EdData II

PhilEd Data: Strengthening Information for Education, Policy, Planning and Management in the Philippines

Component 2: Early Grade Reading Assessment Results
(Grades 3 English and Filipino, Grade 1 Ilokano)
Preliminary version

EdData II Technical and Managerial Assistance, Task Number 17
Contract Number: AID-492-M-12-00001
April 25, 2013

This publication was produced for review by the United States Agency for International Development. It was prepared by RTI International.

PhilEd Data: Strengthening Information for Education, Policy, Planning and Management in the Philippines

Component 2: Early Grade Reading Assessment Results
(Grades 3 English and Filipino, Grade 1 Ilokano)
Preliminary version

Education Data for Decision Making (EdData II)
Task Order No. 17

Prepared for
Mirshariff Tillah, Contracting Officer's Representative (COR)
USAID/Philippines Office of Education
mtillah@usaid.gov

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

RTI International is a trade name of Research Triangle Institute.

The author's views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

Table of Contents

List of Tables

List of Figures

Abbreviations

ARMM	Autonomous Region of Muslim Mindanao
BEAM	Basic Education Assistance for Mindanao
BEIS	Basic Education Information System
BPA	Blanket Purchase Agreement
CALABARZON	Cavite, Laguna, Batangas, Rizal, and Quezon (Southern Tagalog Mainland)
CAR	Cordillera Administrative Region
CEM	Center for Educational Measurement
CI	confidence interval
COR	Contracting Officer's Representative
cwpm	correct words per minute
DepEd	Philippine Department of Education
E/F	English and Filipino
ECARP	Every Child a Reader Program
EdData II	USAID Education Data for Decision Making II
EFA	Education for All
EGRA	Early Grade Reading Assessment
ELSA	Education Livelihood Services Alliance
EQuALLS 2	Education Quality and Access for Learning and Livelihood Skills project
IEA	International Association for the Evaluation of Educational Achievement
K–12	kindergarten to grade 12
L1, L2, L3	first, second, or third language
MANCOM	DepEd Management Committee
MIMAROPA	Occidental Mindoro, Oriental Mindoro, Marinduque, Romblon and Palawan (Southern Tagalog Islands)
MIT	Massachusetts Institute of Technology
MLE	multilingual education
MMPO	Metro Manila Program Office (Save the Children)
MT	mother tongue
MTB-MLE	mother-tongue-based multilingual education
NAEP	National Assessment of Educational Progress
NAT	National Achievement Test
NCES	U.S. National Center for Education Statistics
NCR	National Capital Region
NETRC	National Educational Testing and Research Center
ORF	oral reading fluency
PA	phonemic awareness
PhilEd Data	Strengthening Information for Education, Policy, Planning and Management in the Philippines (EdData II Task 17)
PISA	Program for International Student Assessment \
PRIME	Philippines' Response to Indigenous Peoples' and Muslim Education

RTI	RTI International (trade name of Research Triangle Institute)
SCMPO	South Central Mindanao Program Office (Save the Children)
SEAMEO INNOTECH	Southeast Asian Ministers of Education Organization's Regional Center for Educational Innovation and Technology
SES	socioeconomic status
SOCCSKSARGEN	South Cotabato, Cotabato, Sultan Kudarat, Sarangani and General Santos City (Central Mindanao)
STRIVE	Strengthening Implementation of Visayas Education
TEPT	Test of English Proficiency for Teachers
TNS	Taylor Nelson Sofres
USAID	United States Agency for International Development

Acknowledgments

This study was carried out under a technical assistance contract awarded to RTI International from USAID under the Education Data for Decision Making (EdData II) Blanket Purchase Agreement (BPA) mechanism to support the Philippine Department of Education (DepEd) Education to strengthen their ability to generate and use education data. The Strengthening Information for Education Policy, Planning and Management in the Philippines (PhilEd Data) project began in June 2012 and includes support to DepEd's ongoing efforts to improve analysis and use of existing assessment results and develop strategies for new national assessments aligned with the curricular reforms under way.

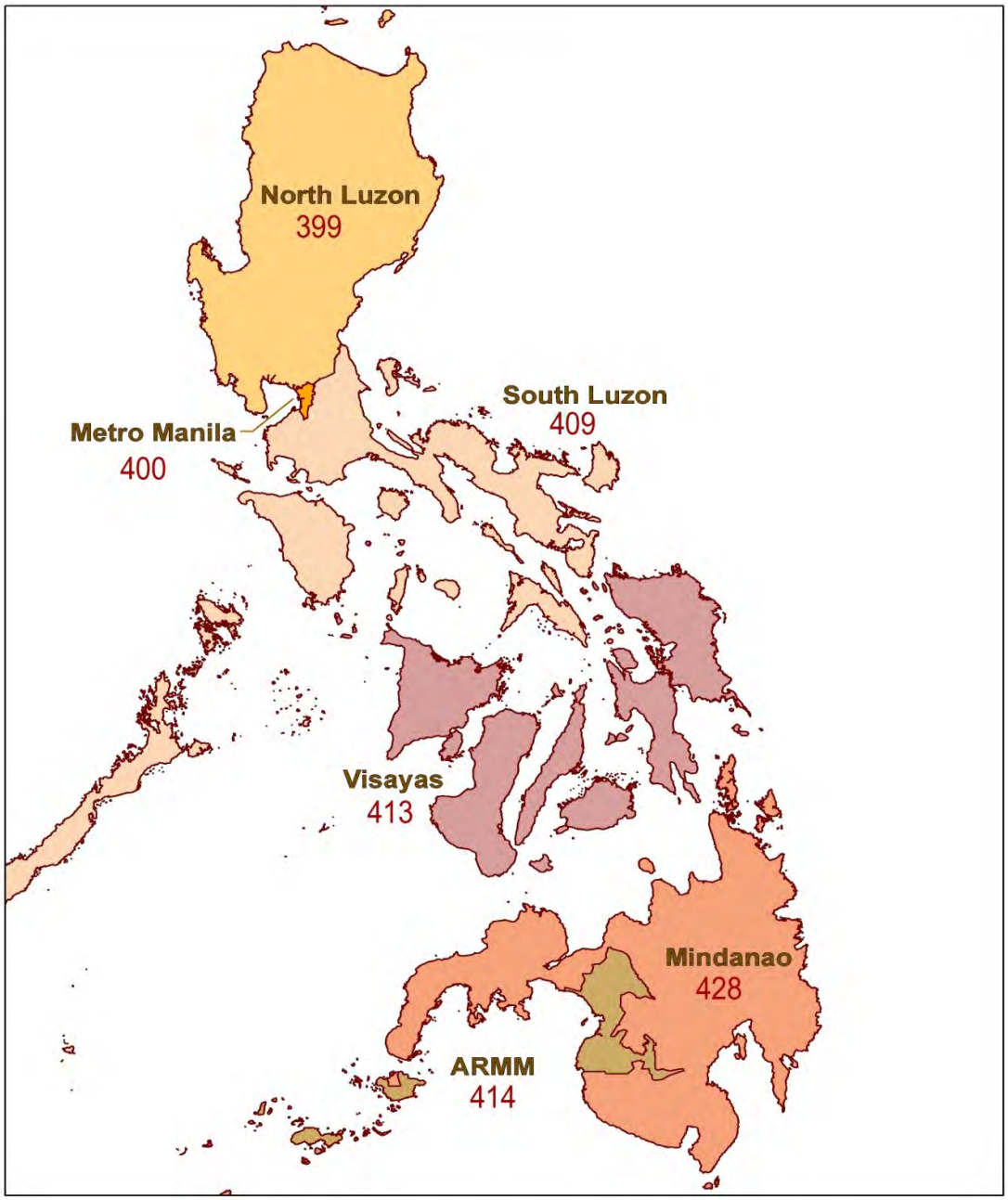
The PhilEd Data team consists of three core institutional partners: RTI International, the Southeast Asian Ministers of Education Organization's Regional Center for Educational Innovation and Technology (SEAMEO INNOTECH), and the Center for Educational Measurement (CEM)—along with Taylor Nelson Sofres (TNS) Global, which managed the field research for this early grade reading assessment.

RTI gratefully acknowledges the support of the previous Undersecretary for Programs and Projects, Dr. Yolanda S. Quijano, who was instrumental in supporting Early Grade Reading (EGRA) activities, mobilizing support from regional DepEd officials, and ensuring cooperation from sampled schools. Additionally, we welcome the new Undersecretary, Dr. Dina Ocampo, and look forward to working together to discuss the results of this activity and the policy implications. We particularly recognize the individual members of the DepEd EGRA team, who also supported planning and implementation of this research; and the experts who contributed to instrument adaptation and training and to the finalization of this report, including the TNS program manager for this study.

