MYANMAR EARLY GRADE READING ASSESSMENT FOR THE YANGON REGION

2014 RESULTS REPORT

JUNE 4, 2015

GEDDR
EAST ASIA AND PACIFIC

WORLD BANK GROUP   EGRA MYANMAR   Australian Aid

MYANMAR EARLY GRADE READING ASSESSMENT FOR THE YANGON REGION

JUNE 4, 2015

GEDDR
EAST ASIA AND PACIFIC

PREFACE

In the 14 years since the Republic of Myanmar committed to the Millennium Development Goals (MDGs), access to and enrollment in primary school has increased significantly. This is a notable achievement. The next important steps for the Government of the Union of Myanmar are to establish learning achievement baselines for its primary school students and to build an internal capacity in the Ministry of Education to monitor and report on learning improvements over time. This snapshot (from schools within the Department of Basic Education, DBE 3) of students' current levels of reading ability in early grades—and factors associated with reading acquisition—is part of this effort to establish measurable learning standards and build institutional capacity to continuously monitor learning achievement.

This document briefly summarizes the state of education in Myanmar, the information needed by stakeholders, and the way the Early Grade Reading Assessment (EGRA), administered in Yangon, Myanmar (DEB 3) addresses these needs. It reviews the most relevant education theories on reading acquisition in the early grades of primary education and how they inform the development of the Myanmar EGRA tool.

This document describes the results of this study and the insights they offer about the characteristics of schools and teachers that can affect students' success in learning to read. The final section proposes recommendations to the Ministry of Education on how to link EGRA results to current policy discussions on language and literacy in the early grades of primary education.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

ABBREVIATIONS

CESR	Comprehensive Education Sector Review
CFWPM	Correct Familiar Words Per Minute
CIS	Correctly Identified Sounds
CIWPM	Correct Invented Words Read Per Minute
CLCQ	Correct Listening Comprehension Questions
CLNPM	Correct Letter Names Identified Per Minute
CLSPM	Correct Letter Sounds Identified Per Minute
CPR	Continuous Personal Record
CRCQ	Correct Reading Comprehension Questions
DBE 3	Department of Basic Education 3
DEPT	Department of Education Planning and Training
DFAT	Department of Foreign Affairs and Trade
DMER	Department of Myanmar Education Research
EFA	Education for All National Action Plan, 2003–2015
EGRA	Early Grade Reading Assessment
GOM	General Outcome Measures
MDGs	Millennium Development Goals
OECD	Organization for Economic Cooperation and Development
ORF	Oral Reading Fluency
TEO	Township Education Officer
WSC	Words Spelled Correctly

"ACCESS, ENROLLMENT, AND NOW QUALITY STANDS AT THE CENTER OF THE EDUCATION POLICY DISCUSSION IN MYANMAR."
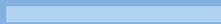
EXECUTIVE SUMMARY

Access, enrollment, and now quality stands at the center of the education policy discussion in Myanmar. As a measure of its commitment to the Millennium Development Goals (MDGs), the Republic of Myanmar has significantly increased access to and enrollment in primary school[1]—a notable achievement. The Myanmar government continues to work to improve the quality of education through its Country Education Sector Reform (CESR) initiative.

Recently, the Myanmar Ministry of Education conducted a rapid education assessment, as part of the CESR initiative, which identified the lack of reliable information on how well students are learning as a significant obstacle to improving the quality of education. It recognized that problems with learning in many schools—whether in other developing countries or in high income countries—often begin during the first year of a child's schooling. Consequently, it supported an Early Grade Reading Assessment (EGRA) in 2014, adapted to the Myanmar context and language, to measure how well primary school children are acquiring fundamental literacy skills (such as learning to read) as one of its priorities. In addition to assessing reading ability (initially in Myanmar's Yangon region), the survey should also provide the Ministry with crucial information needed to develop benchmarks to measure and compare students' performance.

The EGRA project had four main objectives: design a valid EGRA tool for Myanmar, building capacity in assessment, defining core learning reading standards for early primary grades, and identifying factors associated with different levels of reading skills in order to guide policy design. The EGRA project successfully adapted and tested the EGRA tool in Myanmar,[2] although it revealed two test areas that may need to be redesigned, plus some resolvable data and sampling issues. The data gained from EGRA offers clear directions for building domestic capacity to track, report, and analyze students' progress in reading. The Ministry of Education's goal of defining core learning (reading) standards for early primary grades is concretely supported by the EGRA findings. Despite some caveats, the study data identified potential explanatory factors for differences in EGRA scores among schools and students. And tactically, the Ministry of Education will be able to use the EGRA findings to guide it in prioritizing future interventions that best help primary grade students improve their reading skills.

The completion of the 2014 EGRA (data collection and analysis) is a significant accomplishment in this phase of Myanmar's CESR initiative and on-going and future research activities, building upon this work will continue to nourish thinking around the issue of quality. The Ministry of Education continues its resolve to collect data on student reading levels via the (just completed) 2015 EGRA to four more states/regions. This, and further research to come, will deepen and improve the insights provided by this initial study.

This report does not assume expert statistical knowledge by the reader, although the annexes rigorously detail the methodology and statistical analyses. It attempts to present, in plain language, the work to adapt the EGRA tool to the Myanmar language, administer it (and its questionnaires) in the Yangon region, and analyze the collected data. It focuses on answering these four questions.

1. Can EGRA be adapted and used to measure reading skills in the Myanmar language?

Yes, although some changes should be considered that can further strengthen the usefulness of the tool. The EGRA tool comprises eight subtasks, which were specifically developed for the Myanmar language. The findings presented here suggest that the tool, as designed and tested in the Yangon region, worked for six of the eight subtasks. Considerations should be given to redesigning (or dropping) Subtask 2 ("initial sound identification") and Subtask 3 ("letter sound knowledge") in future versions of the EGRA tool. The data from the six successful subtasks, however, still presents a reliable picture of early grade reading skills.

2. What is the status of early grade reading skills in a sample of schools in the Yangon region?

Despite performing relatively well, compared to other countries (in terms of zero scores on the oral reading fluency subtask), the Myanmar EGRA findings are worrying and indicate that too many students have not learned to read or read well. Although the mean scores on all the subtasks improved with grade level, the

---

[1] Ministry of Education, Union of Myanmar, 2013, "Millennium Development Goal Report" (Nay Pyi Taw, Myanmar: Ministry of Education).
[2] The 2014 EGRA study was conducted in the Yangon region, with 176 teachers and 1,681 students participating.

Overview of EGRA scores per grade

Percent

Grade 1: 37, 77
Grade 2: 10, 27
Grade 3: 9, 12
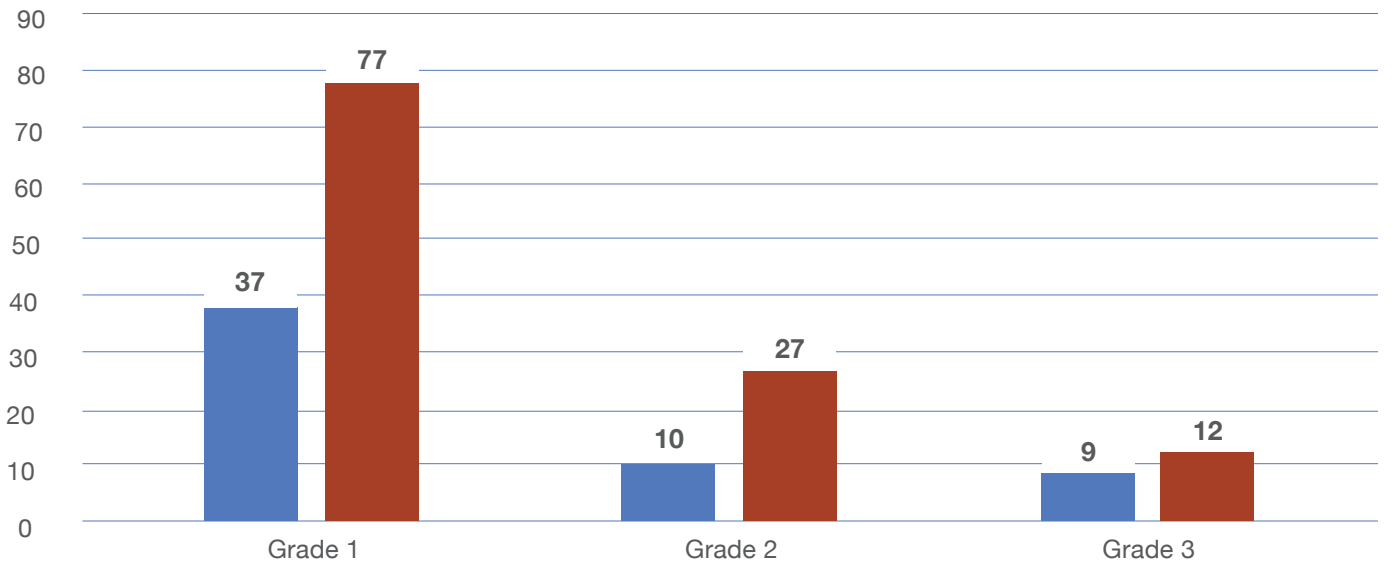
■ % of students who could not read a single word

■ % of students who could not answer a single question correctly about a text they just read

"THE MINISTRY OF EDUCATION'S GOAL OF DEFINING CORE LEARNING (READING) STANDARDS FOR EARLY PRIMARY GRADES IS CONCRETELY SUPPORTED BY THE EGRA FINDINGS."

substantial proportion of students who could not answer one item correctly, and the low mean scores on various subtasks by the end of Grades 2 and 3, remain worrying. For example, 10 percent of second graders cannot read a single word. And, of the second graders who can read, 27 percent did not understand what they had just read (see figure). Furthermore, analyzing the proportion and performance of good and poor readers indicates that students are not meeting curriculum expectations in all three grades. Finally, the reading comprehension results suggest that primary students in Myanmar are capable of understanding a text when it is read aloud to them, but find it more difficult to understand the inferences in the text when they read it themselves.

3. Is it possible to identify factors that affect students' performance and explain the differences in EGRA scores? If so, what are they?

The results of this analysis not only indicate that progress in learning to read may be affected by factors outside the school but also that some schools are better than others at offsetting or ameliorating these effects. Based on the data, a considerable proportion of the differences in students' early reading skills across schools is determined by non-school factors, such as the socio–economic status of students, help with homework, etc.[3] These unfortunately have little to do with educational practice and cannot easily be addressed by

---

[3] As shown by the differences in average EGRA scores between schools.

schools, but this does not mean that such factors cannot be overcome. Indeed, the Myanmar data indicates that the effects of student characteristics, despite being relatively consistent, vary significantly from school to school. Some schools are able to provide environments that offset or ameliorate the effects of gender, family background, and home educational support.

This study does suggest that Myanmar's Ministry of Education may do well to prioritize a few specific interventions, targeting characteristics of school environments which are statistically correlated with differences in test scores while controlling for students' demographics and/or initial skill levels. Interventions with the potential to make a difference include helping teachers, students, and even parents or other adults use the Myanmar exercise book more effectively (by developing and disseminating a set of best practices for it), addressing the need for teachers to take on other jobs, and investing in school libraries or book corners.[4] Further insights into these potential interventions could be obtained from the analysis of the second round of EGRA in 2015 and in the context of future studies such as impact evaluations of such interventions.

But the results of this 2014 Myanmar EGRA are not conclusive in many areas, so further research is needed. One particularly puzzling—counter-intuitive—finding was the inverse relationship between the use of both summative (e.g., use of chapter- or term-end tests) and formative (e.g., feedback to teachers on effectiveness of their instruction) assessments. The use of assessments is expected to help teachers pinpoint where (and which) students need help and/or to fine-tune their instruction methods and so should be positively related to performance. The negative correlation between the use of assessment and EGRA scores should be investigated further. In addition, the study was unable to conclusively distinguish the effectiveness of specific educational techniques used in the schools.

4. What are the recommendations for policy-makers who are interested in improving early grade reading skills?

Potential actions for the Ministry of Education, policy-makers, and development partners to consider can be organized into three main categories.

First, the Ministry of Education could consider expanding the research and linking it to

complementary research and analysis. This can include improving the EGRA tool, research protocols, and reporting efforts. Also recommended is building more capacity within the Ministry to carry out surveys, analyze findings, expand coverage of the research, adapt EGRA and its administration to more ethnic languages, broadly share EGRA findings, and add a math component to the assessment.

Second, the Ministry of Education could use the findings to begin establishing measurable standards to assess reading in the early primary grades. Much value can be added—and knowledge increased—by developing short diagnostic tests and training teachers to measure and report progress toward the standard.

Finally, the Ministry of Education could consider using the findings to help design and select specific interventions aimed at improving reading outcomes. For example, the Ministry of Education and other government agencies could capitalize on EGRA for a number of initiatives:

- Launch a public awareness campaign to promote reading, for example, that encourages parents (or adults) to read to children every day and gives authors incentives to write more children's books.

- Make more reading materials available to students (to complement the basic textbooks) by helping schools add or enlarge a library, or set up reading corners in classrooms via education grant.

- Improve the quality of instruction by helping teachers make better use of Myanmar exercise books, developing best practices for teaching with textbooks, and establishing pre- and in-service programs for teachers on early reading.

- Start the process of defining standards (or competencies) that can be linked to schools' curricula, beginning with reading performance standards for students in early grades.

- Integrate standards into pre- and in-service teacher training programs.

- Improve the monitoring and evaluation system for how well schools (and students) meet performance standards and compare to other schools.

All of the above initiatives could also be rigorously evaluated using impact evaluation methodology combined to pre- and post-EGRA.

---

[4] The potential of these interventions is suggested by the observed statistically significant correlation between the related variables and the EGRA test scores in multivariate hierarchical model (see subsection 5.4).

"WHEN CHILDREN DO
NOT LEARN TO LEARN TO
READ AND UNDERSTAND
A SIMPLE TEXT IN THE
EARLY GRADES, THEY ARE
MORE LIKELY TO FAIL IN
SCHOOL, REPEAT MORE
GRADES, AND DROP OUT
OF SCHOOL."

1. INTRODUCTION

Recently, the Myanmar Ministry of Education conducted a rapid education assessment, as part of its Country Education Sector Reform (CESR) initiative, that identified the lack of reliable information on learning outcomes as a significant obstacle. In addition, education authorities in Myanmar are working in a fragmented system, where regionally-based ethnic and language groups are not cooperating (or not cooperating effectively) with the Ministry of Education or other regional counterparts. Government authorities and donor partners are attempting to promote national reconciliation and collaboration in many fields, including education. They are proposing that these different regional groups work together to develop standards for reading in the early primary grades—and the means to measure them—as a vehicle for collaboration in the education system.

Problems with learning in many schools—whether in developing countries or in high income countries— often begin during the first year of a child's schooling. When children do not learn to learn to read and understand a simple text in the early grades, they are more likely to fail in school, repeat more grades, and drop out of school. Ultimately, it reduces their opportunities in life and increases the incidence of social ills. Yet few developing countries give priority to measuring how well children acquire reading skills in the early grades. The Early Grade Reading Assessment (EGRA) measures a child's progress in learning to read. While the general framework and approach of the EGRA is similar across countries, its specific components can vary.

This report presents the work to adapt the EGRA to the Myanmar language and to administer it (and its questionnaires) in the Yangon region. The results of this exercise were analyzed to assess how valid and reliable this adapted tool is for the Myanmar language. Other aims of the analysis were to define core learning standards in early primary grades, and identify potential explanatory factors for differences in EGRA scores among schools. The specific information from EGRA can help guide the Ministry of Education in prioritizing future interventions that best help students improve their reading skills in early grades, as well as support the development of measurable, comparable benchmarks for students' progress.

The report is organized into six sections, including the introduction. Section 2 briefly reviews the reforms and goals of the education sector in Myanmar and the major challenges that helped define the research questions for the EGRA. Section 3 describes the most relevant education theories about learning to read in the early grades and the way these are reflected in the design of the EGRA. Section 4 outlines the design and implementation of the study and tools, and how they took into account the information needs of education stakeholders and the specificities of the language and education sector in Myanmar. Section 5 presents the methodology and results of the study for non-experts, with the methodology and statistical analyses detailed in annex 5. Section 6 concludes with a summary of recommendations.

"RECENT EDUCATION REFORMS HAVE SET THE LONG-TERM VISION, DIRECTION, AND GOALS FOR THE COUNTRY'S DEVELOPMENT."
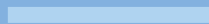
2. OVERVIEW OF THE EDUCATION SECTOR IN MYANMAR

In the last decade, the government of the Union of Myanmar has made great efforts to improve access to basic education, but gaps in knowledge and strategies remain around issues of learning. Recent education reforms have set the long-term vision, direction, and goals for the country's development. The government has also defined specific sector development plans (for immediate deployment) and key areas for immediate attention. Planning efforts recognize that some of these areas represent new challenges and questions. Solutions and answers require more education research, which also need greater technical and financial support. One of these questions—the focus here—asks about the current state of literacy, or how well students are reading, in early grades in Myanmar and the factors associated with it.

2.1 Educational reforms in Myanmar

The government has made access to quality education at all levels a central point of education reform discussions in Myanmar. The framework for Myanmar's most recent plans for educational reform is detailed in its 30-year Long-Term Education Development Plan, 2013-2031.[5] The 10 objectives of the plan assign to the education sector the important role of leading the country's development and modernization. Issues of access to quality education occupy most of the reform agenda, with stated goals of improving basic education,[6] increasing opportunities for pre-vocational and vocational education, and expanding non-formal education at all levels.

One key target in Myanmar's plan is to achieve universal basic education by 2031. Furthermore, to ensure progress toward the plan's goals, it adopted the Education for All (EFA) National Action Plan, 2003-2015.[7] In addition to the objectives of the Millennium Development

Goals (MGDs) for education, the action plan incorporates the six EFA goals:

- Expand early childhood care and education.
- Provide free and compulsory primary education.
- Promote learning and life skills for young people and adults.
- Increase adult literacy.
- Achieve gender parity.
- Improve the quality of education.

As it started to implement reforms, the Ministry of Education recognized that it needed a deeper understanding of the current status of education, especially students' performance.[8] In February 2012, it launched a new initiative, the CESR. The CESR is a participatory process led by the Ministry of Education, which also relies on contributions from a wide range of education stakeholders, including other ministries and development partners (namely, the United Nations and other multi-lateral, bilateral, non-governmental, and civil society organizations). Although the government had provisionally identified a number of priority areas,[9] the CESR's mission started with improving the knowledge base of the strengths and challenges of the Myanmar education system.

The initial rapid assessment and first phase of the review establishes urgent priorities and issues, specifically a quantitative analysis of the access, equity, quality, efficiency, and effectiveness of the school system in Myanmar. Its results provided a baseline for the in-depth education analysis and phase 2 of the CESR, and also identified sources of additional data.

---

[5] Ministry of Education, Union of Myanmar, 2008, "Country Report: Myanmar Education Development Strategy Focusing on Inclusive Education" (Nay Pyi Taw, Myanmar: Ministry of Education). http://www.ibe.unesco.org/National_Reports/ICE_2008/myanmar_NR08.pdf

[6] Basic education encompasses primary school (Grades 1-5), lower secondary school (Grades 6-9), and upper secondary school (Grades 10-11). Ministry of Education, Union of Myanmar, 2004, "Development of Education in Myanmar 2004" (Nay Pyi Taw, Myanmar: Ministry of Education).

[7] Ministry of Education, Union of Myanmar, 2012, "Education for All: Access to and Quality of Education in Myanmar," Conference on Development Policy Options with Special Reference to Education and Health in Myanmar, 13-16 February 2012, Nay Pyi Taw, Myanmar. http://yangon.sites. unicnetwork.org/files/2013/05/Education-for-All-in-Myanmar-Final-2012-FEB-2.pdf.

[8] CESR.org, 2012, "Terms of Reference for the Myanmar Comprehensive Education Sector Review, 2012," http://www.cesrmm.org/index.php/en/.

[9] These focus areas aim to review legislation and enact new laws, reorganize departments in accordance with decentralization, restructure the basic education sector and review the basic education curriculum, improve the quality of teacher education and strengthen the capacity of education personnel, and reform quality assurance systems that assess students.

Among the needs identified, the lack of reliable data on students' performance was a particular limitation for phase 2 and the education sector plan for 2014-2016 (phase 3). Consequently, the CESR Steering Committee expressed great interest in this study in order to collect valid and representative data on the state of early-grade reading in the Yangon region of Myanmar.

2.2 Education sector indicators

Reading, writing, and mathematics are the most important skills for primary students to learn, and represent key indicators and standards of quality for primary schools and teachers. The international standard for classification of education defines primary education as providing students with fundamental skills in reading, writing, and mathematics.[10] In Grades 1-3, Myanmar teachers are advised to devote 25 percent of classroom time to oral and written language instruction, close to the OECD average of 30 percent.[11] In Myanmar and in other countries across the world, available statistics indicate that reading is the subject that students in early grades spend the most time studying.

How well children learn to read in Grades 1-3 has a wide reaching and critical impact on their education: early reading skills are correlated with children's overall learning outcomes, including other subjects. Thus, learning to read is not just an indicator of how effectively schools teach these fundamental skills but can be a proxy for the general quality of education that children receive.

Given the low primary completion rate (54 percent), data on students' achievements in early grades is crucial for decision-makers in the education sector. In Myanmar, about 8.5 million students—with roughly 5 million in primary school—are receiving basic education. The latest national statistics point to a net primary enrollment rate of 88 percent.[12] Among them, 75 percent of first graders reach the fifth (and last) primary grade, but only 54 percent complete primary education.[13] Further analysis reveals that most students leave school in the first two years of primary schooling.[14] Because children in the early grades are likely to quit before completing their basic education, it is vital to determine the actual level of education and skills they achieve. But, such data is rarely available to decision-makers.

The availability of data on students in early grades— in particular on learning—is even more important, given that the low completion rate may be strongly influenced by the quality of schools. The information above suggests that one of the major obstacles to universal basic education in Myanmar is indeed the drop-out rate that occurs early in the educational process. This phenomenon is most often viewed as an issue of access and enrollment,[15] but student retention is also strongly influenced by school quality.[16] Robust research has shown that a student is much less likely to remain in school if that school is of lower quality.[17] Because such little information exists on school quality, it can only be assumed that Myanmar's low completion rate[18] results in part from primary school students having difficulty learning grade-appropriate skills at their school.

[10] UNESCO Institute for Statistics, 2011, International Standard for Classification of Education (ISCED). http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx.

[11] Ministry of Education, Union of Myanmar, 1999, "Instruction and Guide for Teachers," in Myanmar Language Textbook (Nay Pyi Taw, Myanmar: Ministry of Education); and Japanese International Cooperation Agency, 2013, "Data Collection Survey on Education Sector in Myanmar, February 2013" (Tokyo: JICA).

[12] Net primary enrollment rate is defined as the percentage of students aged 5–9 years who attend primary school. From World Bank staff calculations based on the Integrated Households Living Conditions Survey (IHLCA) 2009-2010. UNDP, 2011, "Integrated Households Living Conditions Survey in Myanmar (2009-2010): Poverty Profile" (Yangon, Myanmar: IHLCA Project Technical Unit).

[13] Ministry of National Planning and Economic Development, and the Ministry of Health, Union of Myanmar, 2011, "Multiple Indicator Cluster Survey 2009-2010: Final Report" (Nay Pyi Taw, Myanmar: Ministry of National Planning and Economic Development and the Ministry of Health), http://www.childinfo.org/files/MICS3_Myanmar_FinalReport.pdf.

[14] M. Hayden and R. Martin, 2013, "Recovery of the Education System in Myanmar," *Journal of International and Comparative Education* 2 (2): 47-57.

[15] Ministry of Education, "Millennium Development Goal Report 2013."

[16] S.R. Khandker, V. Lavy, and D. Filmer, 1994, "Schooling and Cognitive Achievements of Children in Morocco: Can the Government Improve Outcomes?" World Bank Discussion Paper, no. 264 (Washington, DC: World Bank).

[17] E.A. Hanushek and V. Lavy, 1994, "School Quality, Achievement Bias, and Dropout Behavior in Egypt," Living Standards Measurement Study (LSMS) Working Paper, no. LSM 107 (Washington, DC: World Bank), http://documents.worldbank.org/curated/en/1994/12/697806/school-quality-achievement-bias-dropout-behavior-egypt.

[18] C.B. Lloyd, B. S. Mensch, and W.S. Clark, 2000, "The Effects of Primary School Quality on School Dropout among Kenyan Girls and Boys," Comparative Education Review 44 (2): 113-47

The main challenge is that Myanmar does not have a way of reliably measuring and tracking students' progress in reading, writing, and mathematics. Currently Myanmar has no regular assessments of student learning at the end of each cycle. There is, therefore, no information about whether students are mastering the most important skills (e.g., reading and math) at the end of primary or middle school. The only regularly administered examination is the Basic Education High School exam at the end of high school. But it is not clear that these exam results are comparable year to year.

Completion of basic education and access to tertiary education depends on passing the Basic Education High School matriculation examination at the end of upper secondary school[19] and the success rate on the matriculation examination and the tests leading to it varies widely across Myanmar.[20] In primary school, despite a promotion rate of 93.3 percent from Grade 1 to Grade 5,[21] students' progress has been regularly assessed since the reform of 2008. Starting in Grade 4, schools review scores on chapter-end tests, in combination with the Comprehensive Personal Record (CPR), to determine which students can progress to the next grade.[22] In Grades 1 and 2, however, the students are given no test or examination, a practice that is sometimes assumed to reduce student absenteeism and drop-out rates.[23]

---

[19] Ministry of Education, "Development of Education in Myanmar 2004."

[20] Hayden and Martin, "Recovery of the Education System in Myanmar."

[21] Ministry of National Planning and Economic Development, and the Ministry of Health, "Multiple Indicator Cluster Survey, 2009–2010."

[22] The CPR for primary school tracks nine components. The student must: 1) have 75 percent school attendance record; 2) regularly take the chapter tests; 3) obey school rules and regulations with no history of social crimes; 4) fulfill obligations to the school, teachers, parents, and the community, and help take care of younger students; 5) Participate in making the school yard verdant and green by helping grow trees and plants; 6) assist in parent's livelihood; 7) participate in sports and physical activities; 8) participate in arts education, such as literary activities, music, singing, dancing, and painting; and 9) keep neat and tidy. The components are divided into six areas, which combined have a highest potential score of 100. In addition, students can earn up to a mark of 100 for each subject in the chapter test. To be promoted to the next level, students must obtain at least 40 percent in both academic and school activities.

[23] Ministry of Education, "Development of Education in Myanmar."

"THE RATE OF LEARNING
TO READ IN THE PRIMARY
GRADES IS A STRONG
DETERMINANT OF
LATER LITERACY: GOOD
READERS BECOME MORE
FLUENT, WHILE POOR
READERS GET WORSE
OVER TIME."

