA PERDOMO AND FERNANDO RUBIO[2]
Juarez and Associates, USAID Lifelong Learning Project[3]

ABBREVIATIONS

CTT	Classical Test Theory
DIDEDUC	Departmental Directorate of Education
DIGEBI	*Dirección General de Educación Bilingüe Intercultural* (General Directorate for Bilingual and Intercultural Education)
USAID	United States Agency for International Development
UVG	*Universidad del Valle de Guatemala*

1. INTRODUCTION

An assessment designed to determine oral proficiency of one or more languages by a student is a key element teachers can resort to when planning the language of instruction and the reading and writing strategy to be used in early bilingual grades.

Guatemala has strived to consistently assess the oral proficiency of students' language skills. A first assessment effort in the early 1990s was administered by official school teachers using graphic stimuli and answer sheets to assess students whose mother tongue was a language other than Spanish (Magzul, 2015). Regrettably, by the late 1990s, the assessment was discontinued (ibid). From 2000 to 2003, an improved version of this assessment was implemented by the Ministry of Education's Bilingual Education Department (currently the General Directorate for Bilingual and Intercultural Education, DIGEBI) in schools located in the Mam, K'iche', Kaqchikel, and Q'eqchi' areas (ibid). However, in 2004, this updated version was also discontinued by the new Ministry (ibid).

The USAID Lifelong Learning Project, aware of the importance of measuring the oral proficiency of languages to which students may be exposed, resumed these assessment efforts and developed the Linguistic Profile to assess beginning students. This assessment instrument has been used in official schools located in seven municipalities of the K'iche' linguistic area and five municipalities of the Mam area. The assessment has been administered to students attending pre-primary and Grade 1 in schools located in regions where the Bilingual Education Model has been implemented. The Bilingual Model targets pre-primary to Grade 3 students, and focuses on reading and writing.

2. PURPOSE OF THE LINGUISTIC PROFILE

The purpose of the assessment is to guide teachers in the usage of the instruction language, whether Spanish or Mayan (K'iche' or Mam), based on the

1 The opinions expressed in this publication rest solely with the authors and do not necessarily reflect the views of the USAID or the Government of the United States.

2 The authors would like to acknowledge the following participating staff from the USAID Lifelong Learning Project: Raquel Montenegro, Sophia Maldonado Bode, Justo Magzul, Gabriela Núñez, Ventura Salanic, Hipólito Hernández and Felipe Orozco.

All authors are employed by Juarez and Associates, a management consulting firm that has been working in Guatemala in the education and health sectors, and has been contracted by USAID to provide technical assistance for the current Lifelong Learning projects.

evidence obtained from their students' assessment results. The initiation of the school cycle is an ideal time for teachers to obtain a linguistic profile of beginning students. This entails determining the language used by children to communicate with their parents, other children and the community at large.

Familiarisation with the Linguistic Profile of students provides teachers with an opportunity to create an effective communication climate in the classroom while promoting understanding, confidence, linguistic self-sufficiency and an appreciation for the students' ethnic and linguistic identities. This assessment will also help teachers better plan their instruction and tailor it to the linguistic characteristics of their students.

The Linguistic Profile assessment allows for the classification of students in one of the following skill levels:

- Mayan-monolingualism (only K'iche' or Mam spoken)
- Spanish-monolingualism (only Spanish spoken)
- Incipient bilingualism (either Spanish or Mayan language spoken as a first language and some level of knowledge of either Spanish or a Mayan language)
- Bilingualism (speaking two languages without apparent difficulty).

3. CHARACTERISTICS OF THE MAYAN LANGUAGES AND LINGUISTIC INTERFERENCE

Measuring the oral proficiency of a beginning student who speaks one or more languages is particularly important given the possibility of linguistic transference or interference between the spoken languages. It is therefore essential that teachers identify potential cases of linguistic interference and re-direct their teaching techniques towards reinforcing the students' proficiency in both languages.

'Transference' refers to the use of certain elements of a language (usually the mother tongue) and their application to a different language. During the course of learning a second language, the student associates new information with previous knowledge in order to facilitate the acquisition of the new language.

'Interference', on the other hand, may refer to 'negative transference'. Interference has to do with errors made in the second language resulting from exposure to the mother tongue. Frequently, errors made during language acquisition are the result of interference triggered by the mother tongue.

There are two linguistic patterns of transference and/or interference among speakers whose mother tongue is a Mayan language and who are in the process of learning a second language. It is critically important for teachers to realise that these phenomena may occur in their classrooms and it is crucial that they understand that Mayan languages are structurally significantly different from Spanish.

To optimise students' mother tongue and second language skills, teachers must consider the likelihood of errors stemming from linguistic interference during the learning process. These errors reflect the children's strong linguistic ties to the mother tongue and do not necessarily represent personal shortcomings. Teachers must immediately and directly address these errors in a way that does not threaten the physical, emotional and cultural well-being of the child. Also, students must be made to understand that this interference is a normal part of the process of learning a second language that is structurally different from their mother tongue. For example, Mayan-speaking students tend to omit the final vowel since these sounds do not exist as such in Mayan languages (e.g. the word *pelota* (ball) is pronounced 'pelot'). Other examples of sounds (phonemes) in Mayan languages that are not present in Spanish are q'(found in the word *q'ij* meaning sun), tz'(found in the word *tz'i'* meaning dog) and ch'(found in the word *ch'ajch'oj* meaning clean). These examples help demonstrate how knowledge and understanding of the implications of common errors can guide language instruction. For instance,

teachers can choose to incorporate these common errors into their phonological awareness activities as a pre-requisite for beginners learning Spanish.

4. ACTORS INVOLVED IN THE LINGUISTIC PROFILE TOOL

Assessing language skills using the Linguistic Profile tool relies on the coordinated efforts of several hands-on actors: the teacher, the pupils, the school principal and parents. The role of teachers is to administer the assessment to each and every student, (ideally at the beginning of the school year) and to develop a strategy for teaching reading/writing adapted to the linguistic characteristics of their students. The role of school principals is essential since in addition to endorsing the process and the teachers driving it, they are promoting education quality. In cases where the teacher involved is not bilingual, the recommendation is to request the principal's or another teacher's assistance in administering the assessment in K'iche' or Mam. Parents are asked about their children's usage of the language in a family setting. The interview is conducted by the teacher and should ideally take place at the time of student registration. Including parents in the process is a novel feature of this assessment since it opens communication channels on the subject of bilingual instruction between parents and the individuals directly responsible for the education of their children. Finally, the pupils provide information on the language they prefer to use with their parents, classmates and teachers in the course of the interview.

Local Ministry of Education authorities, specifically Teacher Coaches and supervisors are indirectly involved in assessment initiatives. While they are not directly associated with the assessment procedure per se, they do play a major role in terms of supporting, announcing and planning visits, and are thus key actors in the successful roll out of this type of assessment.

5. ASSESSMENT DESIGN AND CONSTRUCTION

5.1 Overall plan

As previously mentioned, the Linguistic Profile can be defined as a diagnostic assessment tool designed to help pre-primary and first grade teachers. The purpose of this assessment is to determine the student's level of oral proficiency in Spanish and the Mayan languages (Mam or K'iche'). This will allow early grade teachers to place emphasis on the language of instruction and plan their strategy for reading/writing activities.

Assessments are individually administered (i.e. each student will do the exercises independently), allowing the teacher to create individual profiles associated with the level of proficiency exhibited by each student in the languages spoken in his/her community. The Linguistic Profile tool includes: a) a student test administered in Spanish and K'iche' or Mam; b) a test administration guide in each of the three languages; c) an interview guide for parents; and d) an interview guide for students.

The assessment development process includes several stages: 1) defining the content of the assessment; 2) elaborating the test and item specifications; 3) item development; 4) assembling the test; 5) layout of the test; 6) defining test administration parameters; 7) performing a pilot test; and 8) scoring and ranking of students. Each of the stages is described in this section.

i) Defining the content of the assessment

The construct to be assessed is linguistic oral skills in K'iche', Mam and Spanish in pre-primary and Grade 1 students exposed to bilingual contexts. This construct includes the following sub-constructs: a) oral interaction; b) oral comprehension and ability to follow directions; c) lexical accuracy; d) lexical production; e) phonology or correct pronunciation; f) grammar; and g) oral expression. Each of these sub-constructs is specifically monitored during the assessment. **Table 1** provides an outline of each sub-construct.

TABLE **1**

Definition of the assessment construct

Construct	Sub-construct	Domain indicators
Linguistic skills	Oral interaction **1st Series**	Understands and answers questions concerning self and his/her environment.
	Oral comprehension and ability to follow directions **2nd Series**	Listens to and follows simple directions.
	Lexical accuracy **3rd Series**	Identifies the correct meaning of words.
	Lexical production **4th Series**	Articulates words in response to the images being shown.
	Phonology or correct pronunciation **5th Series**	Reproduces correctly language-specific sounds in Spanish (pronounces vowels at the end of words in addition to sounds like f, g, ñ and d) and Mayan language, as the case may be.
	Grammar **6th Series**	Uses number and gender agreement between nouns, adjectives and verbs correctly.
	Oral expression **7th Series**	Goes into a long, uninterrupted monologue which narrates coherently a story based on visual stimuli. Uses the elemental/logical language sequence of the language (Mayan: verb, subject, object; Spanish: subject, verb, complements) orally and contextually.

Source: Magzul et al., 2015

ii) Test and item specifications

The first (oral interaction) and second series (oral comprehension and ability to follow directions) include five items each. Their objective is to determine whether the student understands what he/she is being asked and his/her ability to follow directions in the assessed language. The third and fourth series (lexical accuracy and production) include 25 and 10 items, respectively, relating to vocabulary. The purpose of the fifth series is to identify the correct phonology or pronunciation of sounds that are characteristic of the Spanish language. The items used in the assessment were designed with a focus on words ending in vowels (particularly the sounds a, e and o) and specific consonants, such as f, g, d and ñ. In addition, at least three items were developed for each sound. The sixth series (grammar) includes 10 items and attempts to determine the range of expressions associated with gender and number. The seventh and last series (assessment of oral expressions) is made up of five items designed to identify the logical expression of ideas arising from visual stimuli associated with two events or stories.

Table 2 provides a breakdown of test specifications with the number of items per series.

Click here for a list of the item specifications that were developed for each of the dimensions of the assessment

TABLE **2**

Test specification table

CONTENTS	Number of items	Series (%)
Oral interaction	5	6
Oral comprehension and ability to follow directions	5	6
Lexical accuracy	25	31
Lexical production	10	13
Phonology or correct pronunciation	20	25
Grammar	10	13
Oral expression	5	6
Total	**80**	**100**

Source: Magzul et al., 2015

iii) Item development

In line with the specifications described above, reading and language specialists drafted the items

to be administered in Spanish. Members of the project's regional bilingual K'iche' and Mam team were responsible for the items administered in these two Mayan languages.

Prior to item development, the drafting team responsible received training in the following areas: a) purpose of the assessment; b) correct grammatical structure of the items; and c) test specifications. The training process also involved the participation of Mam and K'iche' language specialists. Furthermore, during the revision of items, local verifiers from San Juan Ostuncalco, Quetzaltenango, Comitancillo, San Marcos and Huehuetenango assumed the task of identifying dialect variants of Mam and K'iche' languages.

Most items contain images as part of the stimuli. The image production and selection process was led by the item drafting team. The first version of the assessment, had only black and white images with the twofold purpose of preventing colours from interfering with the students' responses and keeping the images as realistic as possible. However, during the first pilot with teachers, full colour images were suggested. The current version is in full colour.

iv) Pilot testing of items

First pilot test

A first pilot test allowed for an analysis of the validity of the test (i.e. finding evidence that the test does assess what it is supposed to assess). At least four aspects were verified during this first stage: a) understanding of the administration manual; b) functioning of images used in the items; c) teachers' understanding of the administration procedure; and d) administration timeline. Thirty children participated in this first pilot test—20 from an urban school catering to Mayan-speaking students and 10 from a rural bilingual school.

During the pilot test, the project team was able to determine that the administration of the assessment takes between 20 and 25 minutes per student. It also allowed fine-tuning the Linguistic Profile both in terms of stimuli and its manual. Most initial corrections had to do with the images since some of them might not reflect specific contexts of the regions where Mam and K'iche' are spoken. For example, the images associated with the graphic stimuli for 'sir' or 'madam' (i.e. *señor* or *señora*) included in the grammar series were typical of urban areas but did not account for the young people of the high plane regions. In the series associated with language-specific sounds, the gender and number of some words were modified in an effort to avoid the use of localisms while other items incorporated the dialect variants of the municipalities where the USAID Lifelong Learning Project has had significant influence. The dialect variants used in the Totonicapán and Quiché municipalities were evaluated with the support of bilingual professionals. In addition, the manual was adapted to reflect what would be considered correct and incorrect in each section.

Second pilot test

The purpose of the second pilot test was to verify item performance based on methods set forth in the classical test theory (CTT) (Crocker and Algina, 1986). Further validation of items will be done at a later date using item response theory (ibid). This pilot test also launched efforts to classify the level of proficiency exhibited by students in the assessment languages—information that, in turn, would facilitate the creation of the ranking standards. The pilot test yielded 339 response sheets from pre-primary and Grade 1 students. The data was collected in schools operating in the Western Highlands—areas where the Lifelong Learning Project has had a greater presence. In general, more Grade 1 students (211) than pre-primary students (128) took the test. In the K'iche'-speaking area, 185 students took the test. Most of these students attended schools located in different municipalities of Totonicapán, particularly Santa María Chiquimula, San Bartolo Aguas Calientes and Momostenango. Some students lived in Quiché, in the municipality of Joyabaj. In the Mam-speaking area, 154 students took the test with a majority attending schools in Chiantla and Santa Bárbara in Huehuetenango and Concepción Chiquirichapa in Quetzaltenango.

It should be noted that the sample used in this pilot test was not randomly selected. The teachers and

students who participated in this pilot test were beneficiaries of the teacher training programmes subsidised by the USAID Lifelong Learning Project in the region. Therefore, it is not unlikely that the results obtained may contain a self-selection bias.

vi) Item analysis

All items were analysed under the assumptions of true score and measurement error associated with the CTT. Specifically, item reliability and item difficulty analyses were conducted on each of the test's series. Item reliability was determined using coefficient alpha. Coefficient alpha is a reliability index that indicates the internal consistency of a measure. The index varies from 0 to 1; ideally, indexes over 0.8 are desired for all scales of this assessment. **Table 3** shows the reliability for each series of items. Broadly speaking, the test administered to children in K'iche' and Spanish is more reliable than that administered to children in Mam and Spanish. In the case of the latter group, the reliability of the Spanish test was low compared to the Mam test where the indices of reliability were satisfactory. Low reliability indices reflect the low degree of difficulty posed by the Spanish test for Mam-speaking children whose proficiency in Spanish is high.

Item difficulty was considered using the percentage of correct answers as per CTT (item difficulty). In Spanish, this percentage differed depending on whether the items were administered to K'iche'— or Mam-speaking students. Items that did not correspond to the child's first language were expected to be harder to interpret. In the K'iche' speaking area, the Spanish items proved much harder than in the Mam speaking area. Results are consistent with the data obtained from the student questionnaire. In the K'iche'-speaking area, 67% of the students report using K'iche' versus 40% who report using Spanish. In this area, 12 students (7%) report using either language indistinctly. In the Mam area, however, 64% of the students report using Spanish while 33% would rather use Mam for communicating. No students in this area report using both.

Rankings based on the results of this first round of item analysis should be approached with caution given the limitations inherent to the Classical Theory of Item Analysis and the restrictions of the available sample of teachers/students (Crocker and Algina, 1986). In future, analyses using item response theory will be performed using a different sample of teachers as part of on-going efforts to further evaluate the test.

Click here for Item difficulty table

v) Test layout and assembly

The test is administered by the teacher to students on an individual basis. Pre-primary and Grade 1 teachers administer the test at the beginning of the school year to help them plan their instruction strategy for language as well as reading and writing.

The test consists of the following components: a) a flip-chart booklet made of sturdy material that presents the stimuli to the students and the directions to the teacher simultaneously; b) a response sheet to record the students' answers; c) an administration, scoring and interpretation manual; and d) a parent

TABLE **3**

Item reliability reported as coefficient alpha

Series	No. of items	K'iche'	Spanish	Mam	Spanish
Oral interaction	5	0.969	0.992	0.904	0.53
Oral comprehension	5	0.976	1	0.976	0.654
Vocabulary (accuracy)	25	0.977	0.996	0.996	0.83
Vocabulary (production)	10	0.942	0.977	0.944	0.714
Language-specific sounds	20	0.97	0.98	0.951	0.765
Grammar	10	0.928	0.936	0.898	0.743
Oral expression	5	0.876	----	0.862	

Source: USAID Lifelong Learning Data

questionnaire on language exposure. The response of the parent questionnaire on the language exposure sheet includes questions intended to identify a student's level of exposure to the different languages.

vi) Test administration

The test is administered by the teacher to each one of his/her students. Normally, the teacher will ask a student to sit in front of him/her so the stimuli are readily seen. Teachers register the students' answers in the response sheet. Several items require teachers to evaluate the answer given by the student. For example, if the student replies 'parasol' instead of 'umbrella' then the teacher should consider this as a correct answer.

Teachers responsible for administering the test received training on administration guidelines, scoring and the interpretation of results. In addition, an assessment administration manual was developed exclusively for their use. This manual was implemented during the first pilot test to verify that it would be clearly understood by teachers. Currently, the section on the interpretation of the results is still being fine-tuned. More data should be collected from teachers active in the project's target areas to facilitate a more accurate elaboration of the criteria required to rank students. This document contains emerging criteria based on a limited number of cases (339 students in both regions).

vi) Scoring of responses and interpretation of results

Each item is scored using the binary method. In other words, students are awarded '1' for each correct answer and '0' for each incorrect answer. A non-response (the child remains silent) is considered incorrect. Because teachers must evaluate the child's response to every item, the manual had to include examples that would ensure consistent scoring by teachers.

As mentioned previously, during the second pilot test, the assessment and its component items were analysed to determine the consistency of results and item difficulty according to the Classical Theory of Tests. Because developing appropriate criteria requires a broader base of data that is not currently available, teachers have been asked to classify students into four linguistic categories: a) Spanish monolingual; b) Mayan monolingual; c) Incipient bilingual; and d) Bilingual. Consequently, classification criteria are expected to be coherent with the teachers' own perception of bilingualism.

6. DESCRIPTION OF IDENTIFIED PROFILES

On average, students in the K'iche'-speaking region answered 32 items correctly in the Spanish test and 47 correctly in the K'iche' test (out of a total of 80 items on both tests). **Table 4** shows the distribution of results obtained from this population. Notably, while 52% of the students did not answer any question in Spanish, this was the case for only 21% of the student taking the test in K'iche. On the other hand, it should be noted that variability in K'iche' is less marked than variability in Spanish in this region.

Teachers were asked to classify students into one of four linguistic profiles. In this region, however, many students were not classified in any linguistic profile at all (114/185) while 13% were classified as Mayan monolingual—a figure that does not match the high percentage of students who were unable to complete the Spanish assessment (see Table 4). Nonetheless, 9% of the students assessed were classified as Spanish monolingual—a percentage that does not

TABLE **4**

Distribution of correct answers among children tested in K'iche' and Spanish

	n	Minimum	Maximum	Possible range of scores	Median
Total number of correct responses in K'iche'	185	0	80	0 to 80	58
Total number of correct responses in Spanish	185	0	80	0 to 80	0

Source: USAID Lifelong Learning Data

correspond to the 20% of students who did not answer any questions in K'iche' (*see* ***Table 5***). The results reflect the great difficulty teachers encounter when classifying students based on such scant information. It is likely that teachers classify students according to their individual perception rather than relying on the data. This should be a topic covered in the next teacher training initiative on the subject of test administration.

In the Mam-speaking region, students answered an average of 65 items correctly in the Spanish test and 47 items correctly in the Mam test. **Table 6** shows the result distribution for this population. It should be emphasized, however, that 53% of the students did not answer any question in Mam while no such cases were reported for Spanish. Variability was greater in the Mam test than in the Spanish test. Results suggest that, in this region, Spanish represents the first language and Mam the second language.

In the Mam-speaking region, teachers were also asked to classify students into one of four linguistic profiles. However, 18% of the students were not classified in any linguistic profile at all while 56% were classified as Spanish monolingual and only 3% were classified as Mam monolingual (*see* ***Table 7***).

Results yielded by the second pilot study suggest the need to broaden the data base in order to look more closely at the distribution of children who report exposure to two languages in both regions.

It should be noted that the test is still in its construction stages and that an improved version will be produced based on recent results. The new version will be field tested in 2016.

7. TEACHER PARTICIPATION IN THE PILOT TEST

As mentioned earlier, the purpose of the Linguistic Profile assessment is to identify the mother tongue of pre-primary and Grade 1 students attending Guatemala's official schools. The information derived from this assessment is essential as it will allow teachers to gain insight into the students they will be teaching during the school year and, above all, help teachers adapt teaching/learning strategies according to the students' mother tongue. As described in the National Curriculum (Currículo Nacional Base, CNB) for Grade 1, "….teaching should commence in the mother tongue…learning in other languages facilitates the transference of linguistic skills". Additionally, the National Curriculum suggests incorporating a linguistic diagnosis to

TABLE **5**

Students in the K'iche'-speaking region classified by their teacher using four linguistic profiles

	Frequency	Percentage	Valid percentage	Accumulated percentage
Did not rank the student	114	62	62	62
Mayan monolingual	24	13	13	75
Spanish monolingual	17	9	9.2	84
Incipient	29	16	16	100
Bilingual	1	1	1	100
Total	185	100	100	

Source: USAID Lifelong Learning Data

TABLE **6**

Distribution of correct answers for children exposed to Spanish and Mam

	N	Minimum	Maximum	Median
Total number of correct responses in mam	154	0	76	26
Total number of correct responses in Spanish	150	32	80	65

Source: USAID Lifelong Learning Data

TABLE **7**

Students in the Mam-speaking region classified by their teacher using four linguistic profiles

	Frequency	Percentage	Valid percentage	Accumulated percentage
Did not rank the student	27	18	18	18
Mayan monolingual	4	3	3	20
Spanish monolingual	84	55	55	75
Incipient	20	13	13	88
Bilingual	19	12	12	100
Total	154	100	100	

Source: USAID Lifelong Learning Data

determine the level of bilingualism in students' as an activity, with a view to "stimulating learning in the areas of Communication and Second Language".

So far, no standard linguistic diagnostic tool is available to teachers. While DIGEBI designed a tool for this type of assessment, its scope and high cost proved unrealistic and it is no longer used. As a result, assessments usually consist of giving students general directions in the Spanish or Mayan language and depending on whether or not they follow directions, familiarity or lack of familiarity with the language is assumed. However, this type of assessment deviates greatly from a standard procedure and does not provide accurate information on the students' language proficiency. Thus, a decision was made to involve teachers in the pilot tests. Teacher participation offers the additional advantage of evaluating if the exercise meets teachers' needs (the final users) while simultaneously capturing item performance inputs.

The teachers who participated in the second pilot test worked in official schools located in 12 municipalities in five departments where the USAID Lifelong Learning Project has a strong presence. K'iche' is spoken in seven of these municipalities while Mam is spoken in the other five municipalities. The invitation to participate in this second test was issued by the Departmental Directorate of Education (DIDEDUC) of the five departments and was led mostly by municipal Teacher Coaches and supervisors. Out of the 352 teachers who participated in the test, 75% (254) were women. The selection criteria required that teachers: 1) taught pre-primary and Grade 1; 2) were interested in the teacher training process; and 3) were bilingual (in Mam or K'iche', and Spanish). The second criterion is particularly relevant since combining the formative process with appropriate teaching tools builds a richer experience. The teacher training was launched in July 2015 and is subsidised by the USAID Lifelong Learning Project and run by the *Universidad del Valle de Guatemala* (UVG). It consists of a Diploma in Reading and Writing, targeted to bilingual and intercultural contexts. Regarding the third criterion, only 10% (30) of the teachers are not bilingual. However, given their expressed desire to participate, the enormous benefits associated with learning about reading and writing in Spanish and considering how important it is for students to learn this subject in their mother tongue, they were allowed to participate in the teacher training programme. They were advised to ask a bilingual colleague or the principal to conduct the assessment in the Mayan language in order to ensure that the Linguistic Profile assessment was carried out in both languages. During the teacher training programme, participants will receive recommendations on how to instruct bilingual students.

Table 8 shows the number of participating teachers by municipality and indicates their linguistic status (mono or bilingual).

Training activities took place in March 2015 and were led by two USAID Lifelong Learning Project pedagogical bilingual coaches and three consultants with vast experience in bilingual education. Out of these five professionals, two are Mam-Spanish speakers and three are K'iche'-

Spanish speakers. The training session lasted approximately 4 hours and involved some 20 teachers. For the training sessions, Teacher Coaches arranged for the use of classrooms made available by several principals of public and private schools. In addition to the administration process and Linguistic Profile assessment materials, information was provided on the Diploma in Reading and Writing programme, and on the importance of bilingual education, specifically in terms of reading and writing acquisition. The methodology used in the administration of the assessment instrument included these four steps:

1. Context

Teachers were given a brief description of what has been defined as a student's 'linguistic profile', placing emphasis on the fact that this is an initiative promoted by the national education system.

2. Distribution of materials

For training purposes, each teacher received a set of the materials they would be using. These materials included i) general guidelines; ii) an administration booklet in Mayan and Spanish; iii) response sheets of student interviews as well as Mam/K'iche' and Spanish tests.

3. Modelling

To illustrate the instrument's administration process, the facilitator described the basic content of the administration guide and subsequently, invited one of the teachers to play the role of a student. The instrument was first administered in Mayan and afterwards in Spanish. Following the presentation, there was a round of questions intended to clarify any doubts and ensure that the procedure was fully understood, before moving on to practice in pairs.

4. Pair practice

Teachers were paired up to perform professional exercises under the supervision of a facilitator. Bilingual teachers worked in both languages while Spanish monolingual teachers were restricted to only one language. During this session, teachers noted that each assessment took between 12 and 15 minutes per student per language, for a total of 24 to 30 minutes.

TABLE **8**

Teachers participating in the pilot test

No.	Department	Municipality	Number of teachers			Linguistic profile of teachers	
			Total	Men	Women	Bilingual (k'iche'/mam-Spanish)	Monolingual Spanish
1	Quetzaltenango	Concepción Chiquirichapa	24	2	22	24	0
		San Juan Ostuncalco	33	1	32	30	3
2	Huehuetenango	Santa Bárbara	21	7	14	21	0
		Chiantla	20	3	17	4	16
3	San Marcos	Concepción Tutuapa	35	10	25	35	0
4	Quiché	Joyabaj	35	14	21	29	6
		San Pedro Jocopilas	34	17	17	33	0
5	Totonicapán	Totonicapán	31	8	23	30	1
		Momostenango	30	10	20	28	2
		San Bartolo Aguas Calientes	30	7	13	26	4
		Santa Lucía La Reforma	30	13	17	28	2
		Santa María Chiquimula	30	7	23	30	0
Total			**352**	**98 (28%)**	**254 (72%)**	**318 -90%**	**34 -10%**

Source: Magzul et al., 2015

© USAID Lifelong Learning Project, Guatemala

Following pair practice, each teacher received assessment materials according to the number of students enrolled in their class. According to the work plan agreed on, administration of the assessment was to be completed on April 30, 2015. Each teacher developed his/her own work plan, which depended basically on the number of participating students under their responsibility. For example, some teachers decided they would devote two hours every morning to the test, some chose to administer the test during the hour of recess while others opted to work with at least two students a day until the total number of students had been assessed.

Following administration of the tests, teachers agreed that the material received met their expectations as it was high-quality, practical, accessible, dynamic and relevant to the context. Furthermore, there was consensus that the assessment experience had proven most interesting since it provided them with an accurate picture of their students' potential. Based on this feedback, it could be surmised that teachers valued this exercise highly.

A novel aspect of this assessment process is the interview conducted with parents at the beginning of the school year, specifically when they first register their children in school. The interview attempts to determine how the student uses the language (or languages) spoken at home. This information will help the teacher enrich the information derived from the Linguistic Profile assessment. Additionally, the interview gives teachers an opportunity to show parents the importance that assessments have in their daily professional practice and emphasise the fact that learning should take place in the student's mother tongue. Tutors participating in the Diploma in Reading and Writing programme will accompany teachers when they give parents the results of the interview and the Linguistic Profile assessment.

8. CHALLENGES FACED AND LESSONS LEARNED

This section highlights some of the challenges encountered throughout the process of administering the pilot test of the Linguistic Profile assessment and the preventive measures adopted. The section concludes with a synthesis of the lessons learned.

8.1 Difficulties encountered with the roll out of the assessment

The following challenges were faced in the administration of the Linguistic Profile assessment and were countered with these potential solutions:

- Despite the fact that invitations to the training initiative were issued through Teacher Coaches or supervisors, not all the invited teachers attended. Various reasons for absences were offered—among them: overlapping activities, procedures related to payments, audits conducted by the Bureau of the Comptroller General, work-related meetings or previous appointments with physicians. The USAID Lifelong Learning Project pedagogical accompaniment specialists implemented a follow-up plan intended to include these absent teachers in the process. This often required visiting teachers at the schools where they taught. While these efforts have been fruitful for the most part, involving every teacher convoked by the Ministry of Education in this training initiative continues to pose a challenge.

- There were unexpected internal changes in the roster of participants. Although initially teachers showed interest in pursuing the Diploma programme, some had to desist due to reassignment to a different grade or incompatibility with other professional development programmes (e.g. the PADEP-D).

- During pair practice in the teacher training, it became evident that some teachers were experiencing difficulties reading the names and words associated with K'iche' or Mam images. In this regard, a recommendation was made to practice individually or with the assistance of bilingual colleagues.

- Uncertainty about the initiation of the Diploma programme posed another challenge. Originally scheduled for January 2015, the programme did not begin until July 2015, a situation that caused some discontent among teachers. This prompted the UVG, Teacher Coaches and project stakeholders to undertake joint follow-up actions to keep teachers informed.

8.2 Lessons learned

- In general, teachers supported the idea of a formative process that went hand in hand with tools designed to improve their professional practice as this represents a link between theoretical knowledge and practical daily work.

- Despite the interest shown by participating teachers, when the time came to collect the answer sheets it became evident that many of them had yet to administer the assessment. One of the reasons given was the excessive amount of time that had to be allotted to administering the instruments. This will be an important issue to emphasise when finalising the technical support procedure as teachers must be made aware of the importance of having this information readily available in the time given.

- The strategy of involving local Ministry of Education authorities in planned activities constitutes a way of guaranteeing their successful implementation.

- In Chiantla and Santa Barbara municipalities in Huehuetenango, training was conducted in the afternoons (i.e. out of working hours). In general, this was regarded as a positive outcome. This was perceived as both an endorsement of the projected training schedules and as a potential policy that could be adopted to prevent unforeseen scheduling issues.

REFERENCES

Crocker, L. and Algina, J. (1986). *Introduction to Classical and Modern Test Theory.* New York: Holt, Rinehart and Winston.

Magzul, J. (2015). Oral Proficiency evaluation in Guatemala. (Interview with L. Rosales).

Magzul, J., Maldonado, S. and Montenegro, R. (2015). *Reporte de Perfil Lingüístico.* Guatemala: USAID Leer y Aprender.