Dr. Rosalina Villaneza
Dr. Paraluman Giron
Mr. Galileo Go
Ms. Angelika Jabines
Ms. Joycelyn Tuguinayo
Ms. Ligaya Ilagan
Ms. Yammie Tolentino
Dr. Roderick Aguirre
Ms. Marie Christine Reyes
Mr. Firth McEachern

We are, of course, most grateful to all of the DepEd and TNS assessors, and TNS field managers too numerous to name who gave so much of their time to conduct this important fieldwork on time despite the challenges of national data collection in a country so geographically complex. Additionally, we thank SEAMEO INNOTECH, and particularly Ms. Sylvette Gunigundo, for hosting the assessor training workshops in Manila and San Fernando (La Union) and ensuring that everything ran smoothly. RTI is represented locally by Mr. Rufino Jamilano, who was invaluable during the coordination and implementation of the events.

Map of the Philippines



Sampled grade 3 students by grouped (super-)region..

I. Introduction

The development of education in the Philippines has reached a critical juncture. The Department of Education (DepEd) is planning to build the classrooms and hire the additional teachers needed to make the final push toward reaching its Education for All (EFA) targets and Millennium Development Goal for Education by 2015. At the same time, the 2011 Sector Assessment[1] documented the recent deterioration of quality and a persistent dropout problem evidenced by declining cohort survival rates. Rising to the challenge of improving quality, DepEd has launched a complete restructuring of elementary and secondary education that includes adding two years to the secondary cycle and overhauling the curricula for all subjects from kindergarten to 12th grade (the K to 12 reform program). One major area of curriculum reform involves introducing mother tongue-based, multilingual education (MTB-MLE) as an approach to improving student acquisition of literacy in the early grades of elementary school.

The MTB-MLE program was designed by executive order in 2009 to institutionalize mother-tongue instruction in the early grades of elementary education across the school system. A Strategic Plan for implementation was completed in 2010. The Strategic Plan laid out specific activities to be undertaken, including advocacy; pre- and in-service teacher training; materials development; policy development and resource mobilization; and assessment, monitoring, and evaluation.[2] The implementation of MTB-MLE means that students have to establish strong reading skills in grades 1 through 3 so that those skills can be transferred into English and Filipino, which are the official languages of instruction after grade 3.

Language and Literacy in the Philippines

The Philippines is a multilingual and multiethnic society with between 120 and 171 distinct languages in the country. The Tagalog-based Filipino is widely understood and used in all domains of life, but Filipino is considered a second language for many, while Tagalog is considered a mother tongue. Census data suggest that while no language enjoys a majority advantage in the country, 65 out of 76 million Filipinos are able to speak Filipino as a first or second language.[3] Another eight languages are considered "major" languages (based on census data and numbers of speakers). These are (1) Bikol, (2) Cebuano, (3) Hiligaynon, (4) Ilokano, (5) Kapampangan, (6) Pangasinan, (7) Tagalog, and (8) Waray or Samarnon. Although many indigenous languages are spoken at home and in communities, printed materials are primarily in English (commerce, advertising, bureaucracy) and Filipino (newspapers, public service messages, entertainment).

The promotion of Tagalog as a national, unifying language dates back to 1937, but it wasn't until 1957 that Filipino was officially declared the national language. The distinction between Tagalog and Filipino is constantly evolving, since Filipino is meant to be enriched by

[1] Buendia, R., Gregorio, J., Molera, R. A., Flor, B. G., Vergel de Dios, B., Ganibe, J.W., Balonkita, A. G., Dawang, C., & Mirandilla, N. (2011). *Philippine Education Sector Assessment Project.* Report prepared for USAID by International Technology Management Corporation. Manila, Philippines.

[2] *Mother-Tongue-Based Multilingual Education Strategic Plan.* Manila: Department of Education, Republic of the Philippines, 13 February 2010.

[3] Nolasco, R. M. (2008). *The prospects of multilingual education*. Report of the Komisyon sa Wikang Filipino [Commission on the Filipino Language]. Retrieved from http://www.seameo.org/_ld2008/doucments/Presentation_document/NolascoTHE_PROSPECTS_OF_MULTILINGUAL_EDUCATION.pdf

vocabulary from other indigenous languages. In practice, there is very little difference in lexicon and no difference in grammar[4] between the two languages, but Filipino is treated as a language that is distinct from that which is spoken by the Tagalog ethnic group. While Filipino is designated as the "national" language, the 1974 bilingual education policy institutionalized English and Filipino as languages of instruction and the 1987 Constitution declares both Filipino and English as "official" languages. In school, this meant that students—regardless of home language—began their primary education in a dual-language immersion model where English was used as the language of instruction for science, math, and technology subjects; and Filipino for social science, civic education, sports, and the arts. Use of local mother tongues was tolerated as necessary in order to facilitate understanding of the subjects. Yet English is a second language for most Filipinos. A 2008 survey by Social Weather Stations indicated that about three fourths of Filipino adults said they could understand and read English, while less than half (46%) said they could speak it and 8% said they were not competent in any way.[5]

Recognizing how difficult it is for a child to learn when the language(s) of instruction is(are) not their mother tongue, at least a dozen mother-tongue pilot programs have been implemented and evaluated in the Philippines over the past 60 years.[6] These as well as international experiences have demonstrated the benefits of early grade instruction in mother tongue, and have confirmed that literacy skills in a mother tongue transfer to second (L2) and third (L3) language reading. As a result, the current policy[7] stipulates the use of mother tongues in grades K–3 as media of instruction and languages of literacy development in grades 1 and 2. English and Filipino (E/F) are gradually introduced orally and then literacy development transitions to E/F in grades 2 and 3, preparing children for fully bilingual E/F instruction by fourth grade. The government has identified an initial 16 major languages to use as mother tongues in the early grades and is investing heavily in teacher training, curricular reform, and materials development. With the advent of MTB-MLE, students will need to successfully learn to read in mother tongue by the end of grade 1 or early in grade 2 so that they can transfer that skill to English and Filipino in grade 3 and beyond.

Reading Reform Programs

Alongside the major curricular reforms being undertaken by the government under the K–12 program and MTB-MLE initiatives, the government has, for many years, been supporting education quality reform initiatives aimed at improving literacy outcomes. Some of these are large-scale, donor-funded programs and others are smaller, school-based initiatives. For example, the Every Child a Reader Program (ECARP) provides catalyst funds to schools to support literacy improvement programs across a range of proven strategies:

- *Reading remediation activities* such as specialized training to teachers who then offer on-on-one assistance to struggling readers outside of class time.

[4] Nolasco, R. M. (2007, August 24). Filipino and Tagalog: Not so simple. *Dalityapi Unpoemed.* Retrieved from http://www.dalityapi.com/2007_08_01_archive.html

[5] Nolasco (2008), *The prospects of multilingual education.*

[6] See, for example, the presentation *Mother-tongue-based multilingual education (MTB-ME),* prepared by the DepEd Management Committee (MANCOM), September 2009. Retrieved from http://www.depedro7.com.ph/uploadedFiles/file/Whats%20New/EXECON%209-17-09/Mother%20Tongue-based%20Multilingual%20Education%20%28MT-MLE%29.pdf

[7] See Republic Act 10157, known as the "Kindergarten Education Act," of January 20, 2012. Retrieved from http://www.lawphil.net/statutes/repacts/ra2012/ra_10157_2012.html

- *Specific classroom instructional strategies* such as word walls, pair and group reading, read-alouds.
- *Increasing time for students to practice reading* such as Drop Everything and Read initiatives in various forms, in-classroom mini-libraries, all-school reading times 15 minutes each day, weekly library hours.
- *Encouraging parents to be engaged in their children's learning* by reviewing assessment results together, inviting parents and other community members to come into the school for storytelling and read-alouds, and encouraging parents to have their children read at home.

"Read-a-thon," "Read more, be more," National Literacy Month (November), and others are smaller advocacy initiatives encouraged by DepEd but led by schools themselves or local nongovernmental organizations to encourage children to spend more time reading age-appropriate books in school and out of school through a variety of fun and competitive activities.

Additionally, many initiatives address specific education challenges in the regions. Examples include USAID's Education Quality and Access for Learning and Livelihood Skills (EQuALLS2) Project which focuses on improving access to and quality of education in grades 1-3 and out-of-school youth, specifically in the ARMM and Regions 9 and 12 in Mindanao; the Basic Education Assistance for Mindanao (BEAM), Strengthening Implementation of Visayas Education (STRIVE), and Philippines' Response to Indigenous Peoples' and Muslim Education (PRIME) are regional initiatives supported by AusAID. These and other programs too numerous to mention here reflecting an awareness of the need to provide targeted interventions to specific subpopulations with different needs.