3. THEORY AND MEASUREMENT OF READING DEVELOPMENT IN THE EARLY GRADES

3.1 Theory of early grade reading

Research shows that students who experience difficulties learning to read in the early grades never really catch up with their peers.[24] The rate of learning to read in the primary grades is a strong determinant of later literacy: good readers become more fluent, while poor readers get worse over time.[25] Longitudinal data on reading in early grades in the United States indicates that gaps in achievement in early grades tend to be magnified over time. Children not only lose out on learning opportunities early but they are set on a trajectory of underperformance that worsens over time. Figure 1 (reproduced from a 1998 study in the United States[26]) plots the trajectories of children with varying literacy skills. At the end of the first grade, good readers begin to separate from children having more difficulty and the gap continues to widen through to the end of second grade.

The scientific literature identified a fixed number of skills and strategies deemed fundamental to becoming a good reader. In alphabetic languages, the initial stages of learning to read include four basic skills (described in more detail as follows):

1. Demonstrate the ability to hear and manipulate the sounds of the language (phonemic awareness).

2. Be able to recognize and sound out letters (and letter groups) of the alphabet (alphabetic knowledge and awareness), and blend the sounds into words (phonological decoding).

3. Accurately and fluently identify and connect written words into text without effort (reading fluency).

4. Understand both oral language and written texts—when text is read out loud to students or they read written text on a page (comprehension strategies).[27]

Phonemic awareness is typically defined as the ability to produce and manipulate the sounds of a language (phonemes). Phonemic awareness is believed to help with reading. Before young children learn to read, they are not "aware" that words can be broken down into their component sounds (or are capable of doing so). Thus, the ability to match letters and sounds (referred to as grapheme-phoneme correspondence) depends on a child's ability to become aware of phonemes. This, in turn, helps them build language decoding skills.[28]

---

[24] K.E. Stanovich, 1986, "Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy," *Reading Research Quarterly* 21: 360-406.

[25] The study called this phenomenon the "Matthew effect," referring to the observation in the Gospel of Matthew (in the Christian Bible) that the rich get richer and the poor get poorer.

[26] R.H. Good, D.H. Simmons, and S. Smith, 1998, "Effective Academic Intervention in the United States: Evaluating and Enhancing the Acquisition of Early Reading Skills," *School Psychology Review* 27: 45-56.

[27] G.H. Good, D.C. Simmons, and E.J. Kame'enui, 2001, "The Importance and Decision-Making Utility of a Continuum of Fluency-Based Indicators of Foundational Reading Skills for Third-Grade High-Stakes Outcomes," Scientific Studies of Reading 5: 257-88; M.J. Adams, 1990, Beginning to Read: Thinking and Learning about Print (Cambridge, MA, USA: MIT Press); National Reading Panel, 2000, Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction—Reports of the Subgroups (Bethesda, MD, USA: National Institute of Child Health and Human Development); and D.C. Simmons and E.J. Kame'enui, eds., 1998, What Reading Research Tells Us about Children with Diverse Learning Needs: Bases and Basics (Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc.)

[28] L. Sprenger-Charolles, P. Colé, and W. Serniclaes, 2006, "Reading Acquisition and Developmental Dyslexia" in Essays in Developmental Psychology (Hove, UK, and New York: Psychology Press.

Figure 1: Early grade reading progression
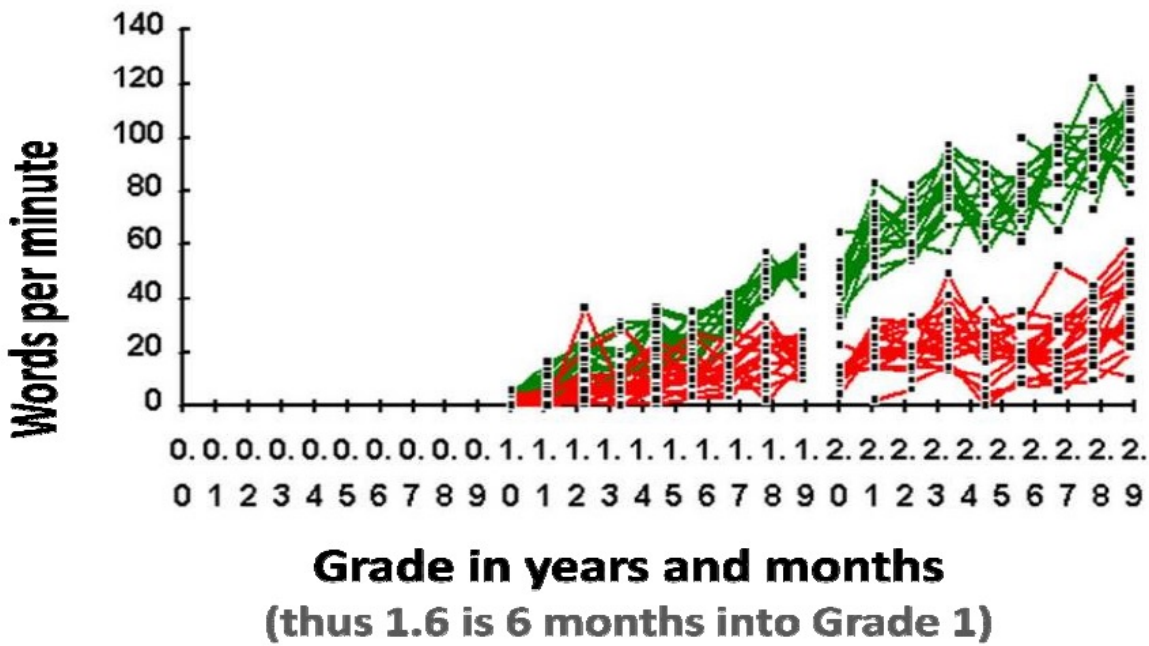
Source: Good, Simons, and Smith (1998).

"GRAMMATICAL KNOWLEDGE, VOCABULARY, AND WORKING MEMORY (LINGUISTIC FACTORS) ALL CONTRIBUTE TO COMPREHENSION."

Phonemic awareness and knowledge of the alphabet must be acquired together to learn to read. They are reliable predictors and diagnostic measures of reading readiness[29] because young children learning to read cannot connect the letters making up a written word with the sounds underlying the spoken word unless 1) they are consciously aware of both and 2) have the intent to learn the relationship between the two (the alphabetic principle). Thus, if children know the letters and know there is some relation between the letters and the spoken word, but they do not know the sounds underlying the spoken word, then they will not be able to figure out the relationship between oral and written words.[30]

Alphabetic knowledge **is also crucial because it emphasizes the interconnection between oral and written language.** It helps students understand that when they read, the letters in written words represent the sounds in spoken words. Thus, in the same way, when they write, the sounds in spoken words can be turned into written words. The purpose of alphabetic knowledge is to help children understand that written words are "sounded out," and not memorized.[31] Knowledge that spoken words are made up of sounds, and written words are made up of letters is not sufficient for developing good decoding skills, or the ability to correctly pronounce words (in the child's language).

[29] D.L. Share, 1999, "Phonological Recoding and Orthographic Learning: A Direct Test of the Self-Teaching Hypothesis," Journal of Experimental Child Psychology 72: 95-129.

[30] W.A. Hoover, 2002, "The Importance of Phonemic Awareness in Learning to Read," SEDL Letter, Putting Reading First 14 (3), http://www.sedl.org/pubs/sedl-letter/v14n03/3.html.

[31] S. Wren et al., 2000, "Cognitive Elements of Reading" in Cognitive Foundations of Learning to Read: A Framework (Austin, TX, USA: SEDL), http://www.sedl.org/reading/framework/elements.html.

Decoding, **the ability to translate text into speech, is considered a powerful strategy that helps beginning readers learn to read words effortlessly and automatically.** It involves matching letters to sounds to form syllables and words until this process becomes automatic. The importance of decoding in early reading has been formalized in the self-teaching theory.[32] This states that each successful decoding of a new word gives children an opportunity to gather word-specific information that makes it easier for them to efficiently and automatically recognize words. Decoding skills are particularly important and effective for beginning readers, who are the most likely to meet unknown or new words in the texts they read and need to develop fluent and accurate readings.

Reading fluency **is an important part of being a proficient reader.[33]** When reading aloud, fluent readers sound natural and read at a conversational pace. Fluent readers are accurate and quick, and use proper expression, while non-fluent readers may read slowly, make errors, and sound monotonic or unnatural.[34] Being able to recognize words easily (word identification fluency, a task that measures automatic word recognition) has been repeatedly shown to be an excellent index of early reading development and an early precursor of fluent text reading, which is necessary for good reading comprehension.[35]

Reading and listening comprehension **is composed of two equally important components: decoding (or translating text into speech) and language comprehension (or the ability to understand spoken language).** Each component is necessary, but is not sufficient on its own to ensure that children understand fully the text they are reading. (Multiple studies have demonstrated that all struggling readers have difficulty with either language comprehension or decoding, or both.) However, what these component skills of reading contribute to reading comprehension differs, depending on the reader's expertise.

When children begin to learn to read, automatic word recognition, decoding and phonological awareness[36] all help them develop efficient word-identification skills and lead to fluent reading (which in large part

occurs by reading itself). However, these skills alone are not sufficient to support and further develop good reading comprehension. When a child does not know the meaning of words and sentences in a text, it becomes a bottleneck and slows down progress in reading comprehension. Grammatical knowledge, vocabulary, and working memory (linguistic factors) all contribute to comprehension, and their contribution increases with age.[37] An essential aspect of language comprehension hinges on the ability to draw inferences and appreciate implications; thus it is important to understand both the explicit and implicit messages contained in language.

The ultimate goal of learning to read is to understand the meaning of text. Thus, a large body of research has examined the particular strengths and difficulties of students with different levels of comprehension skills to establish that the quality of the mental representation built by the reader (i.e., thoughts about things, events, and surroundings that help people create meaning or significance for something) matters significantly. Deep comprehension relies on the construction of a situation model, based on the reader's ability to understand literal information, as well as the ability to draw inferences that bridge elements in the text and support its coherence.

Among all possible inferences that can be drawn from a text, both text-based and non-text-based inferences are most necessary. The reader must make text-based inferences in order to make a text coherent, but often the reader uses non-text-based inferences, such as causal inferences, to connect actions in a story. For example, in this text—"Michael took the drink out of the bag. The orange juice was very refreshing."—the reader makes text-based inferences that the drink in the bag (first sentence) was an orange juice (second sentence). In this example—"Martha and Julius are playing in the sand and swimming."—the reader must make a non-text-based inference that they are at the beach. The reader uses knowledge about the outside world to make the connection that a place with sand and water to swim in is the beach. In addition to processing text literally and through inference, children with good comprehension skills are sensitive to story structure and are able to monitor comprehension.

[32] D.L. Share and A.F. Jorm, 1987, "Segmental Analysis: Co-requisite to Reading, Vital for Self-Teaching, Requiring Phonological Memory," *Cahiers de Psychologie Cognitive/European Bulletin of Cognitive Psychology* 7: 509-513.

[33] S.L. Deno, 2003, "Curriculum-Based Measures: Development and Perspectives," *Assessment for Effective Intervention* 28 (3-4): 3-12.

[34] Sprenger-Charolles, Colé, and Serniclaes, "Reading Acquisition and Developmental Dyslexia"; and L. Sprenger-Charolles et al., 2003, "Development of Phonological and Orthographic Processing in Reading Aloud, in Silent Reading, and in Spelling: A Four-Year Longitudinal Study," *Journal of Experimental Child Psychology* 84 (3): 194-217.

[35] S.L. Deno, P.K. Mirkinand, and B. Chiang, 1982, "Identifying Valid Measures of Reading," *Exceptional Children* 49: 36-45.

[36] Phonological awareness is a general term for meta-linguistic awareness of any of the phonological characteristics of language, including phonemic units, syllables, rimes, and words. This is different from the term phonemic awareness, which refers to the ability to consciously manipulate language at the level of phonemes. Hoover, "The Importance of Phonemic Awareness."

[37] V. Muter et al., 2004, "Phonemes, Rimes and Language Skills as Foundations of Early Reading Development: Evidence from a Longitudinal Study," Developmental Psychology 40: 665-81.

Writing and spelling skills also play a role in developing reading skills in early grades. Writing in preschool and kindergarten has significant predictive relations with later reading ability;[38] there is also a direct causal relationship to reading from spelling. In the early stages of learning to read, children are still mastering the relations between letters and sounds, which explains why multiple studies show that spelling contributes greatly to decoding. Additional evidence suggests that writing also provides the opportunity for children to build word-specific orthographic (letters and spelling) knowledge. In writing, they must select the correct word-specific grapheme (the letter or letters that "spell" a sound in a word) associated with each phoneme (sound in a word) and thereby gain a deeper exposure to words.[39]

In languages with an alphabetic script—like the Myanmar language—research has identified certain basic mechanisms. If these are present, they help students learn to read skillfully; if absent, they make it harder for students to read.[40] These indicators, which can predict reading success in early grades, include phonemic awareness and letter knowledge—both necessary to establish good decoding skills. Decoding skills are a sine qua non and a self-teaching mechanism that enables students to automatically recognize words and read text fluently. Linguistic abilities, along with fluent word- and text-reading abilities are key components of reading comprehension.

3.2 Measuring early grade reading skills

Among different research tools used to evaluate performance and skills, General Outcome Measures (GOM), a tool created in special-needs research, is widely used to improve educational decision-making in a variety of contexts.[41] GOM was created to be a simple, reliable, and valid set of measurement procedures that teachers could use frequently and repeatedly to measure their students' progress in the basic skills of reading, spelling, and writing.[42] As a complement to high-

stakes standard-based measures (that can determine future course), GOM has been recognized as a useful tool for diagnosing reading difficulties in early grades. It incorporates a continuum of fluency measures for reading, which gives teachers essential feedback on the effectiveness of their instruction.

Over the last 25 years, the GOM framework has been used in a variety of ways, such as:

• establishing norms for identifying students who need special-education services,

• evaluating the effectiveness of educational programs,

• reintegrating students with disabilities into general-education classrooms,

• monitoring students' progress and helping teachers plan instructions in general-education classrooms, and

• identifying potential candidates for special-education evaluation using a dual-discrepancy model.[43]

The GOM framework also helped develop the EGRA, a system diagnosis tool to advance the efforts of developing countries to improve the quality of reading outcomes in the early grades of primary education. The structure of the EGRA tool focuses on assessing the degree of *automaticity*[44] students have developed, on average. In other words, it shows how accurately and quickly students are learning basic reading skills in the first grades of primary school. Applied in more than 40 countries and in 70 languages, the EGRA tool has effectively provided useful baseline data to countries that are planning preventive measures for early-reading difficulties, establishing benchmarks to track development targets, and measuring the impact of reforms and literacy programs.[45]

The full set of the EGRA test includes nine modules (or nine subtasks); eight are used in this study.[46] These are the eight subtasks used for this study:

[38] D. Shahar-Yames and D.L. Share, 2008, "Spelling as a Self-Teaching Mechanism in Orthographic Learning," *Journal of Research in Reading* 31: 22-39.

[39] Ibid.

[40] M.S. Seidenberg, 2013, "The Science of Reading and Its Educational Implications," Language Learning and Development 9: 331-60.

[41] S.L. Deno, 2003, "Curriculum-Based Measures: Development and Perspectives," Assessment for Effective Intervention 28 (3–4): 3-12.

[42] Ibid.

[43] P.M. Stecker, L.S. Fuchs, and D. Fuchs, 2005, "Using Curriculum-Based Measurement to Improve Student Achievement: Review of Research," *Psychology in the Schools* 42: 795-819.

[44] RTI International, 2009, Early Grade Reading Assessment Toolkit, prepared for the World Bank Office of Human Development (Research Triangle Park, NC, USA: RTI International; and Washington, DC: USAID). https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&id=149.

[45] USAID and Eddata, "Early Grade Reading," https://www.eddataglobal.org/reading/.

[46] The subtask that was not used in this study is "orientation to print." It measures students' ability to orient themselves in the text, direct the reading, and understand where a new line begins.

1. Letter name knowledge

2. Identification of initial sounds in words (a measure of phonemic awareness)

3. Letter sound knowledge

4. Familiar word reading (a measure of automaticity in word identification)

5. Invented word reading (a measure of decoding)

6. Oral reading fluency with comprehension

7. Listening comprehension

8. Dictation

Table 1 below describes each subtask in the full set of the EGRA tool used in Myanmar, the skill associated with it, and its measure and indicator. (See Annex 1 for how the eight subtasks are presented in the EGRA tool, plus the actual subtask stimuli.) It is important to note that the use of the EGRA tool and EGRA-like assessments do not require that countries administer the full set of subtasks. Each country should determine which subtasks are most relevant for their particular purpose and context.

Table 1: EGRA instrument structure and early skills tested in Myanmar

Subtask	Skill	Ability demonstrated by students mastery of skill	Measure and indicator
1. Letter name knowledge	Letter recognition	Provide the name of upper- and lower-case letters distributed in random order.	The correct letter names identified per minute (CLNPM)
2. Identification of initial sounds	Phonemic awareness	Segment words into 2-5 phonemes. Identify words with different beginning or ending phonemes.	The number of correctly identified sounds (CIS)
3. Letter sound knowledge	Phonics	Provide the sound of upper- and lowercase letters distributed in random order.	The correct letter sounds identified per minute (CLSPM)
4. Familiar word reading	Word reading	Read simple and common one- and two-syllable words.	The correct familiar words read per minute (CFWPM)
5. Invented word decoding	Alphabetic principle	Make grapheme-phoneme correspondences (GPCs) by reading simple invented words to test decoding skills.	The correct invented words read per minute (CIWPM)
6. Oral reading fluency with comprehension	Oral reading fluency	Read a text with little effort and at a sufficient rate.	The correct words read orally per minute in a narrative passage (ORF)
	Reading comprehension	Respond correctly to different types of questions, including literal and inferential questions about the text they have read.	The percentage of correct answers to reading comprehension questions (CRCQ)
7. Listening comprehension	Listening comprehension	Respond correctly to different types of questions, including literal and inferential questions about the text read to them.	The percentage of correct answers to listening comprehension questions (CLCQ)
8. Dictation	Alphabetic principle	Write, spell, and use grammar properly in a dictation exercise.	The percentage of overall early writing skills (spelling and basic conventions)—words spelled correctly (WSC)—with a weighted score.

Note: Adapted by the authors, based on RTI International, 2009, Early Grade Reading Assessment Toolkit, prepared for the World Bank Office of Human Development (Research Triangle Park, NC, USA: RTI International; and Washington, DC: USAID).

"MYANMAR IS COMMITTED
TO PROMOTING
GENDER EQUALITY AND
HAS TARGETED THE
ELIMINATION OF GENDER
DISPARITY IN ALL LEVELS
OF EDUCATION BY 2015."
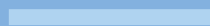
4. SURVEY DESIGN AND IMPLEMENTATION

4.1 Implementation arrangements

Implementation of the EGRA in Myanmar was the result of a collaborative effort between the Ministry of Education, the World Bank, Australia (Department of Foreign Affairs and Trade, DFAT), Save the Children, and representatives of DBE 3 and teacher training colleges. The Ministry of Education led the work financed by the World Bank and DFAT. The World Bank provided complementary technical assistance, Save the Children was contracted to carry out the field work and data entry, representatives of DBE 3 supplied data for the sampling and facilitated logistics, and teacher training colleges assigned teacher trainees to be EGRA enumerators.

To oversee the whole process, the Ministry of Education appointed a steering committee, and to work out the technical details and the implementation, the EGRA Technical Working Group was formed. The Steering Committee comprised a representative from each Department of Basic Education, the Department of Education Planning and Training (DEPT), the Department of Myanmar Education Research Bureau (DMER), and the CESR Task Team. The EGRA Technical Working Group involved the participation of Myanmar language scholars and the representatives from education colleges. Finally, the organizations advising the Steering Committee—DFAT (AusAID at the time), the World Bank, and Save the Children—were invited to send a representative in a funding and implementing capacity. These organizations gave progress reports at each key step in the study during regularly held steering committee meetings. Education experts from the Steering Committee and the EGRA Technical Working Group suggested that the EGRA framework be used to measure early grade reading skills.

4.2 Guiding questions

Prior to designing survey and assessment tools, the Steering Committee defined the scope and research questions to guide the study, which focused on the relation between early skills and gender, language, school remoteness, and the characteristics and behaviors of other students and teachers. More precisely, the Steering Committee defined these four core research questions, which best supported the current reform:

1. *Is reading performance different with gender?* Myanmar is committed to promoting gender equality and has targeted the elimination of gender disparity in all levels of education by 2015. Available rates of enrollment and literacy among young people 15-24 years old indicate that there is no gender disparity in Myanmar. Girls and boys are expected to be treated equally within the education system; therefore, no significant difference in reading performance should be observed between boys and girls in DBE 3.

2. *What effect does a student's language background have on reading performance?* The Myanmar language is the official language of instruction in classrooms and in all textbooks for Grades 1-3, including the reading texts. However, as many as 100 languages are spoken by the people of Myanmar. As in many multilingual countries in Southeast Asia, students entering primary schools are not always familiar with the official language used in schools and textbooks. The Steering Committee wished to investigate whether students speaking mostly another language are reading below the level of their peers who are fluent in the Myanmar language.

3. *Does a school's remote location affect reading performance?* The study looked at whether the remoteness of a school, as measured by the official administrative isolation classification, affected student literacy. In other words, do hard-to-reach schools perform less well than schools within easy commuting distance of the Township Education Office (TEO)?

4. *Do other student characteristics and teachers' behaviors affect reading performance?* The Steering Committee suggested that the study identify and investigate whether other factors influence students' reading performance.

4.3 Development of the EGRA tool in the Myanmar language

If an EGRA tool does not exist in the language of a country, a new tool (never a translation of another tool) must be developed, following a process that depends on the availability of fundamental statistics on the written language of assessment. The official language in Myanmar and its educational

system is the Myanmar language (historically sometimes referred to as the Burmese language); consequently, it is also the language selected for this EGRA exercise. The solid research base supporting the EGRA tool provides authoritative guidelines on the selection and development of the test items in each subtask.[47] International languages, such as English or French, are well researched and statistical information can be readily accessed in scientific literature and scholarly databases from academic research institutions.[48] However, to date, the Myanmar language and its orthographic system have not received the same degree of research attention and lack the statistical data needed to develop an EGRA tool.

The development of test items for EGRA Myanmar therefore started with the research and development of basic linguistics data and statistics on the writing system. Two statistics were needed: (i) The frequency of the graphemes in literature for children in early grades, which is used to design subtasks for letter and phoneme reading or analysis; and (ii) A list of some of the most frequently used words in children's literature, which is used to design word-level reading. To establish relevant statistics, the team first identified the most prevalent books used by students in Grades 1-3. The Ministry of Education and educational experts confirmed that the official classroom reading textbooks were the main reading sources for students in Myanmar.[49]

The statistical analysis of the results was treated separately for each grade level. This is particularly important because the year-end learning goals for some of the most basic reading skills differ in early primary grades. Thus, two different forms of the EGRA were designed, based on specific grade textbook statistics and curriculum expectations: one for Grade 1 and another for Grades 2-3. Consequently, Grade 3 students are expected to perform better than Grade 2 students on each subtask. In Grade 1, the children learn a portion of the letters of the Myanmar alphabet. Thus, the EGRA subtasks for the Grade 1 assessment only contain those letters covered in the curriculum. In Grades 2 and 3, students learn and know all letters (and possible combinations), so the complete alphabet is included in the EGRA subtasks for both grades.

After reviewing the draft version of the EGRA tool tailored to the Myanmar language (the language of reading instruction), the Steering Committee agreed that all designed subtasks would be applied in Grades 1 through 3.

Additional instruments (questionnaires) were also designed to collect information on student and teacher characteristics and behaviors. These instruments were designed by modifying the standard EGRA questionnaires and adapting them to the Myanmar context. The information collected was analyzed to determine how different environmental and behavioral factors could influence students' reading performance.

4.4 Sampling

The study subjects were a random sample of the target population. The target population of the survey consisted of students in Grades 1, 2, and 3 at schools in the Yangon region of Myanmar.[50] The focus on Yangon was not justified on the basis of its representativeness of the rest of the country but, given that this is the first round of EGRA ever carried out, on the basis of cost and easiness of logistics. Furthermore, it was estimated that the reliability of the EGRA tool for the Myanmar language could be tested even if the sample was not nationally representative. A two-stage probability sample, drawn from all 2,609 schools in the Yangon region (which make up the DBE 3), was intended to support the reporting on reading outcomes by gender, school remoteness, class size, and first (or main) language.

First, the schools were ranked by five classifications of remoteness (an administrative isolation classification) and assigned equal probabilities of selection within each stratum. These are defined below:

A—Schools that are located in the same place or within one hour of the Township Education Office (TEO);

B—Schools that are farther from the TEO (more than one hour traveling time) and have easy access by roads and railways;

C—Schools that are located 1-2 miles from roads and railways, where travel is somewhat difficult or requires walking, but still within one day's travel of the TEO;

D—Schools that can be reached from the TEO within a day, but are more difficult to travel to than group C schools, have higher traveling costs, or are remote enough that a round trip from the TEO occurs on school holidays;

E—Schools that are in difficult locations, where travel from the TEO can occur only once (maybe more) per month.

Second, 10 students from each grade in each school were sampled with equal probability, using a systematic sampling technique. They were selected from the roster of students in attendance on the day of data collection. In other words, enumerators obtained the roster for all classes in Grades 1, 2, and 3. For a given grade, the number of students were counted and divided by 10 (target sample size) to obtain the sampling interval ("skip

[47] Deno, "Curriculum-Based Measures."

[48] J.C. Ziegler, A.M. Jacobs, and G.O. Stone, 1996, "Statistical Analysis of the Bidirectional Inconsistency of Spelling and Sound in French," Behavior Research Methods, Instruments, and Computers 28: 504-515; and R. Peereman, B. Lété, and L. Sprenger-Charolles, 2007, "Manulex-Infra: Distributional Characteristics of Grapheme-Phoneme Mappings, Infra-Lexical and Lexical Units in Child-Directed Written Material," *Behavior Research Methods* 39: 579-89.

[49] Pali words (used in the study of Buddhist scriptures), which are sometimes used in writing, were excluded from this analysis and the assessment.

[50] Monastic (Buddhist) schools, which constitute around five percent of the school population in Myanmar, were not included in this sample.

identifying" the Nth student). Starting from the top of the roster, every Nth student among those present was selected for participation in EGRA. The final sample includes 176 teachers and 1,681 students.

The main issues with the sampling strategy is that its implementation limits the validity of the analysis and the interpretation of results with regards to gender and language. This caveat should be kept in mind when reading the results section below. With gender, because the typical class roster in Myanmar lists all boys first and then the girls, in small classes the probability of selecting a girl is lower than for selecting a boy. This bias was further exacerbated because the field team rounded up the skip, which meant that often the last girls on the roster had zero probability of being selected.

For language, despite the intention of reporting on students' language background as a factor in learning to read, this information on students was not ultimately available to apply to the sample frame. Only the information collected in the student interviews, particularly on the languages the students speak at home, can be used to explore the effect of language, but the results cannot be considered highly robust.

4.5 Field work

During the field work, two types of data were collected: 1) information on students, teachers and schools through questionnaires, and 2) students' reading performance through the administration of the EGRA tool. Both the questionnaires and EGRA tool were first field-tested in November 2013 and then administered in January 2014.