Use of Literacy Assessment Results to Improve Reading Comprehension in Nicaragua's National Reading Campaign

VANESSA CASTRO CARDENAL
Instituto para el Desarrollo de la Democracia (IPADE)

ABBREVIATIONS

AMCHAM American Chamber of Commerce
BDF *Banco de Finanzas* (Bank of Finance)
CAPRI *Centro de Apoyo a Programas y Proyectos* (Support Center for Programs and Projects)
CESESMA *Centro de Servicios Educativos en Salud y Medio Ambiente* (Centre for Education in Health and Environment)
CODENI *Coordinadora de Niñez* (Childhood Coordinator)
EDUQUEMOS *Foro Educativo Nicaragüense* (Nicaraguan Education Forum)
EGRA Early Grade Reading Assessment
ELI *Evaluación de Lectura Inicial* (Initial Language Assessment or the Spanish version of EGRA)
FAS *Fonético Analítico Sintético*
GDP Gross domestic product
IDEL *Indicadores Dinámicos del Éxito en la Lectura* (Dynamic Indicators of Reading Success)
IPADE *Instituto para el Desarrollo de la Democracia* (Institute for the Development of Democracy)
MINED *Ministerio de Educación Cultura y Deportes* (Ministry of Education, Culture and Sports)
NGOs Non-governmental organizations
ORF Oral reading fluency
PRIDI *Proyecto Regional de Indicadores Desarrollo Infantil* (Regional Project of Child Development Indicators)
RACN *Región Autónoma Caribe Norte* (North Caribbean Autonomous Region)
RTI Research Triangle Institute
WCPM Words correct per minute

1. INTRODUCTION

This article presents a case study of the impact of literacy assessment on the national reading campaign in Nicaragua, *Vamos a Leer, leer es divertido* (Let's Read, reading is fun), promoted since 2010. The article describes Nicaragua´s current situation in the field of education, summarises the context and purpose of the reading campaign, and addresses the work of three organizations that work in *Vamos a Leer* to improve literacy in Grade 1. These cases were chosen to illustrate the effects of the campaign in the participating schools. Finally, the article presents key conclusions and describes future plans to improve literacy.

2. BACKGROUND

Nicaragua has a per capita gross domestic product (GDP) of US$2,000—the lowest in Central America after Haiti. The country has approximately 9,000 primary schools, of which 67% are multigrade (schools that are centres which have one teacher for several grades). Among these 9,000 schools, 35% do not offer a complete primary education.

These multigrade schools are mainly located in rural areas or on the Caribbean coast, which is home to 15% of the population of the country for whom Spanish is a second language. Although the country has a net enrolment rate of 97% in primary education (UNESCO Institute for Statistics, 2015), less than 60% reach the end of primary education; this has been the case for more than 10 years, and represents some of the lowest completion rates in Central America (EDUQUEMOS, 2013). Retention or keeping children in the education system is a challenge and students with fewer resources have the highest dropout rate, generally dropping out between Grade 2 and 3 (MINED, 2015). In total, only 21% of students who come from families that live in poverty manage to complete primary school (IEEPP, 2012).

Primary teachers have been educated in traditional teacher-training schools that do not offer a higher education degree. In these institutions, literacy methods have not been modernized and are not based on recent research so there is a tendency to overlook findings on how literacy skills develop. To counter this situation, a phonics method, the *Fónetico Análitico Sintético* (FAS) was adopted in January 2015 to teach literacy and teachers have been receiving in-service training on how to use it. However, there are serious qualification issues among teachers—about one third of primary school teachers teach without any formal training (Laguna, 2009). According to two literacy skills diagnoses conducted in 2008 and 2009 using the Early Grade Reading Assessment (EGRA), the most outstanding and highly motivated teachers interviewed expressed that they aspired to quit teaching and envisioned themselves in another profession outside the educational system within the next five years (Castro et al., 2009; 2010). This sentiment is linked to lack of incentives and social recognition for the profession and wages that do not exceed US$226 per month. Indeed, Nicaraguan teachers receive the lowest salaries in Central America and they cannot afford more than 60% of what is known as the *canasta básica*—a list of 56 basic items for living (Rogers, 2012).

Nicaragua has lacked an evaluative culture until recently and thus measuring the results of the learning process has not been part of the instruction cycle. To assess language and mathematics, national written tests have been applied to children in Grade 4 and Grade 6 and samples have been taken every four years for nearly two decades. However, by Grades 4 and 6, a significant percentage of students have already dropped out of school (MINED, 2015).

In 2007, after the Research Triangle Institute (RTI) International piloted an EGRA in Spanish, the Ministry of Education decided to use this oral instrument to evaluate reading abilities in early grades, referring to it as the Initial Language Assessment (ELI in Spanish). The assessments that began in 2009 had two limitations. First, the assessment used the same story to evaluate students as the pilot for more than three years so many students had by then memorised this story. Second, the evaluation plan was too ambitious. The ELI or EGRA were applied to a sample of approximately 50,000 children per year, which prevented the data from being processed, analysed and used in a timely manner.

3. CONTEXT OF THE *VAMOS A LEER* (LET´S READ) CAMPAIGN

Since 2010, a coalition of a large group of organizations from the private sector and non-government organizations (NGOs) begun to promote reading abilities among children in Grade 1. The organizations currently participating in this effort are: AMCHAM, *Banco de Finanzas* (BDF), *Café Soluble*, CAPRI, *Carlos Cuadra Publicidad*, CODENI, COMASA, *Semillas del Progreso, Comunica, Cuculmeca,* CESESMA, EDUQUEMOS, *Fe y Alegría, Grupo Pellas*, IPADE, *Nicaragua Lee*, *Rayo de Sol*, Save the Children, *Vicariato Apostólico de la Pastoral de Bluefields* and World Vision (previously *Visión Mundial*) and these foundations: *Impulso, Libros para Niños, Pantaleón, Uno, Coén, Telefónica* and *Vientos de Paz*. Of these organizations, 13 work directly with schools in poor municipalities while the

rest support promotional efforts through advertising, communication and/or donations in cash or kind.

The decision to launch this campaign was originally embraced by 18 organizations as a consequence of the dissemination of the results of the children's reading outcomes from the assessments done in Grade 1 and Grade 4 (Castro et al., 2009; 2010) along with the disclosure of statistical information on early dropout rates that were affecting the country. The creation of this coalition was supported by an international context that favored literacy and in which the United Nation's Education For All initiative and a group of donors were strongly prioritizing the task of improving education quality. Students' lack of mastery of literacy skills detected in the 2009 and the 2010 studies (Castro et al., 2009; 2010) and the high dropout and repetition rates in Grades 1, 2 and 3—of which is highest in Grade 1—contributed to the coalition´s decision to concentrate its efforts on improving literacy in Grade 1. When the campaign, *Todos a Leer,* started in 2010, it took advantage of the valuable social capital and work experience of the participating foundations and NGOs that had spent years supporting schools in both rural and urban areas in extreme poverty. The coalition, which currently comprises 27 foundations and organizations (including sponsors and the organizations that implement the campaign in the schools), began its work to support literacy with a single initiative—a contest among Grade 1 classes in the country. The contest was based on the results of an EGRA literacy evaluation of fluency and comprehension. This short version of the EGRA measures oral fluency using a story of no more than 65 words and comprehension by posing five questions based on this story.

The EGRA short test used for the National Reading Campaign can be accessed here

To win the contest, at least 80% of students in a given Grade 1 class had to read the EGRA passage fluently and comprehensively. All Grade 1 classes entering the contest also had to meet the goals for each contest period set by the committee coordinating the campaign. The goals set were in compliance with the Dynamic Indicators of Reading Success (IDEL in Spanish) guidelines. In addition, the Grade 1 classes had to meet a few basic requirements at the time of registration to participate in the contest:

1. The teacher should have attended 90% of the calendar days in the school year,

2. The teacher must have taught an average of 22 hours per week,
3. The teacher should be motivated to teach the Grade 1 students how to read.

All of these requirements were and are currently checked with the school principal, the coalition organizations or associations of parents and student organizations, as available.

In 2013, the reading campaign changed its name to *Vamos a Leer, leer es divertido!* ('Let's Read, reading is fun!') and began its sixth cycle in 2015. *Vamos a leer* works in a decentralized manner—however, a coordinating committee (made up of representatives from the NGOs working in the territories to promote reading in schools) sets fluency goals at each contest stage and makes methodological decisions with the advice of a literacy specialist. During the reading campaign, the literacy specialist was in charge of dealing with methodological issues and preparing the short versions of the EGRA used at each stage of the campaign. Each stage of the campaign uses three short EGRA versions featuring three different stories (stories are varied at each stage to deter memorisation). The first stage of the competition begins in August and each Grade 1 classroom enrolled in the campaign participates. The second stage takes place in September at the municipal level during the second semester of school. The schools that participate at this stage are those in which 80% of the students in a given Grade 1 classroom have achieved the set goals of correct words read per minute and answering correctly the questions based on the reading (about two questions). The closure of the contest and the campaign takes place in October in the capital city of Managua where all the winning schools from the municipal contests congregate to participate. Each year the campaign uses at least nine iterations of the short version of the EGRA. The literacy specialist trains the test administrators annually to ensure that they are prepared to administer the EGRA short test. These test administrators are generally volunteers and the number of people administering the tests in a given area depends on the number of participating students. The goal is to administer the tests to all Grade 1 students enrolled in the contest and who present themselves on the day that the EGRA short test is administered. The test is unannounced. The EGRA short test is sent to each participating organization 24 hours before the contest to prevent its early disclosure. Each NGO or foundation is responsible for applying the EGRA short test and processing the resulting data using their own funds. For the final phase of the contest, a group with experience working on the EGRA studies is hired to ensure the accuracy of the results and that students are assessed in a timely fashion—this has become necessary as up to 250 students made it to the finals in the last three years of the contest.

In 2014, 425 Grade 1 classes with a total enrolment of 11,435 students participated and the winners at the municipal level totaled 250 students from 10 schools. These students then went to Managua to compete in three different categories: rural multigrades (i.e. school that are centres with one teacher for several grades), regular schools with one teacher per grade and church schools whose students pay a small fee to enrol.

In 2012, after detecting a severe shortage of children´s storybooks in schools, the campaign began a process of fundraising to provide each participating school with mini libraries of about 30 books each. The largest book donor has been the international foundation *Vientos de Paz*, which has awarded annual grants of US$25,000 to the campaign since 2012. Thanks to this foundation's support, 800 mini-libraries with a total of approximately 33,000 books have been distributed, benefiting an average of 250 schools per year.

Each organization contributes its experience and capabilities, creating an environment that not only facilitates discussion on common issues encountered but also drives an agenda to find solutions to promote comprehensive reading. A great number of initiatives have stemmed from the specific experiences of each organization and the analysis of data generated by the annual administrations of the EGRA, including training for

teachers and student leaders as well as community work to promote their participation in this initiative.

During the first three years of the campaign´s implementation, the EGRA short version assessment of students' progress was only a means to motivate teachers, principals, parents and students to improve reading. The contest generated enthusiasm as it awarded significant prizes[1] to winning classrooms, schools and children. In the first years of work, the opportunity to process and use the data to improve the learning process was missed. However, in the fourth edition of the campaign (2013) with the knowledge that that the awards had motivated teachers and students' families to strive harder in Grade 1, a training process was launched to use the EGRA short test results for decision-making. The process continued with a stronger emphasis in 2014 and that year, most organizations began to use the results of the annual assessment to improve teaching and promote initiatives to foster reading.

4. CONTRIBUTIONS TO INSTRUCTION AND READING PROMOTION FROM THREE CASES

Save the Children and World Vision have participated in the reading campaign for several years. Save the Children is a founding member of the campaign and has worked to support Nicaragua's education since 1987. World Vision has worked in Nicaragua since 1989 and joined the campaign in 2012. The working methods of these NGOs differ. Save the Children sponsors the campaign, working with five different partners (four NGOs and one foundation)—together these organizations serve 97 schools. Save the Children's counterparts in the campaign are CAPRI that focuses its effort in poor urban neighborhoods of Managua, CESESMA that works in rural municipalities in Matagalpa, Cuculmeca that focuses its work in a rural municipality in Jinotega, IPADE that works with schools in the Northern Region of the Caribbean Coast, and the foundation *Fundación Libros para Niños* that promotes recreational reading in communities in the municipality of Tuma La Dalia.

World Vision has made donations to finance the campaign's awards and also works in schools to promote education directly. It has a presence in 535 communities and 531 schools. In 2014, Save the Children's counterparts had 61 schools enrolled in the campaign comprising 143 Grade 1 groups while World Vision participated with 62 schools and 68 Grade 1 classes. This section documents the experience of two Save the Children's counterparts (CAPRI and IPADE) as well as World Vision's experience.

The Support Center for Programs and Projects (CAPRI in Spanish) was founded in 1990 and has worked with the campaign's coalition since 2010. It provides assistance mostly to regular schools located in poor marginalised neighborhoods in Managua. In these schools, the average number of Grade 1 pupils per teacher varies between 45 and 50 students.

The Institute for Development and Democracy (IPADE in Spanish) was created in 1993 and has worked since 1999 in the municipalities of the so-called mining triangle (Siuna, Rosita and Bonanza) as well as the municipalities of Prinzapolka and Puerto Cabezas, and more recently, in the municipality of Mulukukú—all located in the Autonomous Region of the Northern Caribbean (RACN in Spanish). The goal of this institution has been to contribute to agro ecological development and democratic governance through educational and community projects to help build citizenship, environmental governance and improve educational quality. The IPADE also works with indigenous communities in these territories. Most of the schools in the RACN are multigrade while a small number

1 Thanks to the sponsorship received from private companies and other organizations, a backpack containing learning materials, school supplies, shoes and clothing is distributed to each student competing at the final level of the national contest. Teachers and principals from the winning schools are also awarded with a cell phone, an endowment of books and a household appliance. In accordance to the place won in the contest, a winning school can receive computers, library furniture, books and television sets with DVD players.

© Margarita Montealegre, Nicaragua

of centres offer bilingual intercultural education in Miskito and Mayagna.

World Vision works in marginalised and impoverished rural communities in 10 of the 17 municipality departments in the country. A high percentage of the communities where World Vision works are located in a Nicaraguan territory known as the 'dry zone' where extreme poverty is very high because the main livelihood of the inhabitants is based on basic grain agriculture in forest lands with the lowest rainfall in the country.

4.1 The campaign and creating a culture of assessment

Thanks to the widespread use of the EGRA short version in the reading campaign, NGOs, teachers and Ministry of Education, Culture and Sports (MINED) officials have become familiar with this literacy assessment that can be easily and quickly administered by volunteers who are not trained as researchers. After several years of working with the EGRA short version and much training done to manage and use the data generated by the test applications, a large number of people have since developed good analytical skills. Thus, the local contest results of hundreds of schools can now be used to assess how children are reading as well as detect learning difficulties. The use of evaluation as a key feature in the instruction cycle has led many of the participating organizations to create literacy baselines and develop monitoring systems, and even hire consultants and/or specialised agencies to provide greater depth and coverage in their evaluation efforts.

As a result of the poor performance in the contest by a large group of schools covered by Save the Children's counterparts that work directly with schools (i.e. CAPRI, CUCULMECA, CESESMA, IPADE), Save the Children performed a qualitative study of eight schools. The schools were chosen purposely to include a sample of four urban and four rural schools. To better understand the educational factors underlying the excellent or poor results achieved by some schools during the reading contest, the sample included four centres that performed above average and four that performed below average. The results of this study yielded important information on positive teacher and community practices that can enhance literacy. In terms of teacher's planning and organization of their language classes, poor planning was usually linked to poor student results and lack of teacher training in literacy instruction (O'Connell, 2012). In 2015, two of the organizations participating in *Vamos a Leer*,

Save the Children International and the *Asociación de Familia Padre Fabretto,* submitted tenders to hire international organizations to administer a full version of the EGRA to students in Grades 1-3. According to the information provided to the reading campaign's committee, their goal was to evaluate precursory skills to literacy development while continuing to measure fluency and comprehension.

A report prepared for Save the Children by an external consultant (Rivera, 2014) based on a study performed in schools attended by the CAPRI in Managua found that the campaign had contributed to enhancing the coordinated efforts between the CAPRI and the Ministry of Education, and fostered "an evaluative culture of learning with an emphasis on reading comprehension". The report further states that:

- The creation of a short version of the EGRA has simplified its application, facilitating its widespread use as an assessment tool to evaluate learning in Grade 1.

- The application of the EGRA in participating classrooms and the sharing of these results with teachers and principals has helped them to appreciate the usefulness of evaluating as they use the data provided to improve lesson planning. The data collected have helped to identify students' difficulties and to implement measures to overcome them. Among other initiatives, the MINED (with the support of the CAPRI) is also providing individual attention to students who have greater difficulties.

- EGRA results have also helped principals and vice principals provide more effective and efficient monitoring of their Grade 1 teachers while providing support to their lesson planning and pedagogical decisions.

- The CAPRI uses the information from the EGRA's application to organize their plans and set priorities. This has resulted in a more efficient collaboration with MINED as well as improved teacher training by basing the plans on students' learning needs and teachers' shortcomings.

© Margarita Montealegre, Nicaragua

During 2014, World Vision performed evaluative studies with two purposes in mind. The first was to identify literacy promotion practices used in schools. The study assessed a total of 573 families, 410 teachers and 693 students from 52 schools. The second was to create a baseline with the EGRA results. The baseline study was conducted in 39 schools and included a sample of 1,604 children from Grade 1 to Grade 3 (World Vision report, 2014). The results of these investigations allowed World Vision to identify indicators that are now used in their monitoring system. These include:

- Percentage of parents and/or guardians who promote reading at home

- Percentage of children who read fluently and comprehensively in Grades 1, 2 and 3

- Number of boys and girls who participate in book clubs

- Percentage of children who have developed reading habits in book clubs.

According to the World Vision's report in 2014, the EGRA's data can be used to "systematically reflect if the actions implemented with our partners are generating significant results".

The IPADE also created a baseline with two indicators: fluency and reading comprehension. The EGRA's short version was administered in 2014 in 20 schools in the municipality of Siuna located in the RACN. The study covered 323 students—all the children present in their Grade 1 classrooms when the test was administered.

The baseline indicated that 42% of students achieved the reading fluency goal of 28 words correct per minute (WCPM) in September—three months before the end of the school year. The assessment showed that 33% responded correctly to the questions related to the passage read. This gap of nearly 9% between the percentage of students able to achieve fluency and the percentage of students who comprehend the passage has been a constant during the campaign.

The assessments conducted during the five years of the campaign have yielded consistent results showing that the number of students reaching the fluency goal always exceeds the number of students who understood what they read. When the campaign committee analysed this gap (9% or more per year), it drew the conclusion that comprehension limitations are probably linked to a knowledge gap (Snow, 2015) associated with several factors, including low levels of education of children´s mothers, scarce enrolment in preschool and malnutrition. Therefore, these children enter Grade 1 with weaknesses in their oral vocabulary, precursory literacy skills and have difficulties making meaning from stories not related to their context. Lending credence to this assumption, the reading skills diagnosis performed in the Caribbean in 2010 found that the percentage of preschool attendance among Miskito children was 4% (Castro et al., 2010). Therefore, it is possible that the literacy efforts that began six years ago are helping to improve children's decoding skills. Despite improvements, greater efforts should be made to close this knowledge gap, expand children's vocabulary and train them in the use of strategies to improve comprehension. The results found in a study performed by the Regional Child Development Indicator Project (Verdisco et al., 2014) in four Latin American countries also underscores the need for greater efforts to improve learning. This study measured, among other things, language and communication development in children aged 2 to 4 years using the PEABODY test (IPTV) and the Engle Scale. Nicaragua scored below the regional average and was the country with the lowest average score in the area of language and communication development.

4.2 Quality of education should be everyone's concern: community and family participation

The research by the external consultant for Save the Children demonstrated that the CAPRI´s work over several years in the *Vamos a Leer* campaign has promoted a qualitative leap in the participation of families (Rivera, 2014). The campaign has helped teachers involve parents in reading and telling stories to their children. This initiative by the CAPRI and other NGOs was sparked during the 2011 competition when parents borrowed books from their friends and families to prepare their children for the contest. The involvement of families in the learning of children has driven parents to overcome the tendency of only participating in logistical tasks at the school, such as preparing meals and cleaning the educational centres. A mother interviewed by Rivera stated, "my role as a mother is to help them study, help them with homework at home [...] the teacher told us that families can support [children] by reading, accompanying them when they're doing their homework tasks".

World Vision has also been promoting parent training to support inclusion and diminish drop out rates in schools that enrol children living in extreme poverty. In 2014, 288 parents and 888 children were trained. Another interesting initiative to promote community and family participation has been the creation of *Caminos Letrados* (letter roads). Considering that there are no advertising signs in rural areas and that children's exposure to printed letters is almost nonexistent, World Vision designed a strategy to encourage parents to create cardboard signs with mottos and educational sentences for children to read when walking to or from school.

In 2014, the IPADE managed to create and run five reading circles for children at risk of illiteracy due to their reading difficulties. These circles operated in 43 homes in the neighborhoods and communities near the 15 schools served by this NGO. The IPADE also formed a volunteer brigade with young people willing to promote recreational reading in schools, communities and neighborhoods. About 50 teenagers read stories to students in the early primary school grades during 2014. To ensure greater family support for students, a radio advertising campaign was launched that encouraged reading at home. Other initiatives include the organization of fairs and cultural camps in which more than 1,500 children participated in 2014. These fairs and camps gave children the opportunity to demonstrate their artistic skills and to enjoy activities such as dance, theater and puppetry.

4.3 The reading campaign's effects on instruction

The reading campaign has provided participating foundations and NGOs with valuable data on how students read. In 2012, there were 6,000 students assessed and this number increased to 8,000 in 2013. In 2014, more than 9,000 students among the 11,435 children enrolled in Grade 1 were evaluated with the short version of the EGRA. Sharing these data with MINED authorities has allowed the Ministry and these organizations to implement various important measures to improve students' literacy skills.

One of the measures has been to improve MINED advisors' capacity to influence classrooms and provide support to teachers in the planning and implementation of their Language and Literature classes.

Learning difficulties identified in the first two years of the campaign, led Save the Children to organize a training course for its NGO counterparts. The course was offered over three years starting in 2012 and ending in 2014. The course served MINED's delegates and technical advisors in charge of the education system in territories where Save the Children's NGO counterparts operate—key personnel of these NGOs were also enrolled in the course. The course was taught in Managua over 30 days with two to three day sessions for a total of ten days (80 hours in the classroom per year). The training course covered the following topics:

- Process and methodology to teach and learn literacy skills

- Reading difficulties
- Phonological awareness and alphabetic code
- Fluency, vocabulary and reading comprehension
- Teaching strategies through play-like activities to expand vocabulary and improve comprehension
- Strategies to improve reading habits and develop a love for reading
- Reading process evaluation
- Management and use of the EGRA's data
- How to develop education plans with appropriate indicators using the EGRA's data to improve the teaching and learning of literacy skills in kindergarten and the first three grades of primary school.

In 2015, Save the Children is tendering the repetition of the training course—this time to be offered to primary teachers in some of the municipalities attended by its counterparts.

Three of CAPRI's officers took the course in Managua along with six MINED advisers working in the districts of Managua where the CAPRI provides its services. The IPADE participated with four representatives (two of its officials and two MINED advisors for Siuna).

This course and other trainings received by CAPRI officials have played a key role in expanding training opportunities of "educational counselors and MINED teachers" (Rivera, 2014). One of CAPRI's central activities is to support the team of MINED educational advisers in Managua, providing them with the opportunities to train in various literacy teaching topics. Following each training session, CAPRI and MINED departmental advisers and technicians reproduce the workshop among district advisors in Managua who in turn reproduce these training efforts with principals and teachers.

Rivera quotes (2014) a pedagogical adviser stating, "teachers are more encouraged to teach, thanks to the teaching strategies and methodologies provided to develop the content of school literacy curricula". Others teachers quoted stated that "the trainings sponsored by CAPRI have provided us with tools to strengthen the teaching of literacy ... as well as to seek students' involvement in learning" (Rivera, 2014).

The mother of a child in a class of 52 students who won the 2014 contest, placed a high value on the role played by the teacher stating "the teacher played a key role in children learning how to read. She was very organized, identified the children who were having difficulties, sorted them by groups and explained what they did not understand [...] they identified letters, divided syllables, did spelling, participated in the board, and motivated them to read."

The training also contributed to a positive atmosphere that encouraged reading in the classroom. For example, in the Grade 1 classes of the school *República de El Salvador* (multiple winner in the reading campaign's contest) in Managua, the three classrooms visited by Rivera (2014) with over 50 students each were in classes called *aulas letradas* (literate classrooms) and all three classes featured letters, numbers, images, colors, clippings from magazines, newspapers and posters on their walls. According to the teachers interviewed, this setting helped the students identify vowels and consonants as well as form syllables and words.

The CAPRI has also trained librarians to help foster a fun literacy learning culture, encourage the use of libraries and maximise the number of children's books available. The goal of these trainings was to encourage children to borrow books to take home as a means of recreation and avoid the practice of locking books away to prevent them from being damaged (Rivera, 2014).

World Vision reported in 2014 that it has been training teachers in the use of various methodologies to improve literacy learning. In 2014, World Vision provided teachers with several courses on the development of preliterate preschool skills. One of the courses called "playing, talking and drawing to learn how read" was attended by 28 preschool teachers. The focus of another course, "let's venture into the wonder of reading", was to encourage children's reading habits and was taught to 101 teachers of

Grades 1-3 and to 27 reading facilitators working within communities. A third course called "I read, comment, imagine and think" aimed to improve reading comprehension and enrolled 539 teachers (World Vision Report, 2014). These reading promotion methods are very simple, which enabled the course attendees to replicate them with other teachers, with children's relatives and older primary students.

World Vision's work in the classroom also involves the continuous evaluation of how these teaching strategies are being implemented to encourage reading. Teachers and principals often make adjustments based on these evaluations and incorporate elements into their teaching strategies to better suit the children depending on where they live.

In a report entitled "Achievements of literacy promotion in 2014," the IPADE reported that 30 Grade 1 teachers strengthened their lesson planning capacities and their ability to use strategic tools to promote reading through active learning methodologies. During these training courses, an action plan was designed and implemented to help improve children's literacy learning in 15 public schools in Siuna. These trainings also helped to improve the classroom environment through the creation of 46 reading corners (*rincones de lectura*) in the 15 schools. "The corners have used the donation of 70 children's books to promote the formation of spaces for pleasurable reading for children and adolescents" (IPADE, 2014). Another important initiative is the promotion of "good practice fairs" among teachers. These fairs, implemented by Save the Children, allow teachers to exchange positive experiences in literacy development and thus provide opportunities for horizontal learning.

4.4 Results of the reading campaign six years after its implementation

i) Progress according to collected data

The contest results[2] are the first evidence of the campaign's impact. In 2010, 93 schools and 119 classrooms participated in the campaign. In 2010, there were only two winning classrooms, representing 2% of the total number of classrooms enrolled. In 2013, 217 schools participated, 310 classrooms and 22 winning Grade 1 classrooms in which 80% of the children reached the stipulated goal, representing 7% of the total number of classrooms. In 2014, 310 schools and 425 classrooms participated, and 7% of the total number of classrooms enrolled achieved the stipulated goal. The 5% increase in the number of classrooms in which 80% of students achieved the fluency and comprehension goals is an important step in the right direction—although much remains to be done.

At the end of the 2009 school year (November 2009), the students evaluated in the RACN were reading an average of 17 WCPM (Castro, et al, CIASES/RTI, 2010). These results differ from the IPADES's findings in October 2014 in the same territory. Five years later, in 2014, two schools tended by the IPADE located in Siuna won the national reading contest. In one of these two schools, 89% of Grade 1 students had an oral reading fluency (ORF) of 46 WCPM while in the other school (a multigrade school located in a rural community), 88% of the Grade 1 students read 60 WCPM. When comparing the ORF results from the 2009 study with the data from these two schools, the differences seem to be large (i.e. 29 WCPM in the first study and 43 WCPM in the second). These differences indicate that although poverty and low maternal educational attainment are a strong setback in literacy, they can still be overcome through hard work and support from the school and the community.

According to data provided by the CAPRI, there have been vast improvements in reading in Managua. In 2011, 33% of the Grade 1 students who were administered the EGRA short version in the middle of the school year were reading more than 25 WCPM (*see* ***Table 1***). By 2013, that percentage had increased to 85%. There was some progress in reading comprehension as well, although not as much as in fluency.

2 The data collected using the contests have not been analysed for statistical significance to determine standard deviations.

TABLE **1**

Comparative results collected by the CAPRI in 2014 in Managua

	EGRA data from three initial applications in Districts VI and VII, Managua						
	2011		**2012**		**2013**		
Indicators	**n**	**%**	**n**	**%**	**n**	**%**	**TOTAL**
Schools	30		41		41		112
Teachers	30		68		74		172
Grade 1 students enrolled	1,297		1,195		1,662		4,154
Grade 1 students taking the test the day the EGRA was administered	877		1,185		1,162		3,224
Grade 1 students reading 25 WCPM or more in August (four months before the school year ends)	288	33%	520	43%	993	85%	1,802
Grade 1 students who answered the two required questions correctly	203	23%	192	16%	426	37%	821

Note: Statistical significance of results presented are unknown.
Source: CAPRI, 2014

TABLE **2**

National results of the 2013 reading campaign

School	**Grade 1 students' WCPM scores**	**Reading Comprehension**	
		Percentage of students who answered questions correctly	**Percentage of questions answered correctly**
Primavera Mulukukú RACN (IPADE/STCH)	70	82% of the students	63%
Mixta San Lorenzo Boaco (WV)	63	81% of the students	61%
Las Américas MGA (CAPRI/STCH)	51	82% of the students	53%

Note: Statistical significance is unknown.
Source: Castro and Laguna, 2013

TABLE **3**

National results of the 2013 reading campaign

Year the study was conducted	**Grade 1 WCPM scores**	**Grade 2 WCPM scores**	**Grade 3 WCPM**
2007 a	28	66	91
2009 b	17	60	90
2010 c	30	68	104
2011 d	13	49	74

Note: a) EGRA piloting in Oct 2007 with data from 42 schools; b) EGRA Caribe in Oct 2009 with data from 23 schools; c) EGRA in 2010 in 10 schools attended by an NGO; d) EGRA in Sept 2011 with data from 38 schools attended by two national NGOs. Statistical significance of results is unknown.
Source: Castro and Laguna, 2013

The 2013 data also shows progress in the results of Grade 1 students from winning classrooms when compared with the results obtained by Grade 1 students in other studies performed in the country (*see **Table 2** and **Table 3***).

As shown in Table 2, students from the winning schools in the contest attained a higher ORF than their peers. The differences appear large (note: statistical significance is unknown). Comparing the highest result obtained by the Grade 1 students in the 2010 study (30 WCPM) with the lowest result attained by the winning Grade 1 students in 2013 (51 WCPM), a difference of 21 WCPM is observed.

According to Save the Children´s report, the schools tended by its counterparts have achieved a 3% improvement in comprehension (note: statistical significance of the results presented are unknown) (Rivera G, 2014). This is not an optimal number but if we consider the reading difficulties evidenced in the assessments, there is hope for further improvement. However, the difference between the percentage of students reading fluently and the percentage reading with comprehension remains large and is a problem to address in the near future.

ii) Testimonies

The campaign's effort has also had a positive impact on the lives of many children. The testimony of Marcia Isolde Urbina, a Grade 1 student from a school tended by the CAPRI, illustrates this impact. Marcia Isolda, whose father has no formal job and sells cold water at the traffic lights in the streets of Managua, read 109 WCPM in the 2014 final contest and answered all questions about the story correctly. Marcia Isolde who is now 7 years old and in Grade 2 said in an interview:

"I like to read short stories and books. In my school we have a 'traveling children's library' that comes to my classroom. When the library arrives, I am delighted to discover new things in books. I also like to read stories to children in my neighborhood. I read to them Cinderella, Snow White, Little Red Riding Hood and other tales so they learn to enjoy reading too.

Last year I was the best reader in the National Reading Contest *Vamos a leer, Leer es divertido*. That made me feel good and motivated me to read more. My mom and my teachers supported me a lot".

5. CONCLUSIONS AND PROSPECTS

The joint effort by organizations already committed to improve Nicaragua´s education has paid off. These efforts are still insufficient in light of the many difficulties encountered by schools in communities facing poverty and considering the educational lag experienced by most of the children entering Grade 1.

However, it is important to note that a new path to improved literacy has been opened. The joint effort between civil society, the MINED and some private companies demonstrates that unity is strength and that the negative elements and practices of our educational culture can be changed through sustained efforts. This is possible when a noble cause such as improving reading comprehension in the early grades arouses sympathy and support.