Purpose of the 2013 Early Grade Reading Assessments

To support quality implementation of the abovementioned reforms, the Department of Education recognizes the importance of measuring reading performance and progress over time. The MTB-MLE Strategic Plan established benchmarks for performance and stressed the need to "establish achievement, persistence, dropout and other baselines for all schools where MTB-MLE is implemented" as well as "comparative achievement studies" that would document how well schools are performing one year after implementation of MTB-MLE. The 2013 EGRA studies described here had two basic objectives supporting the Strategic Plan:

- **Objective 1:** To establish a scientific, national baseline of reading performance in English and Filipino in third grade. The cohort of children tested in school year 2012–2013 reached third grade prior to the implementation of K to 12 and MTB-MLE. So this assessment provides a "pre K to 12, pre MTB-MLE" baseline.
- **Objective 2:** To assess mother-tongue reading acquisition in a region in the country where the MTB-MLE approach had been piloted since 2011–2012. This would provide some initial data on how well children were being taught to read in a mother tongue supported by the national curricular reforms.

Although previous efforts have been made to institute classroom-based measures of reading ability (e.g., the Philippine Informal Reading Inventory), there is currently no systematic way of measuring and aggregating early reading outcomes across schools in the Philippines. The

National Assessment Test (NAT) begins measuring English and Filipino achievement in grade 3, and then again in grade 6, but the NAT is more of a language/language arts test than a reading test. Moreover, since it is delivered as a written test, it automatically assumes some reading ability on the part of the children. As a result, there is no reliable measure of what early literacy skills children are acquiring.

The present report describes the results of two reading assessments that were carried out to measure performance of a representative national sample of children reading E/F at the end of grade 3 (Objective 1, above) and a regional sample of children reading Ilokano in grade 1 (Objective 2, above). The first included a national sample of 2,463 students, selected in a stratified, random framework to be representative of six geographic areas: North Luzon, Metro Manila, South Luzon, Visayas, Mindanao, and the Autonomous Region of Muslim Mindanao (ARMM). The second was a regional sample of 494 children in 50 schools selected randomly in Region 1 from all schools who reported having begun implementation of Ilokano instruction in school year 2012–2013 or 2011–2012. Additional details about the sample and the sampling methodology appear in ***Annexes 1 and 2.***

Although the purposes are different, the basic methodology was the same for both studies. In neither case was the study comparative or experimental, i.e., comparing schools with MTB-MLE against schools without it, or comparing performance at two different points in time. However, as an initial baseline it will allow such comparative studies to be undertaken in the future. For Ilokano, the assessment measured how well the current curriculum and its implementation were giving children the foundational skills needed to acquire literacy, and as such, whether established benchmarks were realistic and aligned to the language context. For E/F, the grade 3 assessment provided a reference for the extent to which the existing system of bilingual immersion had prepared children to "read to learn" in the higher grades using these two standard languages of instruction. In the future, follow-on assessments conducted with the same rigor can compare results to this baseline to understand whether beginning to read in their mother tongue leads to similar levels of proficiency in E/F.

In both cases, the exercise also served to build the capacity of the Department of Education to develop early grade reading assessment instruments and conduct rigorous, sample-based assessments, and is expected to support their efforts to conduct additional evaluations of early grade reading.

About the Early Grade Reading Assessment Tool

In November 2006, USAID and the World Bank supported RTI and a team of reading experts to develop a 15-minute, individual oral assessment of five core skills that are considered to be the building blocks of reading: phonemic awareness, alphabetics, fluency, vocabulary, and comprehension. The purpose was to give low-income countries a tool for systematically measuring how well children read in the early grades. The resulting tool became known as the Early Grade Reading Assessment, or EGRA, which translates the five core skills into observable tasks: recognizing letters of the alphabet, distinguishing sounds within words, reading simple words, and understanding sentences and paragraphs. Because the assessment components are aligned with essential and teachable reading skills, the results provide clear guidance for changing instructional methods and offer hope of improvement. Measurements of how quickly and accurately children can read a text out loud, and how much of it they understand, also align with a scientific and a popular understanding of what it means to be

able to read. The key differences between EGRA and other kinds of national diagnostic tests are that:

- The assessment is aligned to essential and teachable components of reading and to the expectations of the local curriculum, but it is curriculum independent. The same instrument and items can be used to compare results across different curricula, instructional methods, or grades.
- It is orally administered by a trained assessor to one child at a time.
- There are timed subtests that provide measures of fluency (the number of items answered correctly per minute).

An adaptation process conducted in the Philippines, involving the Department of Education and language specialists from around the country, served to ensure that the instruments used for this exercise:

- were culturally appropriate
- corresponded to the grade-level material students were expected to be able to read
- reflected the way the language was being used locally in reading materials for the target grade level
- reflected the way that reading is typically taught
- took into consideration the linguistic structure of the language
- addressed issues of mother-tongue interference, local dialects, and pronunciation.

The instruments developed were pilot tested with more than 150 students and analyzed using "Rasch analysis" to ensure that items measured a range of skills and were neither too easy nor too hard for the target population.[8] The final instruments included the subtests summarized in ***Table 1.***

Table 1: EGRA instrument subtests in the Philippines

Subtest	Skill	Description The child is asked to…	Grade(s) implemented
Initial-sound segmentation (Filipino and English only)	Phonemic awareness	…say the first sound heard in a word pronounced by the assessor. For example, "Tell me the first sound you hear in the word 'bat'" (Answer is "b").	Grade 3
Differentiating initial sounds (Ilokano only)	Phonemic awareness	… say which word, of three pronounced by the assessor, begins with a different sound. For example, "Tell me which word begins with a different sound: bat, ball, car." (Answer is "car")	Grade 1
Letter-sound identification	Alphabetic principle—letter-sound correspondence	… say the sound each letter makes, while looking at a printed page of 100 letters of the alphabet in random order, upper and lower case. Timed.	Grade 1 Grade 3
Familiar word reading	Fluency—automatic word recognition	… read a list of 50 common words printed on a page. Timed. For example, "stop," "the," "was."	Grade 1 Grade 3
Nonword reading	Alphabetic principle—	… read a list of 50 non-words printed	Grade 1

[8] For more information or data on the Rasch analysis for this study, please contact spouez@rti.org.

Subtest	Skill	Description The child is asked to…	Grade(s) implemented
	letter sound correspondence and Fluency—automatic decoding	on a page. Timed. Words are constructed from actual orthography, but are not real words. For example "jaf," "tob."	Grade 3
Oral reading fluency	Fluency—automatic word reading in context	…read a short story out loud (up to 60 words in length) printed on a page. Timed.	Grade 1 Grade 3
Prosody (Filipino and English only)	Comprehension	… complete the oral reading fluency subtest while the assessor scores expressiveness and size of word groups read.	Grade 3
Reading comprehension	Comprehension	… respond to 5 questions that the assessor asks about the short story.	Grade 1 Grade 3
Listening comprehension	Language comprehension and vocabulary	…listen to a story that the assessor reads out loud, then answer questions about the story.	Grade 1 Grade 3
Dictation (Filipino and English only)	Alphabetic principle	… write a short sentence that the assessor says out loud.	Grade 3

These subtests can be characterized as "fluency" measures based on whether they are timed or untimed. A timed subtest is administered over a one-minute period in which the child reads as many letters, words, or sentences possible. The assessor notes which letters or words are read incorrectly, and at the end of a minute, the assessor notes how many letters or words total were read. The result is calculated and expressed as "correct items per minute." Nonfluency subtests include untimed sets of questions.[9] The results are calculated and expressed as "percentage of items correct out of total possible."

Each third grade child was assessed in both English and Filipino, whereas grade 1 children were assessed only in Ilokano. In addition to the reading assessment, children were asked a series of questions about their home and school environment. These questions were designed to gather contextual information that can help explain variations in reading ability. A short classroom and school snapshot also gathered basic information about school enrollment, class sizes, availability of textbooks, and progress in the reading curriculum at the time that the EGRA took place. For full methodological details related to data collection—including assessor training, fieldwork procedures, and quality control—please refer to a separate report prepared by TNS, *Methodological Report for the Early Grade Reading Assessment* (April, 2013).[10]

II. Outcomes Summary

Overview

The ultimate purpose of any reading assessment is to answer a seemingly simple question: "Are students learning to read and understand text appropriate to their grade level?" To answer that question, EGRA uses research-based measures of oral reading fluency (accuracy

[9] Although the child was given a limit of time in which to answer the questions—for example, 3 or 5 seconds—this is not considered a timed subtest resulting in a measure of fluency. The time limit serves only to limit the overall duration of the test, and to create consistency among assessors in administering the test.