Save the Children oversaw training, pilot-testing and collection of the field data by specifically trained teams, comprised of supervisors and enumerators from Save the Children and education colleges. A group of six data quality supervisors from Save the Children,

six supervisors from DBE 3, and 42 enumerators from three education colleges (Thingankyun, Yankin, and Hleku) in the Yangon region were trained in early January 2014 by the World Bank Specialist, local consultants, and Save the Children. The training lasted eight days and included the following activities: (i) Practice presenting each survey to the participants, one by one and question by question; (ii) In-classroom practice with peers (role-play); and (iii) 3-days of in-school practice for inter-rater reliability tests.

All the questionnaires were pilot-tested in a sample of 18 schools in three townships (Dagon North, Tharkayta, and Mingaladon) with 600 respondents (540 students and 60 teachers).

For the field work, there were six data-collection teams, each with one data-quality supervisor from Save the Children, one supervisor from DBE 3, and seven enumerators from the education colleges. Each team was assigned one school per day and was responsible for 10 schools. At the schools, the team interviewed all Grade 1, 2, and 3 teachers, assessed the reading skills of a randomly selected subsample of the students, and then interviewed these same students. An enumerator from the team administered the EGRA task to each sample student individually, away from the distraction of the on-going classes and other assessment activities.

Save the Children carried out data processing with oversight by the World Bank Specialist, using a rigorous process for entry and verification of data. Excel templates were developed prior to the data-processing phase and included efficient checks with clear error messages for inconsistent or out-of-range data. Data clerks trained for three days at the Save the Children office. Data was entered once and random verification checks were carried out twice a day. Save the Children staff conducted spot checks and an extensive process of revising datasets occurred over two months. The final datasets were submitted in May 2014 and are described in Table 2.

Table 2: Description of datasets

Dataset	No. of participants	Unit	Description of key information
Teacher questionnaire	176	Teacher	Demographics, pedagogical practices, years of teaching, pre- and in-service training, materials, assessments and homework, class enrollment, and absenteeism, etc. (Information came from 53 Grade 1 teachers, 50 Grade 2 teachers, 46 Grade 3 teachers, and 27 multi-grade teachers.)
EGRA results	1,681	Student	Reading assessment results on the different subtasks (Information came from 560 Grade 1 students, 577 Grade 2 students, and 544 Grade 3 students)
Student questionnaires	1,681	Student	Socioeconomic status, preschool, homework, absenteeism, learning materials, and outside tutoring, etc. (Information came from 560 Grade 1 students, 577 Grade 2 students, and 544 Grade 3 students)

"IN ALL GRADES, AS EXPECTED, ORAL READING FLUENCY WAS SHOWN TO BE A STRONG CONTRIBUTOR TO READING COMPREHENSION."

5. ANALYSIS AND RESULTS

The analysis of the assessment and survey results has three main objectives related to gauging the validity of the EGRA tool, describing student reading performance, and explaining differences in EGRA scores. First, since this is the first occasion that the EGRA tool was adapted and administered in the Myanmar language, subsection 5.1 describes the characteristics of teachers and students in the Yangon region, and subsection 5.2 assesses the appropriateness of this specific EGRA tool for this study measuring early reading skills in the Myanmar language. Second, subsection 5.3 gives a picture of student performance in reading by looking at means, zero scores, and performance groupings. Finally, subsection 5.4 presents the investigation into potential explanatory factors for differences observed in early reading skills (as requested by the Steering Committee). Note that, as described in subsection 4.3, because the EGRA subtasks differ for Grade 1, the analysis presented below is sometimes split between Grade 1 and Grades 2 and 3.

5.1 Who are the students and teachers in Yangon?

The first objective of the analysis is to provide an overview of the characteristics of students and teachers in Grades 1, 2, and 3 in the Yangon region (DBE 3). It attempts to answer who are the students and teachers in early grades in DBE 3. The main findings indicate that the target population possesses the following general characteristics: (i) A majority of teachers and students are in non-remote schools; (ii) Students from the poorest households attend the most remote schools; (iii) Most students and teachers speak and use the Myanmar language in the classroom; (iv) Student absenteeism is high; (v) Availability of basic teaching and learning materials is adequate.

The discussion below presents the main characteristics of students and teachers (weighted, according to the sampling strategy, to be representative of the Yangon region) as reported on the interview questionnaires. (See Annex 4 for the detailed results.) It focuses first on the characteristics related to the primary research questions selected by the Steering Committee, followed by other characteristics on which information was collected. When (and only if) statistically significant, the difference across grades, gender, or school remoteness is reported.

Remoteness and wealth

Most schools are classified as non-remote and the remoteness index is correlated with wealth. The distribution of schools according to the remoteness classification (see subsection 4.4, Sampling) is: 54 percent in A (least remote), 23 percent in B, 12 percent in C, seven percent in D, and 3.5 percent in E (most remote).[51] Regarding wealth, the student questionnaires collected data on several durables owned by the student's household (e.g., a radio or bicycle). On the basis of the answers to these questions, wealth variables—one continuous, one categorical—were constructed using principal component analysis (PCA).[52] Not surprisingly, the analysis shows that wealth is significantly different according to school remoteness: students in non-remote schools have more wealth than those in remote schools.

Gender and class size

Gender parity is not an issue and average class size (47 students) is relatively high, especially in non-remote schools. Classes are significantly larger in non-remote schools (49 students) than in remote ones (24 students), but class size does not vary significantly across grades. On the day the EGRA study cohort visited, about 50 percent of the students were girls (the computed parity ratio is 1.03). Among students tested, 46 percent were girls, but this is potentially a consequence of the bias in the random selection (as described in subsection 4.4, Sampling).

Language

Variation in the several language variables collected was extremely low: most students, parents, and teachers speak and use the Myanmar language. Although the proportion of students speaking the Myanmar language in school is slightly lower in remote schools (89 percent), overall in DBE 3 almost all students use the Myanmar language (94 percent). The second most frequently reported home language was Kayin (eight percent of remote school students and two percent overall). A similar pattern applies for the ability of family members to read the Myanmar language: 85+ percent for mothers, fathers, and siblings; and 69 percent for grandparents overall. (Percentages are slightly lower in remote schools.) The same pattern emerges with teachers too. Teachers of 89 percent of students speak the Myanmar language as

---

[51] To simplify the classification for this discussion, a remote school is defined as being in group D or E, as opposed to A, B, or C.

[52] The absolute level of this wealth variable has no concrete meaning and is therefore not reported. Its only purpose is to order students according to household wealth.

their first language; the rest of the teachers are equally split between speaking Kayin and Rakine languages.[53]

Interestingly, languages other than Myanmar are more rarely used in Grade 1 than in Grades 2 and 3. In other words, more Grade 1 teachers reported never using a language other than Myanmar in their classes. There are no significant differences regarding the remoteness of the school.

Teacher demographics, class types, qualifications, training, and years of experience

The representative teacher is a woman, relatively inexperienced, who reports using the Myanmar language almost exclusively for teaching. As expected, almost all students (98 percent) have a female teacher. The majority of students are taught by a class teacher, rather than by subject. Thus, only 24 percent of students have a teacher who teaches the Myanmar language exclusively as a subject. Overall only 14 percent of students in the DBE 3 sample are in multi-grade classes. Multi-grade classrooms, however, are significantly more common in remote schools (43 percent of students) than in non-remote schools (10 percent). Most students have relatively inexperienced teachers: 68 percent of teachers have 0-2 years of experience teaching, 20 percent have 3-5 years, and 11 percent have six years or more experience.

The academic qualifications of teachers are relatively high, but the types of teaching certifications are quite diverse. In less remote schools, 94 percent of students are taught by teachers with a bachelor's degree; in remote schools, 80 percent are. The most common teacher certifications are the Certificate for Junior Assistant Teacher (30 percent), Diploma in Teacher Education Competency (26 percent), Certificate in Teacher Education (24 percent), and Post-graduate Diploma in Teaching (15 percent).

Training in methods and pedagogical approaches specific to teaching reading is uncommon. Teachers of 93 percent (pre-service) and 67 percent (in-service) of students report they have never had specific training to teach students to read. Grade 1 students seem to be slightly more likely to have a teacher who had such pre-service training: 10 percent in Grade 1, five percent in Grade 2, and none in Grade 3. (In-service training does not show such a pattern.) Remote schools appear disadvantaged regarding teachers with in-service training in reading: only 65 percent had training versus 84 percent in less remote schools.

According to their responses, teachers live in proximity to the school where they teach and a sizeable number has another income-generating job. Most teachers (67 percent) live within 30 minutes of the school where they teach. Significantly more teachers live closer in remote schools (83 percent) than in non-remote schools (66 percent). Possibly this is because remote areas require that teachers live there or the teacher's home village may be near the school. Worryingly, 18 percent of students are taught by a teacher who has another income-generating job. This situation was more than twice as likely in remote schools (42 percent) than in non-remote schools (15 percent).

Students' prior enrollment in preschool

A significant proportion of students declared they had been to preschool (72 percent overall), but this opportunity was more rarely available to (or taken advantage of) students in remote schools, relative to non-remote schools. Students in non-remote schools were almost twice as likely to have attended preschool (74 versus 46 percent). Conditional on reporting preschool attendance, 40 percent of students said they went to preschool for one year and 13 percent for at least two years. Field reports, however, indicated that students had trouble recalling how long they attended preschool, so this information is unlikely to be reliable.

Student absenteeism

Reported student absenteeism is high. When asked whether they were absent from school at least once in the past week, a majority of students (55 percent) responded affirmatively. The most frequent reason was illness, which explained 59 percent of absences. Furthermore, according to school registers (which are kept universally), on the day of the survey, the student absenteeism rate averaged 14 percent.

Teaching and learning materials

Ownership of the Myanmar reader textbook and a Myanmar exercise book is almost universal, but use of this material and the availability of other reading resources varied. Almost all students (90 percent) owned a Myanmar reader textbook, with slightly higher ownership in non-remote schools. All teachers also reported using the Myanmar reader textbook and finding the material useful. Almost all students owned a Myanmar exercise book (96 percent overall, with no difference according to grade, gender, remoteness, or language) and most had used between one-quarter and three-quarters of it at the time of the survey (as observed by the enumerators).

---

[53] Further, 89 percent of students are taught by a teacher who reports never using a language other than Myanmar in class, while 2 percent of students are in classes where only another language is spoken.

Figure 2: Proportion of assessment types

Assessment type	Value
Oral test	0.78
Lesson recitation	0.41
Chapter end test	0.38
Other	0.10
End of term evaluation	0.04
Homework	0.03
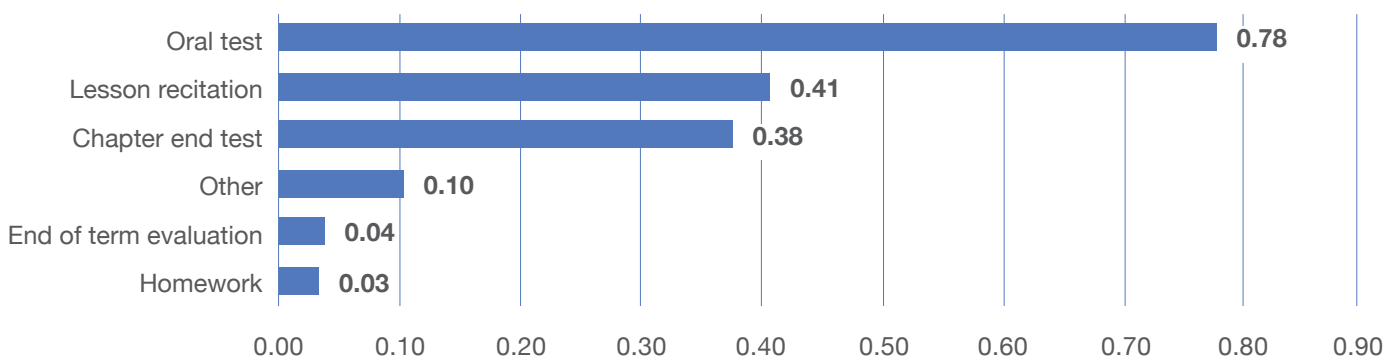
Enumerators were also asked to check the proportion of pages in the exercise books that had markings by teachers, as a proxy for how much the teachers monitored the students' work. Although teachers' markings were visible throughout the exercise book (30 percent of pages), it was not unusual to see no markings (12 percent) in students' books). Finally, enumerators also asked students to show them original work (not copied from a book or blackboard, or dictated by the teachers), but this pedagogical exercise was never used.

Surprisingly, 65 percent of students reported having other reading materials at home (more so in non-remote schools). Furthermore, 90 percent of students went to a school or had classroom with a library, a reading corner, a book box, or book shelf, especially in non-remote schools (92 percent versus 74 percent in remote). In most cases (85 percent), they used these facilities more than once a month. Only 23 percent of students were taught by teachers using teaching and learning materials other than the Myanmar reader: 23 percent used Du Won, 25 percent used Pan Taing,[54] and 77 percent intriguingly used other unspecified material.

Homework, family involvement, and outside tutoring

The frequency of homework assignments varied significantly from the students' versus the teachers' perspectives. Only 63 percent of the students reported having homework assignments in the week prior to the survey. On the other hand, only 20 percent of students are taught by teachers who said they assigned homework less than once per week. The frequency of homework, according to teachers, varied according to the remoteness of a school: 40 percent of non-remote schools reported daily homework being assigned versus only 18 percent of remote schools.

A majority of students (71 percent) reported benefiting from some support with their studies at home, mostly from parents/guardians or siblings (54 and 40 percent, respectively). However, according to teachers, the review of homework by parents is infrequent. In more than 50 percent of cases, teachers reported that only some homework is reviewed by parents and that involvement of parents is lower in remote than in non-remote schools: six percent of parents never review homework in non-remote schools versus 16 percent in remote schools.

Getting outside tutoring (paid or not) in reading skills or in the Myanmar language is quite common for students and more so in non-remote schools: 59 percent of all students reported seeking academic help outside school (62 percent in non-remote schools compared to 37 percent in remote schools).

Assessments

The most frequent methods used by teachers in the study to measure students' progress in reading and writing are oral (77 percent) and written tests (54 percent), shown in Figure 2. Homework or end-of-term tests are almost never used for this purpose (less than 4 percent). Teachers use oral tests significantly more often in non-remote schools (80 percent versus 56 percent), homework is used more often in remote schools (13 percent versus 2 percent), and chapter-end tests are used less often in Grade 1 (19 percent) than in Grades 2 (40 percent) and 3 (56 percent).

Assessment results (all types grouped together) are not often used to inform parents or administrators about student performance (28 percent) or to categorize students and create instructional groups (28 percent)—as shown in Figure 3. Assessments are mainly used to identify children who need help (47 percent), grade students (45 percent), and adapt teaching methods to better suit students' needs (41 percent).

---

54 Du Won and Pan Taing (and Pimya) are privately published exercise books, work books, and guide books, based on the standard (government-issued) text book lessons.

Figure 3: Proportion of uses of assessment results

Identify children who need help	0.47
Grade students	0.45
Adapt teaching to better suit students' needs	0.41
Evaluate students' understandihg of subject matter	0.30
Categorize students (create instructional groups)	0.28
Inform parents/administration of pupils performance	0.28
Other	0.05

0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50
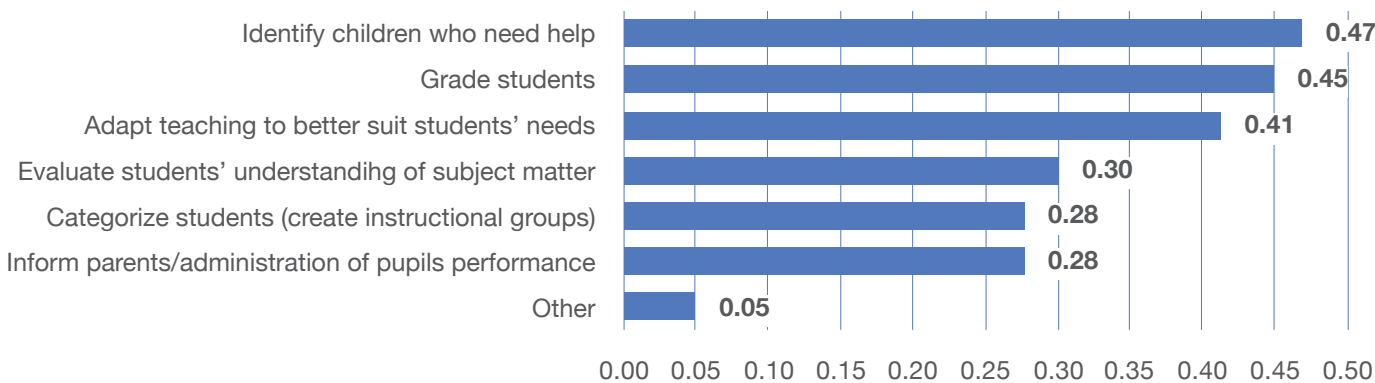
Activities and pedagogical approaches to teach reading

Repetition is the most common activity used in reading lessons. Figure 4 shows the frequency of the different activities reported by the teachers surveyed. Activities involving repetition are relatively common, while sounding out letters; reading for comprehension; reading new words, sentences, or texts; or constructing new words are relatively rare.

There are few differences in reading methods across grades and degree of school remoteness:

• Some comprehension activities are encountered less frequently in remote schools than in non-remote schools (30 percent versus 50 percent). For example, the teacher provides questions on text comprehension and the answers first, and students answer questions on the text after.

• Letter-related activities are more frequent in lower grades than in higher grades. For example, the teacher models and shows a letter first, and then the students repeat the sound of the letter or give the name of a letter without modeling.

• Constructing new sentences by using newly learned words is reportedly more frequent in Grade 3 than in Grades 1 and 2.

The most common pedagogical methods for teaching students to read are "look and say," "questioning," and "pictorial story telling"; least frequently employed are "sound-based," "role-play," and "playing with words" activities. In addition to describing the activities they use to teach reading (Figure 4), teachers were asked to identify their pedagogical methods (Figure 5). The traditional method of "thin bone gyi" (reading by spelling word by word) is used significantly less often in Grade 3 (40 percent) than in Grades 1 (97 percent) and 2 (85 percent).

The same is true for "sound-based reading," although to a lesser extent. "Playing with words" is significantly more common in Grade 1 (64 percent) than in Grade 2 (42 percent) and Grade 3 (43 percent). Non-remote schools (versus remote schools) favor "sound-based reading," "look and say," "questioning," "role playing," and "playing with words" methods.

5.2 Can EGRA be adapted for the Myanmar Language?

Since the EGRA tool has not been designed to assess early reading skills in the Myanmar language, the second objective of the analysis is to assess the appropriateness of the tool as developed for this study. In other words, the analysis of this subsection aims to answer whether the adapted EGRA tool is appropriate to measure early reading skills in the Myanmar language.

The short answer is yes, but some changes could further improve the usefulness of the tool. In fairly technical terms, this subsection discusses how the reliability of the tool was assessed and what the specific findings are. The subsection is structured according to the three main characteristics of an appropriate EGRA tool: 1) internal consistency in reliability of scales, 2) high correlations across subtasks, and 3) predictive power from lower level to higher level skills.

Internal consistency in reliability of scales

The EGRA tool and six of eight subtasks (subtask 2, initial sound segmentation; and subtask 3, letter sound knowledge excepted) appear to have sufficiently high internal consistency in the reliability of scales to justify its use in Myanmar. Annex 2.1 presents the estimates for Grade 1 and Grades 2-3 for the three statistics supporting this assessment:

Figure 4: Proportion of reading activities

Activity	Proportion
Teacher explains the meaning of new words	0.96
Students repeat isolated words after teacher models	0.95
Students repeat new words	0.93
Students engage in whole class reading	0.93
Teacher reads first, student(s) read/repeat after	0.92
Students repeat the name of a letter after the teacher models	0.90
Teacher tells a story orally, students retell the story on their own	0.85
Shown a letter, students give the name of a letter without model	0.76
Students read isolated words without model	0.68
Student(s) read a new, unseen text or passage without teacher	0.64
Students write compound words using known stem words	0.60
Teacher provides text comprehension questions and answers	0.55
Students repeat the sound of a letter after the teacher models	0.54
Students answer questions on text after	0.50
Students construct new sentences by using newly learned words	0.50
Teacher tells a story orally, students repeat the story verbatim	0.46
Shown a letter, students give the sound of a letter without model	0.38
Students engage in silent reading	0.35
Don't know/Refuse	0.00

Figure 5: Proportion of pedagogical approaches

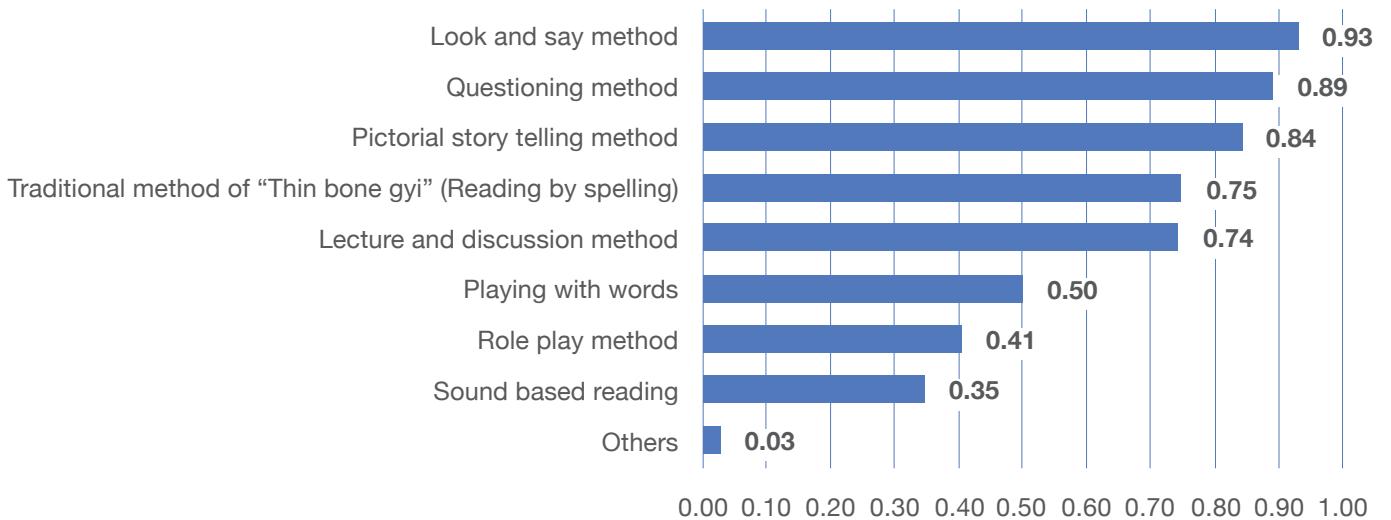| Approach	Proportion
Look and say method	0.93
Questioning method	0.89
Pictorial story telling method	0.84
Traditional method of "Thin bone gyi" (Reading by spelling)	0.75
Lecture and discussion method	0.74
Playing with words	0.50
Role play method	0.41
Sound based reading	0.35
Others	0.03

Figure 6: Predictive relationship between lower and higher level reading skills

Skills or cognitive concept	Higher level skills/ subtasks	Lower level skills/subtasks (predictors)
Text reading	Oral reading fluency ←	- Letter name per minute - Non-word per minute - Familiar-word per minute
Spelling	Dictation ←	- Letter name per minute - Non-word per minute - Familiar-word per minute
Reading comprehension	Correct reading comprehension ←	- Oral reading fluency - Dictation

1. Cronbach's alpha: High average correlation of scores in all subtasks, which is estimated using Cronbach's alpha, indicates high internal consistency in the reliability of several subtasks, which are assumed to measure the same construct.

2. Item-test correlation: High correlation between scores on each subtask and the overall score indicates high reliability.

3. Item-rest correlation: High correlation between the score on a given subtask and the global score, computed using only the other subtasks, indicates high reliability.

The overall Cronbach's alpha (task scale) is 0.77 for Grade 1 and 0.82 for Grades 2 and 3, which indicates that the overall task is sufficiently reliable. At the subtask level, the results indicate that scores on all subtasks have high item-rest and item-test correlation, except for letter-sound fluency (subtask 3), initial sound segmentation (subtask 2), and, to a lesser extent, listening comprehension (subtask 7). Looking at the values of Cronbach's alpha overall (column [v] in Tables A2.1 and A2.2 in Annex 2.1), removing letter sound fluency and initial sound segmentation would lead to substantial increases in Cronbach's alpha. Therefore, these two subtasks, as designed and as administered, are not sufficiently reliable measures of early reading skills in the Myanmar language.

Pair-wise correlations across subtasks

Similar to the observations for internal consistency and reliability, on the basis of pair-wise correlations, most subtasks appear appropriate, except for the initial sound identification and letter sound knowledge (subtasks 2 and 3). Because each EGRA subtask is meant to measure the same overall skill— namely, early reading—appropriate subtasks should be highly correlated to each other. Pair-wise correlations of scores among most of the subtasks of the EGRA battery are significant and positive (see Tables A2.3 and A2.4 in Annex 2.2), except for initial sound identification and letter

sound fluency, which show undersized correlations with other subtasks. Reading and listening comprehension also are significantly correlated with all the other subtasks.

Predictive power of lower level skills to higher level skills

The general expectations derived from the construction of the EGRA tool set were that performances on subtasks measuring lower level skills should be able to predict (or, at least, not be negatively correlated) performance on subtasks measuring higher level skills. Given this, the analysis of the collected data indicates that the EGRA tool set exhibits the appropriate relationship between lower and higher level reading skills. Figure 6 shows the main expected predictive relationships between lower level skills and higher level skills, in text reading, spelling, and reading comprehension. (See Annex 2.3 for the hierarchical multiple regressions run to assess these relationships.)

In all grades, as expected, oral reading fluency was shown to be a strong contributor to reading comprehension. Therefore, the unique contributions of 1) phonetic awareness to reading fluency and spelling, and 2) oral language skills to comprehension, illustrated in A2.5 Annex 2.3, confirms the alignment of the data collected in the Yangon region of Myanmar with the global theoretical and experimental data on reading comprehension.

Based on the statistical analyses performed, it is possible to conclude that, overall, the EGRA tool developed for this study appears appropriate for measuring early reading skills in the Myanmar language. However, two of the subtasks—initial sound identification (subtask 2) and letter sound knowledge (subtask 3)—are either irrelevant to explain reading skills in the Myanmar language or their designs or administration were inappropriate and should be revised for future rounds of assessment.[55] Consequently, the discussion that follows does not refer much to these two subtasks.

---

[53] There are only hypotheses, but no convincing evidence as to why subtasks 2 and 3 appear inappropriate. These subtasks may be particularly difficult for students in Myanmar because of 1) the structure of the Myanmar language, which makes sounding the letter particularly difficult; 2) the pedagogical approaches used in surveyed schools (i.e., very little focus on phoneme awareness and letter sounds); and 3) the administration protocol, which does not provide enough clues as to the expected answer —and which could help soften the challenges as a result of 1 and 2.