There are challenges to improving literacy that have to be faced. The first and foremost is to improve reading comprehension—reading fluency has already improved in recent years thanks to this effort. Efforts to improve literacy include:

1. Continuing to promote initiatives that encourage reading for fun and accompany them with specific measures based on the knowledge that although teaching how to read is not an easy task, there are scientific components from research done worldwide that can provide important guidelines that teachers must master.

2. Strengthening key skills among teachers on how to teach reading and how to plan a Grade 1 language and literature class in an efficient yet entertaining way.

3. Expanding the campaign's funding to support three main aspects:
   - Assessing vocabulary (even if done in random samples of classrooms) using the IDEL as a reference for this EGRA extra section.
   - Promoting strategies that work with phonological awareness, vocabulary and listening skills in preschool.
   - Providing more in-depth training to MINED advisers and NGO officials to improve the quality of their advice and support.

4. Improving the culture of assessments by providing continuous follow-up of key indicators for literacy and systematically recording the data produced to create databases that can predict instruction and learning problems or enable these problems to be managed efficiently.

5. Improving statistical analyses to encourage funding for the *Vamos a Leer* campaign to continue promoting education quality through literacy improvement among at-risk children.

REFERENCES

Castro, V., Laguna, J. and Vijil, J. (2010). *Informe de Resultados ELI 2009 Caribe*. Managua, Nicaragua: RTI/CIASES. (In Spanish).

Castro, V., Laguna, J. and Mayorga, N. (2009). *Informe de Resultados: EGRA 2008*. Managua, Nicaragua: RTI/CIASES. (In Spanish).

Castro,V. and Laguna, J. (2013) *Formación de docentes en servicio en Nicaragua: la experiencia de los TEPCE y su influencia en la lectura.* Managua, Nicaragua: TEPCE (In Spanish).

CIASES (2011). *Informe de Línea de Base de Escuelas patrocinadas por Programa Alianzas de USAID*. Mimeo. Managua, Nicaragua: RTI/CIASES. (In Spanish).

EDUQUEMOS (2013) *Informe de Progreso Educativo Nicaragua.* Managua, Nicaragua: PREAL/IBIS. (In Spanish).

IEEPP (2012). *Niños, niñas y adolescentes que se desgranan de la Educación Primaria*. Managua, Nicaragua: IEEPP. (In Spanish).

IPADE (2014) *Logros en el trabajo de Promoción Lectura en RACN*. Managua, Nicaragua: IPADE. (In Spanish).

Laguna, J.R. (2009). *Análisis de la Situación Docente en Nicaragua 2008*. Documento Borrador. Managua, Nicaragua: MINED/CETT. (In Spanish).

Ministerio de Educación Cultura y Deportes (MINED) Database (2015).

O´Connell, S. (2012). *Teaching Reading in Nicaragua.* Managua, Nicaragua: Save the Children.

Rivera R. G., (2014) *Sistematización de experiencias de lectoescritura comprensiva en programas de Save the Children Internacional en Nicaragua*. Managua, Nicaragua: Save the Children. (In Spanish).

Rogers, T. (2012) "Impoverished teachers, poor schools". *The Nicaragua Dispatch*.

UNESCO Institute for Statistics Database (2015). http://www.uis.unesco.org/datacentre (Accessed February 12, 2016).

Verdisco A, Cueto S, Thompson J, Neuschmidt O, PRIDI, (2014) *Urgency and Possibility*: *First Initiative of Comparative Data on Child Development in Latin America.* Washington D.C.: Inter-American Development Bank.

World Vision (2014) *Informe 2014*. Versión Electrónica. Managua, Nicaragua: World Vision. (In Spanish).

The Yemen Early Grade Reading Approach: Striving for National Reform

JOY DU PLESSIS, FATHI EL-ASHRY AND KAREN TIETJEN
Creative Associates

ABBREVIATIONS

AQAP	Al-Qaeda in the Arabian Peninsula
AR	Action Research
CLSPM	Correct letter sounds per minute
CLP	Community Livelihoods Project
EGRA	Early Grade Reading Assessment
ERDC	Education Research Development Center
FOI	Fidelity of implementation
GCC	Gulf Cooperation Council
MFC	Mother Father Council
MOE	Ministry of Education
MSA	Modern standard Arabic
NER	Net enrolment rate
NGO	Non-government organization
RTI	Research Triangle Institute
SMS	Short message service
T'EGRA	Teachers' Early Grade Reading Assessment
T'EGWA	Teachers' Early Grade Writing Assessment
TIMSS	Trends in International Mathematics and Science Study
ToTs	Training of Trainers
TPOC	Teacher Performance Observation Checklist
WCPM	Words correct per minute
YEGRA	Yemen Early Grade Reading Approach

1. INTRODUCTION

In early 2012, Yemen's Ministry of Education asked the Community Livelihoods Project (CLP), a USAID funded development programme implemented by Creative Associates International, to support the development of a new approach to teaching reading in the primary grades. Low rankings on international education assessments combined with the Ministry's own assessments and monitoring pointed to a serious problem with the teaching of reading in Arabic. Optimism ran high post-Arab spring as Yemen had a transitional government that replaced the 30-year rule of Ali Abdullah Saleh and a national dialogue to determine a new constitution. Seizing the opportunity for positive change, the Ministry of Education prioritised early grade reading reform with a view to set a foundation for overturning years of underdevelopment in the education sector.

Yemen's recent history has been characterised by conflict. The Transitional National Government was formed after elections in February 2012 with the expectation that a national dialogue would result in a new constitution and general elections in 2014. In September 2014, however, Houthi rebels from Yemen's northeast launched a takeover of the government of Sana'a. The Yemeni president and other leaders eventually fled to neighboring Saudia Arabia. In March 2015, a Saudia Arabia-led coalition started aerial and ground strikes, destroying infrastructure and escalating the conflict.

Yemen has a rich history of learning. In the 7th Century, Yemen was home to some of the earliest writings on Islam. The Great Mosque of Sana'a is believed by Yemenis to have been designed by the Prophet Muhammed. With a population of 25 million people, scarce water resources, a reliance on imported goods and food, a limited industrial base and high unemployment, Yemen's economic and social indicators have stagnated or deteriorated over the years.

In recent years, Yemen has made progress in education particularly in improving primary school enrolment rates. The primary net enrolment rate (NER), or the percentage of primary school age children enrolled in primary school, increased from 57% in 1999 to 86% in 2012. During the same period, the gender parity index (ratio of girls to boys in school, with 1 denoting an equal numbers of boys and girls) improved from .58 to .84—although there is still a considerable way to go to attain gender parity, with the NER at 94% for boys and only 79% for girls (World Bank, 2012). In addition, nearly 600,000 primary school age children remain out of school in Yemen (UIS, 2015).

Yemen to date has consistently ranked near or at the bottom on a number of international measures of education. Yemen ranked the lowest of 36 participating countries in the Trends in International Mathematics and Science Study (TIMSS) assessments of Grade 4 and Grade 8 students in 2003 and 2007, possibly in part because of low literacy skills. Save the Children conducted a baseline study of early literacy skills in 2011 under its Literacy Boost programme and found that 52% of Grade 2 students and 28% of Grade 3 students could not read a single word in a passage (Gavin, 2011). This is consistent with the 2011 Early Grade Reading Assessment (EGRA) results under the USAID funded EdData II programme, which found that 27% and 42% of Grade 3 and Grade 2 readers, respectively could not read a single word of a grade-level text in Modern Standard Arabic (Collins and Messaoud-Galusi, 2012). Yemeni students in the early grades have not acquired the basic skills for reading. And with the conflict raging from late 2014 onwards and more than half of Yemen's children now not attending school, the situation has worsened. Clearly, for Yemen and the Ministry of Education (MOE) there has been no shortage of bad education news in the last few years.

In 2012, in response to the findings of the EGRA and Literacy Boost assessments, the MOE made the improvement of early grade reading a national priority and included it in the new education strategy. At the time, teaching and learning materials and the Grade 1-3 curriculum focused on reading to learn without including a learning to read component. This paradigm is more suited to learners who come to school with some knowledge of letters and sounds. Most learners in Yemen do not come to school with this knowledge and therefore struggle with basic texts. In addition, teachers did not have the skills needed to teach non-readers to read. The MOE, USAID and the Creative Associates-implemented CLP collaborated to develop the Yemen Early Grade Reading Approach (YEGRA), a learning-to-read programme in Modern Standard Arabic for Grades 1-3.

In the 2012-2015 post-Arab Spring period in Yemen, with military action to rout Al-Qaeda in the Arabian Peninsula (AQAP) from the Abyan Governorate and a transitional national government in place, a national dialogue to develop a new constitution was followed by the subsequent breakdown of the government with the Houthi take-over of Sana'a in September 2014—which resulted in the Saudi Arabian-led coalition air and ground strikes. Despite the tumult, a curriculum reform in early grade reading was designed, trialed, improved and scaled up nationwide to include all 6,000 primary schools. Oral reading assessments carried out during the various stages of the reform beginning in 2012 showed improved learner performance in reading Arabic, which helped to galvanise already nascent political will, foster widespread educator and community commitment, improve the capacity and the public perception of the MOE as well as teachers, and possibly provide a unique opportunity for unity and (at least temporary) stability in a country in disarray.

This article explores how oral reading assessments were used in Yemen to improve the reading ability in

Arabic of Grade 1-3 students. It also tells the story of how these oral reading assessments are linked to the start of an education reform in a country undergoing transformation amidst on-going conflict. The authors examine four modes of oral reading assessment used in the programme, each with a different purpose. Further, this article explores how the oral reading assessments influenced education policy; curriculum reform; system strengthening within the MOE; political will; teacher performance and professionalism; parental engagement in education; donor involvement; and stability in the nation. For a primer on the Arabic language, please refer to **Box 1** and for an overview of the lessons learned from conducting a reading assessment in Arabic, please refer to **Box 2**.

Box 1: Arabic language

- Arabic is often considered a 'diglossic' language, denoting the existence of a higher and a lower register used in semi-exclusive contexts (Ferguson, 1996). The higher register is sometimes referred to as *fusHa*, classical Arabic, standard Arabic or modern standard Arabic (MSA). The lower register will be referred to simply as vernacular Arabic, which is used for day-to-day communication and is seldom codified.
- Arabic is an alphabetic language with a primarily consonantal system. MSA has 28 consonant phonemes and 6 vowel phonemes, consisting of 3 short vowels and 3 long vowels.
- The language of instruction across the Arabic-speaking countries is MSA and children do not use this language to speak at home. Research shows that early oral exposure to MSA through storytelling was associated with gains in literary language development and reading comprehension (Abu-Rabia, 2000).
- The vowels are mostly depicted by diacritical marks presented below, above or inside the consonants. Arabic is a shallow orthography when the diacritical marks are present, and a deep orthography when short vowel diacritics are not presented. The unvoweled orthography is the norm in adult reading material while the voweled is used in literary texts and in beginning reading materials.
- Letter shape changes per position in the word (no manuscript letter shape is present in Arabic).

Box 2: Lessons learned from the reading assessment in Arabic

- Consistent with the nature of the Arabic language orthography, Ministries of Education in some Arab countries (e.g. Egypt, Yemen) test the ability of students in the early grades to manipulate short syllables—the consonant with the attached vowel (diacritics/harakat)—rather than the manipulation of phonemes, which is the target of the EGRA in other alphabetic languages (e.g. English).
- Some EGRA subtests (e.g. invented word decoding) are not welcomed by some Ministries in the Arab countries. The ministries have made the case that 'familiar word decoding' and 'oral reading fluency' subtests are enough to test the ability of children to decode words. Indeed, most of the words in these subtests are not 'familiar' but 'real' words that the vast majority of the children being tested have not yet seen before.
- The MOE in Yemen adopted a more strict approach in testing the children in the familiar word reading and oral reading fluency passage by insisting on the right pronunciation of all short vowels (diacritics/harakat) attached to the consonants of every word. Typically, while all words used in the tests are marked with these diacritics/harakat vowels, the common practice for scoring EGRA subtasks is to not hold the students in early grades accountable for the accurate pronunciation of every diacritic in the word. This is because the skills of attaching these diacritics to their consonants to provide the accurate pronunciation take time to develop. In the other Arabic-speaking countries where EGRA was conducted, for example, the word would be considered 'right,' if the student missed the last diacritic on the last letter of the word or any other diacritic that does not affect the meaning of the word.
- Unlike the emerging evidence that points to the importance of oral reading fluency in predicting reading comprehension in some orthographies (e.g. English), improved mechanical reading fluency alone is not particularly associated with predicting future reading comprehension in Arabic. This is attributed to the diglossic nature of the Arabic language and the subsequently reduced exposure of language users to the written Arabic code (Saiegh-Haddad, 2003).

The Yemen Early Grade Reading Approach

In early 2012, the MOE appointed 30 educators with qualifications in primary education, curriculum, Arabic and literacy to work with the CLP and its technical experts to design a phonics-based reading programme. Working intensively for six months the MOE/CLP team was able to produce: a scope and sequence for Grade 1; teachers' guides; student readers with decodable text linked to the scope and sequence; training manuals; teaching and learning materials; as well as coaching and supervision manuals.

The Yemen Early Grade Reading Approach (YEGRA) includes 117 systematically organized lessons focused on Grade 1, starting with one lesson per day and accelerating the pace to two or three lessons per day over the year as students gain literacy skills. Each lesson consists of 45 to 70 minutes of reading instruction and includes 25 minutes of direct, systematic Arabic phonics with three to four lessons per letter/syllable. Review and catch up lessons are included to allow teachers to focus on topics not mastered. The phonics progression is aligned with Gulf Cooperation Council (GCC) research on reading. The GCC research (and a similar study by Save the Children) found that a letter order sequence that progresses from the most frequently used letters to those less (or least) frequently used, enhances student learning of Arabic. Each lesson is tightly focused, scripted and includes letter-sounds, reading comprehension, and writing components. The student readers (one for each semester) contain independent reading stories that follow the scope and sequence of the lessons. Guided writing exercises are also included in the readers. The teacher's guides (one for each semester) include teacher continuous assessment tools—the Teachers' Early Grade Reading Assessment (T'EGRA) and the Teachers' Early Grade Writing Assessment (T'EGWA)—to provide feedback to teachers and students, school directors and supervisors to monitor student progress. The teacher's guides also contain the scope and sequence of reading skills introduced and reviewed, a calendar for implementation of the lessons and key messages to Mother Father Councils (MFCs) as well as parents to support the reading of their children—even for parents who are not literate.

The YEGRA is designed for administration in Grades 1-3 but is also focused on instruction in Arabic for the entire school year in Grade 1. In Grades 2 and 3, the YEGRA is used at the beginning of the year as a review and support for emerging readers and as the main Arabic teaching package for non-readers. Grade 2 and 3 teachers transition from using the YEGRA to existing MOE Arabic textbooks after using the YEGRA approach in the beginning of the school year. Grade 2 and 3 non-readers have also used the YEGRA for longer periods. The teachers were also trained in how to apply elements of the YEGRA to the teaching of the regular MOE textbooks, helping to transfer skills learned from the new Grade 1 curriculum to the teaching of Grades 2 and 3.

In addition to the teacher's guides and student readers, the following supporting materials and media have been developed and used in the YEGRA:

- big books for whole class teaching
- a pocket PowerPoint on the YEGRA method for teachers' reference
- poster boards for MFCs and parents' training
- training manuals for master trainers, Training of Trainers (ToTs) and supervisors, including how to structure meetings and give constructive feedback (coaching), the Teacher Performance Observation Checklist (TPOC), the action research cycle, interview instruments and sampling methods
- summary report templates for regular reporting on what the supervisors and master trainers learned from their visits
- an orientation guide for district and governorate education officials
- Facebook communities of practice for teachers, administrators, trainers and MOE officials.
- short message service (SMS) through mobile phones with tips and information for teachers and trainers.

Click here for an example of parent training materials

Instructional innovations in the YEGRA included asking learners to read independently, all at the same time as well as in low voices at their own rates. This ensures that all children have daily reading practice that far exceeds the amount of practice when only one child reads at a time. It also allows teachers the time to stop and listen to several children reading each day, noticing progress and common errors to use as an opportunity to reteach. Five T'EGWA and T'EGRA assessments are included throughout the 117 lessons as classroom based formative assessments so that teachers, headmasters, social workers and parents can monitor students' performance and modify teaching to improve student achievement.

2. ORAL READING ASSESSMENTS IN YEMEN AND THEIR USES

This section examines four types of oral reading assessments used in Yemen to design, evaluate, provide feedback to teachers and students, or to inform the revision and further development of the programme. For each type of assessment, we identify who develops and uses the assessment, what the assessment measures, when the assessments are/were carried out, the purposes of the assessment and the successes (and sometimes unintended outcomes) and challenges of each type of oral reading assessment.

2.1 EdData II EGRA and Save the Children Literacy Boost (2011)

Two recent studies in Yemen using oral reading assessments document difficulties with fundamental reading skills in the early grades of primary school. The Save the Children Literacy Boost study (Gavin, 2011), which covered students in Grades 1-3 in Aden, Lahj and Abyan governorates, found that, given one minute to read, 69% of Grade 1, 52% of Grade 2 and 29% of Grade 3 students could not read any of the words in a short, grade-appropriate story. Similarly, the EdData II EGRA study, conducted for students in Grades 2 and 3 in Amran, Lahj and Sana'a governorates, found that 43% of Grade 2 and 25% of Grade 3 students could not read a single word of text (Collins and Messaoud-Galusi, 2012). Among the students who could read one or more words, the EdData II study found that on average, Grade 2 students read 11 words per minute and Grade 3 students read 16 words per minute.

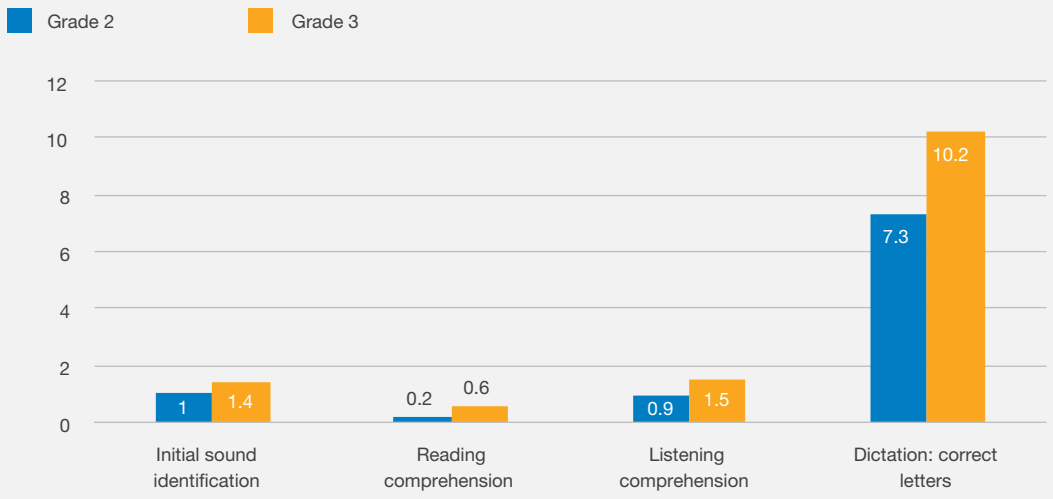
In the EdData II study, students' performance on four untimed EGRA subtasks (initial sound identification, reading comprehension, listening comprehension and dictation) showed similar results. Reading comprehension scores were very low, with 0.2 total correct answers out of 6 in Grade 2, and 0.6 correct in Grade 3. The listening comprehension scores were somewhat higher but still low, with an average of 0.9 correct answers out of 6 for Grade 2 and 1.5 correct for Grade 3. Finally, students had some success spelling some of the individual letters contained in the dictated words, with 7 to 10 letters spelled correctly (*see* ***Figure 1***). However, the average dictation scores in Grades 2 and 3 showed that students were unable to spell any words correctly.[1]

The Literacy Boost study (Gavin, 2011) found a strong association between students' letter knowledge and their word reading ability, suggesting that increased focus on alphabetic awareness may lead to improved reading outcomes, particularly for children with the lowest levels of reading skills.

The EdData II study identified a number of factors associated with student reading performance. For instance, student absenteeism was associated with reduced reading fluency while students with opportunities to read at school and who received corrective feedback from teachers correlated with improved reading performance. In addition, students who missed one day of school the week before the survey "identified fewer correct letter sounds, read fewer words in lists and in the passage, and were less accurate in spelling the three dictated words" (Collins and Messaoud-Galusi, 2012, p. 4).

1 Sample size: 735 students from 40 sampled schools (16 from Amran, 8 from Lahj and 16 from Sana'a). Data was collected in 2011. No significance scores were reported.

Figure 1. Total correct responses on the four untimed EGRA subtasks by Grade 2 and Grade 3 students, 2011



Source: Collins and Messaoud-Galusi, 2012

The two studies were used by MOE officials and the CLP to gauge the state of early reading instruction and student performance in Yemen. EdData II results were discussed with MOE officials, non-government organizations (NGOs) in education, UNICEF and other stakeholders through video conference[2] in May 2012. Initial reactions from MOE officials were that the study was not representative of the country and most officials seemed surprised that there even was a study. In essence, the low results were not just contested but more fundamentally, there had been a lack of engagement by MOE officials in the assessment. There was little ownership of the study.

The Literacy Boost study was disseminated widely and highlighted issues of diglossia (i.e. two different forms of the same language used by a community) in Arabic (i.e. MSA or *fusHa* vs vernacular Arabic or *Ammyya*) and the impacts this has on learning to read. In particular, the study highlighted the need for phonological awareness in MSA (or *fusHa*) as many of the sounds in MSA may be new to children in the early grades because they do not have a high exposure to MSA in the home. The study also presented analyses of the MOE Grade 1 textbook which, among other findings, had no complete sentences for students to read in the first year.

The CLP team working with MOE officials assigned to develop the new early grade reading approach used the study findings to design the materials, approaches and assessments. In particular, the design of parent engagement materials, the student book and continuous assessment corresponded to the three most significant findings from the 2011 EGRA. **Table 1** shows the relationship of the studies' findings with the YEGRA programme design.

In addition, both studies indicated that teacher skills were weak and there was a need for structured lesson design and individual student readers/writers. The teacher's guide includes 117 structured lessons based on the scope and sequence for the entire school year. Each student had one leveled text for each semester. The text included guided writing and was based on the scope and sequence.

2.2 YEGRA impact assessments

The CLP team and the MOE designed an impact assessment plan to assess the effectiveness of the YEGRA. The impact assessment included three

2 The security situation at that time did not allow for the EdData II staff to travel to Yemen.

TABLE **1**

Relationship between the findings of the 2011 EGRA and Literacy Boost studies and the 2012 YEGRA programme design	
2011 EGRA and Literacy Boost studies' findings	**2012 YEGRA programme design**
Children who have regular attendance do better in reading.	The national media campaign and parent training messages included this statement: "Getting your children prepared for school in the morning and on time everyday helps student learning".
Children who practice reading more, do better in reading.	All children have individual daily in-class reading.
Children who are read to at home or have books in the home perform better than those who don't.	Training for parents in making the home a print rich environment, reading to children at home and ensuring they have opportunities to read outside the home (i.e. at mosques, libraries, shops and other places with public texts).
Regular corrective feedback to students is correlated with increased early grade reading scores.	Five T'EGRA and T'EGWA assessments included in the teacher's guide. One assessment administered after approximately every 20 lessons.
Student's phonological awareness in MSA is weak likely leading to poor uptake of letter sound recognition.	Teacher guides include focus on phonemic awareness with daily interactive practice for students.

rounds of data collection and analysis over a three-year period (2012-2015), covering students' EGRA performance, teacher practices, school director and parents' support, among other topics. For each round of assessment, a sample of students attending CLP schools (the intervention group) and those attending comparable, non-CLP schools (the control group) were assessed. Baseline data were collected in October and November before the implementation of each phase, and endline data were collected in April and May—the traditional end of the school year in Yemen. However, because of unrest in Yemen, particularly during 2014-2015, the programme was interrupted and many schools were closed to protect students and staff. Ultimately, it was not possible to collect endline data in 2015. In addition, due to the increasing and unpredictable violence in Yemen, USAID suspended the CLP programme in May 2015 as a result of the continuous deterioration in security and the conflict. Therefore, exposure time for students to the YEGRA interventions was only about four to five months in each of the three years, interrupted by school closings and school safety issues due to instability.

The research approach was a quasi-experimental design to ensure high accuracy and validity of the data collected. The study used a mixed methods approach, combining quantitative and qualitative methods to obtain data on the range and prevalence of variables influencing student reading performance. Using stratified random sampling, treatment and comparison schools were selected in six focus governorates in phase 1 and ten governorates in phases 2 and 3. At each school, the data collection team tested and interviewed students, observed classroom lessons and interviewed teachers, school directors and parents about their background, behaviors, attitudes and practices. The population for the study included all Grade 1 and Grade 2 students who were attending school during the 2012-2013 (phase 1), 2013-2014 (phase 2) and 2014-2015 (phase 3) academic school years. In phase 1, the baseline survey targeted 90 government basic schools (45 of these schools were intervention schools in which the CLP was implementing YEGRA and another 45 were control schools). In phase 2, the baseline survey targeted 115 government basic schools, including 50 intervention and 15 panel study schools and 50 control schools. In phase 3, the baseline survey targeted 130 government basic schools, including 50 intervention and 30 panel study schools and 50 control schools. To obtain a random sample of Grade 1 and Grade 2 students, a three-stage sample was implemented by selecting: schools, classrooms and then students.

The impact assessment addressed the following questions:

1. Have student reading-related skills, attitudes and behaviors improved?
2. Have teacher reading-related skills, practices, attitudes and behaviors improved?
3. Are school directors (and/or education supervisors) implementing reading support activities?
4. Has school, community and parental support for reading/writing activities been instituted?

Using the EGRA, students were assessed orally on a variety of essential early grade reading tasks, including initial sound identification (phonemic awareness), syllable reading (letter sounds), familiar word fluency, oral reading fluency, reading comprehension, listening comprehension and writing (dictation). In addition, to obtain a fuller picture of how schools are performing and which school characteristics are associated with EGRA performance, data were collected on school, classroom and teacher characteristics. The EGRA instrument for assessing reading used by the Research Triangle Institute (RTI) and the Yemen Education Research Development Center (ERDC) in 2011 was revised and adapted for use in the study. Other instruments, such as teacher, school director and student interviews as well as classroom observations, were adapted for the Yemeni context from the USAID's Read to Succeed programme in Zambia (implemented by Creative Associates and the RTI). The EGRA instruments were trialed in schools in Sana'a, and modified.

Phase 1 findings indicated that the YEGRA strengthened Arabic language instruction in the intervention schools on all EGRA measures, as evidenced by student reading outcomes. After only four months of improved reading instruction, students in the intervention schools made marked improvement in their reading performance. Many of these students were able to identify initial sounds (phonemic awareness); read letters with diacritics (syllable reading); read words in isolation (familiar word reading) and in context (oral reading fluency); understand—to some extent—what they read or listened to (comprehension); and write letters, words and short sentences. Comparing across EGRA subtasks, the study found that the greatest improvements were in phonemic awareness, letter sound knowledge, familiar word reading and oral reading fluency. For example, on average, Grade 1 and 2 students in the intervention schools read 19.3 correct letter sounds per minute (CLSPM), indicating progression in performance from 6.5 CLSPM after four months of intervention. Similarly, on average, Grade 1 and 2 students in the intervention schools were able to read 9.3 words correct per minute (WCPM), indicating progression in performance from 3.5 WCPM after four months of intervention. However, the improvement in oral reading fluency did not enable students to read enough words to make progress on reading comprehension questions. Intervention schools increased their mean score from 1.1 to 1.7 only. Note that tests of significance have yet to be done on the Phase 1 data reported here.

In addition, based on phase 2 findings, there is convincing evidence that the positive effects of the YEGRA programme are cumulative across the two phases of the YEGRA for schools that participated in the school years 2012-2013 and 2013-2014. Overall, panel study students had substantially higher scores than other students. These findings might be expected at the Grade 2 level in the panel study schools as the Grade 2 students had been exposed to the YEGRA the previous year while enrolled in Grade 1. However, higher scores (especially in familiar word reading and in oral reading fluency) were found among the Grade 1 students at endline. Furthermore, after only one school year of exposure at baseline, essentially all students across all groups had the same EGRA scores, but by endline, Grade 2 students in panel study schools had significantly higher scores ($p < .05$) and higher percentages of change than did either the intervention or the control group. The baseline results of phase 3 also indicated that Grade 2 students exposed to the YEGRA in the previous school year had significantly higher mean baseline scores ($p < .05$) than did the comparison students on the EGRA subtests. These findings suggest that the cumulative effects of the YEGRA in those panel study schools, with those teachers and school administrations as well as parents and communities, made the difference in student

performance. It may also point to less learning loss over the summer vacation for students exposed to the programme over two years.

Indeed, intervention students had statistically significantly higher scores than did control students in several EGRA subtests ($p < .05$). Again, while the groups essentially were the same at baseline, at the endline, intervention school students had significantly higher scores ($p < .05$) than control group students in familiar word reading (Grade 1 only) as well as in initial sound identification, letter sound knowledge and listening comprehension (in all grades). Notably, students in the control group did not have significantly higher mean scores ($p < .05$) on any EGRA subtest when compared with either the intervention or the panel group students (*see* ***Table 2***).

In addition, the programme was able to make changes in teacher practices and attitudes, which likely had an effect on student performance. Teachers in intervention schools have been able to implement the new YEGRA model in their classrooms fairly consistently across most measures when compared with the control school teachers. For example, in phase 1, intervention teachers were able to guide students to pronounce sounds of letters (92%), associate words with letters (81%) and blend letter sounds to form syllables and words (75%). The ability to recognise letter sounds and differentiate among them is a key building block for success in reading, particularly for getting non-readers to begin to read. The skills to teach Arabic phonics, another key early grade reading activity, also improved among intervention teachers when compared to teachers in the control schools. In teacher interviews, 98% of intervention teachers reported teaching phonics on a daily basis at the endline compared with 61% of control teachers. In contrast, control teachers relied on the common practice of reading sentences without attention to the basic skills. This greater concentration on the basic early reading skills in the intervention schools when compared with the control schools is an important distinction resulting from the training and support provided to intervention teachers. In phase 2, at baseline, panel study teachers were more likely than the intervention or control group teachers to use different reading practices in the classroom. Intervention teachers and control group teachers had similar baselines across many measures but intervention teachers showed far greater gains by endline. For instance, at endline, intervention teachers were over twice as likely (89% versus 40%) to guide students to pronounce the sounds of letters and to blend sounds (82% versus 39%) and nearly twice as likely to guide students in reading books

TABLE **2**

Comparison and significance of differences in EGRA mean scores at endline

		Intervention versus control schools			Panel versus control schools		
EGRA measures	Grade level	Intervention	Control	P value	Panel	Control	p-value
Initial Sound Identification	1	4.24	3.32	.00*	4.13	3.32	.00*
	2	4.89	4.54	.04*	6.2	4.54	.00*
	Total	4.56	3.93	.00*	5.16	3.93	.00*
Letter Sound Knowledge	1	10.07	4.89	.00*	12.75	4.89	.00*
	2	12.89	10.16	.04*	27.09	10.16	.00*
	Total	11.47	7.52	.00*	19.88	7.52	.00*
Familiar Word Reading	1	4.09	3.27	.02*	5.01	3.27	.00*
	2	7.77	7.8	0.95	14.34	7.8	.00*
	Total	5.91	5.53	0.23	9.64	5.53	.00*
Listening Comprehension	1	3.17	2.11	.00*	3.12	2.11	.00*
	2	3.59	3.13	.00*	4.28	3.13	.00*
	Total	3.38	2.62	.00*	3.69	2.62	.00*

Source: Community Livelihoods Project, 2015

(68% versus 36%). Note that tests of significance have yet to be performed on these data.