[10] For a copy please contact Yazmin.Tolentino@tnsglobal.com.

and speed) and comprehension. The other subtests, as outlined above, measure underlying skills leading to fluency—phonemic awareness, decoding, alphabetic principle—and can be useful for understanding more specifically what children can and can not do, and thus explaining reading and comprehension outcomes. This section provides an initial broad overview of performance across all three languages, and as an initial introduction refers only to oral reading fluency (ORF) and comprehension outcomes. Section III provides more detailed results on all subtests, for all languages. The instruments are available in the full methodology report for this assessment.[11]

Average scores in reading and comprehension

Mean ORF scores for each language, along with the proportion of children who achieved 80% comprehension or better, are provided in ***Table 2.*** Children were asked comprehension questions based on the amount of text they read in one minute; thus, a child who read slowly—i.e., 20 correct words per minute (cwpm) or less—may have been asked only two questions out of five possible. Only the children who completed the reading passage within a minute were asked five questions. Therefore, there are two ways to describe comprehension: either as a percentage of questions answered correctly out of the total questions possible (five, in this case), or as a percentage of questions asked. Both measurements are shown in Table 2.

Table 2: Mean ORF and comprehension scores

Grade and language	Mean ORF (cwpm)	% achieving 80% comprehension (out of total of 5 questions possible)	% achieving 80% comprehension (out of total attempted)
Grade 3 – Filipino	68	64%	67%
Grade 3 – English	67	17%	18%
Grade 1 – Ilokano	18	8%	14%

cwpm = correct words per minute.

Although English and Filipino oral reading fluency were essentially equivalent, comprehension scores were much lower for English, indicating a lack of oral language and vocabulary background in English. As mentioned in the background section above, English was an L2 for most of these children, and this survey found that only 2.5% of children spoke English in the home (based on students' self-reporting). For Ilokano children, the low comprehension score aligns with the low reading fluency demonstrated.

Note that the differences in the two measures of comprehension were not large for grade 3, since most children read to the end of the 59-word story. For Ilokano, however, most children did not complete the story and the actual number of questions which they were asked varied. Therefore, the first measurement—percentage correct out of total questions possible—likely underrepresents children's actual levels of comprehension. Nonetheless, the low levels of ORF and comprehension in Ilokano are surprising given that 75% of students reported

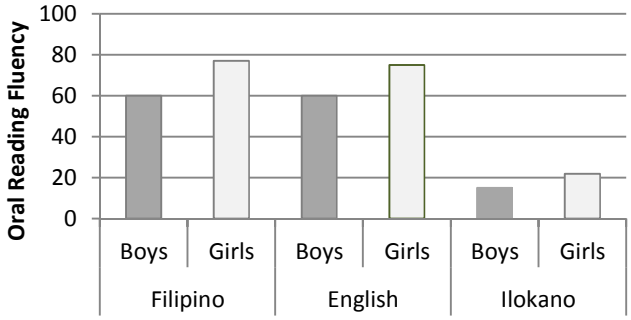
[11] Flores, J., Tolentino, Y. (2013). *Methodological report for the Early Grade Reading Assessment.* Report prepared for USAID PhilEd Data. Manila: TNS and RTI International.

speaking Ilokano as their mother tongue. More details on distribution of scores on ORF and comprehension are presented in the sections that follow.

Gender dimensions

An important and statistically significant finding is that girls were doing better than boys across languages, subtests, and regions. For example, on average compared to boys, girls read about 25% more fluently in English and Filipino in grade 3 and 45% more fluently in Ilokano in grade 1. ***Figure 1*** shows the differences in ORF for girls and boys on all three subtests.

Figure 1: Boys' and girls' performance in ORF



The reasons for this gender gap have not yet been uncovered by this research, but it is useful to note that consistent gender differences in reading and math are found internationally, such that this is not necessarily a reflection of inequities in the system.[12]

Regional variation

The 1987 Constitution emphasizes the right to free, relevant, integrated basic and secondary education for all Filipino citizens, and it sets out a number of strategies for achieving equity and reaching poor, traditionally underserved populations. These include support for nonformal, indigenous alternatives to public education, such as *madrasah* education; equitable investment in school infrastructure; a national system of early childhood education; and financial subsidies for the most underprivileged. Yet this study has shown significant disparities among regional populations in the Philippines, particularly the Mindanao and ARMM regions ***Table 3.***

[12] For example, a meta study of data from the Program for International Student Assessment (PISA) showed that 15-year-old boys scored lower in reading in all nations in all four PISA assessments (2000, 2003, 2006, 2009), whereas boys scored higher than girls in mathematics: Stoet, G., & Geary. D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 years of PISA data. *PLOS ONE, 8*(3): e57988.

A 1990s international reading literacy study including data from the Philippines showed that at the 9-year-old level, girls had higher achievement than boys in all countries for narrative passages, and in 24 countries for expository passages: International Association for the Evaluation of Educational Achievement (IEA). (n.d). *Reading Literacy Study*. Informational website, http://www.iea.nl/reading_literacy_study.html

Table 3: Average ORF and reading comprehension by grouped Region

Region	Filipino		English	
	ORF	Comprehension	ORF	Comprehension
Metro Manila	80	84%	83	40%
Visayas	69	68%	69	38%
South Luzon	69	84%	74	30%
North Luzon	68	73%	67	31%
Mindanao	62	69%	62	28%
ARMM	56	61%	56	25%
National average	68	73%	67	32%

* Comprehension scores are percentage of questions answered correctly out of total possible.

The differences in regional averages were statistically significant. Further analysis of the characteristics of strong versus weak readers[13] indicated that a child in the ARMM region was about five times more likely than his or her peers in Metro Manila (the highest performing region) to be a weak reader in Filipino and six times more likely to be a weak reader in English. In Mindanao, the odds were closer to three times for both languages. Considering the findings of both gender and regional differences, it is important to recognize that boys in ARMM were at a significant disadvantage; whereas the mean Filipino ORF for boys in Metro Manila was 76, the mean for boys in ARMM was 47 and in Mindanao, 54.

This study therefore supports previous findings from other institutions that noted barriers to education for children in ARMM and Mindanao: lack of educational infrastructure, poverty, instability, lack of parental support, and cultural biases.

> *"…the legal bases to expand access to basic education are in place. However, it is noted that dropout and participation rates remain high, and lack of teachers and educational resources are still widespread in Mindanao. These problems are most acute in the provinces of ARMM, especially those affected by conflict"* (p. 40).[14]

A smaller-scale study by Save the Children[15] using an approach similar to EGRA also found that students in the South Central Mindanao region scored much lower than students in Metro Manila on all components of the test, in grade 1 and grade 3 alike.

Achievement according to benchmarks

The Philippines Department of Education has set specific and ambitious benchmarks for each grade level and language using accuracy and ORF as the key indicators of reading progress.[16] ***Table 4*** shows the results of English, Filipino and Ilokano assessments in terms of the ORF and accuracy benchmarks set by DepEd.

[13] For the purposes of this study, a "weak reader" is defined as one who read less than 30 cwpm.

[14] SEAMEO INNOTECH. (2007). *Policy research on access to quality basic education for Muslim learners: Final report.* Prepared for the Education Livelihood Services Alliance (ELSA), USAID Education Quality and Access for Learning and Livelihood Skills (EQuALLS) project. Manila: USAID. Retrieved from http://pdf.usaid.gov/pdf_docs/PNADK756.pdf

[15] Cao, Y. (2009). *Literacy boost assessment: Philippines Country Office, MMPO/SCMPO.* Manila: Save the Children Philippines Country Office. Retrieved from http://resourcecentre.savethechildren.se/node/6862.

[16] Department of Education. (2012). *K to 12 Curriculum Guide: English, January 31, 2012;* and Department of Education (2012). *K to 12 Curriculum Guide: Filipino, January 31, 2012.*

Table 4: Results according to Philippines national ORF benchmarks

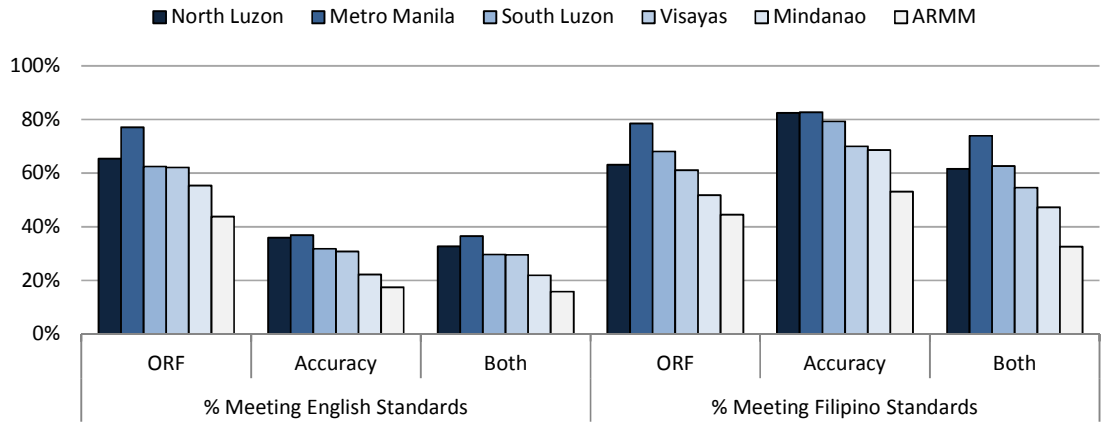
Grade and language	ORF benchmark	% meeting ORF	Accuracy benchmark	% meeting accuracy	% meeting both
Grade 3 – Filipino	80 cwpm	**36%**	95-100%	**74%**	**35%**
Grade 3 – English	60 cwpm	**61%**	95-100%	**30%**	**28%**
Grade 3 English and Filipino		**35%**		**28%**	**20%**
Grade 1 – Ilokano	60 cwpm	**< 1%**	95-100%	**23%**	**<1%**

The majority of children were not meeting the benchmarks, with the exception of English reading fluency (for which the reading benchmark is set lower than for Filipino fluency) and Filipino accuracy. Testing the same grade 3 children in English and Filipino allowed us to determine the number of children who had met all fluency and accuracy benchmarks for both English and Filipino. This was the case for only 20% of children.