Figure 7: Mean scores per subtask, per grade
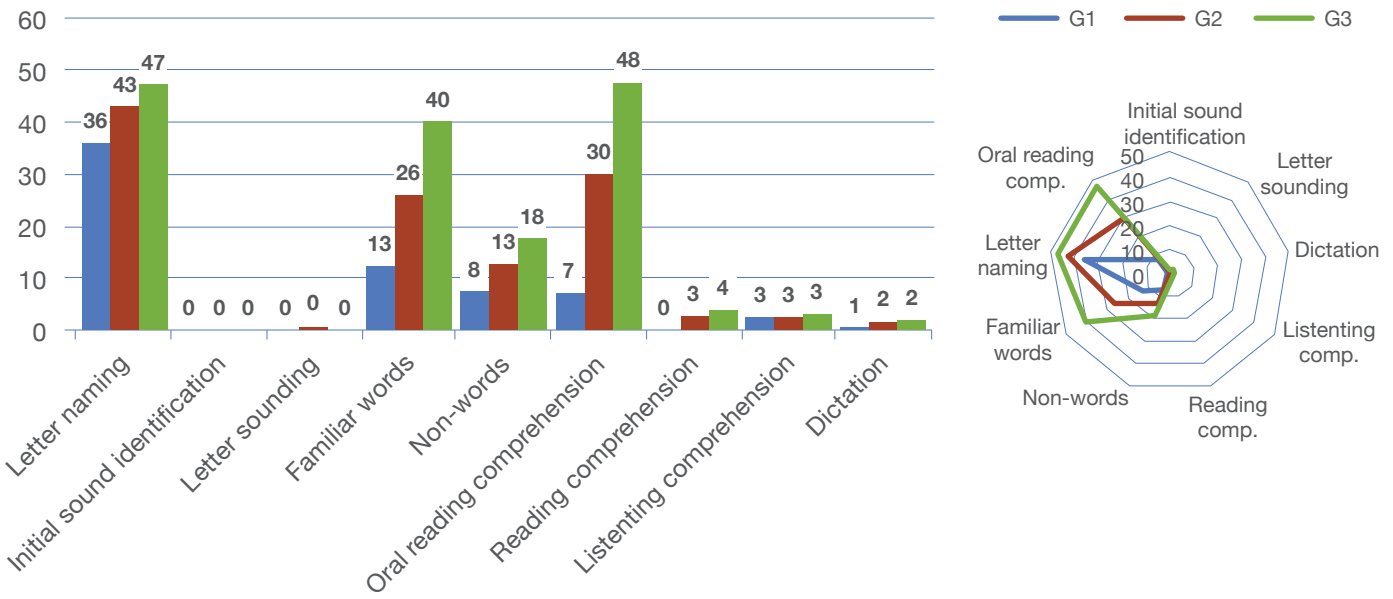
5.3 What does EGRA results indicate about early reading skills in Yangon?

The third objective of the analysis is to assess the appropriateness of the EGRA tool developed for this study to measure early reading skills. In other words, the analysis here aims to answer how, based on the EGRA results, Myanmar students fare in general and, relative to curricular expectations, in terms of early reading skills.

Overall, the study portrays a situation where, despite having a relatively low proportion of non-readers compared to other countries, the low mean scores indicate reading results that fall short of Myanmar curriculum expectations and point to particular difficulties with phonemic awareness, decoding, and inferential comprehension of text. This subsection examines students' performance on the basis of

- mean scores and percentage of zero scores per grade,

- international comparisons in the proportion of non-readers,

- curricular expectations and reading patterns of good and poor readers, and

- performance on questions about inferential versus literal comprehension.

Mean scores and percentage of zero scores per grade

Although mean scores on all subtasks improve with the grade, the substantial proportion of students who could not answer one item correctly and low mean scores on various subtasks by the end of Grades 2 and 3 remain worrying. Figures 7 and 8 illustrate the mean scores and percentages of students who could not answer one item correctly (called "zero scores") on each of the EGRA subtasks in each grade. (See Annex 3 for details statistics, such as standard error, number of observations, etc.) Across grades and for the rest of the skills tested, Grade 1 students had a larger proportion of zero scores, especially in reading comprehension and dictation. Grades 2 and 3 students showed significant improvement on these skills, but the proportion of zero scores remains worrying.

Across subtasks and grades, the largest proportions of zero scores appeared in the two problematic subtasks (initial sound identification and letter sound knowledge), even in Grades 2 and 3, which again suggest that these tasks are either too complex or too unfamiliar for students. Such a pattern points unambiguously to floor effects, which seem to indicate that the tasks fail to capture reliably the cognitive skills targeted. As a consequence, both tasks are excluded from further analyses.

Figure 8: Zero scores per subtask, per grade
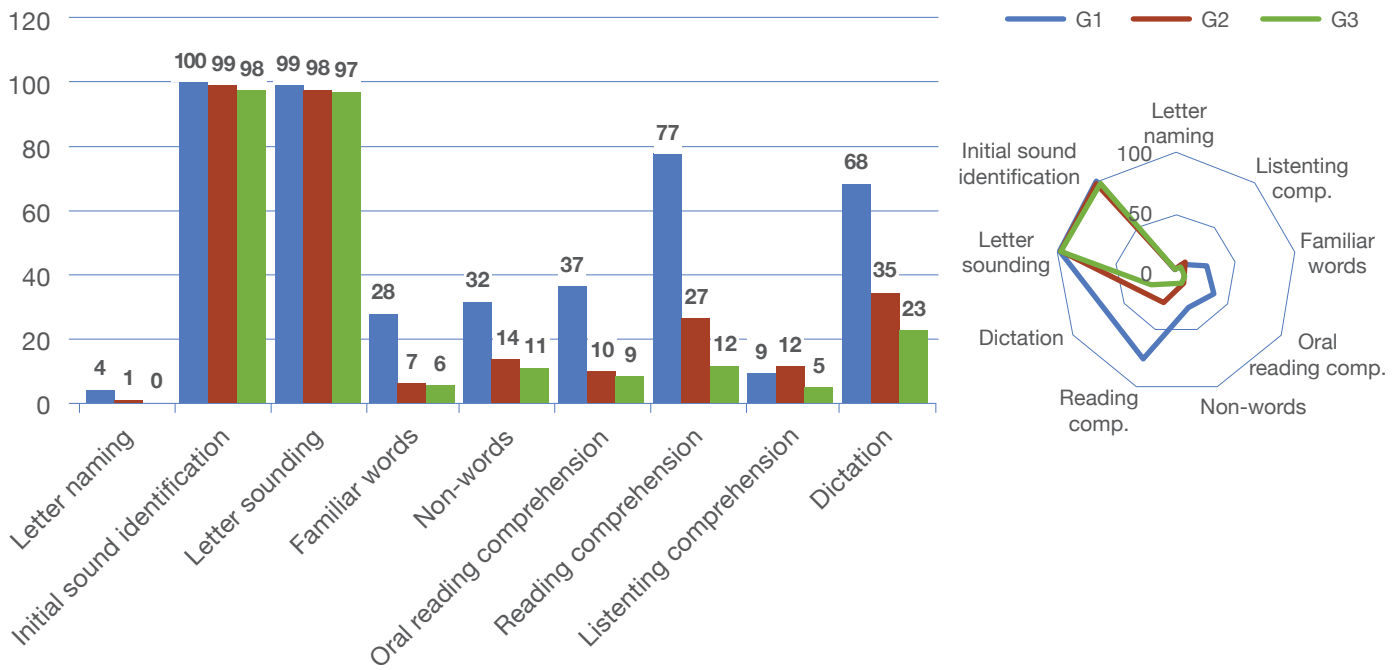
International comparison

When comparing the proportion of non-readers in Yangon with those in other countries, the picture is somewhat encouraging. As a general rule, differences in language structure and complexity (and differences in sampling frame and implementation protocol) introduce variations in the EGRA results that preclude direct comparisons. In particular, students tend to learn to read at a faster rate in languages with shallow or transparent orthographies than in languages with deeper or opaque orthographies.[56] However, because it can be assumed that the inability to read even a single word is relatively comparable across languages, zero scores on oral reading fluency can be compared and reported.

Figure 9 illustrates scores for students in Grades 2 and 3 in a number of countries. It shows that, **although it is in the best quartile, Yangon region is performing worse than the Philippines, Indonesia, Vietnam, Guatemala, Nicaragua, part of Honduras, the central region of Kenya and part of DRC.**

Curricular expectations and reading patterns of good and poor readers

Relative to the Myanmar curriculum in Grades 1, 2, and 3, analyzing the proportion and performance of good and poor readers indicate results well below expectations in all three grades and high correlation across subtasks. Given that no norms or references are available to evaluate and review EGRA scores, the following analysis breaks down and examines students' performance against national curriculum standards. It anchors the interpretation of the scores to Myanmar's local curricular expectations.

On the basis of curriculum expectations, students were assigned to four groups: non-readers, poor readers, readers with limited comprehension, and fluent readers. Currently, the primary education curriculum in Myanmar expects that students acquire these abilities and skills in language and reading:

- For Grade 1, students successfully read and understand a short connected text and correctly answer at least three questions.

- For Grade 2, students successfully read and understand a short connected text and correctly answer most of the six reading comprehension questions.

- For Grade 3, although curriculum expectations are different, students are benchmarked against the grade 2 curriculum.

---

[56] Transparent orthographies have more direct, regular correspondences between their sounds and letters. With deep or opaque orthographies, the correspondence between sounds and letters is less consistent, and letters or certain groups of letters often have different sounds in different words. See P.H.K. Seymour, L. Aro, and J.M. Erskine, 2003, "Foundation Literacy Acquisition in European Orthographies," *British Journal of Psychology* 94: 143-74.

Figure 9: International comparison of zero scores in oral reading fluency

Chart data (bars, approximate values read from axis):

Country/Region	Value
Phillipines - Manilla/Filipino - G3	1%
Vietnam - G3	2%
Phillipines - Manilla/English - G3	2%
Guatemala - Spanish - G3	4%
DRC - Bandundu/French - G4	4%
Indonesia - G3	5%
Nicaragua - Spanish - G2	6%
Honduras - Urban/Spanish - G3	7%
DRC - Equateur/French - G4	7%
Myanmar - Yangon - G3	8%
Kenya - Central - English - G3	8%
Myanmar - Yangon - G2	9%
DRC - Equateur/orientale - G4	10%
Laos - G3	12%
Kenya - Central - Kiswahili - G3	12%
Kenya - Central - Gikuyu - G3	14%
Iraq - G3	16%
Senegal - French - G3	17%
Kenya - Luo-Nyanza - English - G3	18%
Kenya - Luo-Nyanza - Dholuo - G3	19%
Afghanistan - Dari and pashtu - G2	20%
Phillipines - Mindanao/Filipino - G3	23%
Yemen - G3	24%
Kenya - Luo-Nyanza - Kiswahili - G3	25%
Honduras - Rural/Spanish - G2	28%
Guyana - English - G2	28%
Phillipines - Minsdanao/English - G3	30%
Laos - G2	32%
Nicaragua - Acoast/Miskito - G2	34%
Liberia - English - G2	34%
Ethiopia - Orromifa - G3	36%
Vanuatu - G2	37%
Timor Leste - G2	40%
Haiti - Creole - G2	47%
Haiti - French - G2	48%
Uganda, Central/Luganda - G2	50%
Uganda Central/English - G2	52%
Gambia -English - G2	53%
Mozambique - Portuguese - G3	56%
Pakistan - Urdu - G2	66%
Nigeria - Hausa - G3	79%
Nepal - G2	79%
Uganda, Lango/Lango - G2	82%
Mali - Bamanakan - G2	83%
Mali - Songhoi - G2	84%
Uganda, Lango/English - G2	88%
Pakistan - Pashtu - G2	91%
Mali - Fulfulde - G2	91%
Mali - Bomu - G2	93%
Mali - French - G2	94%
Malawi - Chichewa - G2	95%
Malawi - English - G2	98%

Sources: Authors' calculations for Myanmar; World Bank, 2013, "Report on Early Grade Reading Assessment: Grades 1 and 3 in SEQAP Schools (Vietnam), unreleased PowerPoint (Washington, DC: World Bank); M. Machuca-Sierra and S. de Silva, 2014, "Draft Baseline Report for Laos" and 2012, "Lao P.D.R. EGRA Survey Report" (Washington, DC: World Bank, East Asia and the Pacific Education Unit); eddata//, https://www.eddataglobal.org/reading/; and A. Gove and P. Cvelich, 2010, "Early Reading: Igniting Education for All—A Report by the Early Grade Learning Community of Practice" (Research Triangle Park, NC, USA: Research Triangle Institute).

Figure 10: Proportion of students per reading ability grouping

Percent

| | Non-readers | Poor readers | Limited comprehension | Fluent readers |

G1: Non-readers 37%, Poor readers 49%, Limited comprehension 12%, Fluent readers 2%

G2: Non-readers 10%, Poor readers 17%, Limited comprehension 41%, Fluent readers 32%

G3: Non-readers 9%, Poor readers 3%, Limited comprehension 33%, Fluent readers 55%

"LIBRARIES HAVE A SIGNIFICANT, INDEPENDENT POSITIVE INFLUENCE ON STUDENTS' READING PERFORMANCE, ACCORDING TO THE STUDY."

Given these curriculum expectations, the study examined the continuum of students' reading skills assessed by EGRA and assigned students to four groups:

1. ***Non-readers:*** Students who do not read a single word of the connected text (ORF = 0).

2. ***Poor readers:*** Students who read a portion of the connected text (ORF > 0), but cannot correctly answer a single question of the reading comprehension task.

3. ***Readers with limited comprehension:*** Students who read a portion of the text (ORF > 0), but only answer correctly a maximum of two questions (Grade 1), four questions (Grades 2 and 3), or fewer questions of the reading comprehension task.

4. ***Fluent readers:*** Students who read (a portion of) the text (ORF > 0) and answer correctly at least three questions (Grade 1) or five questions (Grades 2 and 3) of the reading comprehension task.

The proportion of fluent readers increased across the three grades, but a significant proportion of students remained in the weak reading ability groups, even in Grade 3. Figure 10 illustrates the percentage of students in each of the reading ability groups. Among first graders, hardly any students met the curricular expectations of successfully reading and understanding the short connected text used in EGRA. Only 0.7 percent of Grade 1 students were capable of reading the grade-appropriate text presented and responding correctly to at least three questions. In Grades 2 and 3, only 23 percent and 48 percent of the students, respectively, reached curricular expectations and could be considered good readers.

Figure 11: EGRA scores for good and poor performers in Grades 1, 2, and 3

Grades 1 / Grades 2 / Grades 3

Legend: — Poor — Good

X-axis categories (all three charts): Letter Name, Familiar-Word, Non-Word, Oral Text, Reading, Listening, Dictation-Word
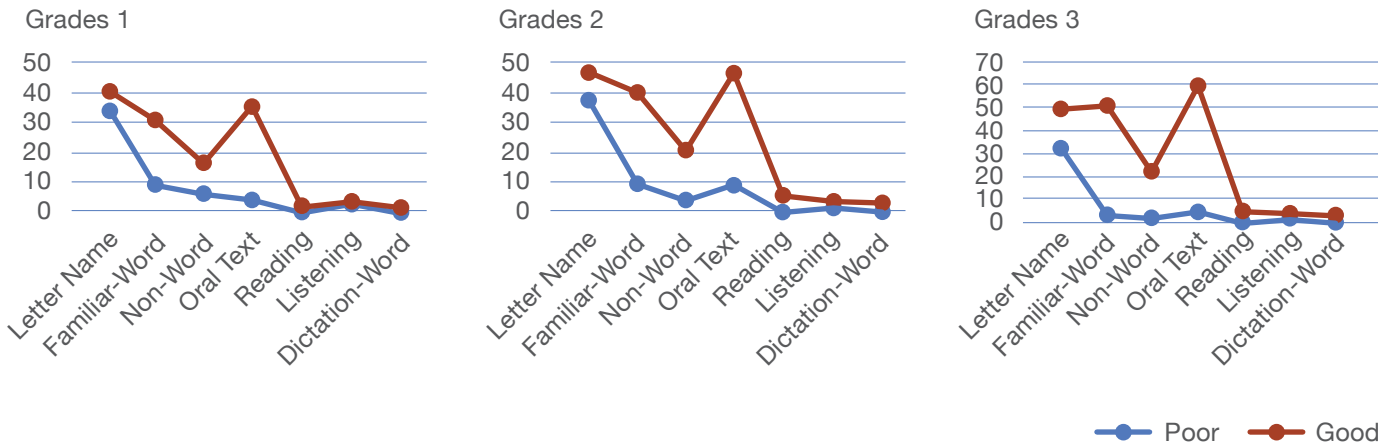
It is also notable that the prevalence of fluent readers versus limited readers versus weak readers (which includes non-readers and poor readers) exhibits different patterns according to grade level. In Grade 1, there were far fewer fluent and limited readers (14 percent) than weak readers (86 percent). In Grade 2, only a minority of students (32 percent) met the criteria for fluent reading performance and 27 percent of students were weak. While all third-graders were expected to perform well against the Grade 2 curriculum, only 55 percent of Grade 3 students met this standard, and 12 percent were weak readers.

Performance of good and weak readers is relatively consistent across the EGRA subtasks, with weak readers performing worse than good readers. With the students classified by reading ability, it is interesting to explore the performance of the average student in these groups on the different subtasks. Figure 11 provides an overview of good readers' performance relative to poor readers and non-readers. (Fluent readers in Grades 2 and 3 were considered as a single category; in Grade 1, however, fluent readers were combined with readers with limited comprehension, since there were too few fluent readers for a meaningful analysis.) In all three grades, students categorized as good readers scored substantially higher than any of the weak reading groups.

Although no inference or causal link between tasks can be made at this stage of the analysis, these results indicate that successful readers who reach or approach curricular standards for their grades performed well on all component skills of reading assessed by EGRA. On the other hand, students identified as the furthest from meeting the standards—because they could not read or could not comprehend the text—scored in the lowest range on EGRA and exhibited difficulties performing the tasks.

Performance on literal versus inferential comprehension

Deepening the analysis by looking into scores on literal and inferential comprehension suggests that students in Myanmar are capable of processing and understanding inferences when language is presented orally, but they find it more difficult when the inferences are text-based. The comprehension subtasks (reading and listening) of EGRA include questions that assess two important processes of comprehension: the capacity to understand and respond to literal questions and the capacity to understand and respond to inferential questions. Table 3 summarizes the students' scores for retrieving literal and inferential information when 1) reading the text themselves and 2) when hearing the text read to them.

Students in Grades 2 and 3 made more errors when drawing inferences from printed text than with retrieving literal information. However, listening comprehension showed a reverse pattern: students were better able to answer inferential questions than literal ones. This result suggests that students in Myanmar are capable of processing and understanding inferences when language is presented orally, but they find it more difficult when the inferences are text-based.

This pattern of results was also observed in students who were presented with all six questions and five of the six questions (see Tables A3.4–3.6 in Annex 3). This suggests that the results are not an artifact from the administration of the EGRA's reading comprehension task. (Only questions about the portion of the text read were asked.)

Table 3: Percent of correct responses in reading and listening comprehension for Grades 1, 2, and 3 (literal and inferential questions presented separately)

	Reading comprehension		Listening comprehension	
	Literal question	**Inferential question**	**Literal question**	**Inferential question**
Grade 1	5%	1%	58%	30%
Grade 2	51%	37%	38%	56%
Grade 3	72%	55%	49%	70%
Total N (%) of questions	4 (100%)	2 (100%)	4 (100%)	2 (100%)

5.4 What factors explain differences observed in early reading skills?

The fourth and final objective of the analysis investigates the potential explanatory factors for differences in EGRA scores in the Yangon region (DBE 3). It attempts to answer what can potentially explain the observed difference in early reading skills.

The results suggest that the Ministry of Education should prioritize interventions for a few specific factors—regardless of the students' original or initial skill levels—when aiming to improve reading skills. Interventions with potential for adding value could include;

- more efficient use of Myanmar exercise books by teachers, students, and even parents;

- development and dissemination of best practices for use with the exercise books;

- the issue of teachers taking second jobs; and

- the presence of a library in schools or a book corner in classrooms.

One of the main puzzles requiring further research is the persistent and strong counter-intuitive relationships between the use of both summative (e.g., use of chapter- or year-end tests) and formative (e.g., use of tests for feedback to teachers on effectiveness of their instruction) assessments of student performance. It is unclear from the findings whether tests students reflect on what they have learned or whether teachers use test scores to guide their curriculum or understand the progress of students. Finally, regarding the general discourse on teaching methods in Myanmar, the data suggests that the research base may not be sufficient to advocate for or against specific techniques.

The text that follows is necessarily shortened and distilled to highlight the results of the bivariate and multivariate analyses for the non-expert reader. Please note that Annex 5 contains the rigorous discussions of the methodology and bivariate and multivariate analyses.

Overview of the methodology

In order to identify potential factors that explain the differences across schools, grades, and students in early reading skills (as measured through EGRA scores), the correlation between EGRA scores and student and teacher characteristics are analyzed first in a bivariate context (as a preliminary and preparatory stage) and second in a multivariate context, using a set of three-level hierarchical mixed models (students nested within grades and grades nested within schools). The first step combined the results of the subtasks for early reading skills into a composite EGRA score. Then, in order to identify potential factors that might explain the differences in early reading skills (measured by EGRA scores) across schools, grades, and students, the correlation between EGRA scores and student/teacher characteristics were analyzed in two ways. First, a bivariate analysis examined school and non-school variables against the EGRA scores. Second, a multivariate analysis looked at several variables together, using groups of variables (A-G) and a set of successive mixed models in a three-level hierarchy (students nested within grades and grades nested within schools).

The bivariate correlation analysis looks at the one-to-one relationship (or correlation) of the variables captured in the study to higher/lower overall reading performance—the composite EGRA score. However, those interventions that can help students' performance or school/teacher effectiveness cannot be directly determined from just the bivariate results. Indeed, interpretation of bivariate results does not take into account that some of the study variables, for example, associated with a school or teaching, may also reflect (and are therefore correlated with) non-school factors, such as students' socioeconomic status or the skills that students have before they start primary school. In other words, bivariate relationships may be the result of many factors and may results from a reversal of causality—a cause and effect that may be the opposite of a commonly held presumption. For example, schools that use chapter- or year-end tests may have introduced them to check student performance. But a common factor could underlie both variables. For another

example, low-performing schools might use a different means of assessing students (e.g., as an intervention to help them monitor students) that not all schools use. However, a bivariate analysis is a crucial step because it helps identify the variables that are consistently correlated with the EGRA scores; and thus are necessary to—and included in—the multivariate analysis that follows.

In order to analyze the correlation among multiple potential explanatory factors (and ensure some control for the relationship), a multivariate analysis with a three-level hierarchical mixed model is also necessary. This included grouping variables into sets (A–G), related to areas important to the study, and establishing five "successive" models to help remove less relevant variables.

The reason for using this hierarchical structure modeling technique is to uncover which groups of "intervenable" variables (e.g., education inputs and instructional practices) can improve the differences in students' reading performance (high/low EGRA scores) that may be due to socioeconomic characteristics and non-school factors. Identifying such variables can help identify and design interventions to improve early reading skills in Myanmar classrooms. This approach can also help explain 1) why schools differ from each other in average reading performance, and 2) why students within the same schools differ from each other in reading performance. The goal of this study and this analysis is primarily concerned with understanding differences between schools to aid the Ministry of Education in identifying productive interventions.

Bivariate analysis

Many of the relationships between the characteristics surveyed (identified in the bivariate analysis) are intuitive; in other words, the relationships seem obvious and based on reasonable presumption. The analysis uncovers intuitive positive correlations (both variables increase or decrease) for female status, wealth, adult support (for reading and homework), prior enrollment in preschool, use of a Myanmar language exercise book, and availability of library. Expected negative correlations (one variable increases as the other decreases—an inverse relationship) appear for teacher travel time and teachers with other jobs. However, the analysis also reveals several non-intuitive (or counter-intuitive) relationships, such as the positive relationship for absence due to sickness and the negative correlations with the use of summative assessment and the use of assessment results for modifying instruction. The characteristics (or variables) discussed below offer the clearest findings in the bivariate analysis. These findings are both positive and negative in relation to the composite EGRA scores. However, some of the results are puzzling and need further investigation.

Remoteness and wealth

Unsurprisingly, being enrolled in a remote or non-remote school is significantly and positively correlated with the EGRA composite score overall, as well as for each grade. Students in less remote schools are more likely to read better and have higher EGRA composite scores than are students going to schools in more remote areas. Wealth is also positively and significantly, although weakly, correlated with scores. The less well off a student's family is, the less likely they will read better.

Gender and class size

Bearing in mind the caveats regarding the sample— more male students are in the study than female students, due to the fact that boys are listed in class rosters before the girls—being a female student appears to be positively and significantly correlated with a higher EGRA composite score. However, this difference disappears in Grade 3, possibly indicating that the boys are catching up or that the gender/ability mix has changed as a result of differentiated drop-outs. Class size is uncorrelated with performance.

Language

Despite the low variation in the language spoken by students, parents, and teachers in the study, the results point to some possible positive correlation between EGRA scores and speaking the Myanmar language at home or being taught by a native Myanmar speaker. The high number of households speaking the Myanmar language (and fewer households using a different primary language) implies that it is difficult to robustly estimate the correlation between students' language and EGRA scores. This is because the proportion of non-Myanmar languages in the sample was quite small, but included numerous languages.[57] This is also the case with language-related variables for teachers.

Remembering that these results are only indicative and explorative, the bivariate analysis points to some positive correlation between speaking the Myanmar language at home and being a stronger reader. Living with an adult (parents, guardian, and grandparents) who can read the Myanmar language fluently also seems to positively relate to higher EGRA scores. In addition, teachers whose first language is Myanmar have higher performing students, which may indicate that these teachers are better at teaching this specific language.

Class types, teacher demographics, qualifications, training, and years of experience

---

[57] In order of frequency, the non-Myanmar languages were Kayin (2%), Shan (0.6%), Chin (0.5%), Rakine (0.4%), Kachin (0.2%), Mon (0.2%), and others (0.2%).

As often seen in education literature, few of the basic teacher characteristics—such as gender, qualifications, experience, etc.—are significantly correlated with student EGRA scores. Pre-service training on how to teach reading is the notable exception: it is positively correlated with EGRA scores, but at a low significance level. In-service training on reading instruction is oddly not correlated with scores, possibly indicating the low quality or relevance of this training. Concern over this finding is heightened when seen in conjunction with another result showing that being taught by a teacher who has another income-generating job is significantly correlated with lower EGRA scores.

Preschool and absenteeism

The results indicate a positive correlation between EGRA scores and preschool (as expected), as well as with student absenteeism (unexpected). Going to preschool and the duration of preschool attendance are both significantly correlated with higher EGRA scores. Absenteeism, as reported by students, is surprisingly not significantly correlated with the EGRA score. However, the absenteeism rate on the day of the study enumerators' visits showed—counter-intuitively—a positive correlation with performance, although at a very low level of confidence. This positive correlation may be an artifact of teachers asking the weakest students to stay at home on the day of the assessment, which was announced ahead of time.