Parental engagement to support students' reading also improved during the study period. More than 23,000 parents in the intervention schools were trained to support their children's reading at home and to prepare children to attend school regularly and on time. When students in phase 1 were asked how often someone at home reads to you or with you, the percentage of students who answered 'never' declined by 62% in the intervention schools by the endline, while the data showed a decline of 2% in the control group. In addition, there was a 58% increase in the number of students who reported someone reading to them or with them at home 'sometimes' in the intervention group compared to only a 13% increase in the control group. Global studies on children's early literacy confirm that these are key home factors that contribute to children's early reading achievement and as noted previously, these home factors were associated with higher reading performance in the Yemen EdData II EGRA and the 2011 Literacy Boost study.

2.3 Continuous assessment: Teacher's Early Grade Reading/Writing Assessments (T'EGRA and T'EGWA)

Effective and regular feedback to students on their performance is a key teaching skill that can improve student learning. Untrained and poorly skilled teachers, however, may not have the knowledge and skills to regularly assess student performance and provide them with the needed guidance on how to improve. Additionally, teachers may have large classes that make it difficult to assess every student regularly, keep accurate records, analyse results and give concomitant feedback to students. The 2011 EGRA confirmed that in Yemen, teachers rarely provide regular and effective feedback to students. One of the key elements of the training and support of teachers for the YEGRA, therefore, was to include continuous assessment for reading and writing. The teacher's guide includes five reading (T'EGRA) and writing (T'EGWA) assessments (*see* ***Box 3*** *and* ***Box 4***) aligned with the scope and sequence of the YEGRA programme. The five assessments were also consistent with the MOE policy of monthly assessments in all subjects (at all grades). Each assessment includes components of the seven steps of the programme and assesses student knowledge and skill based on the curriculum up to the point of the assessment.

Box 3: Teacher's Early Grade Reading Assessments

The five T'EGRA assessments included in the teacher's guide are based on the YEGRA scope and sequence for reading. They are quick assessments administered to all children in a class, individually. They assess students' letter-sound recognition, syllable and word reading (decoding), oral reading fluency and comprehension through reading a passage and answering questions.

Each daily lesson of the YEGRA includes seven steps structured in such a way as to provide practice to students in phonemic awareness, phonics, letter sound recognition, vocabulary, listening comprehension, independent reading, reading comprehension and writing. After approximately 20 lessons, there is a T'EGRA and T'EGWA built into the timetable. The assessment tools for each of the five T'EGRA and T'EGWA are included in the teacher's guide. The T'EGRA and T'EGWA assess four of the seven lesson components, namely: reading a passage, reading words, letter sound recognition and writing words spoken by the teacher. The writing task involved arranging the dictated words in sentences with an increasing number of words added to each T'EGWA (e.g. in the first T'EGWA, students are given three words to write and in the fifth T'EGWA, students are asked to write three full sentences based on some statements made by the teacher).

The assessments are administered individually to students during periods in the school timetable for Arabic and are allocated in the scope and sequence. They are considered low stakes assessments and teachers and principals are aware of their purpose which is to find out what students know, understand and can do. The assessments provide

Box 4: Teacher's Early Grade Writing Assessments

The five T'EGWA assessments included in the teacher's guide are based on the YEGRA scope and sequence for writing. They measure beginning writing skill progress throughout the school year. The early T'EGWAs include letter and syllable writing while the latter assessments ask students to write words and sentences dictated twice by the teacher. The assessments are administered to individual students by the teacher.

students with appropriate feedback to improve their performance. Teachers are also aware that the regular assessments point to areas where teachers need to place more emphasis on the teaching of reading and writing.

Evidence from the impact assessments, action research and other monitoring activities indicate that teachers do use the T'EGRA and T'EGWA assessments, provide feedback to students and adjust their teaching to ensure that students are mastering the requisite reading and writing skills before moving on. It is likely that the use of the T'EGRA and T'EGWA had an impact on student performance although we were not able to make a direct link between continuous assessment and learner performance.

In the context of (2012-2013) post-conflict Yemen, the ability of teachers to validly assess student progress relatively rapidly and understand the implications had powerful effects on teacher motivation and parental engagement in the reading programme. During the initial 10-day training, teachers were taught the assessment techniques in authentic contexts. That is, the teachers being trained at cluster center schools were able to try out the assessment techniques (along with the actual reading methods) with students in their own classrooms with constructive feedback from trainers to get the techniques right. The assessments during the training period provided teachers with opportunities to experience successful implementation of the new reading approach because the results of the oral reading assessments showed student gains in only a few days of teaching. Many of the teachers noted that their own students were reading more within a few days of the training activity than a whole semester with the old approach to reading. Gusky (2002) notes that the change in teacher attitudes and beliefs is essential for new skills and practices learned in an initial training to take hold, and that these changes are strongly linked to whether or not teachers experience improved student learning outcomes. In the training model where continuous assessment is used in authentic contexts, teachers return to their schools after the training with belief in the method and the commitment to use it. Without the assessment practice in authentic training contexts, these attitudes and beliefs might come much later in programme implementation, resulting in a much slower uptake of the approach by teachers.

Despite the apparent success of this model of structured continuous assessment, the administration of the T'EGRA and T'EGWA has posed challenges for teachers. In Yemen, student-teacher ratios in the lower grades can be as high as 100 to 1. Ideally, each assessment carried out with an individual student takes one teacher at least 5 minutes, including the recording of the student marks by the teacher. In most cases, this has resulted in more time spent on the assessment than what was intended as teachers try to ensure that all students get assessed.

In reality, not all students are assessed five times in a school year. Short cuts, such as the timed student reading of the Arabic passage, are used by some teachers as a proxy for reading achievement. Parents and other teachers are sometimes called to help administer the periodic T'EGRA and T'EGWA. These methods, while better than no assessment, do not provide adequate information on student weakness in other areas, such as letter sound recognition, listening comprehension, etc.

Schools that do manage to assess all children in a class with all five T'EGRAs/T'EGWAs are those that suspend all other subject teaching for several days in favor of the assessments. In these instances, more teachers are available to help administer the

tests. In some cases, parents are brought in to support the assessments. This method, however, forfeits valuable teaching time to the assessments.

The importance of low-stakes, regular assessments of student progress and teacher feedback to students on how they can improve their performance are key factors in improving student reading skills. Yet, there are few good models on how teachers can assess all learners (especially in large classes) in a range of reading and writing skills, record the individual results, analyse them for improved instruction and prepare and teach to improve student learning. Technology perhaps holds the most promise but in resource-poor environments, most technologies are not yet affordable for teachers.

2.4 Action research/progress monitoring

In order to gain ongoing feedback on programme implementation that could inform design and ensure the fidelity of implementation, the YEGRA included an Action Research (AR) component. In the first year of the programme, the governorate-based master trainers were trained in AR in addition to their role as trainers of trainers. This training included the AR cycle; how and when to administer the T'EGRA and T'EGWA to a random sample of Grade 2 students; recording and analysing results; how to provide targeted feedback to teachers at the schools based on the results; and reporting on progress and challenges. A team of master trainers at each governorate were assigned two schools in their governorate to visit twice a month during the school year to observe teachers in class; check learner performance of a sample of students using the T'EGRA and T'EGWA; interview administrators, teachers and students; develop reports on their findings; and discuss these with school staff to solve problems and make improvements. Monthly AR reports from each governorate were sent to the programme headquarters.

Using timers, master trainers would carry out spot checks by randomly sampling 10 boys and 10 girls from Grade 2 and administering a T'EGRA and T'EGWA. There are five T'EGRAs and T'EGWAs in the teacher's guide (one after about every 20 lessons) and the master trainers use one of these with the students depending where the students and teachers are in the sequence of the lessons.

The oral reading assessments of the AR component of the YEGRA were important tools for ensuring the fidelity of implementation (FOI) in both instruction and the use of T'EGRA/T'EGWA by teachers, and as a progress monitoring tool for school directors, district officers and the MOE/CLP programme team. In addition to the post assessment feedback to teachers and school directors by the master trainers, the AR provided important data to the MOE/CLP team on how to improve FOI among all teachers—not just the AR school teachers. An SMS (text message) system was set up to provide teachers with specific tips and suggestions based on the results of the T'EGRA/T'EGWAs at the AR schools. In Yemen, all teachers have mobile phones as this is one of the key identifiers for picking up one's pay. After one of the first assessments, many master trainers reported that teachers were not asking the students the meaning of words in the lesson as prescribed in the lessons. As a result of this information, the programme was able to send out a text message to all teachers reminding them:

> "… to ensure the students understand the meaning of new words by asking them what words mean and providing the meaning when they don't know it, especially in step 6".

One unforeseen challenge of providing the five assessments in the teacher's guide was the perception from parents and teachers that teachers could teach to the test and thus inflate student scores. Since the five assessments were spaced out in the 117 lessons, they were directly tied to the content covered during teaching. It is likely that this perception could come from the predominant view that the purpose of assessments are to determine pass and fail rather than to be used formatively for constructive feedback to students. It was clear to the team that teachers and parents needed more

understanding of how oral reading assessments on a periodic basis could be used to improve student performance by helping teachers focus on areas that need strengthening.

3. REPORTING AND USE OF ORAL READING ASSESSMENT DATA

Teachers, students, parents and education officials were not unaware of the problems in teaching reading in Arabic prior to the Literacy Boost and EdData II assessments. Indeed, the Yemeni primary curriculum had not been revised since 1995. Students were dropping out of school in the primary grades with a drop in enrolment at Grade 4 due in part to the lack of attainment of foundation skills in reading and writing. Parents withheld children from school because of perceptions of a lack of quality teaching at primary schools. The oral reading assessments have helped to identify specific problems in the teaching of reading in Arabic in the primary grades in Yemen. This has helped to shape the types of reform and guide the new reading programme design.

Data from each of the four oral reading assessments in Yemen each had their own purposes as described above. This section describes the uses of the oral reading assessment data and the uses of the data for programme design, revisions as well as the larger impacts the assessment has had on policy, system reform and possibly, national stability.

3.1 Parental engagement

Although in the YEGRA, parents were not largely involved in conducting or supporting oral reading assessments, they were involved in understanding how the YEGRA was taught in schools and were provided with training on specific support they could provide to their children to foster their reading. The illiteracy rate, particularly for women is near 60% in Yemen. Yet, with the support of teachers, school directors and MFCs, all parents were able to understand their role in helping their children learn. In some cases, parents learned to read by following along with the students in their readers.

3.2 Policy and system reform

As the Early Grade Reading programme was seen as part of the solution to the problem of poor performance on national and international assessments, the MOE took the lead in the programme from the start. The MOE appointed a core team of education officials representing a wide range of geographic regions to work on materials development and teacher development. Teacher Education, Supervision, Curriculum, Community Mobilization and Gender departments all deployed staff to work on the design, implementation and monitoring of the programme.

Periodic update meetings led by the MOE were held from 2012-2014 to analyse progress and address challenges. These meetings reviewed programme progress as represented by outcomes in terms of teachers trained, students with books, parental training activities held and other programme results. They also analysed the outcomes of oral reading assessments of control and intervention schools in the research strand of the programme as well as the progress monitoring by supervisors. The meetings provided recommendations for improvements of the programme as well as for policy and systemic reform, and decisions were made on the tempo and scope of scaling up. With support from the CLP, the World Bank and the Gesellschaft für Internationale Zusammenarbeit (GIZ), the programme was scaled up nationwide to all Grade 1 schools in the country at the start of the school year in September 2014.

3.3 Programme design and revision

As was discussed previously, the 2011 EdData and Literacy Boost oral reading assessments provide important directions for initial programme design (i.e. providing scaffolding for teachers in the form of scripted lessons, ensuring there was adequate time for reading in the classroom and that each child had their own reader). The assessments also pointed to

difficulties teachers were having with the materials or with the instructional methods. For example, teachers were coming to class without being fully prepared. One aspect of the lesson that was weak, based on the findings of AR and the endline EGRA, was that teachers had not sufficiently practiced the read alouds (stories with questions for listening comprehension and the application of phonemic awareness and phonics). The lack of preparation resulted in mispronunciations, problems with speed and a lack of prosody when reading. Another finding was that some teachers were themselves not fluent in MSA. As a result, training of teachers was modified to place more emphasis on preparation for lessons, practice and feedback from peers on read alouds and a focus on redressing some of the common mistakes teachers make in MSA.

Perhaps more importantly, was the way in which oral reading assessments for monitoring by supervisors, school directors and others helped programme designers and the MOE to get 'quick wins' that were used to generate enthusiasm for the programme. The assessments even after one month of the intervention showed progress in learning to read. Since a phonics-based (or perhaps more accurately, a syllable-based) approach to reading in Arabic had not been officially taught or learned in public schools in Yemen for many years, the improvements were rather dramatic. Parents were turning up to schools to praise teachers because they never expected that their child in Grade 1 could read. While the early results were impressive, there was still a long way to go to get oral reading fluency to a level where it would result in improved comprehension. Nevertheless, sharing of the AR and monitoring results with district, provincial and national MOE representatives in addition to parents and community members helped to garner support from broad stakeholders for the programme.

3.4 Teacher professional development

In Yemen, 51% of the Grade 1-3 teachers are female. This is considerably higher than the national average of 29% female teachers. The concentration of women teachers in the lower grades is not uncommon in many countries. In Yemen, however, these teachers possess the least amount of professional training among all teachers in the system. Many of them have not completed high school. Most have only attended in-service teacher training and if so, sporadically. The primary teachers have the longest hours with no breaks. They have the largest classes and the lowest pay. In short, primary teachers are not highly regarded or valued by the public or the system.

The YEGRA provided these teachers with a platform for successful teaching and helped them thrive. The scripted lessons provide unskilled teachers with the steps and tools for teaching every day of the year. The interactive nature of the (mostly) whole class teaching methods helped teachers manage their large classes effectively. Communities of practice and constructive coaching helped them hone their skills in a collegial and respectful way. The T'EGRA and T'EGWA showed them that they were helping children to learn. If they weren't, they were able to determine how to help those who struggled. Teachers in higher status private schools asked the YEGRA teachers for help and to share their YEGRA teacher's guides and other materials. Student-teacher relations improved and parents reported that their children were reading or were read to more at home and were reluctant to miss school. In essence, early grade teachers became better teachers and the oral reading assessments provided empirical evidence that this was so.

4. CONCLUSION

The Yemen MOE with the support of international donors was able to bring an early grade reading reform to a national scale within two years of a trial programme of early grade reading. The various oral reading assessments played an important role in providing evidence of success in a transparent and objective way. By getting 'quick wins' from the programme through evidence of reading improvement provided by oral reading assessment, the programme was able to gain momentum for expansion during its first phase. As news spread of children reading (or at least decoding words) in

Grade 1 classrooms, teachers in non-intervention schools and private schools borrowed the YEGRA teacher's guides from friends, photocopied them and starting teaching the YEGRA method in their schools. Education officials at all levels understood that the YEGRA was having a positive effect on student learning outcomes.

The EGRA including the oral reading assessment provide compelling evidence that focusing on the quality of reading instruction in early grades can have a positive impact on student's reading performance in a relatively limited amount of time of approximately four months. However, it is clear that designing, implementing and revising an educational reform in early grade reading in a country going through transition while experiencing nationwide conflict is no small feat. Overcoming roadblocks and barriers requires flexibility, creative problem solving and compromise. Establishing new ways of teaching and learning, supervision and support requires new ways of thinking for many education professionals. What the evidence of oral reading assessments provided was opportunities for people at all levels to back the programme on educational grounds. In addition, because the programme focused on the early grades, there was a sense that education at this level is apolitical. Indeed, representatives in the National Dialogue called for a moratorium on all curriculum revision until a new constitution was in place—the only exception was the continuation of the development of the early grade reading curriculum. The results from the oral reading assessments proved to broad stakeholders that the YEGRA was working, and that it offered hope for the future.

The YEGRA programme was on a path of initiating some major paradigm shifts in post-Arab Spring education in Yemen, resulting in improved reading achievement in the early grades during the 2012-2014 period. As the conflict has raged in the 2014-2015 school year, education officials, teachers, parents and community members have made valiant and resourceful efforts to continue to find ways to support children in learning. With peace, safe learning environments and support to schools and communities, the resumption of improvements in the early grade reading and writing programme holds the potential to heal, inspire hope and ultimately, break the cycle of conflict as children get the foundational skills they need to meet their full potential.

REFERENCES

Abu-Rabia, S. (2000). "Effects of exposure to literary Arabic on reading comprehension in a diglossic situation". *Reading and Writing: An Interdisciplinary Journal,* Vol. 13, pp. 147-157.

Collins, P. and Messaoud-Galusi, S. (2012). *Student performance on the early grade reading assessment (EGRA) in Yemen*. EdData II report prepared by RTI International for USAID. Research Triangle Park, NC: RTI International.

Community Livelihoods Project (2015). *Improved Reading Performance in Yemeni Schools.* Impact Assessment Report of the Yemen Early Grade Reading Approach (YEGRA), Phase 2 (2013-2014). Sana'a Yemen: USAID/Community Livelihoods Project.

Ferguson, C.A. (1996). "Diglossia". *Sociolinguistic Perspectives: Papers on Language in Society 1959-1994.* Huebner, T. (ed.). Oxford, UK: Oxford University Press.

Gavin, S. (2011). *Literacy boost: Yemen baseline report*. Sana'a, Yemen: Republic of Yemen, Ministry of Education, Education Section, Inclusive Education Department.

Gusky, T. (2002). "Professional Development and Teacher Change". *Teachers and Teaching: Theory and Practice,* Vol. 8, No. 3/4.

Ibrahim, R. and Eviatar, Z. (2012). "The contribution of the two hemispheres to lexical decision in different languages". *Behavioural and Brain Functions,* Vol. 8, No. 3.

Saiegh-Haddad, E. (2003). "Bilingual oral reading fluency and reading comprehension: The case of Arabic/Hebrew (LI)—English (L2) readers". *Reading and Writing: An Interdisciplinary Journal*, Vol. 16, pp. 717-736.

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

UNESCO Institute for Statistics (UIS) database (2015). http://www.uis.unesco.org/datacentre/pages/default.aspx (Accessed February 2015).

World Bank (2012). World Bank Databank (using UNESCO Institute for Statistics data). http://data.worldbank.org

Assessing Reading in the Early Grades in Guatemala

MARÍA JOSÉ DEL VALLE CATALÁN
Guatemala Ministry of Education

ABBREVIATIONS

Digeduca	*Dirección General de Evaluación e Investigación Educativa*
EGRA	Early Grade Reading Assessment
ELGI	Reading Assessment for Initial Grades (Spanish acronym)
ELI	*Evaluación de Lectura Inicial* (Initial Reading Assessment)
INE	National Institute of Statistics (Spanish acronym)
LEE	Evaluation of Early Reading and Writing (Spanish acronym)
PAMI	Evaluation of Early Mathematics Skills (Spanish acronym)
RTI	Research Triangle Institute

1. COUNTRY CONTEXT

Guatemala represents a blend of various cultures, languages and needs. Occupying an area of 108,889 km^2, Guatemala and neighbouring Belize mark the southernmost boundary of the landmass that separates North and South America, and the southern tip of Central America. Guatemala is divided into 22 departments featuring numerous mountains, volcanoes, lakes and rivers that flow into the Pacific Ocean and the Caribbean Sea, which all contribute to a wide variety of flora and fauna and its rich biodiversity (INE, 2015a).

Guatemala's mild weather is conducive to the year-round production of fruits and vegetables, making agriculture one of the country's main sources of labour. After several years of armed conflict, the signing of Peace Agreements in 1996 paved the way for reforms in the areas of education, economy and agriculture, among others.

Based on the last census in 2002, the National Institute of Statistics has provided some projections for 2014. Guatemala's population is estimated at 15,607,640 inhabitants, most of the population is relatively young with 37.8% falling in the 0- to 14-year-old age group and 37.0% comprising those in the 15- to 34-year-old age bracket (INE, 2013). In Guatemala, 93.3% of the population aged 15-24 years is literate. The last population census revealed that 59.3% of the population is living in conditions of poverty and 23.4% under conditions of extreme poverty. (INE, 2015b) The population is divided almost evenly between those inhabiting rural (76.1%) and urban areas (42.1%) (INE, 2015a).

The country's inhabitants are divided into four populations: indigenous peoples*, Garifunas, Xincas* and *Ladinos*. The largest group, the *Ladinos* (59%), live in urban centres and are mainly in the capital city. The other 41% are indigenous, with 39.3% representing Mayans and 1.7% accounting for *Xinkas* and *Garifunas* (VV, 2009).

While Spanish is the country's official language, 24 other languages are also spoken—22 of which are Mayan followed by Garifuna and Xinca languages spoken by various peoples. Nowadays, while some populations speak two or three languages (Spanish

among them), other groups have only acquired an oral mastery of their mother tongues.

2. EDUCATION IN GUATEMALA

In Guatemala, education represents a challenge as numerous needs have yet to be met. Coverage differs widely among the different educational levels (*see* ***Table 1***). In primary education (children aged 7-12 years), the coverage is 82% while in pre-primary education (children aged 5 and 6 years) where coverage is mandatory, the coverage is only 47% and the required infrastructure is unavailable. Compared to pre-primary coverage, a smaller number of students (between the ages of 13 and 15 years) are enrolled in lower secondary (45%) and even fewer students (aged 16 to 18 years) are enrolled in upper secondary education (24%). Students wishing to pursue this last level must choose between the *Bachillerato* (the equivalent of a high school diploma), secretarial studies, certified accountancy or pedagogy for pre-primary education programmes. Two years ago, the teaching profession was given university status but prior to 2013, a teacher could begin teaching at age 19 years, following the successful completion of a three-year theoretical/practical programme. Currently, individuals who wish to become teachers are first required to earn the equivalent of a high school diploma with a major in education and subsequently, complete university-level courses.

Great efforts have been made to reduce the incidence of school failure. In 2014, school dropouts reached 3% in pre-primary, 4% in primary, 4% in lower secondary and almost 2% in upper secondary. Repetition in 2014 continues to be a recurring issue affecting primary education schools in particular, where the rate has risen to 5% in primary with a 19% repetition rate in Grade 1 and 12% in Grade 2; 2% in lower secondary and almost 1% in upper secondary (Diplan, 2014). One of the major causes of repetition in this grade is student absenteeism, the result of internal migration (Garnier, 2008). Students' poor reading skills and the lack of teaching tools compound the problem. Guatemala's classrooms are also characterised by a substantial number of over-aged students. In fact, it is not uncommon to find 13-, 14- and even 15-year-olds attending Grade 1 of primary education. This is considered one of the main reasons for dropping out (Garnier, 2008).

TABLE **1**

Number of students and teachers at the pre-primary, primary and secondary levels across the country

Level		Age	Percent coverage in 2014	Number of registered students in 2014	Number of teachers in 2014
Pre-primary		5–6	47%	549,350	49,872
Primary	Grade 1	7	59%	489,157	
	Grade 2	8	44%	426,771	
	Grade 3	9	39%	411,145	
	Grade 4	10	36%	391,464	122,738
	Grade 5	11	34%	363,273	
	Grade 6	12	33%	335,618	
	Total	—	82%	2,417,428	
Lower secondary	Grade 1	13	28%	298,034	
	Grade 2	14	24%	249,149	71,663
	Grade 3	15	20%	221,980	
	Total	—	45%	769,163	
Upper secondary		16–17	24%	396,461	52,706
Total				4,132,402	296,979

Source: Diplan, Mineduc, 2015

3. ASSESSING READING SKILLS AT THE NATIONAL LEVEL

The Ministry of Education—the country's governing body in the area of education—is divided into several bureaus that work in close coordination. Towards the end of 2007, the evaluation unit responsible for measuring students' learning skills was formally created. Tests are administered nationwide and cover every education level in the areas of mathematics and reading. Primary education tests measure expected skills in Grades 3 and 6 as set forth in the National Curriculum. At the secondary level, tests are more focused on the broader area of life skills. All tests are comparable among years. Each instrument meets the conditions required for the production of a valid, standardised and reliable test, including the elaboration of specification tables, training of the drafting team, item development, revision and edition, verification by experts, and qualitative and quantitative pilot testing, among others. In generating the results, analyses are performed in Classical Theory and Item Response Theory. Different types of analyses are performed using different software: analysis of discrimination, difficulty and distractors. In addition, the conditions of equating, anchoring and adjustment of the model is verified.

These measurements collect data on the attainment level of each assessed individual (*see* ***Table 2***). Attainment levels indicate the number of students who have developed reading comprehension and problem solving skills equivalent to their school grade requirements. Results allow stakeholders to determine student performance at the national level and make policy decisions geared to improving learning opportunities (Digeduca, 2015).

Tests are comparable among ages but not among grades. However, there is increasing awareness that the number of students who acquire an acceptable level of attainment is steadily declining. The data obtained from these assessments along with other related factors are being analysed in an attempt to understand the full impact of these findings. The same factors seem to surface year after year to different degrees of magnitude (Cruz and Santos, 2013; Cruz and Santos Solares, 2015; Alcántara et al., 2015; Quim and Bolaños, 2015; Bolaños Gramajo and Santos Solares, 2015). For instance, urban students appear to exhibit the best results in mathematics and reading while students who self-identify as *Ladino* (students whose mother tongue is Spanish) seem to obtain the highest scores. Consistently, grade repetition and child labour have a negative impact on learning, which is to be expected.

4. ASSESSING READING SKILLS IN THE EARLY GRADES

In Guatemala, the assessment of reading skills in the early grades was formally launched in 2004. Previous assessments had focused essentially on student promotion to higher grades and involved different languages (Fortín, 2013). The first tests were led by private organizations seeking to determine the entry level of reading proficiency among Grade 1 students for research purposes (Rubio et al., 2005). To this end, the LEE (Evaluation of Early Reading and Writing) test, an adaptation of David K. Dickinson's (Educational Development Centre) and Carolyn Chaney's (University of San Francisco) Emergent Literacy Profile were used (del Valle M.J., 2015). In 2007, the Ginsbur and Baroody's Test of Early Mathematics Ability was also administered. The test, based on the theory

TABLE **2**

Reading attainment levels

Grade	Grade 3 (primary) 2014	Grade 6 (primary) 2014	Lower secondary 2013	Upper secondary 2015
Attainment	50%	40%	15%	26%

Source: Diplan, Mineduc, 2015

of mathematical reasoning, evaluates formal and informal knowledge of mathematics. It is frequently used with groups of children that either cannot read or are in the learning stage (del Valle and Castellanos, 2015).

In 2005, initial efforts began to assess reading and mathematics at the national level. Promoted from within the Ministry of Education and with support of the USAID Education Standards and Research Program, this assessment was intended to evaluate pupils aged 7, 9 and 12 years in the areas of reading and mathematics. However, it is not until 2006 that a formal assessment of Grade 1 students was administered through the Digeduca.

Among other responsibilities, Digeduca provides information on the quality of the education imparted by the National Education System by measuring students' academic performance (Digeduca, 2015b). With this aim in mind, a new test was created to measure the performance of Grade 1 students in the areas of mathematics and reading (language and communication component). The results derived from this assessment, which will be administered upon completion of this level, will be used to diagnose the current state of education. The assessment dealt specifically with reading comprehension as described in the Ministry of Education's national curriculum requirements. Matrix-based tests were used which included all the skills students should develop distributed between different instruments (Digeduca, 2015b).

In 2010, in an effort to identify the reading skills students first entering the educational system possessed, a study was conducted on three Grade 1 primary education student cohorts selected from a nationwide sample of schools. The PAMI (Evaluation of Early Mathematics Skills) and LEE tests with some adaptations to the Guatemalan context were used as a diagnostic tool to gauge school effectiveness. Both tests were simultaneously administered to Grade 1 students orally and individually during the first few months of classes.

These tests were administered for three consecutive years—in 2010 and 2011 in the last months of

school and at the beginning and at the end of school cycle in 2012. Following completion of the school year, these students were re-tested in order to measure academic gains.

The LEE test measures the following concepts: letter identification or familiarity with the alphabetic principle; emergent writing; initial reading or willingness to read; and print concepts. The following results were reported (del Valle M.J., 2015);

1. Letter identification:
- At the beginning of Grade 1, approximately 42% of the students were capable of identifying, unaided, the name or sound of 0-6 letters of the alphabet.

2. Emergent writing:
- Overall, 43% knew the basic rules of reading and writing and are capable of writing their names or a familiar word.
- Three out of 10 pupils held their pencils correctly.

3. Initial reading:
- Overall, 28% of pupils have not yet developed word reading skills or at least the ability to read his/her name.

4. Print concepts:
- Only 60% had had any contact with books, recognized their orientation and could identify their different parts.

A narrow positive relationship was found between the four tasks, the math test (PAMI), and some related factors, such as having completed pre-primary education and speaking Spanish as a first language.

It should be noted that during 2010, 2011 and 2012, pre-primary education coverage steadily declined (from 55% to 48% and to 45%, respectively) and, simultaneously, a large number of official sector primary school students enrolled in Grade 1 for the very first time. Now, primary teachers had to focus on school readiness before teaching letter recognition.

In order to monitor LEE and the findings derived from this test, the Reading Assessment for Initial Grades (ELGI) test was created. Some of the topics assessed in the ELGI are similar to those covered in previous tests. However, this particular version focuses more deeply on learning to read. The ELGI, which was based on the Early Grade Reading Assessment (EGRA), was first administered towards the end of 2010. The ELGI pursued two goals: measuring students reading skills (short-term goal) and providing teachers with a formative assessment tool (long-term goal). The latter goal requires teachers to be aware of the importance of this tool and be familiar with the different versions of the test.

The ELGI, an adaptation of the EGRA, attempts to retain the structure and guidelines of the original test. An adaptation was preferred over a translation since the latter was more likely to produce an error of interpretation of a construct, method or item. The peculiarities of the Spanish language were also taken into account (Rubio and Rosales, 2010). The tool was adapted with the technical support of USAID/Classroom reform and the Research Triangle Institute (RTI), an entity that has provided technical support to efforts aimed at assessing reading in several countries. The ELGI incorporates the same areas included in the EGRA, adapted to the language and the Guatemalan context—namely, knowledge of the name/sound of letters; phonemic awareness; reading of words and sounds; reading of familiar words and pseudo-words; reading of stories; oral/silent reading comprehension; and writing. These sections were classified under six categories: oral language, alphabetic principle, phonemic awareness, reading fluency, reading comprehension and writing. A section containing instructions was also included to serve as an indicator of language mastery.

In 2011, the ELGI was administered nationwide to official sector Grade 2 students (8-year-olds) using a representative sample disaggregated by department. On average, six students were randomly chosen from a class for a total of 1,524 students and 262 schools. As with the

PAMI and LEE tests, specialised personnel were responsible for administering the test across the country. These persons underwent theoretical and practical training for several days, including visits to rural and urban schools, in order to provide feedback on the process prior to the test's final administration.

This initial test produced a report containing several interesting findings, including the following results (del Valle and Cotto, 2015):

1. Oral language:
- When it came to following directions, most students whose native tongue was Spanish performed with excellence, whereas students whose native tongue was a Mayan language did not perform nearly as well.
- After listening to a text, students answered 65% of the questions correctly.

2. Alphabetic principle:
- The names of letters were generally mastered but sound is more elusive. Three quarters of the population cannot identify 'll,' 'q' and 'ch' by name, while 'a,' 'e,' 'I,' 'o,', 'u,', 's' and 'm' posed no problems.

3. Phonemic awareness:
- Most Grade 2 students identified the initial sound of 80% of the test words. When words started with a vowel, the percentage was even higher.

4. Reading fluency:
- Students read, on average, 66 letters per minute when the task consisted of identifying the name of the letter and 33 letters per minute when identifying sounds. Students also read 40 meaningful words per minute and 32 pseudo-words per minute on average. Lastly, students read 71 words per minute on average within the context of a text. This last indicator provides the greatest amount of information.