Only a handful of children in the Ilokano sample met the grade 1 benchmark, set at the same level as for grade 3 English. Benchmarks usually are set by observing data from a large number of readers and studying the relationships between ORF and comprehension for that language. Large-scale assessments such as this one can be used to take a data-based approach to setting targets and monitoring them for all languages. The more detailed results presented in Section III can help determine what Ilokano learners were capable of after one year of instruction, but more longitudinal research will be needed to determine how children perform in grades 2 and 3 and what type of learning progression can be expected. Yet even if the grade 1 mother-tongue reading benchmark were set at 40 correct words per minute, then only 13% would have met the target at the time of this assessment, which is still a considerable minority.

Figure 2 presents the proportions of children, by region, who met the benchmarks in English and Filipino.

Figure 2: Achievement across regions in fluency and accuracy according to benchmarks



As shown by this graph, performance differed greatly by region. These regional differences most likely reflect other contextual factors such as language, socioeconomic status, and

absences, but would require more in-depth analysis. The next section describes some other factors associated with performance in addition to region and gender.

Factors correlated with reading achievement

The Save the Children study referenced above concluded that:

> *"Overall, results also indicate a strong relationship between SES, home literacy environment and literacy outcomes, but these results remain at the correlational and not causal level"(p. 62) and "… girls as a whole perform better than boys. Indeed, the lowest performing girls, from the T'boli group, still perform better than the best performing boys, from the Ilonggo group"* (p. 53).[17]

The findings from this EGRA study—a significantly larger, random, national sample—corroborated those conclusions, and additionally found other demographic and classroom factors affecting reading outcomes. ***Table 5*** describes school and non-school-level factors outside of gender and region of origin (discussed above) that were shown to be significantly associated with higher or lower ORF scores in English and Filipino. These characteristics came from the student context questionnaire (self-reported) and classroom snapshot (assessor's observation) instruments, which were then run through simple linear regression analysis.[18] These are the factors that were associated with statistically significant differences in oral reading fluency. The last two columns of the table show the magnitude of the difference for children who had the characteristic compared to children without the characteristic. For example, we can say that a child who attended Kinder before grade 1 is likely to read 12 words per minute more in E/F than a child who did not attend Kinder; furthermore, 81% of children in our sample had attended Kinder.

Table 5: Factors associated with reading achievement in English and Filipino

Characteristic	% of population affected	Difference in mean ORF scores for students in each category	
		English	Filipino
Factors positively associated with reading			
High socioeconomic status (SES) score *	10%	20	18
Attended Kinder before grade 1	81%	12	12
Has reading textbook at home	79%	8	9
School participates in ECARP	49%	6	8
Teacher *often* asks the child to read aloud	27%	9	8
Teacher assigns reading homework	79%	8	7
Child has nonschool books at home	45%	8	7
Factors negatively associated with reading			
If classroom is multigrade	11%	-9	-10
If teacher assigns no homework (English or Filipino) **	4%	-14	-16

[17] Cao (2009), *Literacy boost assessment.*

[18] As noted above, all instruments can be found in the separate methodology report for this study.

Characteristic	% of population affected	Difference in mean ORF scores for students in each category	
		English	Filipino
Missed 2, 3, or 5 days of school last week	27%	-12	-13
Is over-age	32%	-15	-14
Takes more than an hour to get to school **	2%	-20	-19

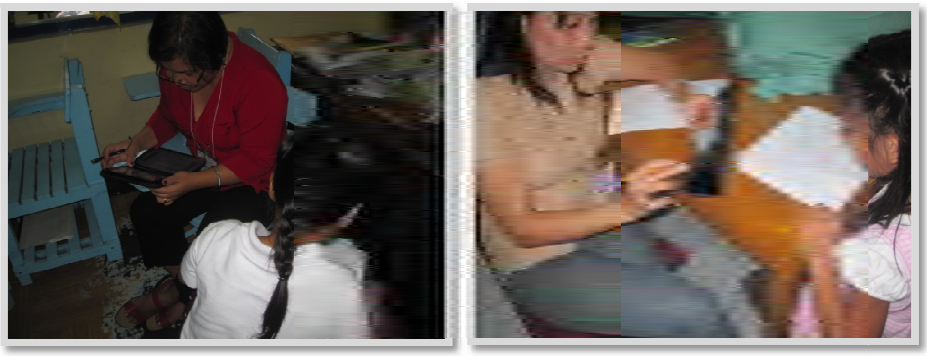
* For SES, the questionnaire asked whether the child had certain items at home that are associated with wealth, such as television, interior toilet, car, etc. For analysis, these were reduced to a value of how many total items the child had, regardless of which ones they were, and this became a proxy for wealth. More precise measures are being calculated and will be incorporated into an addendum to this initial report.

** Note that the actual numbers reported in these categories were small, as indicated by the percentages. N=83xx and N=33 for 'assigns no homework' and 'takes more than an hour to get to school', respectively. So although statistical significance is present, the association should be read with caution.

Of particular interest is the magnitude of differences caused by factors that negatively affected reading. Additionally, some of the factors—for example, the age of children in grade 3—affected a large proportion of the population. The expected age of a grade 3 child in the Philippines is 8 or 9 years old; children who were 10 years old or above yet were still in grade 3 scored 14–15 words per minute less in E/F reading fluency than children who were the appropriate age. Class size was not found to be a significant determinant of outcomes in any region or grade level. These and other factors are discussed in further detail, by language, in Section III.

Socioeconomic status also was correlated with reading achievement in an initial review of basic indicators of wealth; however, more precise analysis is underway and will be reported in the final version of this report. Interestingly, class size was not significantly correlated with reading achievement in any language in this study in grade 3 or in grade 1.

For a complete description of the sample characteristics, see Annex 1.



Practicing EGRA administration during assessor training, January 2013. Photo: S. Pouezevara.

III. Outcome Details, by Language

The data collected by this study from nearly 3000 children in two grades and three languages resulted in a large body of data that can be analyzed in multiple ways, both for each language an comparatively between languages. In an attempt to present the most important information clearly yet comprehensively for this preliminary report, we have separated the discussion of results by language; then within each language by subtest. For each language a broad

overview is presented of results and factors associated with the results. The discussion of each subtest includes some item analysis as well as distribution of scores and discussion of factors that resulted in large differences in real scores. Finally, some discussion and conclusion for each language before moving on to the next language or section. Where interesting comparisons between the languages were found, they are addressed within the discussion of the results. Section IV then concludes with some broad cross-language findings and recommendations.

Filipino Results

Students presented reading stimuli in the Filipino language at the end of grade 3 demonstrated relatively high fluency and comprehension across subtests, with the exception of letter-sound identification. The prevalence of zero scores was low, except for listening comprehension, which registered more than 10% zero scores. Zero scores indicate children who did not respond correctly to any stimuli. Some children obtained zero scores because they just did not say anything at all, but we cannot know if they did not respond because they *could not,* or because they were *unwilling* to do so.

Table 6 presents an overview of results on all subtests,[19] including the standard error (of the mean) and corresponding confidence interval. These two measurements tell us how confident we can be in the precision of our estimate of the mean.

Table 6: Filipino results on all subtests, differentiated by gender

Subtest	Overall Mean	Female mean	95% CI – lower bound	95% CI - upper bound	% zero scores
Initial-sound segmentation (10 items – % correct)	65%	68%	62%	68%	7.4%
Letter-sound identification (100 items, correct per minute)	21	23	19	22	9.9%
Familiar word reading (50 items, correct per minute)	54	58	52	55	1.4%
Invented word reading (50 items, correct per minute)	31	34	30	32	3.3%
Oral reading fluency (57 items, correct per minute)	68	77	66	70	1.4%
Reading comprehension (5 questions – % correct)	73%	78%	71%	75%	6.4%
Listening comprehension (3 questions – % correct)	66%	69%	63%	68%	11.2%

In addition to the low letter-sound identification mean score (21 letters per minute), it is surprising that the listening comprehension mean percent score (66%) was lower than reading comprehension mean percent score (73%), and that the percent of zero scores was higher (11.2% vs. 6.4%, respectively).