Teaching and learning material

Owning a Myanmar reader, having access to other reading material at home, and having an exercise book are all significantly and positively correlated with EGRA scores, as are the extent of teachers' markings in the exercise books. The use of teaching materials (privately published) other than the Myanmar reader (reported by teachers) is negatively correlated with EGRA scores with Pimya materials and positive with Pan Taing materials. This possibly indicates different values of these complementary resources. The presence of reading resources in the school (or library, etc.) is also positively correlated with scores, but not the frequency of use.

Homework, family involvement, and outside tutoring

Getting outside tutoring and help at home was predictably positively correlated with scores, but the frequency of homework does not seem to matter. Given the students' reading skills in the sample, it is quite troubling that the frequency of homework, as reported by students or teachers, is not significantly correlated with scores. If students get help with homework, the result is positively correlated with EGRA scores, but the proportion of parents reviewing the homework (as perceived by the teacher) is not. The variable for students who get outside tutoring (paid or not) in reading or in the Myanmar language was also significantly and strongly correlated with EGRA scores. The possible usefulness of such tutoring sessions hints that they compensate for what may not be adequately taught in school.

Reading activities, pedagogical approaches for teaching reading, and assessments

None of the activities or instructional methods used by teachers is significantly correlated with EGRA scores, which is troubling on several levels, but needs more in-depth investigation before drawing conclusions. Using chapter- and year-end tests (summative assessments) is, intriguingly, negatively correlated with EGRA scores. Use of other assessments is not significantly correlated with EGRA scores. Every purpose for using the assessments is also negatively correlated with scores, but only a few are significant.

Multivariate modelling

For brevity and relevance, the multivariate analysis includes only a subsample of the student, teacher, and school characteristics. The full discussion of the multivariate analysis is in annex 5. Here only the clearest results are included, plus a cursory introduction of the elements of the multivariate analysis. To make selection easier, the explanatory variables were grouped into specific sets, namely, areas important to the study:

Group A—explanatory variables that lie outside the scope of schools or government to affect

Group B—family and home variables that have an effect on school outcomes or variables where the Ministry of Education could have some influence (e.g., through communication campaigns)

Group C—variables that can account for students' initial skill levels when they enter primary education (which can also serve as a proxy for how ready the students are for primary school)

Group D—variables that represent student-teacher interactions in a classroom setting (such as tests, marking exercise books, homework, reading activities, etc.)

Group E—school variables that represent education inputs (use of textbooks and types of textbooks) and other elements affected by education policy

Group F—variables that represent different methods for teaching early reading skills, such as letter naming and repetition

Group G—variables that represent teaching methods appropriate for more advanced reading (at primary level), such as answering comprehension questions and creating new words

The correlations for variable sets A to E are almost identical for all grades, even though these variable groups include a wide range of household, personal, and contextual conditions. This consistency of correlations suggests the presence of commonly observable phenomena. Indeed, as seen with the bivariate analysis, many of the same variable relationships are intuitive, and the same counter-intuitive relationships also appeared, which emphasizes the need to include multivariate analysis.

When analyzing the results for sets F and G (variables related to early versus advanced reading instruction, respectively), it is unclear whether education practices are responding to students' needs, rather than driving student performance (reverse causality). One pattern in the analysis of set F—teaching methods and activities for students learning to read—shows positive correlations for Grade 1 students (as expected), but negative (surprisingly) for Grade 3 students, with Grade 2 students generally in the middle. This pattern is consistent with teachers using instructional methods appropriate to younger students when they are teaching older students with poor reading skills. On the other hand, though, the pattern of results (although less strong) seen with techniques for more advanced readers (set G) have a positive correlation for Grade 3 students and weaker (or negative) correlations for students in Grades 1 and 2. In either case, these patterns serve as a warning against simplified interpretations of the findings and support the necessity for more conclusive follow-up studies or more focused experimental research.

Analysis of the variance decomposition

Deeper investigation of the multivariate analysis (variance decomposition between the different successive models; see Annex 5) indicates that the variation between test scores is explained mostly by student-level variables, a proportion by school-level variables and almost none by grade. Knowing which school a student attends is five times better at predicting a student's reading performance than knowing what grade a student is in. Other results may also be from mixes of school and non-school variables, such as whether the potential benefit of attending preschool is already reflected in student/household characteristics, why the use of a Myanmar exercise book plays a key role, and why teaching methods and reading activities seem to have little effect on scores.

Coefficient analysis

With both bivariate and multivariate analyses, caution is required when interpreting the results. Despite providing statistical explanation, they do not necessarily explain the underlying phenomena. The explanatory variables in the model may simply be highly correlated with a phenomenon that happens to be highly correlated with school effectiveness, for example, without actually having any relationship to school effectiveness. For this reason, it is useful to examine the variable coefficients (a numerical measure of the variables) themselves, as the patterns may suggest what some of the true underlying causes may be. However, in many cases, the variables represent proxies for conditions that the study cannot directly measure, despite a wish to. Accordingly, the magnitude of the effect (size of the coefficient) in the results should not be directly interpreted. Rather, they should

be understood as the lower bounds of estimates that indicate the direction of association between the real world conditions they represent and the outcome of student reading performance.

For the most part, the results of this multivariate analysis reflect the same observations as did the bivariate analysis, although with some differences— namely the lack of correlation between scores and adult help with homework, teacher travel time, and enrollment in preschool. The effects on EGRA scores of all the student-level variables also vary significantly across schools. This indicates that, despite consistency of the effects across grades, there may still be ways for schools to ameliorate the effects of students' background. Being a female and degree of wealth are still relatively strong predictors of higher reading scores. Interestingly, the effects of adults reading and outside tutoring are relatively strong and positive, which may explain why adult help on homework shows little effect. Teacher travel time, which also had a strong (but negative) bivariate relationship with EGRA scores, had little effect when controlling for other variables. The most likely reason is that the effects of travel time are explained by socioeconomic factors related to school surroundings (already captured by student/household characteristics).

Of particular note in all models is the large effect of teachers having another job, which has a strongly negative relation to student performance. The negligible effects of attending preschool indicate that preschool itself is likely strongly correlated with other strong predictors of performance, such as family background. In the model with student, household/family, and education variables, there are relatively strong effects related to use of Myanmar exercise books. Although owning an exercise book has a significant relationship by itself with reading performance, each additional variable describing the use of the book (e.g., quantity of use, teacher feedback, and original stories written by students) is associated with additional performance increases.

Interpretation

The results of this analysis indicate that most of what is considered school effects[58]—or how well a school teaches reading skills to students in early grades—is likely influenced by non-school factors, such as the socioeconomic status of students. These factors have little to do with educational practice and can hardly be affected by the Ministry of Education. Indeed, non-school factors account for half of the variation in average school performance. However, this does not mean that these factors cannot be countered. Indeed, the Myanmar data indicates that the effects of student characteristics, despite being relatively consistent, do vary significantly from school to school. Some schools are providing environments that ameliorate the effects of gender, family background, and

---

[58] In other words, the differences in average EGRA scores between schools.

home educational support. It is also possible one of the factors contributing to a healthy environment for student learning is supporting the teachers. The study showed that many teachers must work other jobs, which predictably has a negative relationship with student performance.

It can be easy to infer a cause where there is none using survey data such as the data from the EGRA questionnaires. As in a number of cases with the Myanmar EGRA data, one-to-one relationships that appear stable and interpretable across distinct groups of students may be false. The clearest example of this phenomenon is the relationship between teachers' instructional practices and students' reading performance. After analyzing the relationships, the data most likely shows the results of teachers reacting to existing student skill levels, rather than students developing skills as a result of the teachers' practices.

Nevertheless, the data indicates several specific directions for future inquiry. The results also point to many factors where the Ministry of Education could introduce interventions to help improve students' ability to read, regardless of their original or initial skill levels.

Notably, it appears that effective use of Myanmar exercise books may be the most promising intervention. However, it is unclear from the data what effect is more prevalent: does the use of the exercise books help students read better or are students more likely to use the exercise books if they already have strong reading skills? In this case, the reality may be a vicious circle, where only the students who read better can take advantage of a learning aid that accelerates their progress. In addition, more benefits are possible from developing and advocating best practices for teachers to follow with the Myanmar exercise book.

Another recommendation suggests that the Ministry of Education look into the issue of teachers having a second income-generating job, in addition to teaching. It would be helpful to determine if and how students are negatively affected by teachers who must work another job in addition to teaching. Also useful is exploring the mechanisms that already exist to address teachers' needs, motivations for taking a second job, and mean of mitigating any negative effects.

Libraries have a significant, independent positive influence on students' reading performance, according to the study. This points to the need for further research on how much the presence of a library benefits student learning and what a cost-effective strategy for distributing this benefit might look like.

The results also indicate several puzzles that require further investigation, mainly the persistent and strong inverse relationships between both summative and formative assessments for student performance. It is unclear whether the tests accurately reflect what the students have learned (how well they can read) and whether teachers apply students' test performance to making their teaching more effective. The data does not suggest what is underlying this phenomenon, other than to confirm that the relationship exists and is likely not explained by non-school factors, such as students' socioeconomic characteristics or initial skills.

Finally, the Myanmar EGRA data suggests that the research base may not be sufficient to say which specific teaching techniques are the most (or less) effective with students. Moreover, given the results of this current study, the survey data may be inadequate to draw such conclusions unless the data also sufficiently captures the initial skill levels of students before they face the specific instructional practices common in the early primary grades.

"THE MINISTRY OF EDUCATION COULD PROFITABLY LOOK INTO BUILDING ITS OWN CAPACITY TO CARRY OUT SURVEYS AND ANALYZE FINDINGS."
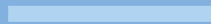
6. SUGGESTED RECOMMENDATIONS FOR POLICY-MAKERS INTERESTED IN IMPROVING EARLY GRADE READING SKILLS

The analysis and findings presented in Section 5 represent a significant step toward systematically measuring and better understanding the learning outcomes (for reading) in the early grades in Myanmar. While specific recommendations based on one round of one type of learning assessment should be made with care, a framework of potential actions can be sketched out and proposed to the Ministry of Education, policy-makers, and their development partners for consideration. Possible responses to the study can be organized into three main categories.

1. **Expanding the research and linking it to complementary research and analysis**

- *The EGRA instrument, research protocol, and reporting need to be improved.* This report has identified specific ways that the EGRA instrument and research protocol can be improved (see Section 5.2). Many of these changes have already been incorporated in the second round of assessing early grade reading that took place in four states and regions in January 2015.[59] Further work is needed to resolve data and sampling issues, and improve the robustness of the reliability analysis by using alternate approaches, such as test-retest, addressing the technical issues with the two subscales that didn't work well for the Yangon exercise. This should help make the reporting process for future rounds of assessment more efficient. But the delays with this first report (just now being finalized 16 months after the January 2014 survey) were not only

due to data issues; the World Bank needs to do a better job of expediting the reporting and analysis.

- *The Ministry of Education could profitably look into building its own capacity to carry out surveys and analyze findings.* The capacity building process is already off to a good start as the Department of Myanmar Education Research (DMER) took the lead—working closely with the Departments of Basic Education, the states, and the regions—in organizing the field work for the second EGRA survey in January 2015. However, DMER would benefit (and more easily accomplish its mission) from developing more capacity—specifically upgrading skills and hiring analysts—so that it (or its successor) can process and analyze data, and draft analytical reports independently.

- *Coverage of the research needs to be expanded.* Given the difficulty with sampling and implementation in the Yangon area (former DBE 3), it is not recommended that the Ministry of Education move rapidly to a nationally representative survey, which would be costly and difficult to manage. A nationally representative survey could best be a medium-term objective (3-5 years) of the Ministry. In the meantime, it has already expanded the research to two townships each in four additional states and regions in Myanmar. Based on the findings from this second EGRA round, the Ministry can decide whether (and when) to expand it to all townships in selected states and regions, and eventually to all states and regions.

---

[59] A second round of assessment was prepared in 2014 and conducted in the field in 2015. Several of technical issues noted in this Yangon assessment were rectified including 1) coverage of areas where language is more diverse to assess its relationship with literacy, 2) modifications to the instructions protocol of subtask 3 to weaken the floor effect, 3) improvement of the questionnaires to capture more variations in most relevant school, teacher, and student characteristics, and 4) addition of a classroom observation module.

- ***The Ministry of Education could roll out EGRA in some of the other ethnic languages by working with local representatives.*** This process has already started. The instructions for the EGRA protocol (not the items of the subtasks yet) were adapted for four ethnic languages in the second round of EGRA in 2015. Because children learn to read the Myanmar language in government schools, the assessment components of EGRA remain in the Myanmar language. Adapting the instrument to other major languages could be a medium-term objective and be carried out in close collaboration with ethnic-based, non-government authorities.

- ***Broadly sharing EGRA findings could benefit the Ministry of Education reform initiative in several ways.*** To date, reports and workshops on the findings have been limited to government officials at various levels and the participants in the survey. The Ministry of Education could work with development partners to make the public more aware of its initiative to measure and promote reading skills, and communicate the importance of reading. Myanmar could also participate in international conferences and workshops that focus on assessing students' progress in early primary grades. School-level EGRA results could also be shared with parents in order to improve school accountability.

- ***The EGRA research can offer further insights if it is linked to other assessments on learning outcomes and research on the effectiveness of teaching.*** An oral survey of early grade reading skills is only one picture of learning effectiveness. Written forms of early literacy administered in groups (such as the "Provinha," the children's literacy assessment used in Brazil and in Mozambique) could help bring down the costs. Moreover, the EGRA findings can usefully be analyzed in terms of whether they are consistent with other assessment initiatives for upper grades (currently carried out by UNICEF) or with the Ministry of Education's own examination system.

- ***The findings could also be linked to measures of teaching effectiveness (for example, classroom observations or assessments of teachers' ability to teach reading skills)*** to investigate the impact of child-centered teaching practices promoted by the Ministry of Education and various development partners. The first attempt has already started in early 2015.

- ***It would be interesting to use the EGRA work as a foundation for creating a more institutionalized assessment system.*** It could also support the assessment-related requirements likely to come from the upcoming post-2015 Sustainable Development Goals (SDG) and Education for All (EFA) goals. These will include one or more learning goals, including literacy and numeracy, and related targets and indicators.

- ***The Ministry of Education can further expand its knowledge base on student performance by launching similar research in the other core primary education subject—math.*** The early grade math assessment that goes along the EGRA could be adapted and used for this purpose.

2. **Using the findings to begin establishing measurable reading standards for the early grades**

- ***Myanmar would be well-served by developing specific standards (sometimes called competencies) for reading in all primary grades.*** One of the main objectives of this research is to support the establishment of specific standards for reading so they can be embedded in Myanmar's curriculum for primary grades. The findings in this report (and those to come) can help the Ministry of Education define, continually measure, and update reading standards in the primary grades. For example, the number of letters or words pronounced correctly in 60 seconds or the number of words of connected text read correctly in 60 seconds could be used by schools to monitor the performance of all early primary grade students in Myanmar. The number of questions answered correctly about a paragraph or text (reading comprehension) could also be monitored and reported on.

- ***The EGRA findings can guide the Ministry of Education in developing specific, short diagnostic tests (or classroom-based assessments) that teachers can use to measure and report on the reading performance of all their students.*** In the design and implementation of standard diagnostic tests, it is recommended that reporting stay school-based at the beginning. Gradually this reporting can be linked into a state-, region-, and national-based education information system as a medium-term project. School-based reporting is also recommended as a first priority, so that school headmasters are aware of (and able to report on) which students are reading at grade level, below grade level, or above grade level, as measured by the Ministry of Education's reading standards.

- **All in-service and new teachers need training in measuring and reporting on reading standards.** Helping all of Myanmar's roughly 300,000 primary school teachers monitor reading performance and report on performance is a medium-term goal. If kept simple and well-designed, training teachers to measure reading performance quickly and efficiently is an achievable objective for Myanmar.

3. **Using the findings to help design and justify specific interventions aimed at improving reading outcomes**

- *The Ministry of Education and other government-actors can use EGRA as the basis of a public awareness campaign to promote reading.* The Ministry of Education can go much further than workshops or conferences in using the basic findings of the EGRA analysis. One idea is to design a multi-media campaign aimed at promoting reading and making the public aware of the importance that all children in Myanmar become competent readers at the end of primary education. Development partners can provide financing and advice, but the initiative is not complex and can be undertaken by the government.

- *The Ministry of Education can immediately start the process of defining standards (or competencies) and linking them to the curriculum.* Establishing reading standards does not require major financial investment or significant external technical assistance. The Ministry currently has the capacity, in cooperation with local academics, to define simple reading standards, design simple diagnostic instruments, and to pilot-test them in schools. This action is highly recommended because it is a prerequisite for training teachers to use and report on these standards—a medium-term investment.

- *Once developed, the standards can be integrated into the pre-service curriculum at education colleges; also the Ministry of Education can consider supporting an off-site program for in-service and school-based training.* Professional development for teachers in reading instruction is most valuable when contingent on the establishment of standards or competencies for reading by grade. Once this process has started (it does not have to be complete), the Ministry of Education and its development partners can invest in designing and establishing off-site training programs and school-based networks aimed at disseminating the standards and diagnostic tools. In-service could be best linked to a new cadre of township-based master teachers, who could be responsible for classroom observation and ongoing teacher support. The integration of standards

(or competencies) and their measurement in pre-service teacher college curriculums is a separate exercise that should be approached in parallel. Development partners can provide financing through Ministry of Education training budget codes and help with the monitoring and evaluation of results.

- *Schools need more reading materials.* In connection with the public awareness campaign, the Ministry of Education can think about launching—with development partner assistance—a program that encourages parents to read to their children daily and offers incentives to children's book authors, especially local authors, to write more books. A competition could be organized and books could be selected by panels for the Ministry of Education to purchase, print, and distribute, or to allow schools to purchase directly from authors or publishers using school grant funds. Gradually, schools should become responsible, using school grants and their own resources, for purchasing and maintaining classroom book corners and school libraries. Finally, before encouraging greater use of any textbooks, it would be useful to review them for quality and appropriateness in order to identify possible issues that need to be fixed beforehand.

- *Let the Ministry of Education lead the development of systematic monitoring and evaluation.* Investing in EGRA is an example of the type of monitoring and evaluation that should be systematic in any modern education system. EGRA, complementary assessments, and teacher observation can best be carried out under the leadership of the Ministry of Education, with the support of its development partners. This is an ongoing investment in reforming the education system in Myanmar through improving learning outcomes, learning how to continuously improve outcomes, and reporting on learning outcomes. One example is to use pre- and post- EGRA results as part of a rigorous impact evaluation of pilot interventions on improving early reading skills.[60]

Table 4 summarizes the suggested recommendations according to the likely timeline of their realization and responsibilities.

---

[60] More than 40 countries have successfully used pre- and post EGRA to rigorously assess the impact of specific interventions on early reading skills, such as Liberia, the Gambia, and Mozambique (www.eddataglobal.org/countries/index.cfm).

Table 4: Suggested recommendations and timelines

Recommendations	Description	Department or institution responsible
Short term (1-2 years)		
Improve EGRA tool, research protocol, reporting	- Resolve data and sampling issues - Improve robustness of reliability analysis - Redesign the two problem subtasks to fit Myanmar context	DMER
Build capacity to conduct surveys and analyze findings	- Already begun with January 2015 survey - Work with development partners to support and design training	DMER, Departments of Basic Education, Development partners
Expand EGRA to other regions	- Already begun with January 2015 survey in 4 different states/regions	DMER, DBE
Share EGRA findings	- Share school-level results with parents to promote accountability - Use findings to encourage government and teacher participation in international conferences and workshops on assessing student progress	Ministry of Education
Start process of developing specific standards (competencies) for reading	- Use findings from first two EGRA surveys to start establishing specific standards for reading - Promote a culture of continually reviewing, defining, and updating standards for reading	Ministry of Education
Begin to link reading standards to curriculum	- Review early grades curriculum for quality and appropriateness in meeting reading standards and change accordingly to link the standards to curriculum	DMER, Development Partners
Develop school-based diagnostic assessments for monitoring student performance in reading	- Design short, periodic assessments for teachers to administer in class - Make reporting by teachers/school officials mandatory and easy	DMER, DBE
Offer professional development to teachers to learn to measure and report on students' reading performance	- Develop off-site training programs and school-based networks for teachers or school clusters - Promote dissemination of tools and information through school-based networks or school clusters	Development partners, Ministry of Education, education colleges, local school networks
Promote a multi-media public-awareness campaign about reading	- Use public-awareness campaign to emphasize importance of learning to read in primary grades - Encourage parents/guardians/adult family members to read to children	Ministry of Education, other government bodies, Civil Societies, local NGOs, DPs
Increase access to reading material	- Offer incentives to authors to write more children's books - Offer grants to encourage schools to purchase materials and maintain libraries and classroom book corners	Development partners, Ministry of Education,
Integrate reading standards into education colleges (pre-service)	- Design and pilot courses (pre-service) and workshops (in-service) to train teachers on best approaches to teaching reading, effective use of textbook, and assessment and reporting to track students' progress	Education colleges

Link EGRA research to other types of assessments	- Consider group-administered written assessments of literacy, such as "Provinha" used in Brazil and Mozambique - Design primary grade assessments to link to—or be consistent with—higher grade assessments - Link assessments of student achievement to assessments of teachers' effectiveness - Link assessment of student achievement with specific teaching approaches/methods - Use EGRA research to support post-2015 Sustainable Development Goals and Education for All goals	UNICEF, development partners, Ministry of Education
Develop research surveys to examine student performance in math	- Design and adapt EGMA to the Myanmar context	Various development partners and Ministry of Education
Medium term (3-5 years)		
Build capacity to conduct surveys and analyze findings	- Upgrade skills of current government personnel to process and analyze data - Hire and develop internal experts to oversee analyses and draft reports	DMER, Departments of Basic Education, Development partners
Offer professional development to teachers to learn to measure and report on students' reading performance	- Develop a cadre of "master teachers" to observe students and teachers in classrooms, and to provide quality mentoring support to teachers	Development partners, Ministry of Education, education colleges, local school networks
Integrate reading standards into education colleges (pre-service)	- Institutionalize courses (pre-service) and workshops (in-service) to train teachers on best approaches to teaching reading, effective use of textbook, and assessment and reporting to track students' progress	Education colleges
Expand EGRA survey to all Myanmar	- Design and conduct nationally representative survey once sampling and implementation difficulties of early surveys are resolved	Ministry of Education
Adapting the EGRA in other languages	- Already started in January 2015 survey, but better to wait until difficulties of initial EGRA surveys resolved	Ministry of Education with regional/local representatives
Develop systematic system for monitoring and evaluation of student achievement in reading	- Link school-based assessments of students into a state-regional-national education information system	Development partners and Ministry of Education

BIBLIOGRAPHY

Adams, M.J. 1990. Beginning to Read: Thinking and Learning about Print. Cambridge, MA, USA: MIT Press.

CESR.org. 2012. "Terms of Reference for the Myanmar Comprehensive Education Sector Review, 2012." http://www.cesrmm.org/index.php/en/.

Deno, S.L. 2003. "Curriculum-Based Measures: Development and Perspectives." Assessment for Effective Intervention 28 (3–4): 3–12.

Deno, S.L., P.K. Mirkinand, and B. Chiang. 1982. "Identifying Valid Measures of Reading." Exceptional Children 49: 36–45.

Eddata//. https://www.eddataglobal.org/reading/.

Good, R.H., D.C. Simmons, and S. Smith. 1998. "Effective Academic Intervention in the United States: Evaluating and Enhancing the Acquisition of Early Reading Skills." School Psychology Review 27: 45–56.

Good, G.H., D.C. Simmons, and E.J. Kame'enui. 2001. "The Importance and Decision-Making Utility of a Continuum of Fluency-Based Indicators of Foundational Reading Skills for Third-Grade High-Stakes Outcomes." Scientific Studies of Reading 5: 257–88.

Gove, A., and P. Cvelich. 2010. "Early Reading: Igniting Education for All—A Report by the Early Grade Learning Community of Practice." Research Triangle Park, NC, USA: Research Triangle Institute.

Hanushek, E.A., and V. Lavy. 1994. "School Quality, Achievement Bias, and Dropout Behavior in Egypt." Living Standards Measurement Study (LSMS) Working Paper, no. LSM 107. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/1994/12/697806/school-quality-achievement-bias-dropout-behavior-egypt.

Hayden, M., and R. Martin. 2013. "Recovery of the Education System in Myanmar." Journal of International and Comparative Education 2 (2): 47–57.

Hoover, W.A. 2002. "The Importance of Phonemic Awareness in Learning to Read." SEDL Letter 14 (3), Putting Reading First. http://www.sedl.org/pubs/sedl-letter/v14n03/3.html.

Japan International Cooperation Agency. 2013. "Data Collection Survey on Education Sector in Myanmar." Tokyo: JICA.

Khandker, S.R., V. Lavy, and D. Filmer. 1994. "Schooling and Cognitive Achievements of Children in Morocco: Can the Government Improve Outcomes?" World Bank Discussion Paper, no. 264. Washington, DC: World Bank.

Lloyd, C.B., B. S. Mensch, and W.S. Clark. 2000. "The Effects of Primary School Quality on School Dropout among Kenyan Girls and Boys." Comparative Education Review 44 (2): 113–47.

Machuca-Sierra, M., and S. de Silva. 2012. "Lao P.D.R. EGRA Survey Report." Washington, DC: World Bank, East Asia and the Pacific Education Unit.

———. 2014. "Draft Baseline Report for Laos." Washington, DC: World Bank, East Asia and the Pacific Education Unit.

Ministry of Education, Union of Myanmar. 2004. "Development of Education in Myanmar 2004." Nay Pyi Taw, Myanmar: Ministry of Education.

———. 1999. "Instruction and Guide for Teachers." In Myanmar Language Textbook. Nay Pyi Taw, Myanmar: Ministry of Education.

———. 2008. "Country Report: Myanmar Education Development Strategy Focusing on Inclusive Education." Nay Pyi Taw, Myanmar: Ministry of Education. http://www.ibe.unesco.org/National_Reports/ICE_2008/myanmar_NR08.pdf.

———. 2012. "Education for All: Access to and Quality of Education in Myanmar." Conference on Development Policy Options with Special Reference to Education and Health in Myanmar, 13–16 February 2012, Nay Pyi Taw, Myanmar. http://yangon.sites.unicnetwork.org/files/2013/05/Education-for-All-in-Myanmar-Final-2012-FEB-2.pdf.

———. 2013. "Millennium Development Goal Report 2013." Nay Pyi Taw, Myanmar: Ministry of Education.

Ministry of National Planning and Economic Development, and the Ministry of Health, Union of Myanmar. 2011. "Myanmar Multiple Indicator Cluster Survey 2009–2010: Final Report." Nay Pyi Taw, Myanmar: Ministry of National Planning and Economic Development and the Ministry of Health. http://www.childinfo.org/files/MICS3_Myanmar_FinalReport.pdf.