5. Reading comprehension:
- Students answered correctly 72% of the questions after reading a text.

© María José del Valle Catalán, Guatemala

TABLE **3**

Statistical significance of associated variables

Variable	Coefficient alpha	Significance
Grade repetition (repeat)	-4.97	0.00*
Type of classroom (multi-grade)	4.19	0.04*
Student´s native tongue (Spanish)	3.43	0.01*
Area (urban)	3.24	0.01*
Teacher´s gender (female)	-3.15	0.05*
Books in home (existence)	1.82	0.03*
Student´s gender (male)	0.68	0.34
Pre-primary (attendance)	0.60	0.45

Note: * Statistically significant.
Source: del Valle and Cotto, 2015

6. Writing:
- Overall, 46% of students can write correctly. Most errors were due to omissions and substitutions.

7. Associated variables (*see **Table 3** for statistical significance*):
- In total, 44% of all students repeated a grade. Of that percentage, 20% repeated Grade 1. Grade repetition was one of the variables that had the most influence.

- Students who were in a multi-grade classroom also reported lower results. This was associated with living in rural areas, which showed similar results.
- Results obtained by students whose mother tongue was Spanish differ markedly from those obtained by students whose mother tongue was a native language. Overall, 75% of students did not learn to read in their mother tongue when it was a native language.
- Not all students in the country possess books so it is unlikely that these students can read.

8. Relationships between areas:

- A multiple regression model was performed to determine which variables predicted reading comprehension (*see* ***Table 4***).

TABLE **4**

Reading comprehension	
Predictors	**Regression coefficient (β)**
Oral language	27.26*
Alphabetic principle	-2.66*
Phonemic awareness	1.06
Reading fluency	11.89*
Writing	5.73*

Note: * statistically significant
Percentage explained by the model (R^2): 52%.
Source: del Valle and Cotto, 2015

These data greatly helped determine the current status of students across the country. At the time, Digeduca was still administering and using the results derived from the reading comprehension test at the national level. Considering that most Grade 1 students exhibited low attainment levels in reading comprehension, a decision was made to switch to a test that could provide more detailed information on the students' actual reading skills. This would enable stakeholders to measure learning at a stage prior to comprehension and fluidity, such as emergent reading. The LEE test was ruled out since it is usually administered at the beginning of the school year when the student still hasn't acquired reading skills.

The test was modified so it could be adapted to Grade 1 students. A decision to introduce improvements was also made based on the results of previous tests. Sections were redistributed into eight subsections with the addition of decoding and rapid automatised naming. The main purpose of decoding, traditionally a Grade 1 test, was to identify students who have not yet acquired reading fluency and to determine whether they are at least capable of reading short words. Rapid automatised naming was included as a cognitive factor predictive of reading (Norton and Wolf, 2012).

Some areas, such as alphabetic principle, were also modified to facilitate reporting all the lowercase and uppercase letters mastered by the student. This highlights the importance of letter identification instruction—unquestionably one of the strongest predictors of reading success (Ziegler et al., 2010).

Phonological awareness also underwent some modifications. Previously, asking students to identity the initial phoneme of words had not produced any significant information—probably because at this age, students had already developed this skill and the exercise no longer posed a challenge (Linan-Thompson and Vaugh, 2007). Consequently, the level of complexity was raised.

New aspects were also incorporated. A list of bad reading habits, quality of writing descriptors and new questions designed to detect the causality of results were added, among others. The number of questions was increased so a more accurate measurement of the reading comprehension construct could be obtained. The same parameters used by Digeduca were implemented to assess this skill.

In 2014, a second nationwide administration of the ELGI was launched. There were three main differences—it was administered to more students, a sample was drawn from each municipality and it was administered among Grade 1 official sector students (i.e. 7-year-olds), which is why more aspects and different difficulty levels were incorporated as previously mentioned. An average of six students was randomly selected from a class for a total of 5,949 students from 1,057 schools.

© María José del Valle Catalán, Guatemala

The new test administration detected the following findings (Digeduca, 2015a):

1. Oral Language:
- The great majority of students (80%) exhibited an acceptable level of comprehension of directions.
- After listening to a text, students answered 44% of the questions correctly.

2. Alphabetic principle:
- Most students (61%) identified vowels by their name while only 52% did so by sound.
- On average, a student knew nine letters of the alphabet by their name and their sound, both lowercase and uppercase. Vowels were the easiest letters to recognise (between 50% and 75% rate of success). Ñ/ñ, Z/z, K/k, C/c, J/j, G/g, Ch/ch, H/h, Ll/ll, Q/q and W/w were the hardest letters to recognise (only between 7% and 20% of the population were successful).
- Overall, 8% of the students were still at a learning stage previous to the recognition of letters and the sound of vowels.

3. Decoding:
- Overall, 50% of students could read one-syllable words.
- Students read, on average, 19 pseudo-words per minute.

4. Phonologic awareness:
- While 60% successfully identified the initial sound, only 35% could segment the sound of words.

5. Rapid automatised naming:
- Students read, on average, 37 letters per minute when the task consisted of identifying the name of the letter and 30 letters per minute when identifying sounds. Roughly, they were capable of reading two letters per second.

6. Reading fluency:
- Students read 22 words per minute and 32 words per minute when they are part of a text. This last indicator provides the greatest amount of information.

7. Reading comprehension:
- After listening to a text, students answered 33% of the questions correctly.

8. Writing:
- Overall, 27% of the students had not acquired Grade 1 writing skills.

- Only 5% of the students wrote sentences using capitals and full stops, as required by the National Curriculum for Grade 1 students (primary education).

These results have contributed to the development of an increasingly more accurate reading assessment. This becomes particularly important when informing the country about the level of performance of official sector teachers and the rate at which students are learning. These data also emphasise the urgent need to improve reading instruction.

Undoubtedly, the lack of clear guidelines and the low priority given to all the areas associated with learning to read represent serious obstacles that students must overcome. Furthermore, it is very likely that an under-developed phonetic awareness hinders their ability to learn and slows down their learning pace. The fact that for many students the language of instruction is not their mother tongue poses a serious disadvantage. Additionally, the lack of exposure to reading at home, due to the low educational level of parents or the non-existence of books, slows down the learning rhythm and, on occasions, translates into reading comprehension problems which become evident when students graduate.

5. NEXT STEPS

This section outlines the next steps for assessing reading in early grades. The following initiatives are recommended to improve reading assessment:

Predictive factors model

All the elements being collected in current assessments will be used to develop a model capable of identifying the different factors that play a part in the reading comprehension of Grade 1 students. While the model will be based on specialised international publications, it will be adapted to a Guatemalan context.

Establishing levels

Various levels of performance should be established if the objective is to generate qualitative information on the areas mastered by students. These levels will help identify the students' actual competencies and provide teachers with tools that are suitable to the needs of their student groups.

Number of words per minute

Currently, there is no such standard in the country for oral reading so cut off points are being established that will allow comparisons between students.

More accurate measuring

Results collected over the years have allowed the quality of testing to improve significantly. Each section feeds off the preceding one, new areas are added or existing ones modified, and comparability is strictly maintained across reports. Consequently, every new test receives a new name as it increasingly becomes a more contextualised version of the country and its needs. The name of the current test is *Evaluación de Lectura Inicial*, ELI (Initial Reading Assessment). Each of these changes has a direct impact on the quality of the administration since the greater the complexity, the more training administrators must receive. Therefore, to ensure consensus when it comes to interpreting or assigning scores, on-going training is of critical importance.

Identifying students with reading difficulties

The next test will focus on measuring rapid automatised naming to determine students' ability in this area and thus identify potential problems (López-Escribano et al., 2014).

Native languages

Language is certainly one of the key factors in developing reading skills. Results show that low-scoring students usually speak a native language as a first language. However, determining if students have developed these skills in their own language is a necessary step and one that calls for administering assessments in native languages to determine whether the students have the required reading skills to begin with. Parallel to the test administered in Spanish, four other tests are being developed in as many major native languages, namely:

- *Xje'lb'il u'jin kye tnejil kol te xnaq'tzb'il*—XJU'TKOLX in Mam
- *Yalb'a'ix chi rix Ilok Ru Hu ut tz'iib'ak, sa' xb'een raqal li tzolok* -YIRUHU in Q'eqchi'
- *Retab'al Etamab'äl Sik'inem wuj pa Nab'ey taq Juna'* - RESNAJ in Kaqchikel
- *Etab'al Etamab'al rech Sik'inem wuj pa ri Nab'e taq Junab'*—EESNAJ in K'iche'

All native language assessments incorporate the same components used in the Spanish test. These assessments will provide insight into the students' mastery of their mother tongue and evaluate how well it is being taught in the classroom. Currently, the assessments are pilot-testing their instruments. So far, the main challenge has been locating schools that use the students' mother tongue as the language of instruction. The modifications made to the Spanish assessment were replicated in the native language assessments to further fine tune the instruments.

The following recommendations are broader steps to be considered in the future to improve reading assessment in Guatemala:

Longitudinal studies

Ten years have elapsed since the Digeduca last administered an assessment in reading. The principal focus has been the school but in future, students will be the main target so we can provide teachers with an idea of student performance.

Teacher training

Further research is needed to produce an in-depth account of the methodology Grade 1 teachers use to teach reading. This information will help determine the kind of material teachers would require if they are to efficiently apply the guidelines issued by the Ministry of Education.

Associated factors

Variables that provide a more profound understanding of results should be further analysed.

Use of technology

Assessments are still paper and pencil exercises, although the idea is that future test administrators can register the students' responses on electronic devices. This would represent an important savings in terms of paper and the time required to enter the data.

Assessing reading in primary education

Efforts should be made towards the implementation of on-going reading assessments in the early grades of primary school. This may imply creating different tests with different levels of difficulty in order to determine the reading level of each student. This will help create new strategies to use in the classroom.

Census taking

Currently, only official sector schools are the object of censuses, which differ from those taken in the private sector. Measurements of both these populations are crucially important since they will facilitate the development of new national standards.

6. LESSONS LEARNED

While there are several measurement instruments available on the market, a country that has developed its own national test holds a contextualised tool that will strengthen the measurement competencies of the team responsible for assessments. Over the years, several experts and experiences have merged to bolster the quality of work undertaken by the Digeduca in the country.

The process of developing this test has been an enriching experience for the country—not only because this is the first time such specific data on reading in the early grades have been captured but also due to the know-how of several specialists acquired in the process. Perhaps one of the biggest rewards associated with these assessments is simply being aware of the current status of reading in the country. This knowledge has facilitated improving the test—not only in terms of measuring techniques but also in language-specific areas.

One of the key features of this test has been the creation of various didactic materials, such as Digeduca's reading enhancement book entitled *El tesoro de la lectura* (The treasure behind reading).

© María José del Valle Catalán, Guatemala

The book addresses topics like emergent reading, reading development stages, associated skills and reading comprehension, among others. The test has been administered in three primary education levels: 4- to 6-year-old students; Grade 1-3 students; and Grade 4-6 students. This material has been distributed to official classrooms throughout the country and is expected to become a positive influence on learning to read (*Digeduca,* 2015a, 2015c). Other resources and research materials are also available that provide in-depth analyses of various aspects of Guatemala's child population.

Thus far, there have been several challenges in assessing early grade reading. One of them is the financial burden individual testing implies. The country's geographical features are not particularly favourable to field testing. Some schools are so remotely located that specialists are forced to arrive at their destination the day before the assessment takes place. Naturally, this requires time and resources that may not always be readily available. As the assessment constantly evolves so do these tests. This allows us to provide critical information to decision-makers to help them concretise key national education improvement policies.

REFERENCES

Alcántara, B., Cruz, A.A. and Santos, J.A. (2015). *Informe de primaria 2013.* Guatemala: *Dirección General de Evaluación e Investigación Educativa, Ministerio de Educaci*ón. (In Spanish).

Bolaños Gramajo, V.Y. and Santos Solares, J.A. (2015). *Factores asociados al aprendizaje: informe de graduandos 2012 y 2013.* Guatemala: *Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación*. (In Spanish).

Cruz Grünebaum, A.A. and Santos Solares, J.A. (2015). *Informe de resultados de la Evaluación Nacional de tercero básico 2013.* Guatemala: *Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación*. (In Spanish).

Cruz, A.A. and Santos, J.A. (2013). *Reporte general de primaria 2010.* Guatemala: *Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación*. (In Spanish).

del Valle, M.J. (2015). *Evaluación de Lectura Emergente.* Guatemala: *Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación*. (In Spanish).

del Valle, M.J. and Castellanos, M. (2015). *Matemática inicial en estudiantes de primero primaria.* Guatemala: *Dirección General de Evaluación e Investigación Educativa del Ministerio de Educación*. (In Spanish).

del Valle, M.J. and Cotto, E. (2015). *Informe de Evaluación de lectura en los primeros grados del sector oficial en Guatemala.* Guatemala: *Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación*. (In Spanish).

Dirección de Planificación Educativa (Diplan) (2014). *Anuario Estadístico.* Guatemala: *Ministerio de Educación*. (In Spanish).

Dirección General de Evaluación e Investigación Educativa (Digeduca) (2015). *Qué es la Digeduca?* http://www.mineduc.gob.gt/Digeduca/documents/Folleto_DIGEDUCA.pdf (In Spanish).

Dirección General de Evaluación e Investigación Educativa (Digeduca) (2015a). *Anuario Digeduca.* Guatemala: *Ministerio de Educación*. (In Spanish).

Dirección General de Evaluación e Investigación Educativa (Digeduca) (2015b). *Construcción de las pruebas de Matemáticas y Lectura de Primaria.* Guatemala: *Ministerio de Educación*. (In Spanish).

Dirección General de Evaluación e Investigación Educativa. (Digeduca) (2015c). *El tesoro de la lectura.* http://www.mineduc.gob.gt/Digeduca (In Spanish).

Dirección General de Evaluación e Investigación Educativa. (Digeduca) (2015d). *Serie de cuadernillos técnicos.* http://www.mineduc.gob.gt/Digeduca (In Spanish).

Fortín, Á. (2013). *Evaluación Educativa Estandarizada en Guatemala: Un camino recorrido, un camino por recorrer.* Guatemala: *Ministerio de Educación*. (In Spanish).

Garnier, L. (2008). Repetir o pasar: ¿y la deserción? Ministerio de Educación Pública. Costa Rica. (In Spanish).

Linan-Thompson, S. and Vaugh, S. (2007). *Research-Based Methods of Reading Instruction for English Language learners, Grades K-4.* United States: Association for Supervision and Curriculum Development .

López-Escribano, C., Sánchez-Hipola, P., Suro, J. and Leal, F. (2014). "*Análisis comparativo de estudios sobre la velocidad de nombrar en español y su relación con la adquisición de la lectura y sus dificultades*". *Universitas Psychologica*, Vol. 13, No. 2, pp. 757-769. (In Spanish).

National Institute of Statistics (INE) (2013). *Caracterización estadística de la República de Guatemala 2012.* Guatemala. (In Spanish).

National Institute of Statistics (INE) (2015a). *Estadísticas demográficas y Vitales 2014.* Guatemala. (In Spanish).

National Institute of Statistics (INE) (2015b). *República de Guatemala: Encuesta Nacional de Condiciones de Vida.* Principales resultados 2014. Guatemala. (In Spanish).

Norton, E. and Wolf, M. (2012). Rapid Automatized Naming (RAN) and Reading Fluency: Implications for Understanding and Treatment of Reading *Disabilities.* Massachusetts: Annual Reviews.

Quim, M. and Bolaños, V. (2015). *Informe Primaria 2014.* Guatemala: *Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación*. (In Spanish).

Rubio, F. and Rosales, L. (2010). *Informe técnico de evaluación de lectura para grados iniciales.* Guatemala: USAID/Reaula. (In Spanish).

Rubio, F., Rego, O. and Chesterfield, R. (2005). *El éxito escolar en medio de condiciones difíciles: finalización del primer grado en el área rural de Guatemala.* Guatemala: USAID. (In Spanish).

VV, A. (2009). *Atlas sociolingüístico de pueblos indígenas en América Latina.* Bolivia: FUNPOREIB Andes, UNESCO. (In Spanish).

Ziegler, J.C., Bertrand, D., Tóth, D.C. and Reis, A. (2010). "Orthographic Depth and Its Impact on Universal Predictors of Reading: A Cross-Language Investigation". *Psychological Science*, Vol. 21, No. 4, pp. 551-559.

Expanding Citizen Voice in Education Systems Accountability: Evidence from the Citizen-led Learning Assessments Movement

MONAZZA ASLAM
UCL Institute of Education

SEHAR SAEED
Annual Status of Education Report (ASER) Pakistan

PATRICIA SCHEID AND DANA SCHMIDT
The William and Flora Hewlett Foundation

ABBREVIATIONS

ASER	Annual Status of Education Report
EFA	Education for All
GPE	Global Partnership for Education
ITA	Idara-e-Taleen-o-Aagahi
MDGs	Millennium Development Goals
PAL	People's Action for Learning
SDGs	Sustainable Development Goals
WDR	World Development Report

1. INTRODUCTION

For the first time in history most children are enrolled in school. Over the past 15 years, thanks partly to the Millennium Development Goal (MDG) for universal access to primary education and the Education for All Framework for Action, governments have taken the responsibility of formulating and implementing universal primary education policies, laws and strategies aimed at ensuring that all children enrol and complete primary school. In some countries, politicians came into office on election promises to ban school fees and ensure all children were able to attend regardless of their financial circumstances. Despite significant progress in getting more girls and boys into school, the most pertinent question is whether children are also acquiring the skills that will equip them to lead productive and meaningful lives in modern societies.

Although most developing countries have introduced national examinations and/or assessments to measure children's progress in learning and some also participate in regional or international assessments, these assessments have not yet generated the same level of accountability for learning as there has been for enrolment. The key gaps with these types of assessments are driven by the fact that: (1) these assessments are school-based and therefore do not measure learning outcomes of children who drop out of school, attend irregularly or go to non-formal schools; (2) these assessments are more often than not measuring outcomes too late (upper primary and secondary) when many children have already fallen behind; and (3) the vast majority of assessment results never reach ordinary citizens, and even if they did, they would be difficult to interpret and understand. What these assessments fundamentally fail to achieve is the meaningful engagement of citizens so that the intended beneficiaries of education services can identify and understand whether schooling is translating into learning.

This paper aims to discuss the extent to which another model of assessment—one that is led by citizens rather than governments, conducted in households rather than in schools and that measures whether or not children have mastered the fundamental building blocks for learning—helps to fill existing gaps in government and service provider accountability for delivering quality education. It examines the ways in which citizen-led assessments can strengthen accountability for learning outcomes based on case studies from the following organizations: Idara-e-Taleen-o-Aagahi

(ITA) that implements the Annual Status of Education Report (ASER) Pakistan; Pratham that implements ASER India; Twaweza that implements Uwezo in Kenya, Uganda and the United Republic of Tanzania; Oeuvre Malienne d'Aide à l'Enfance du Sahel that implements Beekunko in Mali; and the Laboratoire de Recherche sur les Transformations Économiques et Sociales at the Université Cheikh Anta Diop that implements Jangandoo in Senegal. More information on the citizen-led assessments movement and links to the country programme websites can be found at www.palnetwork.org.

© ASER, Pakistan

This article will describe how these citizen-led assessments of learning have addressed accountability and participation in education systems by:

- Generating nationally representative and locally-owned data on children's acquisition of foundational skills that have helped re-orient the debate from school access to improved learning for all.
- Creating new opportunities for citizens to better understand the status of their children's learning so that they can decide for themselves whether governments are delivering on promises related to equity and quality in education delivery.
- Promoting new mechanisms for evidence-based policy, proven programme interventions and actions to improve learning.

2. THEORY OF CHANGE

At the global level, citizen-led assessments have played an important role in reorienting the global education agenda through their assessment findings that have been widely cited and used to support discussions on learning (Bangay, 2015). By producing data that, over a 10-year period, repeatedly highlighted the severity of the learning crisis in children's foundational skills, citizen-led assessments provided evidence that helped to make the case for an inclusive and equitable life-long learning for all goal within the Sustainable Development Goals adopted by the 193-member UN General Assembly in September 2015.

At the national or sub-national level, a variety of groups influence the educational decision-making process and ultimately educational change. In a survey of literature across the developing world, Kingdon et al. (2014) argued that access to resources ultimately affects which groups will be able to effectively negotiate change and concluded that the groups with the lowest access to resources are the ones in the weakest negotiating positions. The World Bank's 2004 World Development Report (WDR) identifies two routes for citizens to place demands on their governments (World Bank, 2003). Citizens following "the long route" communicate demands to the State through voting and similar forms of general political accountability. While important, this long route is insufficient by itself because elected representatives ultimately delegate responsibility for service delivery to actors who do not face voter scrutiny and may behave opportunistically, particularly given the inevitable information asymmetries between them and the citizens they are supposed to serve. The long route for accountability must therefore be supplemented by "the short route", in which citizens mobilize at the local level and interact with service providers directly to express needs and demands and obtain better services.

A growing movement of citizen-led, household-based assessments takes the view that ordinary educated citizens can be mobilized for extraordinary

Figure 1. Citizen-led assessments currently underway

Source: Plaut and Jamieson Eberhardt, 2015

actions empowered by evidence. Currently, these citizen-led assessments are being implemented by seven organizations in nine countries (*see* ***Figure 1***).

One important feature of citizen-led assessments is that they directly engage parents and children—who are typically actors with the lowest resources—to become more empowered in seeking the attention of policymakers and service providers and hence, improve their negotiating positions. These assessments are based on following a very simple yet effective premise of building citizen pressure to hold the education system accountable for its dissatisfactory performance. Essentially, building citizen pressure is achieved through both the long and short route of accountability and can be described by the five important stages in the Theory of Change (see also **Figure 2**):

1. **Collect evidence on the learning levels of children**
   Each of the organizations implementing citizen-led assessments works with a network of partners across their respective countries to mobilise and train volunteers in the use of a very simple tool for effectively measuring children's basic reading and math levels. Citizen volunteers then visit households in a sample of villages and test every child within a given age range (*see* **Table 1**).

2. **Communicate findings**
   The findings are collated to provide estimates of reading and math abilities for children aged 6 to 16 years (or starting at age 5 in some countries and ending at age 14 in others) in every district and/or region/state and for each country as a whole. Considerable emphasis is placed on the communication of findings and the fostering of informed public understanding of and debate on children's learning and what can be done to address learning gaps. The results are widely disseminated through national and local media. In many cases, organizations also work at a local level to share findings with parents and communities during the assessment process itself and/or afterwards through local gatherings that often include local elected officials, education officers, teachers and community members. The message is simple: citizens and governments alike must aggressively and creatively take action to improve the quality of education.

3. **Mobilize communities for accountability and action**
   The information is used to engage community and youth leaders, parents and others to take actions to improve learning on their own and through working with their local schools and leaders to advocate for change. As noted earlier, this is the short route for improving accountability for service delivery (WDR, 2004).

4. **Advocate for government action to improve learning**
   Similarly, the information is used to engage directly with national and sub-national policymakers to encourage the government to take steps to improve learning outcomes. In

TABLE **1**

2014 citizen-led assessment coverage

	Districts/regions	Volunteers	Communities	Households	Schools	Children
ASER India	577	25,000	16,497	341,070	15,206	569,229
ASER Pakistan	165	10,000	4,698	93,096	6,235	279,427
Beekunko (Mali)*	216	975	1,080	21,251	2,259	79,079
Jàngandoo (Senegal)	14 (regions)	450	239	9,928	856	26,014
MIA (Mexico)	21	480	187	2,400	187	3,100
Uwezo						
Kenya	158	9,300	4,521	62,089	4,441	135,109
Tanzania	131	8,253	3,930	78,600	3,688	104,162
Uganda	80	4,800	2,372	34,013	2,353	87,339
TOTAL	**1,364**	**59,258**	**33,505**	**642,447**	**35,225**	**1,283,459**

Note: *Data for Beekunko refer to 2013.
Source: Compiled by the authors from citizen-led assessment reports available through country page links on www.palnetwork.org

many cases, organizations work collaboratively with governments to offer solutions. This is the long route for improving accountability for service delivery (WDR, 2004).

5. **Reset the education agenda to focus on learning**
   Over time, the results are used to highlight trends and persistent gaps to make the case for global and country-level goals, targets and indicators related to learning outcomes. This process of consensus building around global priorities is believed to focus donor and national government resource allocations, policies and programme interventions, and to create a universal accountability framework for tracking progress.

As indicated by the Theory of Change, citizen participation is built into the very design of the assessment process. The tools are designed to be simple so that parents, teachers, schools and community members can both conduct the assessment themselves and understand the findings with ease. The approach is led by local organizations and involves thousands of local volunteers, which embeds another element of ownership and citizen participation. Accountability is part and parcel of the ultimate objectives of these assessment exercises that aim to engage citizens everywhere in understanding their situation and taking action to influence education policy and practice from the ground-up.

Figure 2. Theory of Change underpinning citizen-led assessments

Source: Adapted by the authors from Plaut and Jamieson Eberhardt (2015) and consultations with PAL Network members

The remaining sections of this article examine the extent to which the Theory of Change has played out in practice. It looks at experiences in Pakistan

© ASER, Pakistan

and examples from other countries that shed some light on the extent to which citizen-led assessments have strengthened accountability through: (1) global and national-level agenda setting; (2) putting pressure on national policymakers to be more responsive to citizens needs based on evidence of poor learning results in their country (the long route of accountability); and (3) creating opportunities for civil society organizations and citizen groups to better understand the problem and engage with service providers to develop solutions to improve learning (the short route of accountability).

3. ACCOUNTABILITY OF POLICYMAKERS: THE LONG ROUTE

Citizen-led assessments have helped change the environment from one where learning outcomes have been shrouded in mystery to one where they are visible for all to see. They generate widespread attention to the issue of learning in ways that make it impossible for national policymakers or politicians to ignore. The transparency of learning results creates a credible threat to politicians that begs immediate attention lest citizens exercise the long route to accountability.

There is persuasive evidence to suggest that citizen-led assessments have done a remarkable job of helping change policy agendas to focus more on learning by generating greater accountability between citizens and national policymakers (Plaut and Jamieson Eberhardt, 2015). To achieve this, citizen-led assessments have followed certain principles and practices to establish creditability and have identified specific pathways or opportunities to effectively hold policymakers accountable within their socio-political context. At the global level, their collective efforts also created a growing body of evidence and a groundswell to demand improved learning.

3.1 Establishing credibility, familiarity and reach of citizen-generated data

A recent survey of policymakers on the types of external assessments that influence policy suggests that three principles are important: establishing credibility, familiarity and reach through media. Evidence suggests that focusing on the "agenda-setting stage" of policy increases the likelihood that these external assessments will later influence how policies are implemented (Bradley et al.,

Box 1. Sustainable Development Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all

4.1 By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes

4.1.1 Percentage of children/young people (i) in Grade 2/3, (ii) at the end of primary and (iii) at the end of lower secondary achieving at least a minimum proficiency level in (a) reading and (b) mathematics

Source: UN Economic and Social Council Statistical Commission, 2015

2015). Citizen-led assessments are implemented in alignment with these findings.

The first principle has been to ensure that the findings of the assessment are credible. Some organizations have done this by drawing on academics as well as government experts to engage in the process of developing the assessment tools. The citizen-led assessments in Mali, Senegal and East Africa (Beekunko, Jàngandoo and Uwezo) have invited local elected officials, district education officers and others to participate in their assessments to see for themselves how the data is collected and to hear first-hand the communities' response during the dissemination of assessment results. Engagement during the assessment process has increased government faith in the results (Plaut and Jamieson Ebiehardt, 2015).

Second, the annual cycle of assessment creates a natural pulse of repetition where findings are regularly shared. This builds familiarity among national policymakers, civil society organizations and advocacy groups with the assessment and draws attention to the findings.

Finally, citizen-led assessments have capitalized on media coverage to ensure that many people hear about the findings and (in most cases) lack of progress. This includes not only traditional media, like newspapers and television, but also extended reach through local radio shows and social media. In Senegal, which was ranked by a 2013 McKinsey report as one of the top countries in Africa for digital openness/access, Jàngandoo has started to use social media in its campaign. In Mali, there has been more emphasis on local TV and radio, tapping into a rich cultural tradition of story-telling. In East Africa, India and Pakistan, the Uwezo and ASER results have likewise been covered extensively in national and local print, television and radio media.

3.2 Capitalizing on specific global and national policy opportunities

The credibility, familiarity and reach of citizen-led assessments have engendered policy responses that differ depending on the context.

At the global level, citizen-led assessments are likely to have a role to play in tracking progress towards achieving Sustainable Development Goal (SDG) 4 (*see* ***Box 1***). Because citizen-led assessments rely on data collected at the household-level, they capture the learning levels of *all* children—not just those enrolled in and regularly attending formal schools. These data have already and can continue to make an important contribution to better measuring and understanding gaps in equitable learning that otherwise would go unnoticed (UNESCO, 2015 World Inequality Database on Education).

In 2015, the PAL Network joined with other civil society organizations to successfully advocate for the inclusion of an indicator for measuring learning in Grades 2/3 within the SDG Indicator Framework.

This reorientation of SDG 4, and the early grade reading and math indicator endorsed by the Inter-Agency and Expert Group on Sustainable Development Goal Indicators, can be viewed as a first step in a chain reaction to hold global and

national-level institutions accountable for delivering on the promises enshrined in the 2030 SDGs. If approved by the UN Statistical Commission at its March 2016 meeting, the early grade reading and math indicator is likely to generate increased demand for alternative, low-cost and proven approaches for measuring children's early learning, such as the citizen-led assessments.

In India, the national government dedicated additional resources to what they call "learning enhancement" programmes as well as to conducting a learning assessment of its own. In Kenya, the latest Education Sector Strategic Plan focuses specifically on a strategy to improve early reading and math skills and sites the Uwezo results (Plaut and Jamieson Eberhardt, 2015). Similarly, the East African Legislative Assembly called for improved standards and increased investments in education across the region in direct response to the Uwezo 2014 findings that revealed a lack of progress in equitable learning (East African Legislative Assembly, 2015). In Senegal, a special Presidential Commission on the Future of Education adopted many of Jangandoo's recommendations in its final report (*Assises de l'éducation du Sénégal*, 2014). In all these cases, citizen-led assessments have contributed to a shift in national dialogue and policies towards the prioritization of learning.

ASER Pakistan's[1] household-based assessments have provided evidence of gaps in the implementation of Pakistan's Right to Free and Compulsory Education Article 25-A of the constitution by capturing data on enrolments and learning outcomes for all children (Right to Education Pakistan, 2015). ASER Pakistan initiated a Right to Education Campaign as its tactic to communicate assessment information and establish reach with the goal of creating grassroots citizen demand for the government to deliver on this constitutional promise. The campaign—an alliance with other national and global movements on education—focuses on equipping youth, parents, teachers, champions in government and influential people with facts on the status of children's learning and what can be done about it. In response, many local governments have now passed Right to Education Acts. The first was passed in the Islamabad Capital Territory in December 2012, followed by the Balochistan and Sindh provinces in 2013, and the Punjab province in 2014. The Right to Education Campaign's efforts continue in Azad Kashmir, FATA, Gilgit-Baltistan and Khyber Pakhtunkhwa (Right to Education Pakistan, 2015). Case Study 1 describes how the campaign was organized and how it worked (*see* ***Box 2***).

In some cases, citizen-led assessments have directly engaged government officials in the process of conducting the assessments so that they better understand the nature of the problem and experience a first-hand encounter with their constituencies or clients, invoking empathy for how gaps in children's learning affect families and communities.

For example, in 2012 ASER Pakistan started a campaign called "Politicians Knocking on Doors". The purpose of this campaign was to put education issues on the election agenda, giving citizens a platform to demand better access and quality of education from their political leaders. Twenty-two prominent politicians were approached to participate and the 18 who accepted spoke to potential voters on the importance of education and visited their constituencies to observe first-hand the gaps in learning and service delivery. The politicians were filmed visiting their constituents' households and knocking on their doors to ask about the educational status of that household in terms of whether all eligible children were at school and if they were learning well. If the politician found that children were not attending school, they would take the child to the neighbourhood school for enrolment. This campaign's footage, along with a banner, was delivered to political candidates to ensure education issues were featured in their campaigns. Subsequently, during the ASER 2014 report launch, ASER Pakistan revealed the assessment findings

1 To access an example of the tools used in the ASER Pakistan 2015, please click on the tool of interest: the ASER Pakistan 2015 Survey Booklet, the ASER Pakistan 2015 English Tools, the ASER Pakistan General Knowledge assessment.