[19] With the exception of dictation. These scores are undergoing futher analysis and will be presented in an addendum to this report.

The factors outlined above, in Section II, Outcomes Summary, also were found to influence scores on all of these subtests. Specific, notable, changes in scores due to certain factors appear in the next sections, which provide more details on each of these subtests. Unless otherwise mentioned, all results pertaining to correlations are statistically significant at least the $p<0.05$ (95%) level, but often at the $p<0.001$ (99.9%) level.

Initial-sound segmentation

This subtest measured phonemic awareness (PA), or a general appreciation of the sound structure of language, as demonstrated by the awareness of sounds at three levels of structure: (1) syllables, (2) onsets and rimes, and (3) phonemes. A phoneme is the smallest linguistically distinctive unit of sound in a word. The child was asked to repeat the first sound he or she heard in a word pronounced by the assessor. The words used in the Filipino assessment are presented in ***Table 7,*** with the frequency of incorrect answers.

Table 7: Filipino phonemic awareness –and frequency of incorrect responses

nanay	bahay	repolyo	ikaw	lalaki	dakila	walis	halaman	una	patola
37%	34%	41%	28%	38%	30%	41%	31%	16%	35%

Three to five percent of the time, the child did not say anything, as opposed to saying an incorrect answer. The highest rate of error was for the word "repolyo," and the lowest for the word "una." Among children who got only one item correct, it was most often "ikaw"; and when children got only two items correct, they were most often "ikaw" and "una." This suggests that it was easier to isolate initial vowel sounds than initial consonant sounds, for both assessors and children. Assessors were trained to accept as correct only the isolated consonant sound and not a syllable sound—for example, a clipped /b'/, rather than /buh/ or /ba/—although they should have been tolerant of the slight aspiration of air that accompanies even the clipped consonant sound. This can be a difficult distinction to make.

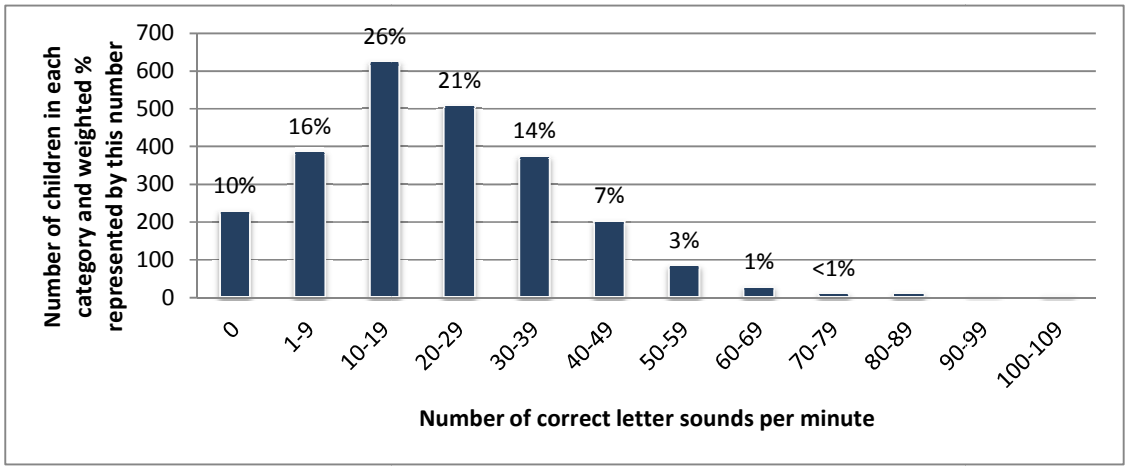
In all, 46% of children demonstrated very good PA skills by answering 8, 9, or 10 items correctly, which indicates very good PA skills, and subsequently these children had Filipino ORF scores averaging 78 correct words per minute—slightly higher than the overall average (68 correct words per minute). There was little difference between girls and boys on this subtest. Since PA is a skill that is not taught in higher grades (and should not be—it is a prereading skill appropriate for early learners), it is not unusual for older children to do poorly on this subtest. In this sample, the over-age children did progressively worse on this subtest. Children in South Luzon—who scored higher in English, according to the summary table (Table 3 in the previous section)—and ARMM also had a difference of about one word lower in achievement on this subtest compared to the highest-performing region (National Capital Region [NCR]). The average score was 59% correct responses in South Luzon, and 62% correct in ARMM. Children speaking Bisaya or Cebuano at home also did better on this subtest (73% and 80% respectively), so characteristics of other local languages may have interfered with the ability to pronounce the Filipino sounds in isolation.

On the other hand, we should expect children in grade 3 to know and use letter-sound correspondences to decode. The results for the next two subtests provide more insight on these two skills.

Letter-sound identification

This subtest involved presenting the child with a grid made up of all of the letters of the alphabet, in upper and lower case, in random order, and repeating according to the frequency with which they are found in the language. The child was asked to provide the sound that the letter makes in Filipino—not the name, but the sound. If the child gave the name, did not provide any response at all, or made an incorrect sound, the item was marked wrong. The subtest was timed to one minute and the result was a fluency calculation of correct letter sounds per minute. Knowledge of letter-sound correspondence has been shown to be a powerful predictor of later reading fluency,[20] particularly in languages with transparent orthographies like Filipino.[21] ***Figure 3*** shows the number and percentage of children who scored within defined ranges of letter-sound fluency (correct letter sounds per minute). The percentages shown above each bar in the graph are the weighted percentages, while the number shown in the vertical axis is the number sampled.

Figure 3: Filipino correct letter sounds per minute – distribution of responses by range of correct words per minute

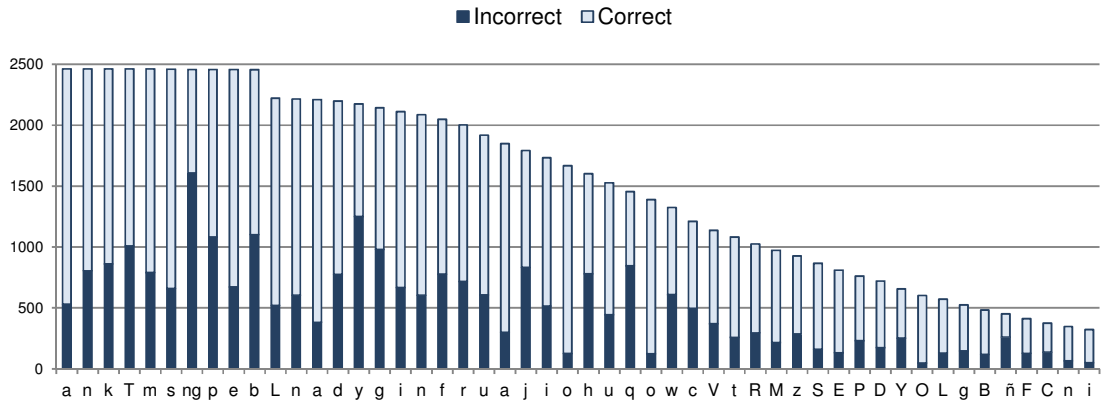


Almost 10% of the children discontinued the subtest because they gave no correct answers (or any answer) to the first 10 letters, resulting in a zero score for that subtest. The average number of letters attempted was 32, so most children got a chance to identify all of the letters of the alphabet, with the exception of x, z, and ñ, which appeared later in the item grid. ***Figure 4*** shows the number of children who attempted each item and the proportion who answered correctly or incorrectly. Items for which more than 40% of children who attempted the item answered incorrectly were: T, ng, p, b, y, g, j, h, q, w, c, and ñ. Letters with higher proportions of correct responses usually were vowels, probably because vowel names and sounds are the same, so children who interpreted the subtest incorrectly and were giving letter names got these answers correct anyway. We can see that all of the children attempted the first 10 items, as required by the administration procedures; but thereafter, the total number of items attempted began to drop according to the rate of responding.

[20] Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, Massachusetts: Massachusetts Institute of Technology (MIT) Press.

[21] A language has a "transparent orthography" when each letter is associated with a unique sound, and vice versa.

Figure 4: Performance on the first 50 Filipino letter sounds



We would expect that by grade 3, children would will know letter names and sounds automatically, without having to reflect or hesitate for an instant. In fact, this is the skill that is required to become a fluent reader. At a rate of 1 letter per second—which is still relatively hesitant—a child should be able to attempt at least 60 letters in a minute. With a 10% error rate, that would still be a letter-sound fluency of 50 letters per minute. The average of this group was only 21 correct letter sounds per minute. The rate of response (60 divided by the number of items attempted, or 32, in this case) was almost one letter every 2 seconds, which shows that either this skill had not been internalized, or children were uncomfortable with the test procedure and were hesitant to answer. Children who answered more than 50 items correctly averaged 91 words per minute on Filipino ORF, and 41 correct invented words per minute, both of which were much higher than the national averages of 68 and 31, respectively. Children who answered fewer than 50 items correctly averaged 67 correct words per minute on Filipino ORF and 31 on correct nonwords per minute, in line with the national averages. Therefore there was a correlation between automaticity of letter-sound recognition and the more complex skills of decoding and reading connected text despite any potential limitations to the test procedure (such as language interference, or assessor error).