Muter, V., C. Hulme, M.J. Snowling, and J. Stevenson. 2004. "Phonemes, Rimes, and Language Skills as Foundations of Early Reading Development: Evidence from a Longitudinal Study." Developmental Psychology 40: 665–81.

National Reading Panel. 2000. Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction—Reports of the Subgroups. Bethesda, MD, USA: National Institute of Child Health and Human Development.

Peereman, R., B. Lété, and L. Sprenger-Charolles. 2007. "Manulex-Infra: Distributional Characteristics of Grapheme-Phoneme Mappings, Infra-Lexical and Lexical Units in Child-Directed Written Material." Behavior Research Methods 39: 579–89.

RTI International. 2009. Early Grade Reading Assessment Toolkit. Prepared for the World Bank Office of Human Development. (Research Triangle Park, NC, USA: RTI International; and Washington, DC: USAID). https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&id=149.

Samejima, F. 1969. "Estimation of latent ability using a response pattern of graded scores." Psychometrika (Monograph Supplement) 34: 100–114.

Seidenberg, M.S. 2013. "The Science of Reading and Its Educational Implications." Language Learning and Development 9: 331–60.

Seymour, P.H.K., M. Aro, and J.M. Erskine. 2003. "Foundation Literacy Acquisition in European Orthographies." British Journal of Psychology 94: 143–74.

Shahar-Yames, D., and D.L. Share. 2008. "Spelling as a Self-Teaching Mechanism in Orthographic Learning." Journal of Research in Reading 31, 22–39.

Share, D.L. 1999. "Phonological Recoding and Orthographic Learning: A Direct Test of the Self-Teaching Hypothesis." Journal of Experimental Child Psychology 72: 95–129.

Share, D.L., and A.F. Jorm. 1987. "Segmental Analysis: Co-requisite to Reading, Vital for Self-Teaching, Requiring Phonological Memory." Cahiers de Psychologie Cognitive/European Bulletin of Cognitive Psychology 7: 509–513.

Simmons, D.C., and E.J. Kame'enui, eds. 1998. What Reading Research Tells Us about Children with Diverse Learning Needs: Bases and Basics. Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc.

Sprenger-Charolles, L., P. Colé, and W. Serniclaes. 2006. "Reading Acquisition and Developmental Dyslexia." In Essays in Developmental Psychology. Hove, UK, and New York: Psychology Press.

Sprenger-Charolles, L., L.S. Siegel, D. Béchennec, and W. Serniclaes. 2003. "Development of Phonological and Orthographic Processing in Reading Aloud, in Silent Reading, and in Spelling: A Four-Year Longitudinal Study." Journal of Experimental Child Psychology 84 (3): 194–217.

Stanovich, K.E. 1986. "Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy." Reading Research Quarterly 21: 360–406.

Stecker, P.M., L.S. Fuchs, and D. Fuchs. 2005. "Using Curriculum-Based Measurement to Improve Student Achievement: Review of Research." Psychology in the Schools 42: 795–819.

UNDP. 2011. "Integrated Households Living Conditions Survey in Myanmar (2009–2010): Poverty Profile. Yangon, Myanmar: IHLCA Project Technical Unit.

UNESCO Institute for Statistics. 2011. International Standard for Classification of Education (ISCED). http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx.

USAID and eddata//. "Early Grade Reading." https://www.eddataglobal.org/reading/.

World Bank. 2013. "Report on Early Grade Reading Assessment: Grades 1 and 3 in SEQAP Schools (Vietnam). Unreleased PowerPoint. Washington, DC: World Bank.

Wren, S., B. Litke (programmer), D. Jinkins, S. Paynter, J. Watts, I. Alanis. 2000. "Cognitive Elements of Reading." In Cognitive Foundations of Learning to Read: A Framework. Austin, TX, USA: SEDL. http://www.sedl.org/reading/framework/elements.html.

Ziegler, J.C., A.M. Jacobs, and G.O. Stone. 1996. "Statistical Analysis of the Bidirectional Inconsistency of Spelling and Sound in French." Behavior Research Methods, Instruments, and Computers 28: 504–515.

ANNEX 1:
OVERVIEW OF EGRA TOOLS AND STIMULI

A1.1 Presentation of the eight subtasks in the Myanmar EGRA tool for Grades 1-3

1. ***Letter name knowledge*** assessed children's automaticity in letter recognition. Each letter was represented proportionally to its occurrence in the grade-appropriate textbook consulted. Letters were presented in a chart containing 10 rows of 10 letters in a fixed random order. Students were asked to identify as many letters they could within one minute,[61] yielding a score of correct letter names per minute (CLNPM).

2. ***Initial sound identification*** assessed children's phonemic awareness. Children were asked to listen to a word and identify the first sound in that word. To ensure that these words were known and familiar to the students, they were chosen from the frequent word list obtained for each grade. After two practice items, children were given 10 test items (no time limit). The final score was the number of words where the children identified the correct initial sound (CIS), with the maximum possible score being 10.

3. ***Letter sound knowledge*** assessed children's knowledge of letter-sound relationships. The test items were selected and displayed following the same procedure as described for subtask 1 (a chart containing 10 rows of 10 random letters). Children were asked to speak the sound of as many letters as they could within one minute, yielding a score of correct letter sounds per minute (CLSPM).

4. ***Familiar word reading*** assessed children's skill at reading high-frequency words. The 50 most frequent words in reference textbooks were presented in a fixed random order to students. Children were asked to read as many words as they could within one minute, yielding a score of correct familiar words per minute (CFWPM).

5. ***Invented word decoding*** assessed children's skill at applying letter-sound correspondence rules to decode non-words (or invented words). For Grade 1, the 50 invented words used the 24 vowel sounds covered in the Grade 1 textbook. For Grades 2 and 3, the invented words had all 50 vowel sounds in the Myanmar language. The children looked at a chart of 50 non-words and were asked to sound out as many words as they could within one minute, yielding a score of correct invented words per minute (CIWPM).

6. ***Oral reading*** assessed children's fluency in reading a passage of grade-level text aloud and their ability to understand what they had read. There were two parts to this subtask:

   a. Oral reading fluency: In Grade 1, all the words included in the text passage included grade appropriate letters and letter combinations. The passage was 44 words long for Grade 1, and 38 words long for Grades 2 and 3. A narrative story was considered appropriate to test connected text reading in those grades. The stories respected the typical features of narrative prose: they had "a beginning section where the characters are introduced, a middle section containing some dilemma, and an ending section with an action resolving the dilemma." Students were asked to read aloud as much of the passage they could in one minute. The oral reading fluency (ORF) score for this task was derived by calculating the number of correct words read per minute.

   b. Reading comprehension: After the children finished the passage or the minute ended, the passage was removed. Children were asked six questions orally: four questions were literal and required students to recall basic facts from the passage read to them, while two were inferential. The reading comprehension score was the number of correct answers, with a maximum possible score of 6 (CRCQ).

---

[60] For all timed subtasks (1, 3, 4, 5, and 6a), two skip rules were established: 1) if the child hesitates or stops on a letter or word for three seconds or more, it was skipped; 2) if the child does not provide a single correct response on the first line (five or 10 items, depending on the subtask), this subtask was discontinued and the child moved on to the next subtask.

7. ***Listening comprehension*** is the ability to make sense of oral language. In this subtask, a short passage was read to children who were then asked six questions orally about that passage. The listening comprehension score was the total correct answers, with a maximum possible score of six (CLCQ).

8. ***Dictation*** assessed children's skill at spelling. Children could spell familiar words from memory or sound out words and apply the sound-spelling correspondences. In this subtask, a short sentence of four words was read to the children, who then attempted to spell the words in writing. The dictation score was the number of words spelled correctly (WSC).

A1.2 Actual subtask stimuli in the Myanmar EGRA tool

Subtask 1: Letter naming

၁	၂	၃	၄	၅	၆	၇	၈	၉	၁၀
က	ဆ	န	င	သ	ပ	တ	c	စ	မ	(၁၀)
မ	က	တ	န	စ	ရ	လ	က	ရ	င	(၂၀)
ပ	သ	လ	အ	ပ	တ	န	မ	င	ပ	(၃၀)
တ	ရ	မ	က	ည	ခ	လ	သ	စ	အ	(၄၀)
သ	န	င	စ	သ	က	မ	အ	ပ	မ	(၅၀)
ရ	အ	ည	ပ	လ	င	စ	န	တ	ယ	(၆၀)
င	တ	သ	ခ	မ	ည	က	ထ	ဘ	ဗ	(၇၀)
လ	င	ခ	ရ	ယ	ဖ	၀	မ	က	ၡ	(၈၀)
ဆ	ပ	ဒ	ဂ	ဟ	၀	ဗ	ကြ	ဓ	က	(၉၀)
ထ	ဖ	ဂ	မ	ဂ	၄	၀	၃	ယ	င့	(၁၀၀)

Subtask 2: Initial sound identification

ဒီစာလုံးရဲ့, ပထမအသံက �’ဘာလဲ"_____"? "_____"? [စာလုံးအား နှစ်ကြိမ် ဖတ်ပြပါ။]				
မာ	/မ/ /m/	○ မှန်	○ မှား	○ မဖြေ
လေ	/လ/ /l/	○ မှန်	○ မှား	○ မဖြေ
ကား	/က/ /k/	○ မှန်	○ မှား	○ မဖြေ
ရေ	/ရ/ /j/	○ မှန်	○ မှား	○ မဖြေ
ပို	/ပ//p/	○ မှန်	○ မှား	○ မဖြေ
သီး	/သ/ /ø/	○ မှန်	○ မှား	○ မဖြေ
တော်	/တ/ /t/	○ မှန်	○ မှား	○ မဖြေ
စိုက်	/စ/ /s/	○ မှန်	○ မှား	○ မဖြေ
ဝါ	/၀/ /w/	○ မှန်	○ မှား	○ မဖြေ
ညာ	/ည//nj/	○ မှန်	○ မှား	○ မဖြေ

Subtask 3: Letter sound knowledge

၁	၂	၃	၄	၅	၆	၇	၈	၉	၁၀	
က	ဆ	န	င	သ	ပ	တ	င	စ	မ	(၁၀)
မ	က	တ	န	စ	ရ	လ	က	ရ	င	(၂၀)
ပ	သ	လ	အ	ပ	တ	န	မ	င	ပ	(၃၀)
တ	ရ	မ	က	ည	ခ	လ	သ	စ	အ	(၄၀)
သ	န	င	စ	သ	က	မ	အ	ပ	မ	(၅၀)
ရ	အ	ည	ပ	လ	င	စ	န	တ	ယ	(၆၀)
င	တ	သ	ခ	မ	ည	က	ထ	ဘ	ဝ	(၇၀)
လ	င	ခ	ရ	ယ	ဖ	၁	မ	က	ၡ	(၈၀)
ဆ	ပ	ဒ	ဂ	ဟ	ပ	ဇ	က	ဓ	က	(၉၀)
ထ	ဖ	က	မ	က	၄	ပ	၃	ယ	၄	(၁၀၀)

Subtask 4: Familiar word reading

နမူနာ အိုးစည်	ဆီမီး	ဆရာ			
၁	၂	၃	၄	၅	
ပွင့်	သန့်	စား	သာ	ထား	(၅)
ရာသီ	မေမေ	လူၣာ	ယူ	ငါ	(၁၀)
နှစ်	ကလေး	နေ	ပူ	သာယာ	(၁၅)
မိဘ	ပဝါ	သစ်ပင်	အစာ	ကစား	(၂၀)
ကင်း	သတိ	ကျောင်း	ပါ	သား	(၂၅)
မမ	ဝါး	ဘာ	ရေကန်	စကား	(၃၀)
သွား	နွား	ငါး	ဘယ်	ပေး	(၃၅)
စာ	နား	ကျန်းမာ	ရှိ	ပြေး	(၄၀)
လူကလေး	အနာ	လာ	ဆရာမ	အနီး	(၄၅)
အသီး	ပညာ	စပါး	အဘိုး	ပျိုး	(၅၀)

Subtask 5: Nonsense word decoding

၁	၂	၃	၄	၅	
ထာ	ေင	ဂါ	ဗား	လို	(၅)
ဘင်း	ဇေး	လွား	ရား	ဆခါး	(၁၀)
နာလီ	ဂျိုး	၀ကား	ဇား	ချန်းမိ	(၁၅)
ကားစ	ေစေစ	ယာသာ	ထူလဂေး	ကူ	(၂၀)
မဘိ	ဓေ	ဒွင့်	ဂါး	စလာဖ	(၂၅)
ဗာ	ဘအိုး	ကပီး	ပဟာ	မား	(၃၀)
သွ	မဟိ	ဘန်	ဒယ်	၀ပါ	(၃၅)
အဘာ	ထာ	အဂေး	၀စ်	စွား	(၄၀)
ရန်ေက	ေဂး	၀	ကျောင်း	စပီး	(၄၅)
နအာ	ပစ်သင်	စျား	ပန့်	မ္ဘာ	(၅၀)

Subtask 6a and 6b: Oral reading fluency and reading comprehension

	စာလုံးအရေ အတွက်	မေးခွန် (အဖြေ)]	မှန်	မှား	မဖြေ
အေးအေးမှာ / အလွန် / ချစ်စရာ / ကောင်းတဲ့ / မိသားစု / တစ်စု / ရှိသည်။	၇	အေးအေးမှာ ဘယ်လိုမိသားစုရှိသလဲ။ (ချစ်စရာကောင်းတဲ့မိသားစု)			
ကိုကို / တစ်ယောက် / ညီမ / တစ်ယောက် / ရှိသည်။	၁၂	အေးအေးမှာညီမဘယ်နှစ်ယောက်ရှိသလဲ။(တစ်ယောက်)			
တစ်နေ့ / မှာ / မိသားစု / ပျော်ပွဲစား / ထွက်ကြသည်။	၁၇	အေးအေးတို့ ဘယ်သွားကြသလဲ။ (ပျော်ပွဲစား)			
ကိုကိုက / ကျောင်း / မှာ / လုပ်အားပေး / ရှိလို့ / မ / လိုက် / နိုင်ပါ။	၂၅	ဘာကြောင့်ကိုကိုကပျော်ပွဲစားမလိုက်နိုင် တာလဲ။ (ကျောင်းမှာလုပ်အားပေးရှိလို့.)			
ပျော်ပွဲစား / မှာ / ကိုကို / မ / ပါလို့ / မ / ပျော်ကြပါ။	၃၂	ပျော်ပွဲစား ဘယ်နှစ်ယောက်သွားကြသလဲ။ (လေးယောက်)			
နောက် / တစ်ခါ / ပျော်ပွဲစား / ထွက် / ရင် / မိသားစု / အားလုံး / အားသော / အချိန် / မှာ / သွားဖို့ / ဆုံးဖြတ်လိုက်သည်။	၄၄	နောက်တစ်ခါပျော်ပွဲစားသွားရင် ဘယ်နှစ်ယောက်သွားကြမလဲ။ (ငါးယောက်)			

Subtask 7: Listening comprehension

မအိမှာ ကြောင်လေးတစ်ကောင်နဲ့ခွေးလေးတစ်ကောင်ရှိတယ်။ မအိက ကြောင်လေး ပူစီနဲ့ ခွေးလေး ပါပီကို နေ့စဉ် အစာကျွေးတယ်။ ပူစီက ငါးစားတယ်။ ပါပီက အရိုးဝါးတယ်။ တစ်နေ့တော့ ပါပီက ငါးစားချင်တော့ လဲစားရအောင်လို့ပြောတယ်။ ပူစီနဲ့ ပါပီဟာ ရန်မဖြစ်ဘဲ ငါးကိုအတူတူမျှဝေ စားကြတယ်။ ပြီးတော့ အတူတူ ကစားကြတယ်။ ထို့ကြောင့် မအိပျော်ပါတယ်။

	မှန်	မှား	မဖြေ
မအိမှာ ဘာရှိသလဲ။ (ကြောင်လေးတစ်ကောင်နဲ့ခွေးလေးတစ်ကောင်)			
မအိနေ့စဉ်ဘာလုပ်သလဲ။ (အစာကျွေးတယ်။)			
ပူစီနဲ့ ပါပီ ဘာစားကြသလဲ။ (ငါးနဲ့ အရိုး)			
တစ်နေ့မှာ ပါပီက ဘာပြောသလဲ။ (အစာလဲစားရအောင်လို့ပြောတယ်)			
ပါပီနဲ့ ပူစီ ဘာကြောင့် အတူတူကစားကြသလဲ။ (ရန်မဖြစ်တဲ့ သူငယ်ချင်းတွေဖြစ်လို့)			
မအိဘာကြောင့်ပျော်သလဲ။ (ပူစီနဲ့ ပါပီဟာ ရန်မဖြစ်ဘဲအတူတူကစားကြလို့)			

Subtask 8: Dictation

	သတ်ပုံခေါ်မည့် စကားလုံးများ	ကျောင်းသား ၏အဖြေ (သက်ဆိုင်ရာအကွက်တွင် ခြစ်ပေးပါ)				ကျောင်းသား ၏အဖြေ
		မှန်	မှား	မဖြေ	သ�க်တာ	မှန်စာလုံးများ
၁။	အ ဘိုးအို					/၃
၂။	လ ယ်တဲကို					/၃
၃။	ကူညီမှီးပေးပါ။					/၅

၄။	စကားလုံးများကြားနေရာလွတ်အရေအတွက် (၂)	_____ ကြားနေရာလွတ်	○ မဖြေပါ။
၅။	ဘယ်မှညာသို့ရေးသလား။	○ မှန် ○ မမှန်	○ မဖြေပါ။
6.	ဝါကျအဆုံးမှာ ပုဒ်မ (။) ချသလား။	○ မှန် ○ မမှန်	○ မဖြေပါ။

ANNEX 2:
DETAILED RESULTS FOR THE RELIABILITY OF THE EGRA TOOL

A2.1 Internal consistency in reliability of scales

The overall Cronbach's alpha (task scale) is 0.77 for Grade 1 and 0.82 for Grades 2 and 3 (Tables A2.1 and A2.2), which indicates that the overall task is sufficiently reliable. At the subtask level, the results indicate that scores on all subtasks have high item-rest and item-test correlation, except for letter-sound fluency (subtask 3), initial sound segmentation (subtask 2), and, to a lesser extent, listening comprehension (subtask 7). It is also observed in column (v) that removing letter sound fluency and initial sound segmentation from the construction of the overall task scale (the statistics in column [v] indicates the value of Cronbach's alpha, if one removes the line subtask) would lead to substantial increases in Cronbach's alpha. Therefore, these two subtasks, as designed and as administered, are not sufficiently reliable measures of early reading skills in the Myanmar language.

Table A2.1: Reliability analysis using correlation and Cronbach's alpha in Grade 1

Subtasks	(i) Obs	(ii) Sign	(iii) Item-test correlation	(iv) Item-rest correlation	(v) Cronbach's alpha
1. Letter name fluency	559	+	0.5951	0.4384	0.7568
2. Initial sound segmentation	559	+	0.321	0.1396	0.7996
3. Letter sound knowledge	558	-	0.184	-0.0146	0.8197
4. Familiar word reading fluency	552	+	0.8838	0.8278	0.689
5. Non-word reading fluency	551	+	0.8271	0.7475	0.7032
6a. Oral text reading fluency	540	+	0.8195	0.7401	0.7068
6b. Reading comprehension	322	+	0.7303	0.6216	0.7433
7. Listening comprehension	559	+	0.4596	0.2801	0.7804
8. Dictation (word score)	550	+	0.6796	0.5549	0.7376
Global EGRA task scale					0.7743

Table A2.2: Reliability analysis using correlation and Cronbach's alpha in Grades 2 and 3

Subtasks	(i) Obs	(ii) Sign	(iii) Item-test correlation	(iv) Item-rest correlation	(v) Cronbach's alpha
1. Letter name fluency	1119	+	0.5098	0.3615	0.8162
2. Initial sound segmentation	1119	+	0.2664	0.095	0.8455
3. Letter sound knowledge	1118	+	0.3019	0.1324	0.8415
4. Familiar word reading fluency	1114	+	0.8858	0.8373	0.7564
5. Non-word reading fluency	1113	+	0.8254	0.7552	0.7674
6a. Oral text reading fluency	1110	+	0.8678	0.8126	0.7598
6b. Reading comprehension	1121	+	0.795	0.7141	0.7727
7. Listening comprehension	1118	+	0.5189	0.3734	0.8146
8. Dictation (word score)	1113	+	0.7647	0.6752	0.7779
Global EGRA task scale					0.8162

A2.2 Pair-wise correlations across subtasks

Similar to the observations for internal consistency and reliability, on the basis of pair-wise correlations, most subtasks appear appropriate, except for the initial sound identification and letter sound knowledge. Because each EGRA subtask is meant to measure the same overall skill—namely early reading—appropriate subtasks should be highly correlated to each other. Pair-wise correlations of scores among most of the subtasks of the EGRA battery are significant and positive (see Tables A2.3 and A2.4), except for initial sound identification and letter sound fluency, which show undersized correlations with other subtasks. In particular,

non-word fluency, familiar-word fluency, and oral fluency in text reading; reading comprehension; and dictation are highly and positively correlated.

Reading and listening comprehension also are significantly correlated with all the other subtasks. Correlations of scores in reading comprehension with scores on non-words, familiar words, and text reading are large in Grades 2 and 3, and moderate to large in Grade 1. With score listening comprehension, these pair-wise correlations are moderate in Grades 2 and 3, and small in Grade 1. Correlations with scores on the letter-naming subtask are significant and moderate to small in Grade 1, while moderate to large in Grades 2 and 3.

Table A2.3: Correlation matrix across all EGRA subtasks in Grade 1

Subtasks	1. Letter name fluency	2. Initial sound segmentation	3. Letter sound knowledge	4. Familiar word reading fluency	5. Non-word reading fluency	6a. Oral text reading fluency	6b. Reading comprehension	7. Listening comprehension
1. Letter name fluency								
2. Initial sound segmentation	0.0193							
3. Letter sound knowledge	0.0817	-0.0053						
4. Non-word reading fluency	0.5325***	0.1419***	0.0323					
5. Familiar word reading fluency	0.5755***	0.11	0.0625	0.88***				
6a. Oral text reading fluency	0.3232***	0.2315***	-0.0186	0.8238***	0.7235***			
6b. Reading comprehension	0.1378	0.1849***	-0.0185	0.6182***	0.484***	0.6709***		
7. Listening comprehension	0.2505***	0.0068	-0.0617	0.2692***	0.2554***	0.197***	0.1587*	
8. Dictation	0.3076***	0.0599	-0.014	0.6216***	0.545***	0.5326***	0.4026***	0.2019***

Note: Significance at * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Table A2.4: Correlation matrix across all EGRA subtasks in Grades 2 and 3

Subtasks	1. Letter name fluency	2. Initial sound segmentation	3. Letter sound knowledge	4. Familiar word reading fluency	5. Non-word reading fluency	6a. Oral text reading fluency	6b. Reading comprehension	7. Listening comprehension
1. Letter name fluency								
2. Initial sound segmentation	0.019							
3. Letter sound knowledge	0.0371	0.2957***						
4. Non-word reading fluency	0.3788***	0.0602	0.1074***					
5. Familiar word reading fluency	0.355***	0.0404	0.0747	0.8585***				
6a. Oral text reading fluency	0.3721***	0.0364	0.092**	0.9202***	0.8229***			
6b. Reading comprehension	0.1378	0.0025	0.0513	0.7292***	0.6499***	0.7434***		
7. Listening comprehension	0.1866***	0.0419	0.0161	0.3409***	0.3048***	0.3293***	0.4198***	
8. Dictation	0.2741***	0.0334	0.0575	0.7051***	0.6502***	0.6844***	0.6444***	0.3614***

Note: Significance at * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

A2.3 Predictive power of lower level skills to higher level skills

To assess the relationships of lower level skills to higher level skills, a few hierarchical multiple regressions (one for each of the three higher level skills) were run separately for Grade 1 and Grades 2 and 3 (Table A2.5). Independent variables are the subtasks measuring lower level skills and a wealth index in order to control (at a minimum) the student's household wealth level and grade (for Grades 2-3). For all three higher level skills, scores on all lower level skills are positively and significantly correlated, except for letter naming (negatively correlated with oral reading fluency).