Box 2. Case study 1: The Right to Education campaign in Pakistan

The Right to Free and Compulsory Education Article 25-A was inserted in the 18th Amendment of the Constitution of Pakistan on April 19, 2010, making it a state obligation to provide free and compulsory education for all children aged 5-16 years. The article was a breakthrough as it codified the state's commitment to fulfilling the fundamental right to education. To promote the implementation of Article 25-A, ASER Pakistan initiated a number of campaign activities.

First, ASER Pakistan spurred the One Million Signature Campaign, undertaken in two rounds over 13 months with two million signatures collected from both in-school and out-of-school children. The signatures were presented nationally and globally to Gordon Brown, UN Special Envoy for Global Education, and the Government of Pakistan in November 2012 and April 2013.

Second, an Education Youth Ambassadors programme is currently mobilizing Pakistan's youth to engage in activism with the goal to yield tangible results in combating the education crisis. The programme builds on and strengthens the emerging worldwide youth movement for global education. It plans to form a network of 500 youth leaders in Pakistan with the passion and dedication to campaign in their schools and communities for actions to get all children into school and learning. To date, 400 youth ambassadors have joined the cause. Over the last year, Education Youth Ambassadors have organized and participated in a number of events, including hosting informative sessions to mark International Women's Day, planning vigils for the Peshawar school attack, and contributing to the National Education Policy by voicing their recommendations and expressing the changes they want to see in education in Pakistan. They have also written stories for the UNESCO Youth Global Monitoring Report as well as countless articles and blogs bringing to light some of the issues they have faced in going to school. The Education Youth Ambassadors have also helped mobilize more than two million petition signatures across Pakistan for A World at School's Up For School campaign – calling on world leaders to ensure that every child is able to go to school, without facing danger or discrimination. The campaign has gathered more than nine million signatures and has been presented to world leaders during the United Nations General Assembly for action on education.

Similarly, the Up For School signature campaign aims to remind governments around the world to fulfil their promise for Universal Education and bring the 58 million children out of school back into school and learning. The Ending Child Marriages programme seeks to create child marriage free zones in Pakistan as a milestone in ensuring the Right to Education for all girls and boys. The campaign is actively involved in the implementation of the Right to Education Act throughout Pakistan with a strong presence and penetration into all the provinces.

and other study results that highlighted various aspects of the political economy of education and learning in Pakistan that required urgent public policy attention.

4. ACCOUNTABILITY OF SERVICE PROVIDERS: THE SHORT ROUTE

Regardless of *how* services fail—through inequitable spending, funding leaks or teacher absences that limit instructional time—there is now evidence that suggests that the effective use of education resources depends on appropriate incentives being in place within the system (Kingdon et al., 2014). In a recent book, Bruns et al. (2011, p. 13) argued that the three core strategies for more accountable education systems are:

- Information for accountability—including the generation and dissemination of information on education inputs, outputs, outcomes and rights and responsibilities.
- School-based management—which works through the decentralization of school decision-making and autonomy to agents at the school level.
- Teacher incentives—which involves policies that link teacher performance to pay and tenure.

This paper focuses on the very first strategy for improved accountability: the generation and dissemination of information. This lies at the core of the Theory of Change developed by citizen-led assessment movements, which posits that information can improve accountability by informing school choice, increasing citizen participation in

school oversight and enabling citizen voice. Thus, in addition to building global consensus and ensuring that equitable learning is on the national-level policy agenda, citizen-led assessments are also experimenting with pathways to influence how these policies actually get implemented where it counts: at the point of service delivery in communities and classrooms. This often starts with ensuring that state and local education officials, school leaders, teachers, parents, youth and others 'see' the problem and are motivated to take action to solve it.

Evidence of whether information campaigns on education influences accountability and learning outcomes is mixed. Where the evidence suggests a positive impact, it shows that information can create a feedback loop between parents (whose access to both information and resources is usually weak) and service providers. For example, national and regional report cards in Central and Latin America provide citizens with short, non-technical summaries that have increased public awareness on the performance of their countries' education systems (Bruns et al., 2011, p. 31). Andrabi et al. (2015) highlights the value of providing citizens with information on their children's test scores, which served to increase subsequent test scores by 0.11 standard deviations, decrease private schools fees by 17% and increase primary school enrolment by 4.5%. However, in another setting, a study in one district of Uttar Pradesh in India found that an intervention aimed at raising awareness of the Village Education Committee's role on improving learning through the use of community scorecards did not impact learning outcomes. The study found that the intervention had not increased parental knowledge nor changed attitudes, suggesting that the intervention was not intensive enough to bring about change (Banerjee et al., 2008).

In the context of citizen-led assessments, early experiments suggest that providing parents with information on their children's learning during the assessment process does not systematically change parental behaviour (Lieberman et al., 2014). In other words, it has proven difficult to animate the short route of accountability through one-time visits to people's homes. That said, organizations conducting citizen-led assessments have been experimenting with other mechanisms for reinforcing the short route of accountability. For example, in Mali, Beekunko has been holding village level meetings to discuss the results and have the community and teachers jointly establish learning improvement plans. In East Africa, Uwezo is sharing the data at a local level more intensively through local champions and experimenting with ways to engage volunteers beyond the assessment process. In India this year, the ASER is running an ambitious *Lakhon Mein Ek*—or "one in one hundred thousand"—campaign that has involved finding volunteers in over 100,000 villages to conduct a complete census of children's learning levels. After completing the census, the results have been compiled into a report card that is shared with the village. Based on these results and with support from the ASER and Pratham, the communities are now debating and deciding the types of actions they might take to improve learning in their village (Mukherjee, 2016).

In both India and Pakistan, there have been promising efforts to work collaboratively with state and local education officials to use assessment results to inform the design of education programmes to accelerate learning. In the Indian state of Bihar, for example, district and state education officials turned to Pratham, the facilitator of ASER India, for ideas on how to improve low learning levels. Pratham first helped state and local authorities improve their own assessment practices: Cluster Centre Resource Coordinators participated in a simple assessment of learners' skills to better understand learning gaps in the schools for which they were responsible. Pratham then introduced practical and proven solutions that these Coordinators could implement along with teachers and school heads. Since the assessments revealed that many children in Grades 3, 4 and 5 were still unable to read at a Grade 2 level, an hour of instruction a day was introduced where children were grouped by learning level and engaged in activities targeted at helping them to advance. In Pakistan, a similar project, Learning for Access, has been tested by ITA/ASER in 560 schools in the Sindh, Balochistan and Punjab provinces. Learning for Access combines

a 60-day participatory learning camp to accelerate learning for children who dropped out, never enrolled or are at risk of dropping out with efforts to upgrade schools to better accommodate and retain newly enrolled students.

These efforts to work directly with local officials responsible for service provision align with some of the recent literature in social accountability which suggests that, in some contexts, creating effective accountability relationships requires shifting from adversarial interactions between citizens and public officials to "co-creation of solutions" (Fox, 2014). Teachers and prospective teachers have also been a part of the co-creation process in India, Pakistan and Mali. For example, ASER India has involved pre-service teachers from the Teacher Training College in its recent survey as data collectors. Similarly, Beekunko has invited the Teachers Union in Mali to join in the national dissemination events, and school leaders and teachers are involved in the local dissemination events and action planning that follows. In Senegal, in response to very low learning levels in Arabic-medium schools, Jàngandoo has worked with school supervisors and the decentralized school authorities in one region to develop, test and introduce remedial education guides that are designed to provide teachers, community volunteers and local organizations working to support children's learning with new instructional approaches and materials.

Case study 2 (*see* ***Box 3***) describes in more detail how ASER Pakistan engaged local officials and teachers in co-creating solutions. In doing so, teachers are involved in the 'change' process and hence become critical agents for ultimate change.

5. WHAT'S NEXT FOR CITIZEN-LED ASSESSMENTS OF LEARNING?

What has been learned and what does it imply for the future of citizen-led assessments? First, it is clear that citizen-led assessments of learning have played a role in building a consensus to shift the global and national agenda towards learning. There is also clear evidence of their influence on national

policies (the long route of accountability). That said, increased attention on learning globally and pro-education policies has not yet translated into learning gains. Furthermore, animating the short route of accountability has proven more difficult than anticipated (Plaut and Jamieson Eberhardt, 2015; Lieberman et al., 2014). In short, citizen-led assessments have strengthened accountability in significant ways but the expected outcomes described in the Theory of Change presented earlier have yet to be fully realized.

5.1 Gaps in the Theory of Change

Bangay (2015) argues that the consistent decline in learning outcomes across 10 years of the ASER in India is testament to the fact that information alone cannot affect change through presumed accountability. He cites the following reasons for the lack of progress in changing the status quo in learning outcomes: the need for more time to actually see change happen for large school-aged populations; citizen-led assessment findings losing their shock appeal after several years of bad news;

Box 3. Case study 2: ASER Pakistan village gatherings

ASER Pakistan introduced the ASER Baithaks, or village gatherings, as a pathway to sensitise and mobilise communities to address the education crisis and create demand for action at the grassroots level. An informal discussion with the community and teachers of a village surveyed under the ASER is an important component of the ASER dissemination.

ITA's teams led by the ASER District Coordinator organize ASER Baithaks (Katcheries, Baithaks or Jirgas for Sindh, Punjab and Balochistan, Khyber Pakhtoonkhwa respectively). These gatherings are organized at school and/or community sites to share ASER findings, mobilize volunteers for education, and decide on actions with community members, youth, parents, teachers and government field officers. They begin with the sharing of objectives of the conversation and reminding the attendees of the survey recently conducted in the village. The results are then shared, underscoring that while the survey is based on a sample of 20 households, the trends represent the whole village. This information then leads to many conversations and reflections in the community as they acknowledge gaps in taking actions for education, lapses in accountability and the role of the village school, its teachers and the larger system of education. The discussions are like a mirror for the community and parents: they begin to see more clearly the state of learning of their own children and education in public and private schools in their community.

Once the discussion reaches a climax, the ASER Pakistan facilitators give a call for action. Who will ensure that children are enrolled on time and brought to the school? Who will take turns to see if the teachers come to school on time? Who is educated at least up to Class (i.e. Grade) 12 and will volunteer to teach an eight to ten week accelerated literacy and numeracy programme for out-of-school children and at-risk in-school children with the lowest learning levels? Who will write to the government with a request to improve the facilities and shortage of teachers?

In this way, the ASER Baithaks provide citizens with a platform to discuss problems of out-of-school children and low learning levels based on information that they can understand and that is relevant to them, and focus citizens on coming up with locally driven solutions. The community is roused into action and a passion for education is unleashed.

communities *knowing* something not necessarily translating into them *acting* on it; and finally, that governments still are not fully engaged or sufficiently accountable to take action.

Moreover, Bruns et al. (2011) argue that while there is no arguing *against* greater information—one of the main objectives of citizen-led assessments—sometimes this intervention leads to unintended consequences like "elite capture", wherein information campaigns can only be understood by more educated groups of parents. A study by Banerjee et al. (2008) in India highlights this as a very real problem and as a challenge that necessitates unique solutions for the increased and sustained success of citizen-led reform movements. Another pitfall that Bruns et al. (2011, p. 73) specifically note with respect to information-for-accountability reforms is the reliance on test scores alone which may be heavily dependent on socio-economic background and other unobserved factors. Not suitably accounting for these can result in misguided interpretations which can undermine the value of such interventions.

5.2 Lessons learned and areas for future experimentation

There are important lessons that have been learned during the assessment process. The process has highlighted the fact that credible assessment can help reset and refocus policy agendas. It suggests the power and potential for expanding citizen participation in monitoring service delivery outcomes, which can provide external checks on government services reliably and cost effectively. Finally, it underscores the importance of local ownership and engaging all education stakeholders at all stages of the process to create real opportunities for change.

Moving forward, there are clear opportunities for leveraging international agreements, such as the 2030 SDGs and the EFA Framework for Action, to

reinforce action at the agenda setting stage and to create agreed indicators and processes for tracking country performance. Citizen-led assessments have a unique role to play in tracking progress for three reasons: (1) they are independent of government assessments; (2) they capture learning for *all* children not just those enrolled in school; and (3) they measure progress on early learning outcomes that are critical for future success.

Carlitz and Lipovsec (2015) identified the remaining challenge of finding new ways to unlock parental action by experimenting with new strategies for communicating information that is relevant and actionable to them. There is also the need to critically engage local elected and education officials, school and community leaders, and teachers as positive agents of change. This could, for instance, be achieved by experimenting more with how to create platforms for parents and other concerned citizens to work together to first jointly diagnose the problem and then create solutions. The citizen-led assessments are also interested in experimenting more with how to involve teacher training institutes to enhance teacher awareness and skills for using assessments to diagnose children's learning status and responding appropriately with strategies for their instruction.

5.3 Expansion and peer learning

The People's Action for Learning (PAL) Network (www.palnetwork.org) brings together seven civil society organizations working across nine countries (and growing) to assess the basic reading and numeracy competencies of all children, in their homes, through annual citizen-led assessments. The PAL Network was formally declared in July 2015 by a group of activists and thought leaders who aspire to create a movement where learning is at the centre of all education endeavours. The network offers a platform from which citizen-led assessments can continue to influence global accountability systems. As the network expands and citizen-led assessments proliferate, so too do the opportunities for promoting accountability for learning in new countries. Perhaps most importantly, the network offers an opportunity for citizen-led assessments to leverage experimentation and learning across many contexts to better understand ways in which these processes can strengthen local accountability relationships. As a result of the diversity of experiences across its network, citizen-led assessments can strengthen and refine the Theory of Change that animates their efforts to give citizens a voice in education system accountability.

REFERENCES

Andrabi, T., Das, J. and Khwaja, A. (2015). *Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets*. World Bank Policy Research Paper 7226. Washington, DC: World Bank.

Assises de l'éducation du Sénégal (2014). *Rapport general: document de travail, 03 août*.

Banerjee, A.V., Banerji, R., Duflo, E., Glennerster, R. and Khemani, S. (2008). *Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India*. Policy Research Working Paper 4584. Washington, DC: World Bank.

Bangay, C. (2015). *Why are citizen led learning assessments not having an impact on home soil—and how can we change that?* Blog posted: June 2, 2015. https://efareport.wordpress.com/2015/06/02/why-are-citizen-led-learning-assessments-not-having-an-impact-on-home-soil-and-how-can-we-change-that/?utm_content=buffer15399&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer

Bruns, B., Filmer, D. and Patrinos, H. A. (2011). *Making Schools Work: New Evidence on Accountability Reforms*, Washington, D.C.: The World Bank.

Carlitz, R. and Lipovsek, V. (2015). *Citizen-led assessments and their effects on parents' behavior. A synthesis of research on Uwezo in comparative perspective*. Twaweza. http://www.twaweza.org/uploads/files/LPTSynthesisFINAL.pdf (Accessed June 8, 2015).

Lieberman E., Posner, D. and Tsai, L. (2014). "Does Information Lead to More Active Citizenship? Evidence from an Education Intervention in Rural Kenya". *World Development*, Vol. 60, pp. 69-38.

East African Legislative Assembly (2015). *EALA wants Standards of Education raised*. May 25, 2015 Press Release. Arusha, Tanzania.

Fox, J. (2014). *Social Accountability: What Does the Evidence Really Say?* Global Partnership for Social Accountability. http://gpsaknowledge.org/wp-content/uploads/2014/09/Social-Accountability-What-Does-Evidence-Really-Say-GPSA-Working-Paper-1.pdf (Accessed June 8, 2015).

Global Partnership for Education (2015). *Global Partnership for Education grants US$235 million to support education in Bangladesh, Mozambique, Nepal and Rwanda*. May 23, 2015 Press Release. http://www.globalpartnership.org/news/global-partnership-education-grants-us235-million-support-education-bangladesh-mozambique-nepal (Accessed June 8, 2015).

Kingdon G.G., Little, A., Aslam, M., Rawal, S., Moe, T., Patrinos, H., Beteille, T., Banerji, R., Parton, B. and Sharma, S.K. (2014). *A rigorous review of the political economy of education systems in developing countries*. Final Report. Education Rigorous Literature Review. Department for International Development. http://eppi.ioe.ac.uk/

LMTF (Learning Metrics Task Force). 2013. *Toward Universal Learning: Recommendations from the Learning Metrics Task Force*. Montreal and Washington, D. C.: UNESCO Institute for Statistics and Center for Universal Education at the Brookings Institution.

Mukherjee, A. (2016). *New Initiative in India Is Mobilizing Communities to Improve Children's Learning, but Will It Work?* http://www.cgdev.org/blog/new-initiative-india-mobilizing-communities-improve-childrens-learning-will-it-work (Accessed on January 12, 2016).

Parks, B., Rice, Z. and Custer, S. (2015). *Marketplace of Ideas for Policy Change: Who do Developing World Leaders Listen to and Why?* Williamsburg, VA: AidData and The College of William and Mary. http://www.aiddata.org/marketplace-of-ideas-for-policy-change.

Plaut, D., Jamieson Eberhardt, M. (2015). Results for Development. *Bringing Learning to Light: The Role of Citizen-led Assessments in Shifting the Education Agenda.* Washington, D.C.: Results for Development Institute.

Right to Education Pakistan (2015). http://rtepakistan.org/about-rte-pakistan/

United Nations Economic and Social Council (2015). Report of the Inter-agency and Expert Group on Sustainable Development Goals Indicators. (Statistical Commission, Forty-seventh session 8-11 March 2016, Item 3(a) of the provisional agenda).

UNESCO (2015). *Education for All 2000-2015: Achievements and Challenges*. EFA Global Monitoring Report 2015. Paris: UNESCO.

UNESCO (2015). World Inequality Database on Education. http://www.education-inequalities.org/ (Accessed on June 8, 2015).

World Education Forum (2015). *Education 2030: Towards inclusive and equitable quality education and lifelong learning for all*. Incheon Declaration. UNESCO, Ministry of Education Republic of Korea, UNDP, UNFPA, UNICEF, UN Women, UNHCR and the World Bank Group. http://www.uis.unesco.org/Education/Pages/post-2015-indicators.aspx (Accessed June 8, 2015)

The World Bank (2003). *World Development Report 2004: Making Services Work for Poor People*. Washington, D.C.: The World Bank and Oxford University Press. https://openknowledge.worldbank.org/handle/10986/5986

© Dana Schmidt/The William and Flora Hewlett Foundation

Chapter 5 Recommendations and Conclusions

This section presents the recommendations for selecting, planning, implementing and using oral reading assessments. The recommendations highlight basic principles that should be applied in the different stages of such assessments and are based on the articles presented in the previous chapters. Although the articles explore a wide range of oral reading assessments conducted on different scales, the following recommendations pertain to practices that can be scaled-up to the system-level.

Recommendations

RECOMMENDATION 1:

Develop an assessment plan for comprehensive reform

➜ Ministers of education or educators must make these three decisions when developing an assessment plan:

- determine the level of assessment or who will be assessed
- its purpose or why it will be administered
- the object of assessment or what knowledge, skills, language level, perceptions or attitudes will be assessed.

➜ When developing an assessment plan, assembling a solid team of partners, ensuring data quality, constructing a vision for what will be done with the results and creating an itemised budget are of critical importance.

Assessment has taken center stage in education reform. Currently, data from oral assessments are used to make system-level programmatic decisions to inform reform efforts or individual projects. If the sustained use of assessments for instructional decision making by ministries of education is the goal, then the current use of early grade assessment measures needs to be expanded and at the same time aligned with the rest of the assessment systems and frameworks within countries.

At the national level, an assessment plan is needed for comprehensive reform. An assessment plan outlines: what data will be collected, by whom and for what purpose; the process for reviewing data, policies and procedures to guide feedback results; and the process for modifying the programme or curriculum. Summative assessments evaluate student learning at the end of a specific instructional period. Interim assessments evaluate where students are in their learning progress and determine whether they are on track.

To improve instruction, ministers of education and educators must make decisions on which assessments will help them develop an assessment plan that will provide data that are not only useful but that can be collected with fidelity. They will have to determine the level of assessment, its purpose and the object of assessment.

LEVEL OF ASSESSMENT

Current oral assessments are individually administered. Individual assessments are preferable when assessing young children, particularly when assessing reading, since children may not be able to read yet and critical skills in assessing reading, such as phonological awareness, are best assessed individually. Alignment across the various assessments is advised to ensure that data provide information on students' progress across time and across measures. Save the Children's IDELA and Literacy Boost assessments provide an example of how measures can be aligned. These assessments measure children's early learning and developing reading skills from age 3 to Grade 3, presenting a continuous assessment framework between pre-primary and primary education. In addition, the range of skills assessed when measures are aligned helps to avoid floor effects.

That is, these assessments pay more attention to what is happening at the lower end of the skill distribution by testing a more basic version of the same skills. This makes them particularly well suited to marginalised populations. Further, they promote the inclusion of a range of continuous indicators, spanning from foundational to higher order skills, and hold the goal of learning to read as the ultimate non-negotiable outcome to measuring comprehension (Dowd et al., 2016).

It should be noted that although individually-administered oral assessments have been the de facto norm, group-administered oral assessments have been used in developing countries. The implementers of the Centers for Excellence in Teacher Training (CETT) (Chesterfield and Abreu-Combs, 2011) developed group-administered oral assessments of early reading skills to measure the impact of the project. More recently, researchers at the Research Triangle Institute (RTI) International have performed psychometric research on the suitability of group-administered oral assessments in developing countries. With the Group Administered Reading Assessment (GARA), the assessor orally administers a reading assessment to a group of students and their responses are collected using paper-and-pencil student response sheets. Since the assessment tool is in the form of multiple-choice questions, children's writing skills are not being tested along with reading—with the exception of the writing dictation subtask. Like the measures developed by CETT, GARA differs from other group-administered reading assessments (e.g. LLECE, PIRLS, PASEC, SACMEQ) in that the test is not reliant on passage reading ability. Since it begins with skills as simple as letter names/sounds, group-administered assessments cover the same range of skills as measured by the EGRA. The goal of group-administered reading assessments is mostly to lower the cost of training as it is time-consuming to train assessment

administrators to conduct oral assessments. The GARA was conducted in Egypt (RTI International, 2014). It is still evolving and being piloted in different countries and contexts.

PURPOSE OF ASSESSMENT

Determining the purpose of the assessment is important due to various implications, such as who will be assessed, how often and where. The plan must also stipulate who will conduct the assessment. Some critical points to consider:

1. What is the purpose of assessment?

This question will drive all other decisions. It is also important to note that the purpose of assessment may change over time. An initial assessment plan may propose including children in the early primary grades to obtain a baseline to determine whether reform is needed nationally or in a specific region. When a ministry of education is implementing a new instructional approach, it may want to assess children more frequently using both formative and summative assessments but only target the children that are part of the intervention. However, if the purpose is accountability and an early grade reading assessment is being integrated into a system-wide assessment plan, then decisions will be driven by curricular factors—for example, the assessment can be administered at the end of the first year of instruction and then again at the end of the first cycle. Another purpose may be to determine the literacy levels of children outside the education system, including children in informal programmes or those who never attended school or who have dropped out.

2. Who will be assessed?

In the case of school-based assessments, this refers to both the grade levels to be assessed and the number of students who will be assessed. Identifying a representative sample of students is key. When assessing students during an intervention, both intervention and control students should be assessed. Whether assessment is part of a national plan or an evaluation of an intervention, it is important to first identify all groups of children that need to be included (e.g. from different language groups, genders, socio-economic statuses, geographic locations, etc.).

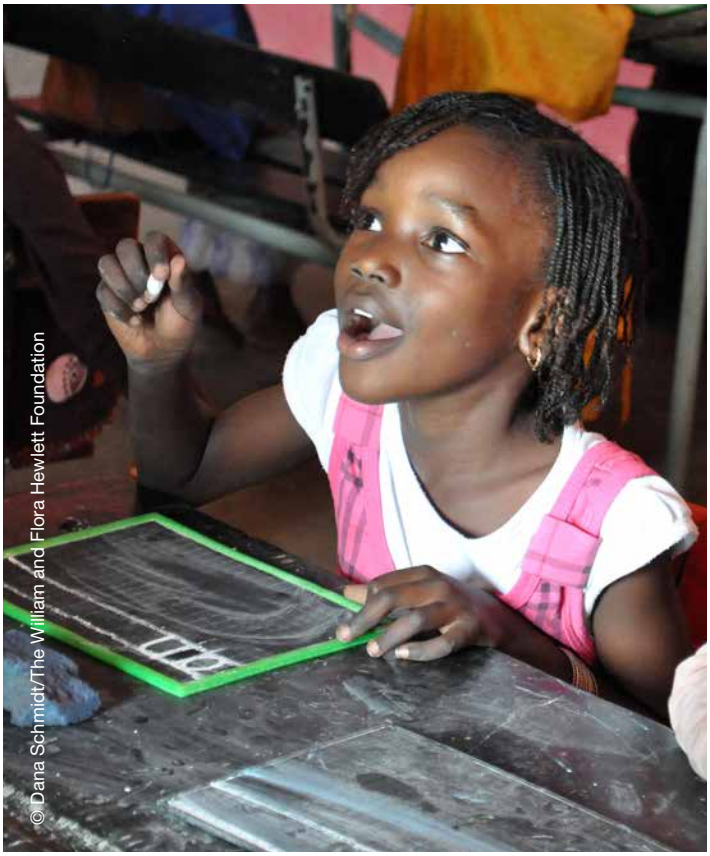
Next, it is key to determine the appropriate proportions for each group. Appropriate representation will ensure that results can be generalised to all students in the education system. These principles also apply to household assessments. In the case of household assessments, the purpose of the assessment may be to determine the literacy levels of preschool-aged children or out-of-school youths. It is also important to find representative samples of children. Although it may be more difficult to identify the eligible population when conducting household assessments, there are formal and informal avenues to collect census information and create sampling frameworks. Official census as well as village, tribal or church registrars can be used to identify participants and to ensure that the sample is representative.

3. How often will children be assessed?

The timing of the assessment is also based on the purpose. For example, when implementing a new reform effort, more frequent assessment may be necessary to monitor implementation. Once the reform is well under way, less frequent assessments may be required or only assessments of students in particular grade levels. Another important aspect of measurement is assessing change and growth in literacy skills. A measurement of change requires at least two points in time. This repeated assessment may not be feasible due to logistics, cost and/or other factors but it is desirable when possible.

4. Where will children be assessed?

If there is an interest in determining what all school-aged children in a particular context know, then a household-based assessment may be more useful, particularly in a context where not all children attend school. However, if the purpose



is to determine how well children in school are learning and/or if all children are in school, then a school-based assessment presents a better option. This assessment could be combined with a targeted assessment of children not in school with a specialised sampling approach to capture data on them. An education system may require information from both school-based and household-based assessments to determine the literacy landscape in the country. However, assessing at both levels may not be possible or sustainable.

5. Who will collect the data and how?

Assessment data can be collected by trained community volunteers, paid trained assessors or ministry of education staff, including teachers. Determining who will assess students will depend on resource availability and capacity. To ensure that the data are collected reliably, the assessors should be fluent in the language of the assessment and should receive adequate training, including (when possible) training on site at the local schools. Additionally, assessors should be assessed before the data collection to ensure that they are well prepared. Measuring inter-rater reliability is highly encouraged to determine assessor readiness.

Oral assessment data can be collected using paper or digital means. Determining which data collection method to use will depend on resources. Capturing the data using digital means will depend on the daily access to electricity to recharge tablets or phones and online connectivity to upload data. Some of the general advantages to collecting data electronically include: rapid availability of assessment data; improved data accuracy and fewer measurement errors due to missing fields, data transcription errors, invalid data types or formats and illegible or incomprehensible data; a reduced amount of paper and supplies required; as well as more simplified logistics to prepare and manage the data collected compared to paper assessments (i.e. no photocopying, sorting, stapling, packaging, etc.). Rapid availability of the data also makes supervision easier and can result in immediate correction of field problems (e.g. quicker online support to an assessor who encounters problems).

OBJECT OF ASSESSMENT

The object of assessment refers to what is assessed. It is important to identify what domains and constructs will be assessed to determine children's knowledge and skills. The focus of oral assessments has been on early reading skills. There appears to be consensus on what to measure in both foundational and higher order skills based on the substantial literature in the field of reading development.

Although oral reading assessments do not vary much in length of administration, they vary in the type of data they provide. For example, while all assessments provide data on students' alphabet knowledge, word-level reading and text reading, the type of information varies. Some instruments assess students on all letters while others assess students on the easiest and/or most difficult. Some assessments are timed while others are not. As a result, these assessments provide different levels or types of data with regard

Box 1. Adapting an existing instrument or designing a new assessment tool

Valid and reliable instruments and tools that have been developed for various purposes can be adapted to different contexts. For example, the Women Educational Researchers of Kenya (WERK) has developed an oral reading assessment for the Maa language based on the EGRA and Uwezo Kiswahili assessment tools (Kinyanjui, 2016). Adapting instruments to each new context requires knowledge of the linguistic structure of the language among the students to be assessed, the context and often some sense of curricular expectations as well as the availability of reading or language textbooks. Only when adapted correctly and applied using proper assessment techniques will the results yield a reliable and valid depiction of skills (Dubeck et al., 2016). Note that when adapting existing tools, piloting is critical. The *EGRA Toolkit* provides detailed guidance on how to develop and adapt an EGRA.

If designing a new oral reading assessment is the goal, then assessment developers must address the following factors:

- Testing economy: how much information do you get from the battery of tests? How many reading constructs will you assess?
- Efficiency and predictive validity: how much information do you get for the effort? Limit the measure to those that are most predictive. Assessments should take no longer than 15-30 minutes.
- Task difficulty: which skills and knowledge will you measure? Are they appropriate for the language and the reading level of the students?
- Developmental validity: how well will the items hold up over time? Have you avoided floor and ceiling effects?

To ensure assessments are reliable, they must be developed though a rigorous process. Overall the tasks to be included in the assessment should be:

- Research-based and capable of assessing critical aspects of literacy;
- Built around contexts likely to be familiar to students in the early years of school;
- Able to be administered by the student's own teacher (when being used by teachers and schools for formative purposes as opposed to outside assessors). In this case, the tasks should be easy for teachers to administer and should be supported with clear and explicit marking and recording guides (Meiers and Mendelovits, 2016).

to children's reading ability that range from basic categories (i.e. letter, word or text) to fluency rates on a number of skills.