The scores on this subtest were not affected much by gender, age, or region. Differences of only three to five letter sounds emerged when the data were disaggregated by various groups.

Familiar word reading

The Filipino familiar word subtest consisted of 50 words, in random order, presented on a page of 10 rows of 5 words each. The words were selected based on their frequency in the Filipino language, according to analysis and research provided by Filipino language specialists. As with other timed subtests, the child was given a minute in which to read as many words as possible (in the order that they appeared on the page). Incorrect answers or no response at all were counted as incorrect, and the final result was a score of "correct words per minute." Children may have read these words through decoding, but we expected that they would recognize them through prior experience with reading them in the classroom. Fluent reading depends on progressing from decoding to acquiring automaticity in word reading, and we should expect to see a higher reading fluency on the familiar word reading subtest than on the nonword reading subtest.

In fact, the average rate of fluency for this grade 3 national sample was 54 correct familiar words per minute—slightly more for females (58) and much higher for children in Metro Manila (62). Both fluency and accuracy were much higher for this subtest than for nonword reading, judging by the low rate of error on the first 10 words, as shown in ***Table 8.*** More than half of the children read all 50 words in the minute given for the subtest (the average was 44 words attempted), and the average accuracy score (number of items correct out of number of items attempted, regardless of time) was 86%.

Table 8: Filipino familiar word reading – items and frequency of incorrect responses (first 10 words)

halos	kami	pero	gawa	mo	tawag	sagot	mabilis	pa	lamang
6%	4%	7%	4%	3%	5%	3%	4%	1%	5%

The only words for which more than 10% of the children who attempted the word read it incorrectly were "kailangan" (14%) and "hangang" (14%). The word with the final "ng" also proved difficult in the nonword reading subtest and the "ng" sound in the letter-sound identification subtest also had a high rate of error. Thus there may be a problem teaching (or pronouncing? or hearing?) this letter-sound correspondence, particularly when it appears in longer words, since the words "lamang," "tanong" and "ang" did not have exceptionally high rates of error (i.e., under 5%).

Over-age children in grade 3 were reading on average 48 familiar words per minute, and children in ARMM or children in multigrade classrooms were reading 47 words per minute. Children who were never asked to read aloud in class were at 46 words per minute on average. Unlike the nonword reading subtest, familiar word reading fluency was more closely related to home language. Children who reported speaking Tagalog or Filipino at home were reading 57 correct words per minute. This is not much more than the average, but it is statistically significant. Similarly, when the community language (as reported by school directors) was Tagalog or Filipino, the average was 58 correct words per minute, possibly reflecting the influence of community media and the print environment on familiar word reading. Interestingly, Cebuano speakers scored even higher than the Tagalog or Filipino speakers—58 cwpm. Where Tausug was listed as the community language, results were much lower (43 cwpm).

The following familiar words also reappeared in the short story that was used for the ORF measurement discussed next: "sila," "naman," "ko," and "ang." The error rate on all of these words was 2% or less.

Nonword reading

This subtest consisted of 50 words, in random order, presented on a page of 10 rows of 5 words each. The words were invented words, with no meaning in Filipino, but with word structure (onset, rime, consonant clusters, etc.) that matched the actual orthography of the language. Thus the words were plausible Filipino words, but we can be certain that the child was attempting to read and decode them in Filipino for the first time. In other words, the subtest measured decoding ability, an important skill for becoming fluent. As with other timed subtests, the child was given a minute in which to read as many words as possible (in the order that they appeared on the page). Incorrect answers or no response at all were

counted as incorrect, and the final result was a score of "correct words per minute." The first 10 words of the subtest are provided in ***Table 9,*** with the percentage of incorrect responses for each word. All children attempted the first five words.

Table 9: Filipino nonword reading – items and frequency of incorrect responses (first 10 words)

paw	sibus	pla	nomi	goong	talis	lat	tahumu	panwa	ayt
8%	13%	20%	14%	29%	7%	5%	13%	10%	25%

The average number of words *attempted* was 36, and more than half of the sample read 39 out of 50 words. In all, 13% of the children attempted all 50 words, while 3.3% discontinued the subtest because they read no items correctly out of the first 5 words. The average rate of fluency (number of *correct items* read per minute) for grade 3 was 31—slightly more for females (34). The average number of items currectly decoded (accuracy) was 61% (number of correct items out of total attempted). Other patterns mentioned previously also were present in this subtest: Metro Manila had the highest fluency at 36 cwpm, while Mindanao and ARMM had the lowest fluency at 28 and 29 cwpm, respectively; over-age children, children in multigrade classrooms, and children who did not attend Kindergarten before grade 1 all scored lower (on average 27 cwpm). Interestingly, children who said they speak Filipino or Tagalog at home had only a slight advantage in this skill—33 correct words per minute on average, which is actually less than the average score of children who reported speaking Ilokano at home (34).

In order to read texts of increasing complexity, children will always face new words that they must decipher using decoding skills. Therefore, this is one of the standard subtests in all EGRA studies. The correlation to ORF is also clear: Children who read nonwords with accuracy of 80% or more had an average Filipino ORF of 97 cwpm, while those reading with accuracy of 20% or less had an average Filipin on ORF 16 cwpm.

Oral reading fluency (ORF)

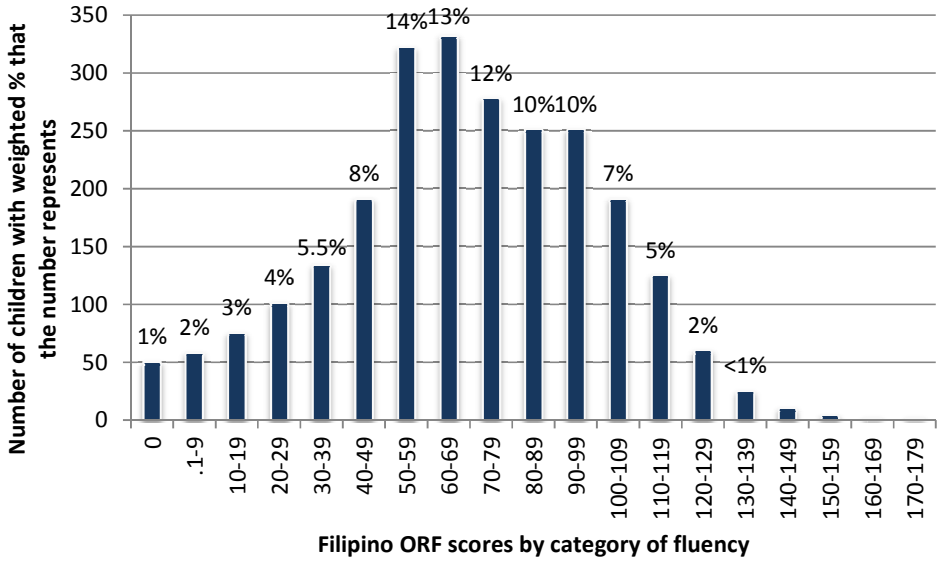
The ultimate goal of reading instruction is not speed, but fluency—a measure of both rate and accuracy—which is subsequently related to comprehension. A sufficiently high rate of reading is important in order to free cognitive resources to focus on word meaning and comprehension strategies; error-prone and hesitant readers apply their short-term memory capacity in decoding and word identification and have little left for synthesizing the meaning of the text from start to finish.

Thus, oral reading fluency was the summative outcome measure that provided an overall picture of whether the child had learned and was able to apply the prereading and early-reading skills described by the subtests above. For the familiar word reading subtest, words were unrelated to one another and out of context. By contrast, this subtest asked children to read a short story so that in addition to decoding and automatically recognizing words, children also could draw inferences and anticipate words based on the context of the sentence and story.

Filipino third graders were reading, on average, 68 words per minute, but with a wide distribution of skills, from zero to 177 cwpm. Nearly half (49%) of students were situated

around this mean, or between 50 and 90 cwpm, as shown by the distribution of scores in ***Figure 5.*** 26% of children read more than 90 cwpm, and 25% less than 50 cwpm (weighted percentages). Again, the percentages shown above each bar in the graph are the weighted percentages,

Figure 5: Filipino ORF – distribution of scores, by range of fluency



The short story that the children were asked to read is copied below ***(Figure 6).*** It consisted of 57 words total. The story was created, along with two others, during the instrument adaptation process involving Filipino linguists, teachers, and DepEd staff. This passage was chosen after pilot testing based on Rasch analysis because it represented the best distribution of difficulty across the range of respondents. Words highlighted in yellow were missed by between 9% and 13% of the children, and the word highlighted in blue was missed by 23% of the children.