Table A2.5: Regression results for predicting higher level skills

Skills or cognitive concept	Measurement (subtask)	Predictors (subtasks)	Grade 1	Grades 2 and 3	% of variance explained
Text reading	Oral reading fluency	Letter name per minute	-0.12***	0.06	
		Non-word per minute	0.26	0.24	
		Familiar word per minute	0.82***	0.92***	
Spelling	Dictation	Letter name per minute	0.0	0.00	
		Non-word per minute	0.02	0.03**	
		Familiar word per minute	0.06***	0.03***	
Reading comprehension	Reading comprehension	Oral reading fluency	0.02***	0.03***	
		Dictation	0.08*	0.09*	

Note: Significance at * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

ANNEX 3: DESCRIPTIVE STATISTICS FOR EGRA SUBTASKS (GRADES 1, 2, 3) FOR SECTION 5.3

Table A3.1: Descriptive statistics for each EGRA subtask in Grade 1
(including mean, standard deviation, standard error, percentage of zero scores and number of students in the sample and the target population)

Subtask	Mean	Standard deviation	Standard error	% 0 score	N (sample size)	N (pop. size)
1. Letter names per minute	36.32	18.71	2.79	4%	557	114,378
2. Initial sound segmentation	0.01	0.22	0.01	99%	557	114,374
3. Letter sounds per minute	0.12	1.89	0.06	100%	556	114,342
4. Familiar words per minute	12.82	10.93	2.04	26%	550	112,121
5. Nonsense words per minute	7.89	6.78	1.14	30%	549	112,095
6a. Oral reading fluency	7.58	9.79	1.77	35%	538	110,323
6b. Reading comprehension	0.34	0.57	0.08	77%	321	69,535
7. Listening comprehension	2.91	1.63	0.13	9%	557	114,331
8. Dictation	0.78	1.08	0.19	68%	548	111,925

Table A3.2: Descriptive statistics for each EGRA subtask in Grade 2
(including mean, standard deviation, standard error, percentage of zero scores and number of students in the sample and the target population)

Subtask	Mean	Standard deviation	Standard error	% 0 score	N (sample size)	N (pop. size)
1. Letter names per minute	43.21	14.51	0.87	1%	577	110,154
2. Initial sound segmentation	0.02	0.48	0.01	99%	577	110,154
3. Letter sounds per minute	0.47	5.64	0.19	98%	576	110,107
4. Familiar words per minute	26.36	16.83	1.84	6%	574	109,213
5. Nonsense words per minute	13.19	9.75	1.05	12%	573	108,389
6a. Oral reading fluency	30.02	19.45	2.23	10%	573	109,802
6b. Reading comprehension	2.81	2.15	0.17	27%	577	110,154
7. Listening comprehension	2.63	1.63	0.14	12%	576	110,077
8. Dictation	1.60	1.26	0.14	33%	574	107,665

Table A3.3: Descriptive statistics for each EGRA subtask in Grade 3
(including mean, standard deviation, standard error, percentage of zero scores and number of students in the sample and the target population)

Subtask	Mean	Standard deviation	Standard error	% 0 score	N (sample size)	N (pop. size)
1. Letter names per minute	47.30	14.34	0.92	0%	542	102,161
2. Initial sound segmentation	0.08	0.38	0.07	98%	542	102,161
3. Letter sounds per minute	0.27	4.89	0.14	97%	542	102,161
4. Familiar words per minute	40.87	21.03	2.63	5%	540	101,308
5. Nonsense words per minute	17.91	11.01	1.28	10%	540	101,308
6a. Oral reading fluency	48.54	23.95	2.75	7%	537	100,663
6b. Reading comprehension	4.00	2.02	0.17	12%	544	102,261
7. Listening comprehension	3.36	1.58	0.11	5%	542	102,161
8. Dictation	2.05	1.37	0.17	22%	539	100,743

Table A3.4: Percent of correct responses to five to six questions in reading and listening comprehension for Grades 1, 2, and 3
(literal and inferential questions presented separately, calculated for students who responded to 5-6 comprehension questions)

	Reading comprehension		Listening comprehension	
	Literal question	Inferential question	Literal question	Inferential question
Grade 1	34%	15%	82%	32%
Grade 2	74%	62%	45%	64%
Grade 3	84%	67%	52%	72%
Total N (%) of Questions	4 (100%)	2 (100%)	4 (100%)	2 (100%)

Table A3.5: Percent of correct responses to five questions in reading and listening comprehension for Grades 1, 2, and 3
(literal and inferential questions presented separately; includes only students asked five questions)

	Reading comprehension		Listening comprehension	
	Literal question	Inferential question	Literal question	Inferential question
Grade 1	27%	2%	70%	0%
Grade 2	56%	29%	32%	60%
Grade 3	67%	41%	53%	67%
Total N (%) of Questions	4 (100%)	2 (100%)	4 (100%)	2 (100%)

Table A3.6: Percent of correct responses to all six questions in reading and listening comprehension for Grades 1, 2, and 3
(literal and inferential questions presented separately; includes only students asked all six questions)

	Reading comprehension		Listening comprehension	
	Literal question	Inferential question	Literal question	Inferential question
Grade 1	34%	15%	82%	32%
Grade 2	74%	62%	45%	64%
Grade 3	84%	67%	52%	72%
Total N (%) of Questions	4 (100%)	2 (100%)	4 (100%)	2 (100%)

ANNEX 4: METHODOLOGY AND DETAILED RESULTS TABLE FOR SUBSECTION 5.4

A4.1 Overview of the methodology

In order to identify potential factors that explain the differences across schools, grades, and students in early reading skills (as measured through EGRA scores), the correlation between EGRA scores and student and teacher characteristics are analyzed first in a bivariate context (as a preliminary and preparatory stage) and second in a multivariate context using a set of three-level hierarchical mixed models (with students nested within grades and grades nested within schools). The bivariate correlation analysis looks at the one-to-one relationship of various factors to higher/lower overall reading performance. However, one cannot directly extrapolate these results to possible interventions, which could positively affect school and teacher effectiveness in imparting reading skills. Indeed, interpretation of bivariate results does not take into account that the investigated characteristics (assumed to be potential factors in making a school or teacher effective) may also reflect (and are therefore correlated with) non-school related factors, such as students socio–economic status or baseline cognitive skills. In other words, bivariate relationships may be the result of many factors, including a reversal of the expected direction of causality.

For example, schools with summative assessment practices may have introduced them as a policy to monitor student performance. But a common factor could underlie both variables, such as a broad-based school intervention targeting low-performing schools with an assessment component that not all schools used. In order to deepen the analysis—and ensure that some control for the correlation among all potential explanatory factors is accounted for—the three-level hierarchical mixed model is used.

To ensure the simplest explanation, instead of using individual subtasks scores, the early reading skills are summarized into one EGRA composite score. The composite EGRA score was created using data from the four higher-level subtasks (familiar words, non-words, oral reading fluency, and reading comprehension). First, for each of the selected four subtasks, item response theory (IRT) was used to create continuous interval-scale scores. The graded response IRT model[62] was used to accommodate the differential accuracy of different items (namely letters, words, or questions) included in each subtask. The IRT scale scores are more mathematically useful than raw scores (used previously to compute mean and zero scores) for analyzing potential explanatory factors. They do not artificially truncate high and low extremes of the score continuum, and avoid the spurious modification of relationships with other variables that results from the truncation.[63] Second, these four scores were standardized within the sample to have a mean of zero and a standard deviation of one, and then averaged to produce the composite score.

The multivariate relationships are estimated using a mixed effects model; the general form of the model is a three-level hierarchical structure, with students nested within grades and grades nested within schools.[64] The purpose of this hierarchical structure modeling technique is to uncover the contribution of various groups of "intervenable" variables (e.g., education inputs and instructional practices) to ameliorate the differences—due to socio–economic characteristics and non-school factors—in reading performance (as measured through EGRA scores). Identifying such variables would help identify and design interventions to improve early reading skills in Myanmar classrooms.

---

[62] F. Samejima, 1969, "Estimation of Latent Ability Using a Response Pattern of Graded Scores," Psychometrika (Monograph Supplement) 34: 100-114.

[63] The letter sound and initial sound subtasks were not used for the composite score, given the reliability issues identified in Section 5.2. The letter naming subtask was not used either, due to the lower psychometric quality of the item responses for it. The responses did not adequately define a single latent dimension, which is necessary for the application of IRT models.

[64] Thus, without any explanatory variables (null model), the score for a specific student is the sum of 1) the grand mean of all students' scores, 2) the deviation of the student's school mean score from the grand mean score, 3) the deviation of the within-grade mean score from the school mean score, and 4) the deviation of the student's score from the within-grade-within-school mean score.

This approach helps explain 1) why schools differ from each other in average reading performance, and 2) why students within the same schools differ from each other in reading performance. Although this modeling approach could also explain why students in different grades differ in performance, the analysis does not focus on normal progression across grades, which is of no interest in policy and intervention design.[65] This analysis is primarily concerned with understanding differences between schools because the Ministry of Education can effectively intervene on this issue.

Finally, the modeling approach follows a successive process, which can remove less relevant variables and prevent unnecessary limits to the degrees of freedom. To statistically explain these differences, the analysis explores the results from these successive models:

- **Null model:** the original statistical model without any explanatory variables

- **Base model:** the model after including variable groups A and B

- **Context model:** the model after including groups A, B, and C

- **Control model:** the model after including groups A, B, C, and D (Although variables in group D reflect activities of teachers and the school environment, they are included in the control model because the variables describe the students' perspectives. They may also reflect student-driven phenomena.)

- **Inputs and pedagogical models (E, F, and G):** the control model plus variables exclusively from groups E, F, or G (To prevent saturation of the data with coefficient estimations, the variables from groups E, F, and G were not carried forward to subsequent modeling stages. In other words, the models using variable sets E, F, and G, only include variables from their respective set in addition to the variables from variable groups A, B, C, and D.)

In this process of successive modeling, all variables in each group were tested for relationships with scores that were likely the result of chance.[66] Before moving to the next variable group, variables with chance relationships were removed from the model. The exception was the control variables (groups A and B), whose effects varied significantly across schools, despite having insignificant main effects. **Conversely, the procedure has fixed student-level predictors, whose effects did not vary significantly between schools.**

A4.2 Bivariate analysis

Many of the relationships identified in the bivariate analysis are intuitive, such as positive correlations for female status, wealth, adult support (for reading and homework), prior enrollment in preschool, use of a Myanmar language exercise book, and availability of library—as well as negative correlations for teacher travel time and teachers with other jobs. However, there are also several non-intuitive relationships, such as the positive relationship for absence due to sickness, the negative correlations with use of summative assessment and the use of assessment results for modifying instructions. These counter-intuitive relationships are a reminder that bivariate relationships may be the result of many factors, including a reversal of the expected direction of causality. The detailed bivariate results are presented below.

Remoteness and wealth

Unsurprisingly, non-remoteness and wealth are positively correlated with the EGRA composite scores. As illustrated in Figure A4.1, being enrolled in a remote or non-remote school is significantly and positively (large coefficient) correlated with the EGRA composite score overall, as well as for each grade. Wealth (as measured through the continuous and categorical variables) is also positively and significantly, but weakly (small coefficient), correlated with scores. It is, however, impossible to confirm the robustness of these correlations in a bivariate setting, especially since remoteness and wealth are highly correlated.

Gender and class size

Bearing in mind the caveats regarding the sample, being a female appears to be positively and significantly correlated with the EGRA composite score (Figure A4.2). However, this difference disappeared in Grade 3, possibly indicating that the boys are catching up or that the gender/ability mix has changed as a result of differentiated drop-outs. Class size is uncorrelated with performance.

Language

Despite low variation in the language spoken by students, parents, and teachers, the results point to some possible positive correlation between scores and speaking the Myanmar language at home or being taught by a native Myanmar speaker. The high proportion of households speaking the Myanmar language (and consequent low variation for this characteristic in the

---

[65] In this model, differences 1 and 2 are represented by the standard deviation of all schools about the base group mean and the standard deviation of the students about their grade means within their schools, respectively. Although the standard deviation of the grade indicates that the school means vary across schools, it will not be modeled in these analyses because they do not concern understanding why students in different grades differ in performance.

[66] For model-building purposes, an arbitrary threshold of $p < 0.15$ was used to identify chance relationships.

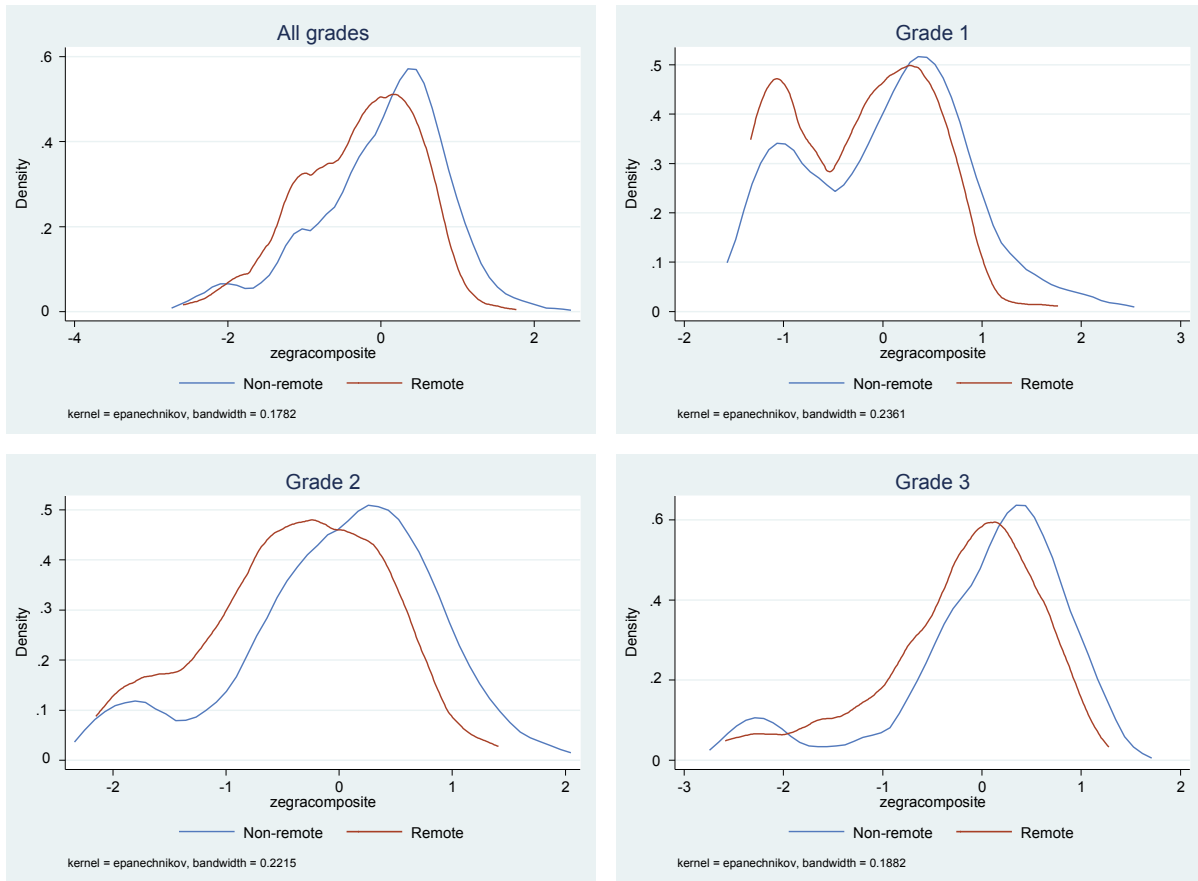## Figure A4.1: EGRA scores per remoteness status and grade

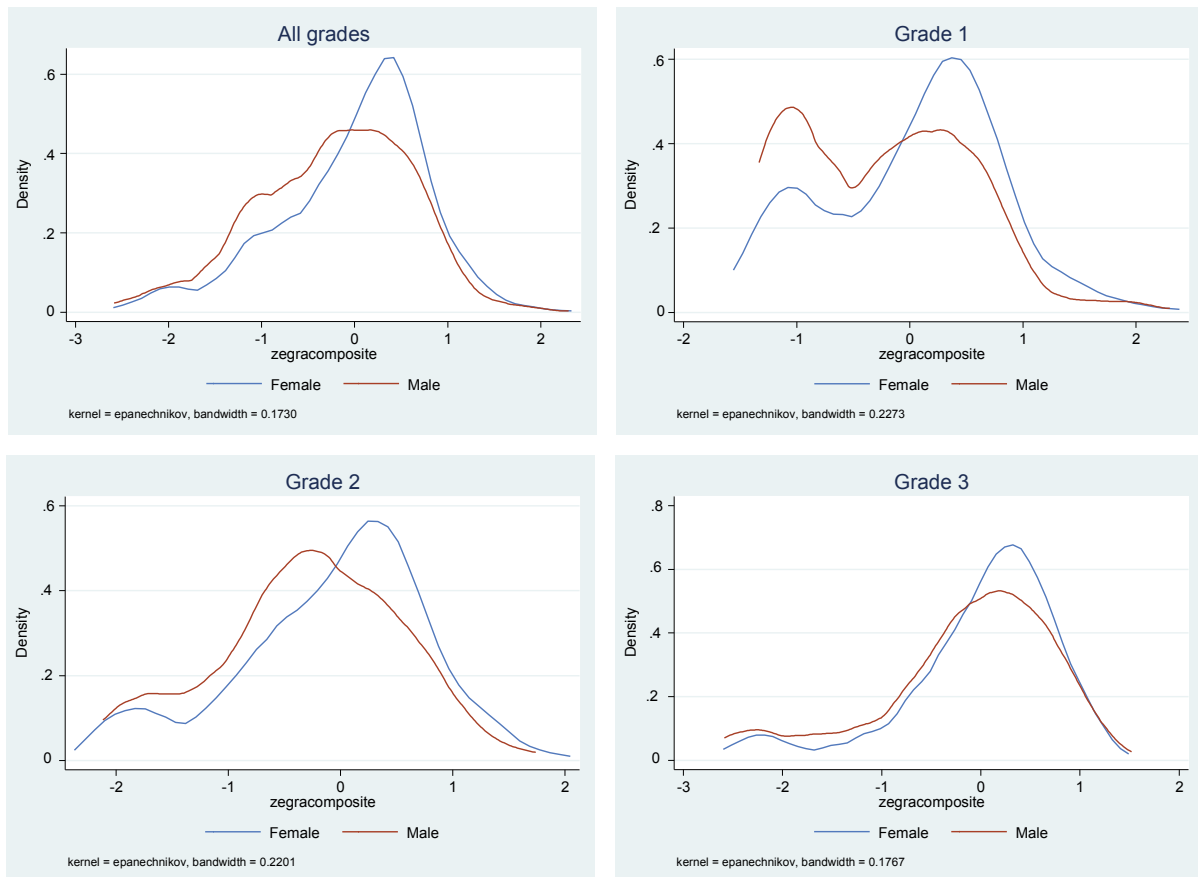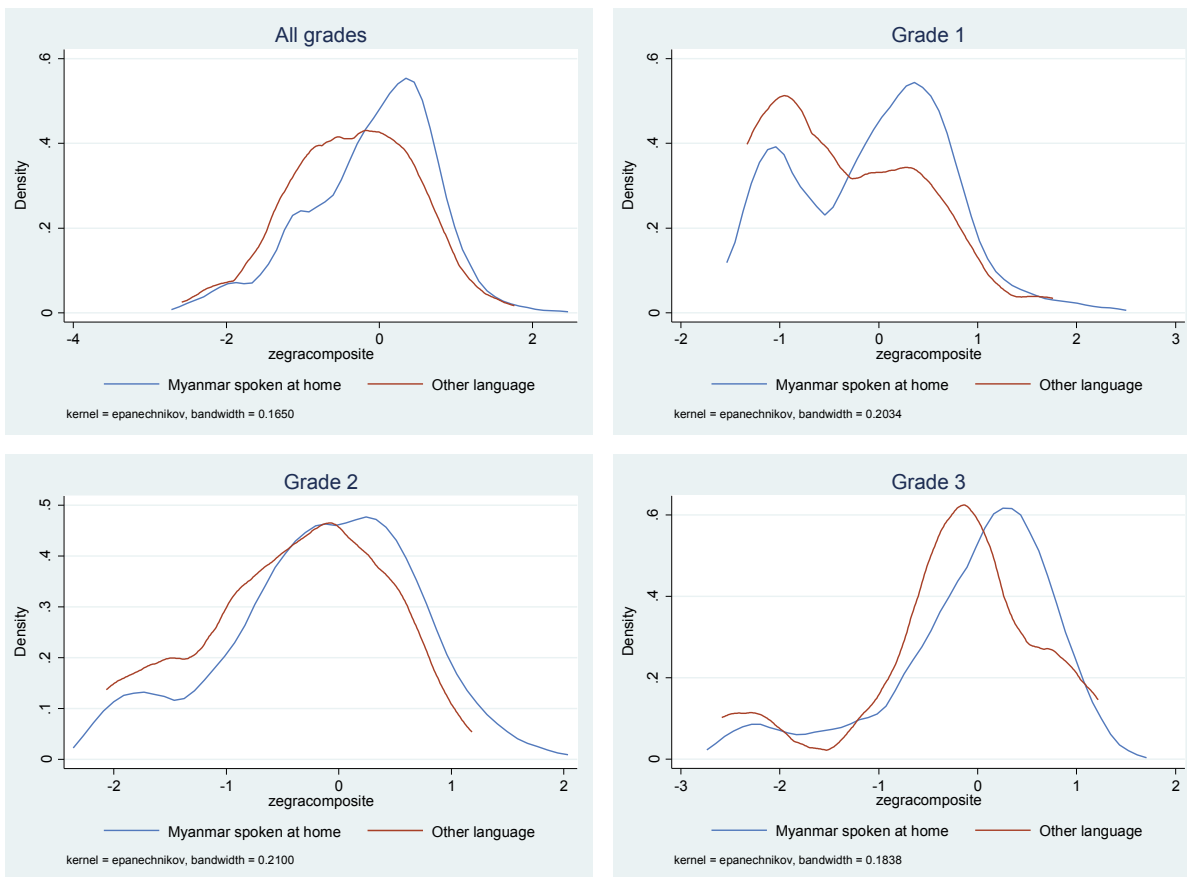## Figure A4.2: EGRA scores per gender and grade

Figure A4.3: EGRA scores per mother tongue and grade

sample) implies that it is difficult to robustly estimate the correlation between students' language and EGRA scores. This is especially the case since this analysis needs to combine all non-Myanmar languages observed, which are quite numerous.[67] This is also the case for language-related variables for teachers.

Remembering that these results are only indicative and explorative (given the low variation in the language spoken at home), **bivariate regressions result points to some positive correlation between speaking the Myanmar language at home and being a stronger reader** (Figure A4.3). Living with an adult (parents, guardian, and grandparents) who can read the Myanmar language fluently also seems to positively relate to higher EGRA scores. Finally, teachers whose first language is Myanmar have higher performing students, which may indicate that these teachers are better at teaching this specific language.

Class types, teacher demographics, qualifications, training, and years of experience

As is generally the case in the education literature, few of the basic teacher characteristics described in subsection 5.1 (gender, qualifications, experience, **etc.) are significantly correlated with student EGRA scores.** Pre-service training on how to teach reading is the notable exception. It is positively correlated with EGRA scores, but at a low significance level. In-service training on reading instruction is not correlated with scores, possibly indicating the low quality or relevance of this training. It is worrying because being taught by a teacher who has another income-generating job is significantly correlated with the EGRA scores.

Preschool

Results indicate a positive correlation between EGRA scores and preschool (as expected), as well as with student absenteeism (unexpected). Attending preschool and the duration of preschool attendance are both also significantly correlated with scores on a bivariate basis. Absenteeism, as reported by students, is not significantly correlated with the EGRA score. However, the absenteeism rate observed on the day of the visits showed—counter-intuitively—a positive correlation with performance, although at a very low level of confidence. This positive correlation may be an artifact of teachers asking the weakest students to stay at home on the day of the assessment, which was announced.

---

[67] In order of frequency, the non-Myanmar languages were Kayin (2%), Shan (0.6%), Chin (0.5%), Rakine (0.4%), Kachin (0.2%), Mon (0.2%), and others (0.2%).

Teaching and learning material

Owning a Myanmar reader, having access to other reading material at home, and having an exercise book were all significantly and positively correlated with EGRA scores, as was the extent of teachers' markings in the exercise books. The use of teaching materials other than the Myanmar reader (reported by teachers) is negatively correlated with EGRA scores when the material is Pimya and positively when the material is Pan Taing. This possibly indicates different values of these complementary resources. The presence of reading resources in the school (or library, etc.) is also positively correlated with scores, but not the frequency of use.

Homework, family involvement, and outside tutoring

Contrary to outside tutoring and help at home, which are positively correlated with scores, the frequency of homework does not seem to matter. With regards to students reading skills, it is quite troubling that the frequency of homework, as reported by students or teachers, is not significantly correlated with scores. Contrary to the frequency of homework, getting help with

homework is positively correlated with EGRA scores, but the proportion of parents reviewing the homework (as perceived by the teacher) is not. The variable for students who get outside tutoring (paid or not) in reading or in the Myanmar language is also significantly and strongly correlated with EGRA scores. The possible usefulness of such tutoring sessions hints that they compensate for what may not be adequately taught in school.

Reading activities, pedagogical approaches to teach reading, and assessments

None of the activities or pedagogical approaches is significantly correlated with EGRA scores, and using chapter-end and end-of-term tests (summative assessments) is, intriguingly, negatively correlated with EGRA scores. Using other assessments is not significantly correlated with EGRA scores. Every purpose for using the assessments is also negatively correlated with scores, but only a few are significant.

Table A4.4 and A4.5 below present a summary of the bivariate analysis between EGRA score and student or teachers variables.

Table A4.4: Bivariate correlation between EGRA score and student variables
(Figures in red are significant at a 5% level.)

Variable	Coeff.	Std. err.	Z	P>|z|	[95% Conf.	Interval]
Non-remote schools (vs. remote schools)	0.29	0.09	3.04	0.00	0.10	0.47
Female	0.19	0.04	4.59	0.00	0.11	0.28
Speaks Myanmar at home	0.26	0.06	4.03	0.00	0.13	0.38
Adult reads Myanmar	0.16	0.06	2.86	0.00	0.05	0.27
Attended preschool	0.10	0.05	2.07	0.04	0.01	0.20
Duration of preschool	0.04	0.02	2.15	0.03	0.00	0.08
Not absent last week	0.06	0.05	1.21	0.23	-0.04	0.16
Absent because sick	-0.03	0.06	-0.50	0.62	-0.15	0.09
Owns Myanmar reader	0.36	0.06	5.55	0.00	0.23	0.49
Other reading material at home	0.17	0.05	3.44	0.00	0.07	0.27
Other reading material (in Myanmar) at home	0.22	0.05	4.25	0.00	0.12	0.32
Owns Myanmar exercise book	0.07	0.02	3.71	0.00	0.03	0.10
Number of pages used in Myanmar exercise book	0.00	0.01	0.40	0.69	-0.02	0.03
Frequency of marking by teachers	0.17	0.02	7.11	0.00	0.12	0.22
Own (self-written) stories in exercise book	0.28	0.23	1.21	0.23	-0.17	0.74
Homework last week	-0.01	0.05	-0.30	0.77	-0.11	0.08
Help with homework	0.14	0.05	2.51	0.01	0.03	0.24
Help (from adult) with homework	0.05	0.04	1.33	0.18	-0.03	0.13
Gets outside tutoring in reading/Myanmar language	0.24	0.05	5.02	0.00	0.15	0.33
Wealth (PCA)	0.08	0.02	4.54	0.00	0.05	0.11
Wealth (sum of durables)	0.05	0.01	4.76	0.00	0.03	0.07
Wealth group (according to SES1)	0.09	0.02	4.41	0.00	0.05	0.13

Table A4.5: Bivariate correlation between EGRA score and teacher variables

(Figures in red are significant at a 5% level, orange = significant at a 10% level, and green = significant at a 15% level.)