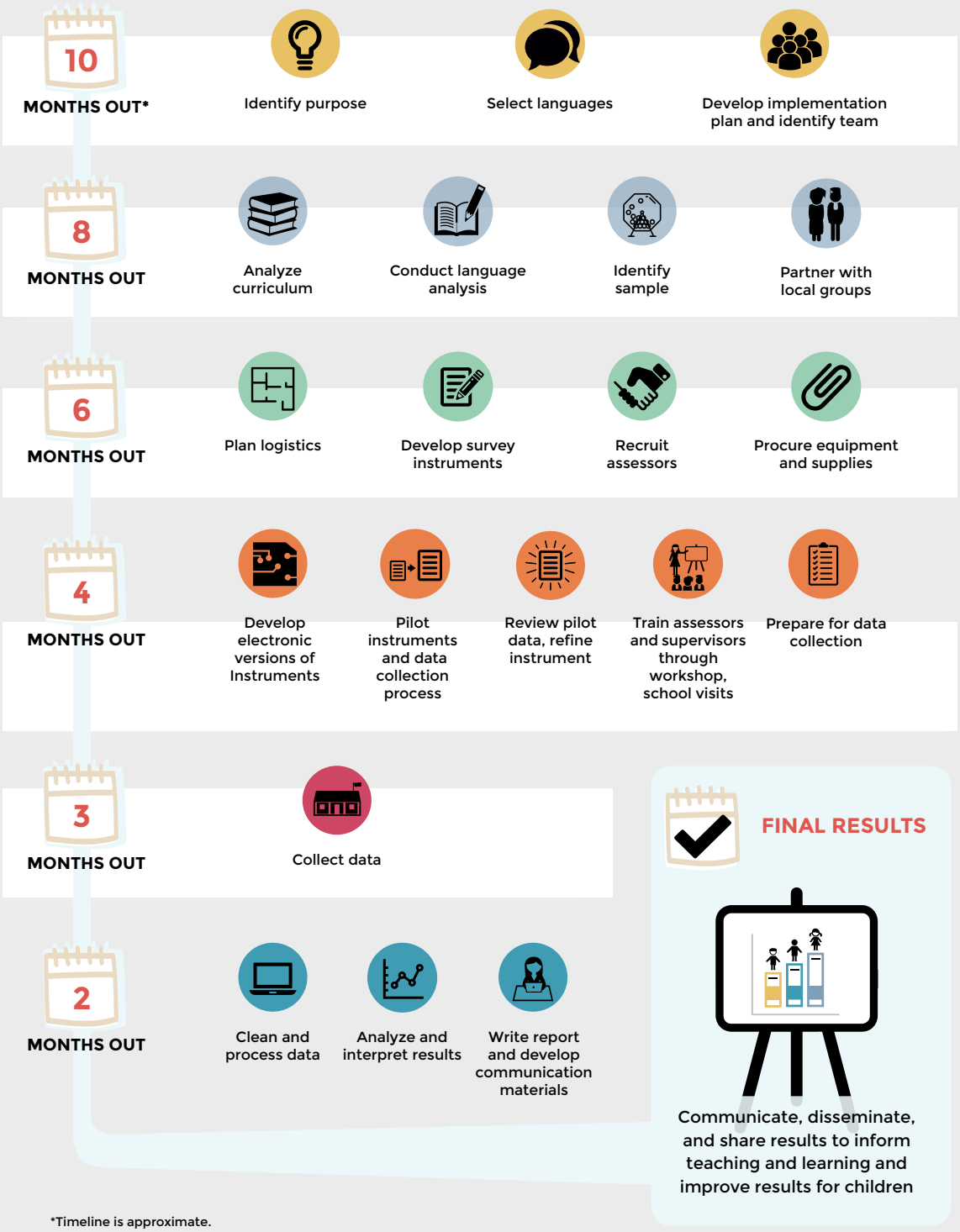
There are also differences in the number of constructs that are assessed. If the purpose of the assessment is to determine what reading skills and knowledge children possess and at what point they developed these, it is helpful to assess foundational skills that are predictive of later reading abilities and higher order skills when planning an intervention. Although not all reading skills are amenable to quick assessments, the data from even brief assessments can provide an index of what students know and can inform reform efforts.

Figure 1 illustrates the steps in planning the EGRA assessment. The timeline is to be used for planning purposes.

IMPORTANT PRACTICAL ELEMENTS TO CONSIDER IN THE ASSESSMENT PLAN

As noted above, an assessment plan outlines what data will be collected, by whom and for what purpose; the process for reviewing data, policies and procedures to guide feedback results; and the process for modifying the programme or curriculum. Aside from the key structural components for developing an assessment plan, there are also two

Figure 1. The Early Grade Reading Assessment Timeline



*Timeline is approximate.

Source: Kochetkova and Dubeck, 2016

important practical elements to consider: building a good team of collaborators and budget planning.[1]

1. A solid team of partners

Strategic partnerships are critical for sharing knowledge and increasing ownership of the assessment results. Partners include donors, ministry staff and technical collaborators. The assessment plan must be shared with the various partners and should be aligned with the country's priorities. It is strongly advised that ministry staff be engaged at all levels of the assessment. Experience shows that those officials who participate in the development and implementation of the assessment will understand its applicability to their national context and are able to advocate for its use when required. Involving local actors also ensures that their skills are built to carry out other assessment activities in the future. Sometimes involving high-level officials in the field work—even for a day—can also prove useful as they can develop an immediate and practical sense of how the assessment works and gain firsthand a sense of children's reading levels.

2. Budget planning

Budget planning depends on how much work has already been done; for example, there are differences in the initial cost of an assessment and a reapplication. In general, costs of oral reading assessments will vary by country and are dependent on sample size, level and number of disaggregation desired, local inputs of labor and transportation and the use of technology for data collection. Of the assessment experiences included in the ebook, the cost per learner ranged from less than a dollar (USD) to a few hundred dollars. When contributors to the ebook were asked to categorise the proportion of funds allocated to each of the assessment stages, almost all allocated the largest proportion

to test application (implementation) and the lowest to dissemination activities. The assessment cost categories were based on a breakdown for budgeting proposed by Wagner et al. (2011) which included test preparation, test application (implementation), processing and analysis, dissemination, and institutional costs.

When planning the initial assessment, donors and non-governmental organizations (NGOs) are encouraged to involve in-country teams who can provide more accurate local costs. It is also prudent to leave a margin in the budget for unexpected consequences/missed expenses. This will ensure that there are sufficient funds for each stage of the process and avoid delays or shortcuts due to insufficient funds.

1 The recommendations in this subsection have been compiled by the UIS based on a questionnaire circulated among a select number of authors. The questionnaire focused on experiences and lessons learned related specifically to budgeting and assessment planning.

Further, good planning for the first assessment can result in reduced costs for future implementations. Taking the time to plan carefully and develop a valid and reliable assessment will help avoid incurring unexpected costs later on. It will ensure that results are valid—the assessment will stand the test of time, reducing the likelihood that it will have to be modified later which will result in additional costs and loss of comparability over time. If there is a desire to measure change over time, then multiple equated assessment forms must be designed from the outset (more information is available in the *EGRA Toolkit*). Careful planning will also reduce the likelihood that training materials will have to be modified down the road. While there are ways to reduce costs, there are areas that are non-negotiable, such as ensuring the assessment is properly developed and piloted during test preparation. Other areas, such as who assesses are negotiable—one way to reduce assessor costs is to train ministry staff or salaried officials as assessors or to use volunteers. Another way to reduce costs is to take advantage of pre-planned activities to disseminate the results of the assessment.

It is advisable to construct a strategy to enhance national capacity in implementing assessments. It should encompass a plan to support ministries in providing oversight and management of the assessment. If strengthening national capacities in analyzing and interpreting data is desired, then planning for activities such as theoretical training sessions on measurement method models, use of special statistical software packages and data analysis techniques should be considered.

© Dana Schmidt/The William and Flora Hewlett Foundation

RECOMMENDATION 2:

Collect additional information to understand the context in which teaching and learning take place

- Data on reading achievement alone are not sufficient to design sound interventions. Additional information is necessary to ensure understanding of the assessment results.
- Choosing what additional information to collect is dependent on the purpose of the assessment and the context. The additional variables can be related but are not limited to: language(s), home literacy practices and quality of instruction.

Understanding the social, economic and political settings of schooling as well as the educational environment in the country in question contributes to appreciating how those contexts impact learning and is critical when interpreting assessment data or designing interventions. Although oral assessments of early reading provide useful data on children's current level of performance, learning occurs in a variety of contexts. Therefore, in addition to decisions related to the type of assessment to be implemented (benchmark, evaluation, national diagnostic, etc.), countries should also collect relevant contextual information to ensure understanding of the assessment results—a necessary step to informing and formulating targeted interventions/policy. The decision on which additional variables to collect is tied to the purpose of the assessment and the context in which teaching and learning take place.

The measurement of the context of early reading skills is complex and there are many variables to consider. These variables include but are not limited to: national language(s); local language(s); language(s) of instruction; instructional practices; organization of the school system and access to formal schooling; teacher training; curriculum; exposure to oral and written languages; funding available for education; the availability of teaching materials and books; gender issues in terms of access to education; home and environmental practices related to literacy (e.g. presence of reading materials in the home, print-richness of the environment); and socio-economic and cultural conditions. It is important to be aware of these variables when assessing the reading skills of an area or country. In some countries or certain communities in rural areas, access to water, sanitation, healthcare and education is quite limited or non-existent. Such variables should be systematically recorded in order to provide a clear picture of the context.

LINGUISTIC CONTEXT

In most countries, children are growing and learning in bilingual or multilingual environments. Children's proficiency in the language of instruction can have an impact on assessment and the effect of interventions. Recently, the use of oral assessments has expanded from assessing reading and mathematics to the assessment of children's oral language proficiency. Although there has been a concerted effort to educate children in their first language or mother tongue, the reality is that many children in low-income countries are learning to read in a second or third language—even if this runs counter to a national policy that the language of instruction in the first few grades should be the children's first or home language. When children are receiving instruction in their first language, they often come to school with very limited oral language skills (Hart and Risely, 2003).

In these contexts, information on children's language proficiency in their first language and the language of instruction is essential for programme planning and to ensure that children are provided instruction that promotes learning success by supporting language acquisition in the language of instruction. The development of measures that provide educators with information on children's

language profile is just beginning. Processes for developing language proficiency measures have been developed in Latin America in Spanish and in indigenous languages in Guatemala (Rosales de Véliz et al., 2016). However, replication in other languages and contexts is needed. If possible, it would be desirable to assess children in all the languages that they speak. This type of multilingual assessment (if relevant) will provide a clearer picture of children's language skills.

QUALITY OF INSTRUCTION

Another area to consider when developing an assessment plan is the quality of instruction. Although collecting data on the fidelity of implementation is essential when introducing an intervention or a reform effort, more general descriptions of instruction are useful when interpreting assessment results.

There is a significant link between teaching quality and student achievement. If possible and when appropriate, teaching quality should be assessed. There are a number of dimensions on which to assess teacher quality. Some of these include the teacher's use of instructional time, instructional strategies, the materials used and the student grouping strategies (individual, small group, entire class). The Standards-based Classroom Observation Protocol for Educators in Literacy (SCOPE-Literacy) was developed to provide information on the quality of instruction. The measure has two dimensions: classroom structure as well as language and literacy instruction. Both areas contribute to quality instruction (Yoon et al, 2007). The results from the SCOPE-Literacy can be used to target professional development and improve instruction (Clark-Chiarelli and Louge, 2016).

CHILDREN'S HOME ENVIRONMENT

In developing contexts, understanding the literacy environment in the home can help explain reading achievement. Measuring the home literacy environment requires collecting and analysing data on: the value placed on reading and the drive to achieve; the availability of reading materials; the frequency of reading to and by children; and opportunities for verbal interaction (Hess and Holloway, 1984). Save the Children further claims that children's motivation and opportunities to read inside and outside both the home and the school should also be considered (Dowd and Friedlander, 2016). Including these elements will enable a better understanding and a broader evidence base that more appropriately represents the rich variety of learning environments in different languages, cultures, physical environments and overall living situations that exist throughout the world *(see **Box 2**).* Measuring the home literacy environment is done through a survey of family members and activities as well as follow-up questions on the motivation for reading and literacy use outside the home.

Save the Children's Home Literacy Environment (HLE) survey complements the data collected in schools by providing data on variables associated with academic success, such as opportunities for verbal interactions or the availability of reading materials at home (*see **Figure 2***). Data can be used to explain the differential effects of interventions which enable implementers (NGOs or ministries of education) to make decisions on how to improve or adapt their programmes for different groups of children. For example, if intervention results show that children who have fewer opportunities to read outside of school have lower reading scores, educators can consider ways to collaborate with other stakeholders to increase those children's opportunities to read both in and out of school.

Box 2. Complex environments

For many years, the international community has supported education in complex contexts albeit mainly through building and rehabilitating infrastructure and providing school supplies and teachers—interventions that improve access to education. However, surmounting the educational barriers found in politically complex environments requires significantly greater initiatives than simply improving access. In reality, the challenges in these environments include overlapping barriers to learning, such as poverty, conflict, gender inequality, low exposure to print, illiteracy of parents and food insecurity. Thus, new approaches to improve learning outcomes are much needed and learning assessments can provide a clearer picture to better inform reform efforts.

Experiences from Somalia and Afghanistan show that there is value in implementing oral reading assessments as a first (and difficult) step in a long process to improve learning. While the administration of the assessment may have been fraught with difficulties due to transport, weather and security problems, the countries have found the efforts to be worthwhile.

In complex situations, there will inevitably be some compromises. However, until assessments are being administered in these contexts, some of the poorest and most vulnerable children will inevitably be left behind. Thus, even small data sets may be useful. Compromises on sample sizes or on supervision of the test administration is acceptable in these cases—although the compromises should be 'principled' (i.e. the limitations they impose should be acknowledged and should be accounted for when the data are reported). When compromises are 'principled,' the claims that are made based on the data have to be explicitly cautious.

Source: adapted from (Shizad and Magee, 2016) and (Beattie and Hobbs, 2016)

Figure 2. HLE survey matrix

Name/initials	Relationship 1-Mom, 2=Dad, 3=Sister, 4=Brother, 5=Grandma, 6=Grandpa, 7=Other Female, 8=Other Male	Seen reading 1=YES, 0=NO	Told/helped you to study 1=YES, 0=NO	Read to you 1=YES, 0=NO	Told you a story 1=YES, 0=NO

Other than at school, did anyone outside your home read to you last week?	__No (0)	__Yes (1)
Other than school, did you read to anyone outside your home last week?	__No (0)	__Yes (1)
Other than at school, did you read alone last week?	__No (0)	__Yes (1)
In the last week, did you use your reading skills outside of school?	__No (0)	__Yes (1)
If yes, where? _____		__Yes (1)
In the last week, have you helped anyone using your reading skills?	__No (0)	__Yes (1)

RECOMMENDATION 3:

Emphasise the relevant skills—be conscious of differences in culture and orthography of the language

- All children should know the names of the letters (in alphabetic languages), be able to read words and pseudo words.
- Across languages, fluent reading contributes to reading comprehension. Yet, when assessing children, it is important to remember that the relative importance of speed and accuracy is dependent on the orthography and culture. Speed should not be pursued for its own sake.
- Comprehension is the ultimate goal of reading and it must be measured, even if done using a limited number of tasks.
- Most oral reading assessments are not designed to be comparable across countries or cultures. However, research shows that there are some skills and areas of development that can be compared.

Reading develops in similar ways across languages and settings. 'Brains are basically the same across languages. Orthographies are not' (Goswami, 2006). Although children require similar skills to become proficient readers, the skills that must be emphasised will vary according to the orthography of the language. The decision to measure specific skills depends on language, script, orthography and instructional methodologies.

ASSESSMENT COMPONENTS

Oral reading assessments generally include basic constructs, such as letter recognition, phonological awareness, word reading and pseudo-word reading as these are the foundations for pre-reading skills and higher order skills (i.e. vocabulary, oral reading fluency, comprehension, etc.). Oral reading assessments that are appropriate for early grades and that vary in the constructs they assess have been developed and implemented in more than 100 languages.

1. Phonological awareness

When determining which constructs to include, consider the role they play in the target language. Certain skills are important precursors to the development of reading. These skills are alphabet knowledge and phonological awareness. Phonological awareness is the ability to hear the sounds within words and to manipulate these sounds. An example of a phonological awareness task is asking a child to say the word 'book' without the /b/ sound and then asking the child to say the sound of the missing letter rather than the name of the letter. Phonological awareness skills will continue to develop as children learn to read.

2. Word reading

Another important precursor to the development of reading skills is the knowledge of letters and their sounds. Across languages, children have to recognise the symbols or graphemes used in the language and to link sounds to the graphemes. This is the basis for decoding words. Although children can learn to decode once they know a few graphemes and their sounds, the total number of symbols that children have to learn will affect how long it takes them to become proficient readers. When developing assessments, select the graphemes that are grade appropriate.

One of the most fundamental skills that should be measured is reading single isolated words when there are no text clues to the meaning of the word. Children need repeated exposure to words and text to develop automaticity in reading. The goal is for all word reading to eventually become automatic. As children get older, and the words they are expected to read are also longer or more morphologically complex, grouping larger units together is a more efficient way to read. Therefore,

the ages of the children who are being assessed informs which types of words should be included in an assessment. Automatic word recognition is necessary for fluent passage reading. Oral assessments of reading should be based, at least in part, on phonics (i.e. knowing the sounds of the letters). Skills in phonics can be assessed by the reading of non-words (technically called pseudo words) that are pronounceable combinations of letters or characters. This task tests the ability to decode print into the sounds of the language. This phonological processing ability is important for decoding new words and names that have not been previously encountered.

3. Fluency

Fluency refers to reading accurately with adequate speed and prosody. Across languages, fluent reading contributes to reading comprehension. Yet, when assessing students, it is important to remember that the relative importance of speed and accuracy is dependent on the orthography of the language. In transparent orthographies, speed is a more important indicator of reading skill but in opaque orthographies, accuracy is a better indicator. This is because in opaque orthographies, the sound/grapheme relationships are less predictable and reading words incorrectly can have an impact on comprehension. One challenging issue is whether or not to measure the speed of reading. Although the ability to read quickly enough to process and store information is important, reading speed can be difficult to measure with any precision in the field. Oral reading fluency is important and often assessed but finding the proper metric that would be comparable across languages is difficult.

4. Comprehension

Comprehension is the most complex of the reading skills and represents the ultimate goal in reading. To comprehend what is read requires word-level skills; vocabulary knowledge; oral language skills; reading with a modicum of fluency; broad conceptual knowledge, thinking and reasoning skills; and specific reading comprehension strategies. There are several reasons why comprehension is difficult to assess well when using a brief measure. In some cases, children may not have the basic reading skills needed to make measuring comprehension feasible. In other instances, there may not be enough items to sufficiently assess comprehension. Although oral reading fluency is correlated to reading comprehension, the correlation can be influenced by factors, such as students reading in a second language in which they are not yet proficient.

Assessments that include a variety of tasks may provide better—but not perfect—clues as to why children have trouble with comprehension. For example, if children clearly cannot recognise common words in their language (as assessed by a word-list task), they will have trouble comprehending when it comes to reading a passage. If they cannot decode pseudo words, for example, they may have trouble processing an unfamiliar word in a passage even if they know the word orally, which in turn slows down comprehension. More complicated comprehension problems, such as the child not having an explicit strategy for comprehending, or not being accustomed to dialoguing around the meaning of text (or even the meaning of a passage read orally to them), are harder to assess even with oral assessments that have quite a few tasks. Overall, more effort needs to be dedicated to measuring reading comprehension in the early grades.

Finally, it is important to note that in many situations having additional reading comprehension questions may not be practical. UNICEF and Save the Children are exploring adding a few comprehension questions to the Multiple Indicator Cluster Survey (MICS), which would provide another measure of early grade reading among children aged 7-14 years in developing countries around the world (Cardoso and Dowd, 2016).

COMPARABILITY OF ORAL READING ASSESSMENTS

Although most oral reading assessments measure the same reading constructs, they are not

© Dana Schmidt/The William and Flora Hewlett Foundation

necessarily comparable across countries and languages. In fact, differences in language structure and complexity make direct comparison of the results impractical, particularly direct comparisons of fluency. Comparing subtask results across countries and languages is therefore not advised—although it is possible to compare the percentages of children obtaining zero scores on specific tasks across languages and countries (Gove and Wetterberg, 2011). Although the inability to complete a task *at all* would not be affected by language structure and complexity, contextual factors such as exposure to print can lead to differences in zero scores.

Assessment results can, however, be used for *indirect* comparisons. For example, by comparing percentages of children reaching locally established benchmarks as opposed to percentages of children who reach a predetermined specific or international benchmark. The use of locally established benchmarks may provide countries and development partners with common ways to measure and discuss progress towards the SDGs related to learning outcomes. Hsieh and Jeng (2016) explain how the government of The Gambia has been monitoring the progress of early grade reading using nationally-set benchmarks for reading in the language of instruction as well as other national languages. Benchmarks should be based on evidence from assessments that demonstrates that the levels of certain skills (or performance on certain metrics) are valid.

Even though most oral reading assessments are not designed to be comparable across countries or cultures, research shows that there are some skills and areas of development that can be compared. These ideas are being put into practice within the International Performance Indicators in Primary School (iPIPS) project, a cross-national oral reading assessment that captures non-cognitive development as well as cognitive skills (Merrell and Tymms, 2016).

RECOMMENDATION 4:

Properly organize the implementation of activities—logistics and monitoring

- ➜ Whether conducted in schools or households, common actions to maximise responses and assure data quality include: engaging a solid team to collect the data, providing adequate training and measuring inter-rater reliability aiming for minimum acceptable levels; providing clear instructions to assessors and properly documenting any actions taken; notifying target schools or villages/households prior to the assessment; and gaining permission to assess the child ahead of time.
- ➜ Timely and consistent monitoring allows the teams to make adjustments during the fieldwork as it may be too late to fix issues that are identified only after the data collection has been completed.

Properly organized implementation of field activities cannot be underestimated. Organizing the implementation of the assessment includes logistics; monitoring the implementation and its progression; and steps to be taken after the assessment has been conducted.

CONDUCTING THE ASSESSMENT: LOGISTICS

Although logistics vary from country to country, there are some common practices that can maximise responses to the assessment, ensure a successful data collection and assure data quality. These include:

1. Engaging a knowledgeable team to collect the data

The individuals who conduct the field operations play a critical role and ultimately, the validity of the data will rely on the quality of their performance. These individuals generally include but are not limited to assessors, supervisors, scorers, data entry staff, drivers and other logistic support staff. They should be trained, have clearly defined roles and should not be expected to perform the impossible. The number of individuals performing the different tasks will vary on a case-by-case basis.

2. Notifying the schools or the villages/households of the assessment

Implementers should contact the sampled schools to confirm their location, that they have pupils enrolled in the grade to be assessed and that the language of instruction matches that of the assessment (Kochetkova and Dubeck, 2016). For assessments to be conducted in the home, common practices include announcing the household visit at the schools visited so that children are aware and wait for the team of assessors to arrive as well as asking the village leaders to inform households of the arrival of assessors prior to the visits (Mugo et al., 2016).

3. Taking into account the weather and terrain conditions

The weather conditions at the time of year that data collection will take place could impact fieldwork. For example, in some countries, the end of the school year may correspond with the rainy season or worsening road conditions and could potentially having an impact on school attendance or operation in certain areas (Kochetkova and Dubeck, 2016).

4. Ensuring clear instructions for actions to be taken when a sample school or child cannot engage in the assessment and carefully documenting and justifying all replacements

Replacement schools should be selected based on their similarity to the originally sampled school, such as location, type (public or private), enrolment, etc. Sampled schools that are located in difficult-to-reach areas should not be replaced simply for convenience—although in some cases, a sampled school will be totally unreachable by assessment teams due to weather or road conditions and will have to be replaced (Kochetkova and Dubeck, 2016).

TABLE **1**

Possible solutions to common challenges encountered in household-based assessments

Challenge	Possible solutions
Parents are unhappy and reprimand children because they cannot read.	▪ Assure parents that with time, children improve if they receive the required support. ▪ Work in pairs so one person can engage the parent in discussion away from the child while the other assesses the child.
Children fear reading in the presence of parents and other people; neighbours and passers-by disrupt the assessment.	▪ Politely request those present to give assessors time with the child alone and tell them that the results will be explained after the assessment. ▪ Ask to take the child away from the crowd for testing with the permission of the child's relatives. ▪ Train volunteers to address disruptions.
Missing the assessment of many children because they cannot be found at home at the time of the visit.	▪ Make callbacks later in the day or the following day. ▪ Announce household visits in the schools so that children are aware and wait for the team of assessors. ▪ Ask village leaders to inform households of the assessors' visits.
Households do not authorise assessing their children.	▪ Take the time to introduce the assessment and the work, and seek consent. ▪ Use village leaders to inform the village prior to visits and if possible, also walk with the assessors during the assessment. ▪ Opt to use volunteers from the community.
Teachers are unaware of the learning issues captured by the assessment.	▪ Share results during education days and visit some schools. ▪ Visit the government schools in the sampled villages to present the purpose of the survey and discuss the previous year's findings with teachers.

Source: adapted from (Mugo et al., 2016)

5. Gaining permission to assess the students/ children

For school assessments, gain the permission and trust of the school personnel, including the administration and the teachers. Negotiating a time to administer the assessments and respecting the wishes of the school is important. In many assessments, the explicit consent of each individual child is sought. In household based assessments, assessors must take the time to introduce the assessment to the household and seek consent from the parents and the children before proceeding (Mugo et al., 2016).

Finally, assessors administering a household-based assessment face a series of challenges during the data collection processes. **Table 1** offers practical solutions to common challenges faced during the administration of citizen-led assessments. This table is based on perspectives from the Annual Status of Education Report-India (ASER-India), ASER-Pakistan, Beekunko, Jàngandoo and Uwezo assessments.

QUALITY CONTROL AND MONITORING

Monitoring entails conducting a series of quality checks to ensure that the data collection is progressing according to plan and that the assessors are administering the assessment in line with the guidelines provided. Monitoring assessor performance throughout the data collection process allows for timely intervention or retraining, which otherwise could go unnoticed until the end of the data collection process. Collecting data via electronic means can also help facilitate the early detection of problems.

In India, the ASER monitoring is done at two levels: one of the assessors by the supervisors or master trainers and the other of the master trainers by the state team (Banerji, 2016). It is also desirable—although not always possible—to measure inter-rater reliability during the fieldwork. Inter-rater reliability requires assessors to pair up to assess one of the selected children together each day. One interacts with the child while the other observes and marks the responses (Kochetkova and Dubeck, 2016). If this proves too expensive, inter-rater reliability can be assessed in training sessions where the trainer makes purposeful mistakes to see how the assessors perform and repeats the exercise until the rate of agreement among all assessors reaches a high percentage. The raters are considered reliable when the scores are the same or very close.

RECOMMENDATION 5:

Cater the analysis and communication of results to the target audience

- ➜ Report on the assessment results in a way that can be easily understood by a wide range of audiences, including non-experts. Descriptive statistics are a good technique for generating strong, easily grasped, readily communicable policy messages. However, results must be interpreted and reported with caution and should respect basic statistical principles.
- ➜ Determine the dissemination of activities at the country level based on two main factors: the purpose of the assessment and the audience. Consider a range of products to communicate the assessment results and use the appropriate language for dissemination, depending on the audience.
- ➜ Contributing to the knowledge base by sharing experiences through international platforms is encouraged; however, using the data generated to serve the country's own purposes and interventions is more important.
- ➜ Media campaigns are not to blame and shame. They should be used to publicise recommendations and strategies for how the system as well as children, parents and teachers can improve learning and disseminate key policy messages.

Analysing and interpreting the results is a crucial part of the assessment process. Additionally, presenting and communicating data from oral reading assessments to the right stakeholders is necessary to enable their use to inform decisions and design targeted interventions to improve reading. In a nutshell, analyses must be conducted and the results reported for different types of users—from policymakers to teachers looking to reinforce their pedagogical approaches and parents who want to work with their children to improve their learning. It is important to use the appropriate language for dissemination, depending on the audience.

ANALYSIS AND INTERPRETATION OF RESULTS

Measurement methods used to analyse and interpret the data depend on the test design of the assessment and the structure of the resulting data. In most oral reading assessments, performance results are analysed using descriptive statistics. This is possible since all children assessed are presented with an identical set of items. The use of descriptive statistics to analyse performance on an assessment typically entails reporting the percentage correct on a set of items. It therefore summarises the results in a format that is easily understood by a wide range of audiences and can be a good technique for generating strong, easily communicable policy messages that can be readily grasped by other education stakeholders and non-experts. Common practices for analysing and reporting results include:

- Disaggregating results by specific characteristics of interest, such as grade, gender and geographical location. However, the number and levels of disaggregation are dependent on the data that has been collected.
- Comparing results against benchmarks that have been set either locally or internationally.
- Reporting results in a way that can easily be communicated and send a strong policy message. Some of the most common ways of reporting the data and the easiest to understand are:
  - mean scores by grade and subtask;
  - percentage of students in a given grade who can read the comprehension passage (within the allotted time, if timed) and correctly answer most or all reading comprehension questions;
  - percentage of students who scored zero by grade and region.
- Understanding how a range of variables interact. This has important policy consequences.

For example, if factors, such as low teacher absenteeism and principal management of student progress characterise good schools, then steps need to be taken to provide these opportunities to schools where the majority of pupils may be performing below the national benchmark. This may entail a combination of actions, such as providing resources and funding but also supporting schools that are accountable and well-managed.

- When possible, consider analysing the associations between different assessment results that are conducted on the same population of interest. For example, Banu Vagh (2016) examines the associations between ASER and EGRA conducted in India, to evaluate the validity of the ASER test. Although the two assessments are administered independently and have some differences, they are comparable in content as they are designed to assess the same abilities or skills.

It is important to remember to respect the basic principles of reporting data and making inferences about the results. For example, when comparing two groups, it is important to include significance tests along with the descriptive statistics (e.g. means and standard deviations) as this is needed to determine whether one can infer a statistically-significant difference between the two groups.

Although analysing data using descriptive statistics can yield valuable results, there are difficulties in comparing students from different grades (i.e. ceiling or floor effects), comparing student abilities over time, and describing the skills of students at specific ability levels. Thus, in some cases, it may be more appropriate to develop a literacy scale—one that describes skills on a continuum and can be used to compare children of different ages or grades or compare children's abilities over time. However, to construct such a scale, a more complicated technique like item response theory (IRT) must be used. An example of using IRT to construct a literacy scale is detailed by Meiers and Mendelovits (2016).

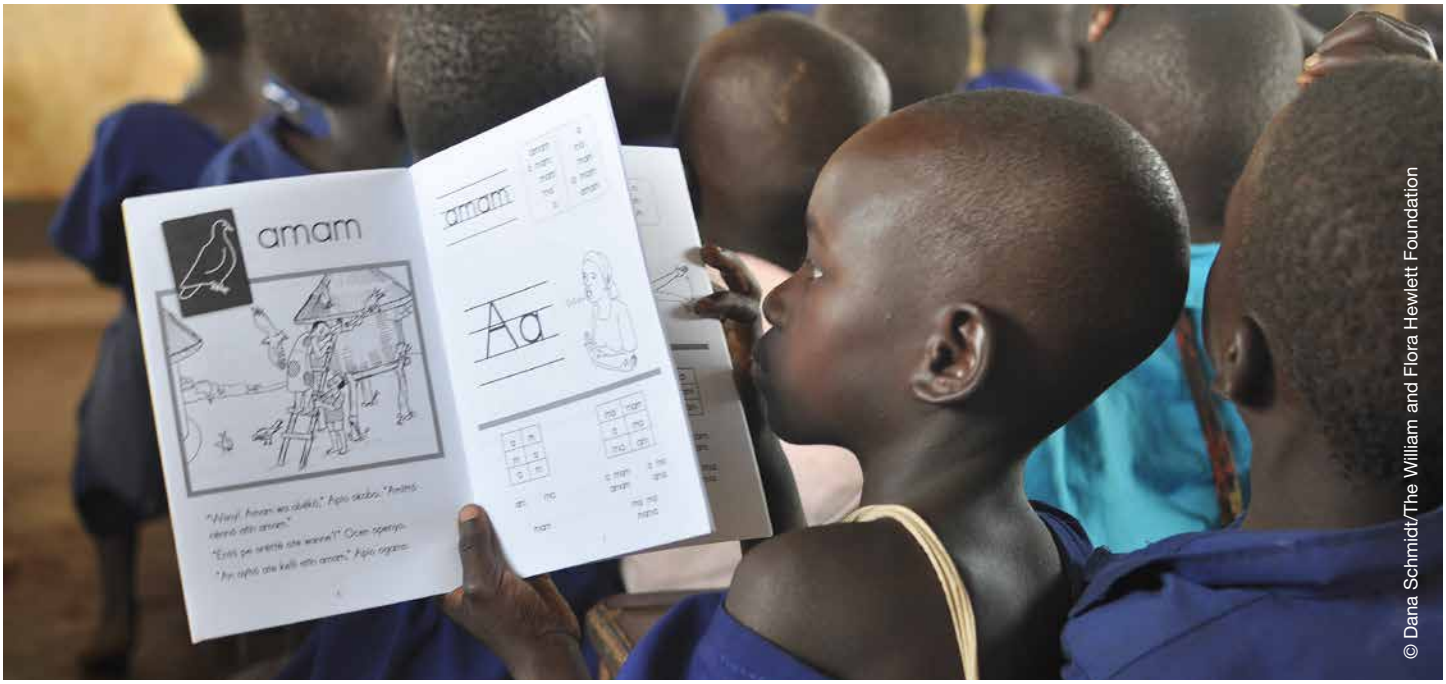
COMMUNICATIONS MATERIALS AND DISSEMINATION STRATEGIES

Once the data have been analysed and interpreted, there is a range of products and formats in which the information could be communicated to the various stakeholders. These include but are not limited to mass media campaigns, policy briefs, data visualisations, short infographics, a national report, dissemination meetings, workshops, journal articles and conference presentations.

In general, the dissemination activities at the country level will be determined by two main factors: the purpose of the assessment and the audience. If, for example, the purpose of the assessment is to serve as a national or system-level diagnostic to design a policy reform, an intervention or a programme, then the audience of interest could be a ministry of education, donors, civil society, community leaders, academics, practitioners and teacher unions. Different activities can be undertaken with different groups, such as policy dialogue workshops, curriculum- or standard-review workshops, social mobilisation or mass media campaigns, project design workshops, policy briefs, press releases, journal articles and conference presentations. Even if the purpose of the assessment was to generate discussion at the national level and to spur ministries into action, the reporting of the results to schools and teachers can complement this promotion of awareness (RTI International, 2009). In India, the ASER Centre prepares a series of slides, presentations and notes for each state. State report cards are printed in a two- or four-page format for large scale distribution at different levels (Banerji, 2016). The media uses these documents to communicate the key findings by state to a wide audience.

INTERNATIONAL PLATFORMS

It is advised to make assessment reports publically available in order to help broaden the knowledge base of experiences in the development and application of oral reading assessments as well as their use. The following international platforms provide a wealth of information to practitioners,



© Dana Schmidt/The William and Flora Hewlett Foundation

international development agencies, governments, teachers associations, academics, civil society organizations, donor organizations, UN agencies, and other stakeholders:

- The United States Agency for International Development (USAID) EdData II website, developed to share experiences and reach a broad range of audiences at the international level.

- The UIS Catalogue and Database of Learning Assessments, compiles information on learning assessments which is organized in succinct formats to shed light on key characteristics of large-scale assessments and public examinations conducted in developing countries.

- The World Bank EdStats (Education Statistics) portal, a data and analysis source for key topics in education. It holds data from various sources which can be accessed through pre-defined data queries.

- The People's Action for Learning (PAL) Network bring together the countries working around the world to assess the basic reading and numeracy competencies of all children, in their homes, through annual citizen-led assessments. The PAL Network website provides relevant resources from the citizen-led assessments, including research reports and assessment tools.

- The Global Reading Network brings together professionals from all over the world to improve children's reading. Resources and evidence-based practices for improving reading skills can be accessed on the Global Reading Network website.

THE MEDIA

How important is the media in disseminating the assessment results? Silvia Montoya, UIS Director, who formerly led several learning assessment initiatives in her native Argentina notes: 'media reports about learning assessment data make me cringe.' She explains in a blog post that the media should not be used to highlight 'bad grades', but rather to publicize recommendations and strategies for how children, parents and teachers can improve learning (Montoya, 2015). This type of media campaign was applied in Yemen and focused on encouraging parents and communities to support children's reading. Mobile service providers supported the campaign by broadcasting key messages to the 9 million subscribers (Creative Associates, 2016).

RECOMMENDATION 6:

Use the data to raise awareness and design interventions aimed at improving teaching and learning

- ➜ Stand-alone analyses may not be sufficient. Common practices to ensure that the data and the recommendations produced are considered for further action include: regular communication to promote community involvement, inviting change agents to participate in the data collection, encouraging local ownership, and consistent and regular assessment.
- ➜ In the process of improving reading skills, time, space and the mechanism must be given to stakeholders to engage with the assessment results. Change will not be immediate—improvements in reading achievements require time, resources and dedicated partners.
- ➜ Improving reading skills can be accomplished through the implementation of various programmes. Practices such as national campaigns as well as inciting parents, teachers and peers to improve teaching and learning have been successful. These practices can be applied in different contexts and settings.
- ➜ A programme design must be informed by the data.

Data from learning assessments can be packaged to serve various audiences, including policymakers and governments, teachers and parents. However, stand-alone analyses may not be sufficient.

PRODUCING LOCALLY OWNED DATA

Some common practices are encouraged to ensure that the data produced are owned by the community and the recommendations are considered for further action:

1. Regular communication with and involvement of the community

Communication should be regular with the different stakeholders throughout the different stages of the assessments. Parents, teachers, school personnel and ministry (government) officials need to be involved in the assessment strategy. The results of the assessment should be shared with parents, teachers and government officials. Teachers and parents should be provided with information on ways to improve reading skills.

2. Local ownership

Local ownership and participation is necessary to build awareness, improve accountability and initiate action towards improving elementary education. Banerji (2016) stresses the importance of local ownership as an important element in the overall architecture of the ASER India. From inception, a key component of the ASER process was to involve local organizations and institutions. The local partners are involved in data collection as well as the dissemination of the results.

3. Consistent and regular assessment

Annual or regular cycles of assessment create a natural pulse of repetition where findings are regularly shared. This builds familiarity with the assessment among national policymakers, civil society organizations and advocacy groups and draws attention to the findings from year to year (Aslam et al., 2016).

IMPROVING READING SKILLS

Most of the organizations that participate in oral assessments emphasise that their purpose in assessing is to encourage the creation of programmes or interventions to improve reading skills and ensure that, through measurement, their chances of success are increased. Whether conducted in schools or households, data from oral reading assessments (i.e. learning achievement

© ASER 2015

data and the accompanying relevant contextual information) can and have been used to design strategies aimed at improving reading skills. The following are examples of good practices for using data from learning assessments to improve reading skills. These practices can be adapted to various settings, although feasibility and cost will differ depending on the context.

1. Reading campaigns

The collaboration among the different stakeholders that use the resulting assessment data to improve reading is just as important as the relationship between the partners during the planning and implementation of the assessment. The *Vamos a Leer, leer es divertido* ('Let's Read, reading is fun') is an ongoing campaign launched in 2010 in Nicaragua. It represents a collaborative effort between government, civil society, NGOs and private organizations to improve literacy in Grade 1 children. Their joint effort has created a culture of assessing reading to spark improvement and has led to an increased number of libraries in schools (following the revelation that there was a shortage of reading materials in schools), helped teachers involve parents in reading and telling stories to their children, resulted in the development of several teacher training programmes, and, most importantly, demonstrated an improvement in early grade reading skills (Castro Cardenal, 2016).

2. Teachers

Although teachers could use the EGRA (or an adaptation of the assessment) and other multi-task assessments in their entirety, this is generally not recommended. More commonly, selected tasks are used as a type of formative assessment to monitor classroom progress, determine trends in performance and adapt instruction to meet children's instructional needs (Dubeck et al., 2016). Oral reading assessments designed to be conducted by teachers have had some positive effects on reading—they have provided important and useful insights into the progress and achievement of

students and has helped teachers adapt teaching/learning strategies to improve instruction. Rosales de Véliz et al. (2016) show how linguistic profiles have been developed and used in Guatemala to help teachers instruct reading in Spanish and other mother tongue languages. Meyer and Mendelovitz (2016) show how longitudinal assessments that yield individually-reported results can provide teachers with a sound basis for planning future teaching strategies to meet the needs of their students.

Citizen-led initiatives—although conducted in households—have also involved teachers in the assessment process. These assessments have used teachers as assessors, enabling them to observe children's weaknesses in the different learning processes and adopt counter measures in the classroom. In Senegal, in response to very low learning levels in Arabic-medium schools, Jángandoo has worked with school supervisors and the decentralised school authorities in one region to develop, test and introduce remedial education guides designed to provide teachers with new instructional approaches and materials (Ba *et al*., 2016).

3. Parents

Parents can play a valuable role in a literacy programme. Where possible, parents should receive training to help their children develop literacy skills. For example, the Yemen Early Grade Reading Approach (YEGRA) programme trained more than 23,000 parents on ways to support their children's reading at home and to prepare children to attend school regularly and on time. The positive gains observed from the programme included improved reading skills in children *(see **Box 3**).* Parents reported that their children were reading or were being read to more at home and were more reluctant to miss school. The programme even influenced illiterate parents to learn how to read with their children (du Plessis et al., 2016). In addition, citizen-led assessments conducted in East Africa have helped shift the thinking of parents from assuming that learning is the sole responsibility of schools and teachers. It has helped raise awareness that parents have a major role to play in their children's academic education. Citizen-led assessments have made a considerable effort to ensure that parents act on the advice provided by the assessors on

Box 3. Response to assessment results and findings

Programmes designed to improve reading skills generally involve various interventions, targeting different 'change agents,' including parents and teachers. The YEGRA programme was designed using the information from two oral reading assessments, EGRA and Literacy Boost.

EGRA and Literacy Boost assessment findings	YEGRA programme responses
Children who have regular attendance do better in reading.	The national media campaign and parent training messages included this statement: 'Getting your children prepared for school in the morning and on time everyday helps student learning.'
Children who practice reading more, do better in reading.	All children have individual daily in-class reading.
Children who are read to at home or have books in the home perform better than those who don't.	Training for parents in making the home a print rich environment, reading to children at home and ensuring they have opportunities to read outside the home (i.e. at mosques, libraries, shops and other places with public texts).
Regular corrective feedback to students is correlated with increased early grade reading scores.	Five assessments included in the teacher's guide. One assessment administered after approximately every 20 lessons.
Student's phonological awareness in Modern Standard Arabic is weak likely leading to poor uptake of letter sound recognition.	Teacher guides include focus on phonemic awareness with daily interactive practice for students.

Source: adapted from (du Plessis et al., 2016)

© Dana Schmidt/The William and Flora Hewlett Foundation

the importance of reading. For example, Uwezo assessors presented families with a calendar that included written suggestions of what parents can do to improve their children's learning (e.g. 'encourage your child to read at home') (Nakabugo, 2016).

4. Peers

Peers, especially those with more advanced skills, can be valuable allies in helping children develop literacy skills. The use of structured peer work has been researched extensively in developed countries (e.g. Dowhowser, 1989; Fuchs et al., 1997) and has been included in reading programmes in developing countries. Peer work can be used to practice a number of literacy skills in the classroom. Outside of school, children in some cultures like to play school, which can help them learn. Although not specifically recommended by the authors, the role of peers is an avenue to explore and a good practice to consider for future interventions or programmes aimed at improving reading skills.

5. Reading materials

Data from oral reading assessments have also provided valuable information on developing reading materials to help improve reading skills. For example, the assessment in Guatemala has helped the development of various educational materials, such as *El tesoro de la lectura* ('The treasure in reading') series that addresses topics, such as emergent reading, reading development stages, associated skills and reading comprehension. The materials were further distributed to classrooms throughout the country and are expected to have a positive impact on learning to read (del Valle Catalán, 2016).

© Dana Schmidt/The William and Flora Hewlett Foundation

Conclusion

To achieve the SDG for education, governments will need more and better data to develop the evidence base needed to identify and effectively address weaknesses while monitoring progress. Early detection of learning gaps and support will be essential to inform remedial action and help realise the ambitious new global education goal of providing every child with a quality education and the foundational skills needed for a productive and fulfilling life. The key advantages that oral assessments provide are:

Timely access to data to inform decision making. Additional background information is often collected to support the results of the assessment and provide evidence to inform policy. The additional information collected depends on specific policy interests.

Early detection of reading weaknesses. Detecting learning problems early—particularly in reading as this will inevitably affect all other learning processes—allows for remedial action to be taken well before the end of primary education when it is often too late.

Viable solutions to measure the reading skills of children who are beginning to read. The assessments help capture the reading skills of children who have not yet mastered the necessary skills to take traditional written tests due to limited mechanical or decoding skills as well as comprehension and/or writing skills.

Means to assess learning for countries that do not participate in cross-national initiatives. The tools to conduct oral reading assessments are mostly open source, which allows practitioners to conduct an assessment at any time without having to wait for the next cycle of cross-national assessments to become available (Gove, 2015). This also holds true for countries that do not participate in cross-national initiatives. However, oral reading assessments (unlike cross-national assessments) are not designed to be comparable across countries and especially across languages. This allows governments and their partners to conduct oral reading assessments at their discretion and without fear of being ranked or compared against other countries. Nevertheless, the accessible open-source availability of tools to conduct oral assessments does present a danger that an organization could apply the assessment carelessly and come to the wrong conclusions. The recommendations presented here have been produced to help address that concern. Along with the present document, many of the open sources for information on oral reading assessments contain detailed guidance that can help ensure that quality data are produced.

There is strong and systematic support from donors for countries measuring oral reading skills as a gateway to improved programmes and policies, stronger advocacy and better use of resources to improve learning outcomes. Further development of the generation and use of data from oral reading assessments must be encouraged through increased dialogue among implementers and practitioners. This will lead to a better understanding of what works and why, within and across countries.

REFERENCES

The primary references for Chapter 5 of the this report are the articles that are published in the ebook *Understanding What Works in Oral Reading Assessments*. The following is the list of articles:

Aslam, M., Saeed, S., Scheid, P. and Schmidt, D. (2016). "Expanding citizen voice in education systems accountability: Evidence from the citizen-led learning assessments movement".

Ba, D., Bèye, M., Bousso, S., Mbodj, A. A., Sall, B. A. and Niang, D. (2016). "Evaluating reading skills in the household: Insights from the Jàngandoo Barometer".

Banerji, R. (2016). "Annual Status of Education Report (ASER) assessment in India: Fast, rigorous and frugal".

Banu Vagh, S. (2016). "Is simple, quick and cost-effective also valid? Evaluating the ASER Hindi reading assessment in India".

Beattie, K. and Hobbs, J. (2016). "Conducting an Early Grade Reading Assessment in a complex conflict environment: Is it worth it".

Cardoso, M. and Dowd, A.J. (2016). "Using Literacy Boost to inform a global, household-based measure of children's reading skills".

Clark-Chiarelli, N. and Louge, N. (2016). "Teacher quality as a mediator of student achievement".

Castro Cardenal, V. (2016). "Use of literacy assessment results to improve reading comprehension in Nicaragua's national reading campaign".

del Valle Catalán, M. J. (2016). "Assessing reading in the early grades in Guatemala".

Dowd, A. J. and Friedlander, E. W. (2016). "Home literacy environment data facilitate all children reading".

Dowd, A. J., Pisani, L. and Borisova, I. (2016). "Evaluating early learning from age 3 years to Grade 3".

du Plessis, J., Tietjen, K. and El-Ashry, F. (2016). "The Yemen Early Grade Reading Approach: Striving for national refor"..

Dubeck, M. M., Gove, A. and Alexander, K. (2016). "School-based assessments: What and how to assess reading".

Hsieh, P, J. and Jeng, M. (2016). "Learning-by-doing: The Early Literacy in National Language Programme in The Gambia".

Kinyanjui, J. (2016). "Utility of the Early Grade Reading Assessment in Maa to monitor basic reading skills: A case study of Opportunity Schools in Kenya".

Kochetkova, E. and Dubeck, M. M. (2016). "Assessment in schools".

Meiers, M. and Mendelovits, J. (2016). "A longitudinal study of literacy development in the early years of school".

Merrell, C. and Tymms, P. (2016). "Assessing young children: Problems and solutions".

Mugo, J. K., Kipruto, I. J., Nakhone, L. N. and Bobde, S. (2016). "Assessing children in the household: Experiences from five citizen-led assessments".

Nakabugo, M. G. (2016). "What and how to assess reading using household-based, citizen-led assessments: Insights from the Uwezo annual learning assessment".

Rosales de Véliz, L., Morales Sierra, A. L., Perdomo, C. and Rubio, F. (2016). "USAID Lifelong Learning Project: The Linguistic Profile assessment".

Shirzad, H. and Magee, A. (2016). "Administering an EGRA in a post- and an on-going conflict Afghanistan: Challenges and opportunities".

ADDITIONAL REFERENCES

Fenton, R. (1996). "Performance assessment system development". *Alaska Educational Research Journal*. Vol. 2, No. 1, pp. 13-22.

Chesterfield, R. and Abreu-Combs (2011). *Centers for Excellence in Teacher Training (CETT): Two-Year Impact Study Report (2008-2009)*. Washington DC: USAID Bureau for Latin America and the Caribbean. http://pdf.usaid.gov/pdf_docs/PDACS248.pdf

Creative Associates (2016). http://www.creativeassociatesinternational.com/past-projects/yemen-early-grade-reading-approach/ (Accessed January 2016).

Dubeck, M. M. and Gove, A. (2014). *The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations*. Research Triangle Park, NC: RTI International. http://ac.els-cdn.com/S0738059314001126/1-s2.0-S0738059314001126-main.pdf?_tid=32ed8cfa-c9e3-11e5-a55a-00000aacb35d&acdnat=1454441515_8f5628dd667cfe473ba82558578c223d

Dowhower, S. L. (1989). "Repeated reading: Research into practice". *The Reading Teacher*, Vol 42, pp. 502-507.

Fuchs, D., Fuchs, L.S., Mathes, P.G. and Simmons, D.C. (l997). "Peer-Assisted Learning Strategies: Making classrooms more responsive to diversity". *American Educational Research Journal*, Vol. 34, pp. 174-206.

Hart, B. and Risely, T.R. (2003). "The Early Catastrophe: The 30 Million Word Gap by Age 3". *American Educator*, Spring 2003 pp. 4-9.

Global Reading Network. https://www.globalreadingnetwork.net/

Goswami, U. (2006). "Neuroscience and education: From research to practice". *Nature Reviews Neuroscience*, Vol. 7, No. 5, pp. 406-413.

Gove, A. and Wetterberg, A. (eds.) (2011). *The Early Grade Reading Assessment: Applications and interventions to improve basic literacy*. Research Triangle Park, NC: RTI International.

Hess, R. D. and Holloway, S. D. (1984). "Family and School as Educational Institutions". *Review of Child Development Research, 7*, 179–222.

Montoya, S. (2015). *Why media reports about learning assessment data make me cringe.* Global Education Monitoring Report: World Education Blog. https://efareport.wordpress.com/2015/06/17/why-media-reports-about-learning-assessment-data-make-me-cringe/

People's Action for Learning (PAL) Network. http://palnetwork.org/

Wagner, D.A., Babson, A. and Murphy, K. M. (2011). "How Much is Learning Measurement Worth? Assessment Costs in Low-Income Countries". *Current Issues in Comparative Education*. Vol. 14: pp. 3-23.

Research Triangle Institute International (2009). *Early Grade Reading Assessment toolkit.* USAID Education Data for Decision Making (EdData II). Washington, D.C.: USAID.

Research Triangle Institute International (2014). *EdData II: Egypt Grade 3 Early Grade Reading—Pilot Group Assessment.* USAID Education Data for Decision Making (EdData II). Washington, D.C.: USAID. http://pdf.usaid.gov/pdf_docs/PA00K7GC.pdf

UNESCO Institute for Statistics Catalogue of Learning Assessments. http://www.uis.unesco.org/nada/en/index.php/catalogue/learning_assessments. (Accessed January 2016).

World Bank EdStats. http://datatopics.worldbank.org/education/

Yoon, K. S., Duncan, T., Lee, S. W. –Y., Scarloss, B., and Shapely, K. (2007). *Revewing the evidence on how teacher professional development affects student achievement.* Washington, DC: U.S. Department of Education, Institute of Educational Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf

Glossary

Accuracy. Ability to perform a skill, such as reading letters or words, correctly.

Additive bilingualism. Occurs in a setting in which the first language and culture are enhanced, such as in dual language programmes. Therefore, the first language and culture are not replaced by the addition of a second language and culture.

Alphabetic principle. Understanding that letters in written words represent sounds in spoken words.

Analogy. An approach to teaching decoding by analyzing letter-sound patterns in previously learned words to read novel words. For example, using the */ight/* in 'right' and 'light' to read the new words 'might' and 'sight'.

Analytic phonics. Teaching letter-sound correspondences through the analysis of words that share phonemes. For example, examining the words 'cat', 'car' and 'can' to learn the /c/ phoneme.

Automaticity. Quick and accurate recognition of letters, sounds and words without hesitation.

Blending. Combining individual sounds or word parts to form other word parts or full words either orally or in print. For example, the speech sounds */c/, /a/ and /t/* = 'cat'. Likewise, the printed word 'cat' can be read by combining the sounds */c/, /a/ and /t/.*

Census-based assessment (or examination). An assessment administered to the whole population of students enrolled at the target grade(s) or belonging to the target age range(s).

Citizen-led assessments. Assessments that are administered by the citizens rather than governments to measure whether or not children have mastered the fundamental building blocks of learning. They are administered in households and target children of primary and lower secondary school age.

Classical test theory (CTT). A measurement theory that consists of a set of assumptions about the relationships between actual or observed test scores and the factors that affect the scores. It is used for measuring and managing test and item performance data.

Coefficient alpha. A measure of internal consistency or how closely related a set of items are as a group. It is considered to be a measure of scale reliability. The coefficient ranges between zero and one. A reliability coefficient of .70 or higher is considered 'acceptable' in most social science research. Coefficient alpha is also sometimes referred to as Cronbach's alpha.

Comprehension. Ability to understand and derive meaning from spoken and written language.

Concurrent validity. The extent to which the results of a specific assessment corresponds to those of an established assessment of the same construct.

Consonant blend. Two or more consonant letters that retain their distinct sounds when read. For example 'mw', 'fl' or 'st.'.

Consonant digraph. Two consonant letters that represent a single sound when the word is pronounced. For example, *th* in 'thin' or *sh* in 'shoe'.

Consonant letters and sounds. All letters and their corresponding sounds that are not vowels (i.e. b, c, d, f, g, h, j, k, l, m, n, p, q, r, s, t, v, w, x, y, z).

Construct validity. The degree to which a test measures what it claims to measure.

Criterion referenced tests. Assessments designed to measure student performance against a set of predetermined learning standards.

Decodable texts. Connected text in which most of the words are comprised of letter-sound correspondences that were previously taught.

Cross-national assessment. A terms used to represent learning assessments that are conducted in more than one country using the same procedures and yielding comparable results. The international assessments PIRLS, TIMSS and PISA and regional assessments LLECE, PILNA, PASEC, SACMEQ and SEA-PLM are cross-national assessments.

Decoding. Using knowledge of letter-sound relationships to read printed words. For example, converting the letters b, and n into the sounds */b/ /i/ /n/* to read the word 'bin'.

Descriptive statistics. Numbers used to summarise data. Examples include simple summaries on the sample and measures used.

Dialects. Regional language variations that may include differences in pronunciation, grammar and/ or vocabulary.

Dominant language. The language used with more proficiency by bilingual or multilingual individuals.

Expository text. Text that is designed to teach or explain a specific topic; also referred to as informational text.

Fluency. Ability to read text correctly, quickly and with expression.

Frustration reading level. Frustration level is often defined as text in which the reader is able to read less than 90% of the words accurately and as a result, comprehension is affected.

Genres. Text structures that are identified by unique sets of characteristics, such as science fiction, mystery and poetry.

Grade(s). A specific stage of instruction in initial education usually given during an academic year. Students in the same grade are usually of similar age. This is also referred to as a 'class', 'standard', 'cohort' or 'year'.

Grade-level text. Text that is appropriate for a specific grade level. The level is based on the time during the school year that a typical student in a specific grade possesses the word recognition, vocabulary and comprehension skills necessary to read a specific text independently.

Grapheme. The smallest unit of written language (e.g. letter or group of letters) representing the sounds in words. For example, the sound /ai / in 'rain' is represented by the written letters 'ai'.

Graphophonemic knowledge. Understanding that there is a relationship between letters and sounds.

Independent reading level. Text level in which the reader possesses sufficient word recognition and comprehension skills to read the text easily and fluently without assistance. The independent level is often defined as text in which the reader is able to read at least 95% of the words accurately.

Inferential comprehension. The ability to draw a conclusion from text based on what one knows or based on judgements drawn from the given information**.**

Instructional reading level. Text level in which the reader possesses sufficient word recognition and comprehension skills to read the text with few errors and only some assistance. The instructional level is often defined as text in which the reader is able to read between 90-94% of the words accurately.

International Standard Classification of Education (ISCED). A classification system that provides a framework for the comprehensive statistical description of national education systems. It also refers to a methodology that translates national educational programmes into internationally comparable levels of education. The basic unit of classification in the ISCED is the educational programme. The ISCED also classifies programmes by field of study, programme orientation and destination.

Inter-rater reliability. The degree of agreement among raters.

Irregular words. Words in which some or all of the letters do not represent their most common associated sounds.

Item response theory (IRT). A group of mathematical models used to relate and predict an individual's performance on a test item to his/her level of performance on a scale of the ability or trait being measured, and the item's characteristic parameters (e.g. guessing, discrimination and difficulty parameters).

L1. The native language or first language.

L2. The second language that a student is attempting to learn. For example, for English language learners, L2 is English.

Language acquisition. The development of language skills.

Language dominance. The measurement of the degree of bilingualism involving a comparison of the proficiencies between two or more languages.

Language proficiency. The degree to which one speaks, understands, reads or writes a language at native-like levels.

Large-scale assessment. A system-wide assessment that is designed to monitor changes and inform policy. It can be thought of in two broad categories: cross-national assessments (see definition) and national learning assessments (see definition).

Letter-sound correspondence. Association between a specific letter and its corresponding sound. For example, the letter 'm' and the sound */mmm/*.

Letter knowledge. Ability to automatically identify the names and the most common sounds of the letters of the alphabet.

Literal comprehension. The ability to identify facts that are directly stated in a passage.

Listening comprehension. Ability to understand and derive meaning from spoken language.

Listening vocabulary. Words that a person can understand when the words are heard.

Longitudinal study. Longitudinal studies collect data from a cohort of individuals on multiple occasions over an extended period of time. They are designed to investigate development in an area of learning, making it possible to study progress over time at the individual level.

Mother tongue. This term is used interchangeably with the term 'native language', referring to the first language one learns to speak, primary language used or one's dominant language.

Morpheme. The smallest, meaningful unit of language. A morpheme may be a word or a word part. For example, 's' as in 'cats', is a morpheme that conveys number.

Morphology. The study of the structure of words**.**

Multilingualism. The ability to speak, understand, read and write three or more languages.

National learning assessment. A nationally-representative assessment of students' learning outcomes at a particular age or grade level. It provides information on a limited number of outcome measures that are considered important by policymakers, politicians and the broader education community.

Narrative text. Text that tells a story and follows a common story structure.

Onset-rime instruction. Use of known word patterns to read unfamiliar words. The onset is the initial consonant or consonant cluster of a one-syllable word (e.g. the 's' in 'sat' or the 'tr' in 'train'). The rime includes the vowel and subsequent letters (e.g. the 'at' in 'sat' or the 'ain' in 'train').

Oral assessments. Assessments that are administered orally.

Orthography. A system of written language, including the formation of letters and the spelling of words.

Out-of-school children. Children in the official primary or lower secondary school-age range who are not enrolled in either primary or secondary school.

Percentage of correct items. The number of test items that a student answers correctly divided by the total number of test items and then multiplied by 100 produces that student´s percentage score on that test.

Phoneme. The smallest unit of sound.

Phonemic awareness. Ability to recognise and manipulate the individual sounds (phonemes) in spoken words.

Phonological awareness. Ability to manipulate the sound system of a spoken language, including words, rhymes, syllables, onset-rimes and phonemes.

Phonology. The study of the sound system of a language and the use of sounds in forming words and sentences.

Print concepts. Skills beginning readers require to understand the concepts of written language. Examples include the concepts of words, sentences and directionality.

Proficiency levels. Refers to the classification of students into categories (or bands) of performance that are identified by a series of cut-off scores on the performance scale. Proficiency levels are commonly used in criterion-referenced tests. Each of these levels should be defined using specific descriptions of what it means to be at that level (in terms of knowledge, skills, attitudes, etc). Each level (or category) represents a degree of mastery of what the test purports to measure. The same levels of proficiency can be expressed with words or letters (e.g. 'below basic', 'basic', 'above basic'; 'high', 'middle', 'low'; or 'A', 'B', 'C').

Progress monitoring. A system of frequent and dynamic assessment to measure student progress in a skill area.

Prosody. Use of appropriate intonation and phrasing when reading; reading with expression.

Public examinations. An exit or end-point standardised examination that is generally set by a central federal/state examination board or department. The examinations are conducted to promote, select or provide certification to all candidates who qualify or are supposed to have formally or informally learned the curriculum of a formal education programme as part of the graduation requirements. Public examinations are generally administered every year at the end of the school year to all students who wish to take the test.

Rate. Speed at which a person performs a task.

Reading comprehension. Ability to understand and gain meaning from written language.

Reading level. Information on the difficulty of a text.

Receptive language skills. Language skills that do not require the student to produce language (i.e. listening and reading).

Reporting level. Refers to the level at which specific reports are drafted and provided to inform stakeholders of the results of the assessment/ examination. Possible levels of reporting include student, school, local, regional and national.

Reporting metrics. Different forms of reporting assessment results (or achievement on a given test), which can be reported for individuals or aggregated for specific groups. The possible forms (metrics) of reporting results include percentage of correct items, scale scores and proficiency levels.

Rhyme. Two or more words that have the same ending sounds but not necessarily the same letters. For example, 'state', 'straight', and 'bait' rhyme because they all end with the same sound unit.

Rime. The part of a syllable that includes the vowel and subsequent consonants (e.g. the 'at' in 'sat' or the 'ain' in 'train').

Sample. The individuals included in a study or assessment.

Scale scores. Ability estimates that are generated from item response theory (IRT) models, which are based on students' response vectors. Scale scores are designed to provide a metric that is consistent for different versions of a test and consistent across time.

Segmenting. Breaking words into individual sounds or word parts. For example, the spoken word '*cat'* can be broken into the speech sounds /c/ /ă/ /t/ *or* /c/ /ăt/.

Sight words. Words that can be read fluently and automatically at first sight.

Significance tests. Measures the probability of observing an effect of the null hypothesis. If the p-value is less than the level, the hypothesis is rejected.

Speaking vocabulary. Words a person uses when he or she speaks.

Story structure. Component parts of a story (narrative text), including characters, settings, events, problems and resolutions.

Summarising. Synthesis of the main ideas in a text.

Syllable. A unit of pronunciation usually containing a vowel.

Syntax. The rules of language to determine the order of words in a sentence.

Synthetic phonics. The systematic teaching of word reading through the blending of known letter-sound correspondences. For example, using the known sounds /c/ /a/ /t/ to read the word 'cat'.

Transparent orthography. A writing system that has a one-to-one or nearly one-to-one correspondence between letters and sounds.

Vowel diphthongs. A vowel sound formed by two combined vowel sounds. For example, /ow/. In 'cloud'.

Word recognition. An approach to reading words.

WCPM. Words correct per minute refers to the number of words read correctly in a minute.

Written expression. The expression of thoughts, feelings and ideas through writing.

Quality education and learning for all is at the center of the Sustainable Development Goal (SDG) for education. Policymakers at the global and national levels clearly recognise that to determine if the quality of education is improving, they must first engage in learning assessments to monitor learning outcomes. Oral reading assessments play an important role in measuring the basic foundational reading skills that are a gateway skill to all other learning.

Today, the use of oral assessments is widespread and while there are some commonalities among the instruments used, there are also differences in purpose, design and administration. In response, the UNESCO Institute for Statistics (UIS) led a collaborative effort with organizations that have been actively financing, designing and implementing oral assessments. Representatives from the participating organizations submitted case studies and position papers to help exemplify good practices in oral reading assessments.

This ebook presents their recommendations for selecting, implementing and using learning assessments as well as basic principles that should be applied at the different stages of oral reading assessments—from planning and design to implementation and use of the resulting data.

As the SDGs become a reality, governments will need more and better data on education to produce evidence, determine the areas of improvement, take corrective action and monitor progress. Early detection of learning gaps will be essential to informing remedial action and securing the ambition of the new goals to ensure that all children gain access to post-primary education. This ebook serves as a unified voice from the community of oral reading assessment practitioners, implementers and donors on the importance of early reading skills to ensure learning for all by 2030.

GLOBAL PARTNERSHIP for EDUCATION

THE WILLIAM AND FLORA HEWLETT FOUNDATION

UNESCO
United Nations Educational, Scientific and Cultural Organization

UNESCO INSTITUTE for STATISTICS