Variable	Coeff.	Std. err.	Z	P>|z|	[95% conf.	Interval]
Female teacher	0.00	0.10	-0.02	0.98	-0.20	0.20
Multi-grade class	-0.09	0.08	-1.09	0.28	-0.24	0.07
Teacher of Myanmar class only	0.13	0.10	1.28	0.20	-0.07	0.32
First language is Myanmar	0.09	0.05	1.74	0.08	-0.01	0.18
Fluent in Myanmar	-0.02	0.10	-0.24	0.81	-0.23	0.18
Class taught in other language than Myanmar	0.01	0.03	0.59	0.55	-0.03	0.06
Academic qualifications	-0.07	0.05	-1.22	0.22	-0.17	0.04
0-2 years of teaching	0.12	0.05	2.26	0.02	0.02	0.22
Pre-service training in reading	0.18	0.09	1.87	0.06	-0.01	0.36
In-service training in reading	0.07	0.07	1.00	0.32	-0.07	0.21
Teaching certification	0.00	0.02	-0.13	0.89	-0.04	0.04
Absenteeism rate (on survey day)	0.00	0.00	1.49	0.14	0.00	0.01
Use of Myanmar reader	-0.06	0.06	-1.01	0.31	-0.18	0.06
Find Myanmar reader useful	0.38	0.02	15.97	0.00	0.34	0.43
Other material used (own)	-0.03	0.11	-0.24	0.81	-0.24	0.19
Other material used (Pimya)	-0.23	0.15	-1.48	0.14	-0.53	0.07
Other material used (Du Won)	0.05	0.09	0.57	0.57	-0.13	0.23
Other material used (Pan Taing)	-0.19	0.09	-2.19	0.03	-0.37	-0.02
School has library, etc.	0.15	0.07	2.23	0.03	0.02	0.28
Frequency of use of the library	0.00	0.03	0.09	0.93	-0.05	0.06
Written tests	0.03	0.06	0.57	0.57	-0.08	0.15
Oral tests	0.00	0.06	0.01	0.99	-0.12	0.12
Lesson recitation	0.04	0.06	0.60	0.55	-0.08	0.15
Homework	-0.03	0.11	-0.31	0.76	-0.24	0.18
Chapter-end test	-0.22	0.05	-4.40	0.00	-0.32	-0.12
End-of-term evaluation	-0.37	0.14	-2.69	0.01	-0.64	-0.10
Other	0.10	0.10	0.98	0.33	-0.10	0.31
Summarize assessment (chapter-end tests or end-of-term tests)	-0.23	0.05	-4.58	0.00	-0.32	-0.13
Categorize students (create instructional groups)	-0.08	0.07	-1.22	0.22	-0.21	0.05
Grade students	-0.07	0.07	-1.04	0.30	-0.20	0.06
Evaluate students' understanding of subject matter	-0.15	0.07	-2.14	0.03	-0.28	-0.01
Identify children who need help	-0.12	0.07	-1.76	0.08	-0.24	0.01
Adapt teaching to better suit students' needs	-0.02	0.07	-0.31	0.75	-0.16	0.12
Inform parents/administration of pupils' performance	-0.18	0.08	-2.41	0.02	-0.33	-0.03
Other	0.15	0.16	0.98	0.33	-0.16	0.46
Assessments used to assess understanding or adapt to needs	-0.11	0.05	-2.09	0.04	-0.22	-0.01

Frequency of homework	0.00	0.04	-0.06	0.96	-0.07	0.07
Proportion of parents who review homework	-0.04	0.04	-1.04	0.30	-0.13	0.04
Students repeat the name of a letter after the teacher models	-0.02	0.06	-0.34	0.74	-0.14	0.10
Students repeat the sound of a letter after the teacher models	-0.01	0.06	-0.21	0.83	-0.13	0.11
Shown a letter, students give the name of a letter without model	-0.08	0.07	-1.17	0.24	-0.21	0.05
Shown a letter, students give the sound of a letter without model	0.04	0.06	0.58	0.56	-0.09	0.16
Students repeat isolated words after the teacher models	-0.07	0.17	-0.39	0.70	-0.41	0.27
Students read isolated words without model						
Teacher reads first, student(s) read/repeat after	0.09	0.11	0.85	0.39	-0.12	0.31
Student(s) read a new, unseen text or passage without teacher modeling	0.04	0.05	0.79	0.43	-0.06	0.13
Engage students in whole class reading	-0.09	0.14	-0.62	0.54	-0.36	0.19
Engage students in silent reading	-0.01	0.05	-0.14	0.89	-0.11	0.09
Engage students in individual reading	-0.29	0.20	-1.41	0.16	-0.69	0.11
Teacher provides text comprehension questions and answers first	0.01	0.05	0.19	0.85	-0.09	0.10
Students answer questions on text after	0.02	0.05	0.33	0.74	-0.08	0.12
Teacher tells a story orally, students repeat the story verbatim	-0.01	0.05	-0.13	0.90	-0.10	0.09
Teacher tells a story orally, students retell the story in their own words	0.06	0.07	0.85	0.39	-0.08	0.21
Teacher explains the meaning of new words	-0.19	0.20	-0.95	0.34	-0.59	0.20
Students repeat new words	-0.03	0.08	-0.32	0.75	-0.19	0.13
Students construct new sentences by using newly learned words	-0.03	0.05	-0.49	0.62	-0.12	0.07
Students write compound words using known stem words	-0.05	0.05	-1.03	0.30	-0.15	0.05
Don't know/Refuse	0.00	(omitted)				
Traditional method of "thin bone gyi" (reading by spelling word by word)	-0.09	0.08	-1.19	0.24	-0.24	0.06
Sound-based reading	-0.01	0.06	-0.15	0.88	-0.13	0.11
Look and say method	-0.03	0.06	-0.48	0.63	-0.15	0.09
Lecture and discussion method	0.00	0.07	0.02	0.99	-0.13	0.13
Questioning method	-0.06	0.06	-0.88	0.38	-0.18	0.07
Pictorial story telling method	0.01	0.06	0.24	0.81	-0.10	0.13
Role play method	-0.07	0.06	-1.07	0.29	-0.19	0.06
Playing with words	0.05	0.06	0.76	0.45	-0.07	0.17
Others	0.18	0.05	3.26	0.00	0.07	0.28
Times it takes teacher to come to school	-0.04	0.03	-1.29	0.20	-0.11	0.02
Teacher has another income-generating job	-0.22	0.06	-3.84	0.00	-0.33	-0.11

A4.3 Selecting variables for multiariate analysis

For the simplest explanation and relevance, only a subsample of the available variables (student, teacher, and school characteristics) is included in the multivariate model below. This subsample of variables was based on similar bivariate correlations with EGRA scores as the one presented above. Without using formal hypothesis testing, bivariate correlations for each of the possible variables (original responses as well as derived or constructed ones) were reviewed to identify those with the highest potential as explanatory variables. Due to the generally weak observed relationships between these variables and the EGRA composite scores, the selection was also influenced by cross-validating the correlation between grades. The variables were retained for later analysis if they showed consistent or predictable relationships with the EGRA composite score across different grades.[68] Table A4.6 lists and defines the explanatory variables examined in these models.

To facilitate the selection, these explanatory variables were grouped into specific sets, namely, areas important to the study.

- **Group A**—variables of explanatory importance that cannot be intervened upon, or control variables at the student and teacher level, which describe factors contextualizing the students' educational environment, but are not directly amenable to educational practice

- **Group B**— students' educational environment at home or family and home variables that affect school outcomes, where the Ministry of Education could have some influence (e.g., through communication campaigns)

- **Group C**—variables that can account for students' initial skill levels when they enter primary education and that serve as a proxy of school readiness

- **Group D**—variables that represent student-teacher interactions in a classroom setting

- **Group E**—school variables that represent education inputs (or education support in schools) and other elements affected by education policy

- **Group F**—variables that represent teaching methods for early reading skills, such as letter naming and repetition, appropriate to non-literate students.

- **Group G**—variables that represent teaching methods appropriate for more advanced reading skills (at primary level), such as answering comprehension questions and creating new words

The first column in table A4.6 describes the variable name, the second column describes the data source, and the third column describes the original question(s) or, for derived variables, the calculation of the variable from the original questions.

---

[68] This criterion is less likely to capitalize on chance relationships, compared to arbitrarily setting alpha levels for significance.

Table A4.6: List of variables used in the econometric analysis and corresponding set

Set		Source	Description
A	Female	Student	[Are you female?]
A	Material possessions	Student	[The number of the durables item in a student's home]
A	Absent (sick)	Student	Were you absent last week [because you were sick]?
B	Adult reader	Student	Does a [parent or grandparent] read?
B	Outside tutoring	Student	Do you get reading or Myanmar language tutoring outside of school?
B	Adult help with homework	Student	[Is there a parent or grandparent at home who helps you in your studies?]
C	Travel time (to school)	Teacher	How long does it take you to come to school?
C	Other job	Teacher	Do you have to work another job to get enough income for your family?
D	Preschool (duration)	Student	How many years did you attend pre-school?
D	Preschool (at least 2 years)	Student	[Did you attend two or more years of preschool?]
E	Exercise book (own)	Student	Do you have a Myanmar language exercise book?
E	Exercise book (use)	Student	[How many pages used in the Myanmar exercise book?]
E	Exercise book (marking)	Student	[How many pages have teachers' marks, corrections, or comments?]
E	Exercise book (stories)	Student	Have you written any of your own stories in your exercise book?
E	Library	Teacher	In your school or classroom, do you have a library, a reading corner, a book box, or book shelf?
E	Assessment (summative)	Teacher	[Without being prompted, teacher reports using chapter-end tests or end-of-term tests.]
E	Assessment (use)	Teacher	[Teacher uses assessments to evaluate student understanding or adapt to student needs.]
F	Isolated words, no model	Teacher	Students read isolated words without model.
F	Letter name, model	Teacher	Students repeat the name of a letter after the teacher models.
F	Lecture/discussion	Teacher	Lecture and discussion method
F	Letter name, no model	Teacher	Shown a letter, students give the name of a letter without model.
F	Isolated words, model	Teacher	Students repeat isolated words after the teacher models.
F	Read/repeat words, after teacher	Teacher	Teacher reads first, student(s) read/repeat after.
F	Repeat story, after teacher	Teacher	Teacher tells a story orally, students repeat the story verbatim.
F	Letter sound, model	Teacher	Students repeat the sound of a letter after the teacher models.
F	Pictorial story telling	Teacher	Pictorial story telling method
F	Letter sound, no model	Teacher	Shown a letter, students give the sound of a letter without model.
F	Read text, no model	Teacher	Student(s) read a new, unseen text or passage without teacher modeling.
F	Retell story (own words), after teacher	Teacher	Teacher tells a story orally, students retell the story in their own words.
F	Silent reading	Teacher	Engage students in silent reading.
F	Sound-based reading	Teacher	Sound-based reading
F	Whole class reading	Teacher	Engage students in whole class reading.
F	Thin bone gyi	Teacher	Traditional method of "thin bone gyi" (reading by spelling word by word)
F	Look and say	Teacher	Look and say method
G	Text comprehension questions, no help	Teacher	Students answer questions on text after.
G	Text comprehension questions, help	Teacher	Teacher provides text comprehension questions and answers first.
G	Write compound words	Teacher	Students write compound words using known stem words.
G	Repeat new words	Teacher	Students repeat new words.
G	Playing with words	Teacher	Playing with words
G	Role play	Teacher	Role play method
G	Construct new sentences, with new words	Teacher	Students construct new sentences by using newly learned words.
G	Questioning	Teacher	Questioning method

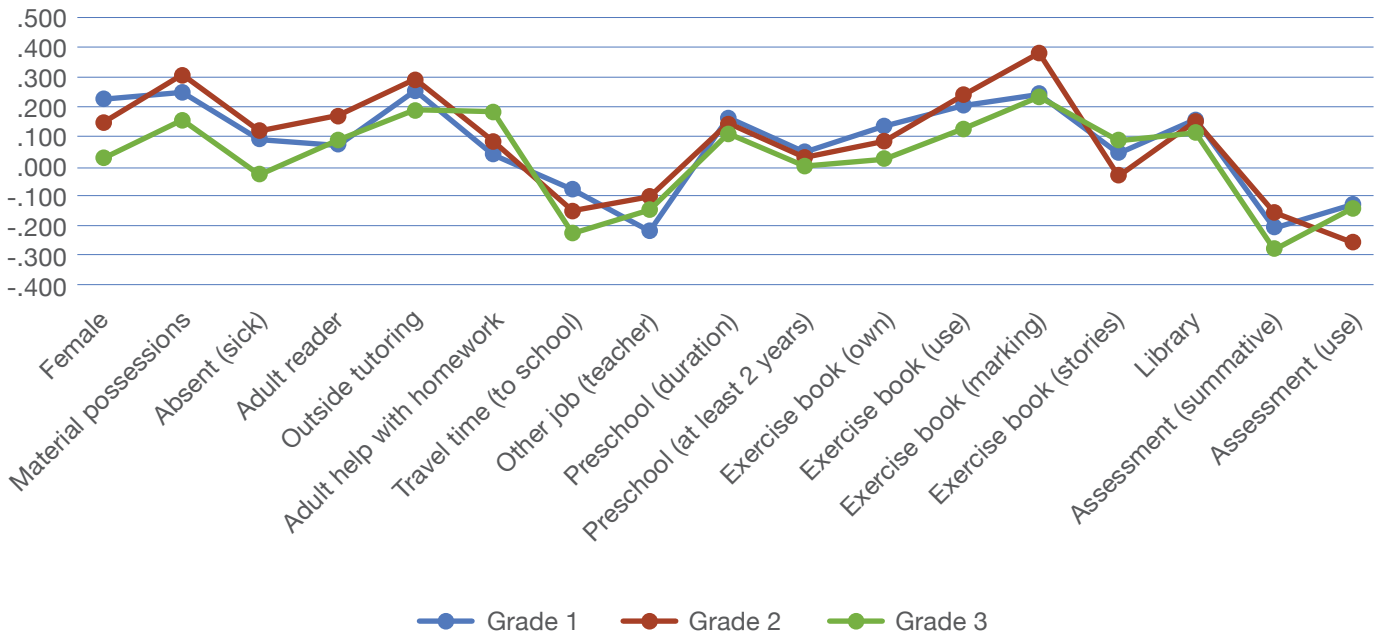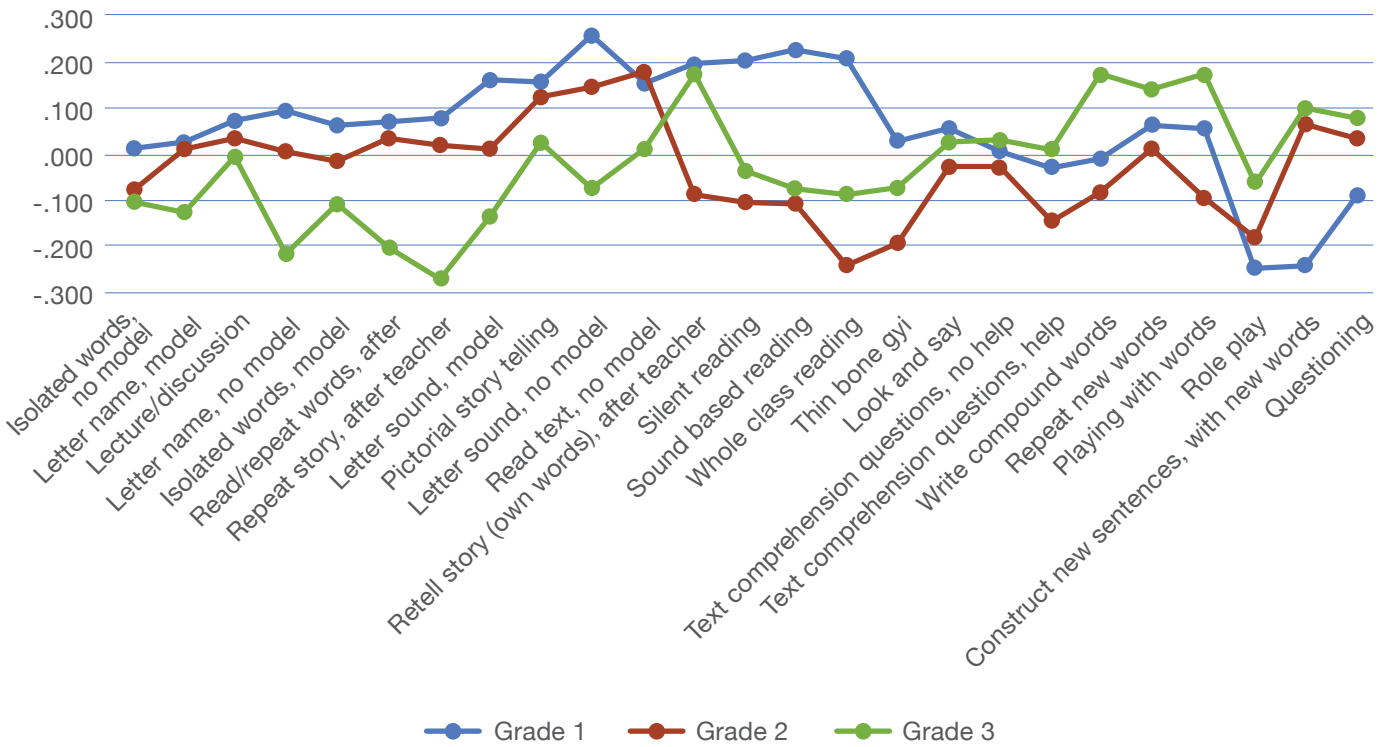## Figure A4.7: Correlations between variable sets A, B, C, D, and E

Figure A4.8: Correlations between variable sets F and G

The simple correlations between the retained variables and the outcomes are presented in Figure A4.7 and A4.8. In each figure, there are three lines, representing Grades 1, 2, and 3, which connect points showing the correlation between each variable and the outcome for each grade. The lines are illustrative only, to emphasize the relative similarity or difference between the results for the different grades. Figure A4.7 displays the correlations for variables sets A, B, C, D, and E. Figure A4.8 displays the correlations for variables in sets F and G. In both figures, the variables have been reordered from left to right to correspond to the variable sets and to provide clear illustrations.

The notable difference between figure A4.7 and figure A4.8 is that the correlations for variable sets A-E are almost identical for all grades, even though these variable groups include a wide range of household, personal, and contextual conditions. Although the ranges of magnitude of the correlations are similar in the two figures, this consistency of the correlations in Figure A4.7 suggests the presence of commonly observable phenomena. Indeed, as also discussed in subsection A5.1 (see subheading "Bivariate analysis"), many of the relationships are intuitive, such as positive correlations for female status, wealth, adult support (for reading and homework), prior attendance in preschool, use of a Myanmar language exercise book, and availability of a library, as well as the negative correlations for teacher travel time and teachers having other jobs. However, there are also several non-intuitive relationships, such as the positive relationship for absence due to sickness and the negative correlations with use of summative assessment and the use of assessment results for modifying instruction. These counter-intuitive relationships are a reminder that bivariate relationships may be the result of many factors, including a reversal of the expected direction of causality.

The possibility of reverse causality—or education practice responding to students' needs, rather than driving student performance—is also a likely factor underlying the pattern of results of figure A4.8. On the left hand side of the figure, the variables correspond to variable group F, the instructional practices more suitable to non-literate students or students just learning to read. The variables on the right side in Figure A4.8 correspond to set G and are typically associated with more advanced students.

For variable set F, the correlations are positive for Grade 1 students and negative for Grade 3 students (with effects for Grade 2 students generally in the middle). This pattern is consistent with teachers using instructional methods appropriate to younger students when they are teaching older students with poor reading skills. On the other extreme, we see a possible similar effect (although weaker in magnitude), where more techniques for more advanced readers have a positive correlation for Grade 3 students and weaker or negative correlations for students in Grades 1 and 2. In either case, these patterns should serve as a warning against simplified interpretations of the findings. As is the case with most survey data analysis, any strong inference of causality should be suspended until more conclusive follow-up case studies or focused experimental research can be conducted.

A4.4 Analysis of variance decomposition

Analysis of variance decomposition between the different successive models, described earlier (null, base, context, control, and inputs), indicates that a high proportion of the variation between schools is explained by non-educational variables. The potential effects of attending preschool may already be covered by student and family background variables, the use of Myanmar exercise book plays a key role, and pedagogical and reading activities have low explanatory power. Table A4.1 summarizes the variance decomposition in each of the successive models. The order of the columns from left to right reflects the sequence of model development: the results for the base model are furthest left, and those for the interpretation models are furthest right. The top three data rows in table A4.1 contain the variance attributable to the school, grade, and student (the main random effects in the model). For each model, there are two columns, "Var." and "%" as shown. The "Var." column describes the unexplained variance of the EGRA scores on its original scale. The "%" column contains the relative size of the variance component. The values are calculated as percentages of the total variance.

Table A4.9: Variance decomposition across models

	Null		Base		Context		Control		Set E		Set F		Set G	
	Var.	%	Var.	%	Var.	%	Var.	%	Var.	%	Var.	%	Var.	%
School	0.132	19	0.060	45	0.062	47	0.038	29	0.022	17	0.047	36	0.046	35
Grade	0.016	2	0.008	47	0.008	51	0.015	97	0.006	35	0.009	56	0.010	64
Student	0.548	79	0.503	92	0.503	92	0.481	88	0.481	88	0.481	88	0.480	88

Because there are no explanatory variables in the null model, the sum of the variance components (in the "Var." column) equals the total sample variance in the EGRA score and the sum of the three percentages (in the "%" column) equals 100. For all other models, the values are calculated as a percentage of the corresponding variance component of the null model and the percentage reflects how well the model explains differences between schools, grades, and students.

The null model indicates that 19 percent of the variation in EGRA scores is a result of school differences and only two percent is attributable to grade. In other words, knowing which school a student attends is five times better at predicting a student's reading performance than knowing what grade a student is in.[69] Accordingly, reviewing the results, the base model alone, which does not include any educational variables, explains over half of the variation between schools. The remaining school level variance is only 45 percent of the original.

In contrast, the base model only explains eight percent of the student-level variance. The context model offers no improvement on the base model because the potential effects of attending preschool are already covered by student and family background variables. Also, the control model again provides substantially more explanatory power of school differences: the additional 18 percent of school-level variance is primarily explained by the use of the Myanmar exercise book. Of the last three models, only set E offers any improvement over the control model, indicating that 83 percent of the school-level variance is accounted for. Sets F and G actually provide worse explanatory power relative to the control model, likely due to lack of degrees of freedom to fit the larger numbers of coefficients with the data.

A4.5 Coefficient Analysis

As with the bivariate analysis results, some caution is required in the interpretation of the variance decomposition since, despite providing statistical explanation, they do not necessarily explain the underlying phenomena. The explanatory variables in the model may simply be highly correlated with a phenomenon that is in turn highly correlated with school effectiveness without actually having any relationship to school effectiveness themselves. For this reason, it is useful to examine the variable coefficients themselves, as the patterns may suggest what some of the true underlying causes may be. However, in many cases, the variables represent proxies for conditions that the study cannot directly measure, despite a wish to. Accordingly, the magnitude of the effect sizes (coefficient) in the results should not be directly interpreted. Rather, they should be taken as lower-bound estimates that indicate the direction of association between the real world conditions they represent and the outcome of student reading performance.

For the most part, the results of this multivariate analysis reflect the observations in the bivariate analysis, although with some differences—namely the vanished correlation between scores and adult help with homework, teacher travel time, and enrollment in preschool. The estimated coefficients from the base, context, and control models are shown in Table A4.2. (The null model is not included because no coefficients are estimated for it.) The effects of all the student-level variables in the base model also vary significantly across schools. This indicates that, despite consistency of the effects across grades, there may still be ways for schools to ameliorate the effects of student background. Female status and wealth are still relatively strong predictors. Interestingly, the effects of adults reading and providing extra lessons (outside tutoring) are relatively strong and positive, effectively explaining away the effect of adult help on homework. Teacher travel time, which also had a strong (but negative) bivariate relationship, also has a null effect when controlling for other variables. This most likely is because the effects of travel time are explained by socioeconomic factors related to school surroundings, which are already captured by student family background.

---

[69] Although the student variance component is relatively large, this is not problematic. Since the purpose of the analysis is to explain differences between schools, we are more concerned with how the models reduce the school variance component.

Table A4.10: Estimated coefficients across base, context, and control models

	Base	Context	Control
Female	0.19 ***	0.19 ***	0.17 ***
Speaks Myanmar at home	0.23 ***	0.23 ***	0.21 ***
Non-remote	0.12	0.11	0.12
Wealth (durable)	0.03 ***	0.03 ***	0.03 ***
Absent (sick)	-0.02	-0.02	-0.02
Other job	-0.19 ***	-0.20 ***	-0.21 ***
Adult reader	0.13 ***	0.14 ***	0.13 ***
Outside tutoring	0.18 ***	0.17 ***	0.15 ***
Adult help with homework	0.03	-	-
Travel time (to school)	-0.04	-	-
Preschool duration	-	0.02	0.02
Preschool for at least 2 years	-	-0.06	-0.06
Exercise book (own)	-	-	0.18 **
Exercise book (use)	-	-	0.03 *
Exercise book (marking)	-	-	0.14 ***
Exercise book (stories)	-	-	0.40 ***

* significant at 1%;   ** significant at 10%;   *** significant at 5%.

Note: Not all variables are included in all models. Where an effect is missing to the right of its first appearance in the table, the variable was removed from subsequent models because they exhibited chance relationship (effect was not significantly different from 0).

Of particular note in the base model is the large effect of teachers having another job, which is strongly negatively related to student performance. The negligible effects of attending preschool indicate that preschool itself is likely strongly correlated with other strong predictors of performance, such as family background. In the control model, there are relatively strong effects related to use of the Myanmar exercise book. Although owning an exercise book has a significant relationship by itself with performance, each additional variable describing the use of the book (quantity of use, teacher feedback, students writing original stories) is associated with additional performance increases.

Table A4.2 summarizes the corresponding results for the three variables groups describing instructional practice. Only group E variables had additional significant relationships to the outcome, consistent with the bivariate relationships observed. The fact that these relationships persist after inclusion of the control variables suggests that these effects are unrelated to school intake and likely have a direct effect on performance, despite the non-intuitive nature of the assessment-related effects. In contrast, the effects of variables in groups F and G are consistently negligible after the inclusion of the control variables, despite the interpretable pattern seen in the exploratory analysis.

Table A4.11: Estimated coefficients across model with variable sets E, F, and G

	Set E Educational Inputs	Set F Pedagogy for poor readers	Set G Pedagogy for good readers
Female	0.17 ***	0.16 ***	0.17 ***
Speaks Myanmar at home	0.22 ***	0.21 ***	0.20 ***
Non-remote	0.11	0.10	0.12
Wealth (durable)	0.03 ***	0.03 ***	0.03 ***
Absent (sick)	-0.01	-0.03	-0.02
Other job	-0.17 ***	-0.24 ***	-0.25 ***
Adult reader	0.13 ***	0.13 ***	0.13 ***
Outside tutoring	0.16 ***	0.16 ***	0.15 ***
sq18_homeworkadulthelp	-	-	-
tq_30travel_time	-	-	-
Preschool duration	-	-	-
Preschool for at least 2 years	-	-	-
Exercise book (own)	0.18 **	0.16 *	0.18 **
Exercise book (use)	0.03 *	0.03 *	0.03 *
Exercise book (marking)	0.13 ***	0.14 ***	0.14 ***
Exercise book (stories)	0.45 ***	0.40 ***	0.39 ***
Library	0.15 ***	-	-
Assessment (summative)	-0.20 ***	-	-
Assessment (use)	-0.11 ***	-	-
Isolated words, no model	-	-0.07	-
Letter name, model	-	0.06	-
Lecture/discussion	-	-0.03	-
Letter name, no model	-	-0.13 **	-
Isolated words, model	-	-0.04	-
Read/repeat words, after teacher	-	0.05	-
Repeat story, after teacher	-	-0.01	-
Letter sound, model	-	-0.04	-
Pictorial story telling	-	0.04	-
Letter sound, no model	-	0.09	-
Read text, no model	-	0.05	-

Retell story (own words), after teacher	-	0.07	-
Silent reading	-	-0.03	-
Sound-based reading	-	0.01	-
Whole class reading	-	-0.09	-
Thin bone gyi	-	-0.03	-
Look and say	-	-0.01	-
Text comprehension questions, no help	-	-	-0.01
Text comprehension questions, help	-	-	-0.02
Write compound words	-	-	-0.07 *
Repeat new words	-	-	-0.06
Playing with words	-	-	0.03
Role play	-	-	0.03
Construct new sentences, with new words	-	-	0.03
Questioning	-	-	-0.08

* significant at 1%;   ** significant at 10%;   *** significant at 5%.

Note: As previously explained, note that although the control variables remain the same for each model, the variables from sets E, F, and G effects are entered and removed in block form.

WORLD BANK GROUP