Additionally, Figure 6 shows the required distance in the text that the child must have reached in order to attempt the comprehension question corresponding to that text. The words underlined were part of the response. Words in bold served as a reminder of frequent words that appeared in the familiar word reading subtest. Note that this was a minimum requirement, so some children who read more text (for example, all the way to the end) had more information with which to answer the questions correctly, even though a correct response was possible having read fewer words.

Figure 6: Filipino ORF text

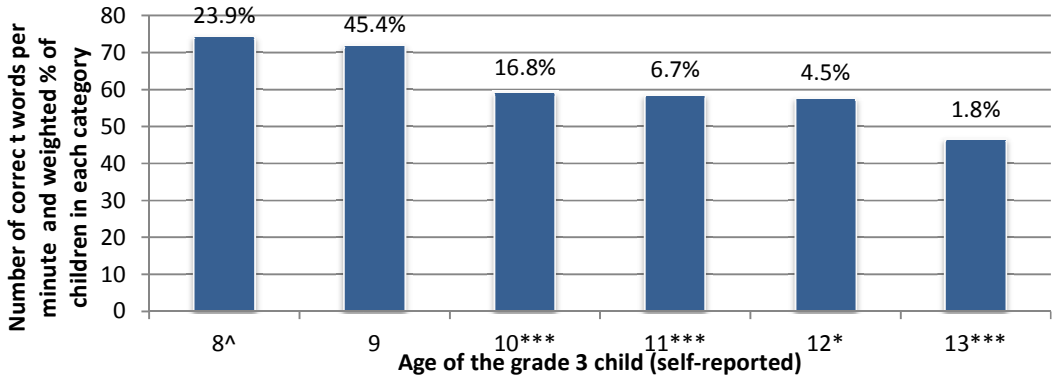
Nasa palengke sina nanay at Perla. [Q1] Bibili **sila** ng isda at gulay. Pumili si nanay ng bangus [Q2] para isigang. Pumili **naman** si Perla ng kangkong, [Q3] labanos, at petsay.

"Naku! Naiwan **ko ang** pitaka ko sa bahay," [Q4] ang sabi ni nanay sa tindera. "Puwede ko bang balikan na lang ang bayad? [Q5] Kukunin ko lang ang pitaka sa bahay."

As indicated by the shaded words above, only five words were missed by noticeably more children than the other words—between 9% and 23%. As noted, the word "isigang" was the most frequently missed word. However, it is important to note that most of these frequently missed words were key for answering the comprehension questions correctly. During the instrument adaptation process, the technical team was careful to work with linguists and then regional language speakers to ensure that all words were common across all regions. The word "isigang" is a particular way to prepare fish, and it may be that despite this review, in some areas of the country, this particular dish is uncommon or referred to by a different term.

Again, factors consistently associated with higher or lower scores included gender, region, age, and type of classroom (multigrade or not; see Section II, Outcomes Summary, for more details). Note that the characteristic of "being over age" was calculated by grouping children into one of two categories: (1) child was 10 years or older, or (2) child was 8 or 9 years old. The difference in the mean between these two groups was 14 words per minute in favor of children who were not over-age. However, this mean masks even larger differences in age ranges when they are not grouped together. For example, a 10-year-old or 11-year-old in grade 3 on average read 15.5 fewer words of connected text per minute than an 8-year-old; a 12-year-old in grade 3 on average read 19 fewer words per minute; and a 13-year-old on average read 27 fewer words per minute. ***Figure 7*** shows the difference in mean Filipino ORF scores for children in different age groups, as well as the percentage of the population in each age category.

Figure 7: Filipino ORF scores for children who were over-age in grade 3



*** Statistically significant to the p<.001 level; * statistically significant to the p<.05

Reading prosody

In the Philippines, for the first time, the EGRA study team used a measure of reading prosody in addition to comprehension. Because this is a new type of analysis, the results are undergoing some additional review and will be incorporated into a later version of this report. However, some basic (unweighted) values are provided in ***Table 10.*** For each of the four categories of prosody, the table shows the number and percentage of students in that category, and the mean ORF for children in that category. By definition, category 1 readers are slow readers. However, it is possible for a fast reader to read without any prosody, so this table only indicates the range of abilities and does not suggest causality.

Table 10: Filipino prosody – distribution of scores

Level	Definition	N	%	Mean ORF
1	**Reads primarily word-by-word.** Occasional two-word or three-word phrases may occur—but these are infrequent and/or they do not preserve meaningful syntax.	273	11%	31
2	Reads primarily in two-word phrases with some three- or four-word groupings. Some word-by-word reading may be present. **Word groupings may seem awkward and unrelated to larger context [i.e., punctuation] passage.**	659	27%	55
3	Reads primarily in three- or four-word phrase groups. Some small groupings may be present. However, the majority of phrasing seems appropriate and preserves the syntax [i.e., punctuation] of the author. **Little or no expressive interpretation is present.**	997	41%	80
4	Reads primarily in larger, meaningful phrase groups. Although some regressions, repetitions, and deviations from text may be present, these do not appear to detract from the overall structure of the story. Preservation of the author's syntax is consistent. **Some or most of the story is read with expressive interpretation.**	487	20%	94

Source: Adapted from Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., Oranje, A., & Goldstein, A. (2005). *Fourth-grade students reading aloud: NAEP 2002 special study of oral reading.* NCES 2006-469. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (NCES), National Assessment of Educational Progress (NAEP). Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006469

We expected that reading prosody would be related to reading comprehension, which is discussed next.

Reading comprehension

After the child had read as much of the Filipino short story as possible in one minute, the assessor noted the prosody score of the child and then proceeded to ask questions related to the story. The first four questions required basic recall of elements in the text, whereas the fifth question was somewhat more inferential and open to different types of answers. Children were asked questions only about parts of the text that they had read. Since most children read to the end of the text, they were asked all five questions.

Regardless of question, the share of children who answered incorrectly was approximately the same, ranging between 12% incorrect on Question 1 and 28% on Question 5. For all other analyses presented in the report, an "incorrect" response was one for which the child either responded incorrectly or just didn't say anything at all. Analysis of truly incorrect reponses compared to non-responses indicates that "no response" was actually a case of children not understanding the question or not knowing the response, and not a case of being inhibited from answering the question.

Students on average answered correctly 67% of the questions they attempted, or 64% of the total number of questions (five); the low difference in the two types of scores indicates that most children completed the reading passage and were asked all five questions. However, only 41% of the population were asked all five questions and answered all five questions correctly; 65% of children answered either four or five questions correctly. Regardless of the number of questions asked (which is, itself, a reflection of fluency and the total distance read in the text in one minute), only about 20% of children answered all questions correctly. There were two notable exceptions. First, among children who were asked only one question, most of them (74%) answered incorrectly. These children had an average ORF of only 5 words per minute correctly, so we can expect they were struggling to decode and probably did not grasp the meaning of the words. Second, of the children who were asked all five questions, a majority (53%) of them answered all five questions correctly.

The level of 75%–80% comprehension began at a fluency rate of about 30 cwpm, which is lower than the level typically found in other languages (40–60 cwpm). However, not all children reading 30 cwpm or more were able to answer the questions asked. In those cases, poor comprehension is not resulting from struggling to decode and fluently advance through text, but may be resulting from poorly developed reading comprehension strategies.

Factors associated with children's scores in Filipino reading comprehension were similar to those listed for other subtests, with the most significant differences associated with gender and region. Of the lowest-performing children (<40% comprehension), the majority were boys; did not have other books at home; and were from Visayas, Mindanao, or ARMM regions. Exit interview questions for this group also indicated that 16% of these children were in classes where Filipino or Tagalog was not spoken in the classroom. The reasons why schools are not consistently using Filipino in grade 3 classrooms should be explored in more detail.

Listening comprehension

The listening comprehension subtest was designed to determine to what extent low reading comprehension scores might be due to lack of oral language proficiency in the language of the subtest. The exercise involved the assessor reading a short story to the child out loud and then asking questions. As with reading comprehension, this subtest was selected out of three

possible subtests based on an analysis of pilot data. The story was shorter than the one used for reading comprehension and there were only three questions, but all children were asked all three questions. Therefore there were four possible scores for this exercise: 0%, 33%, 66%, and 100%. On this subtest, the average score was 66%—or two out of three questions; but there was also a high rate of zero scores—11%. A closer look at the data reveals only a weak correlation between reading comprehension and listening comprehension ($r = 0.43$). While some children who got no correct responses on listening comprehension still achieved high reading comprehension scores, on average, children who did better on listening comprehension did have higher average reading comprehension scores, as shown in ***Table 11.***

Table 11: Comparison of average reading and listening comprehension scores for Filipino

	Listening comprehension score (no. correct out of 3 questions)			
	0/3	1/3	2/3	3/3
Average reading comprehension score	41%	66%	75%	85%

Again, a review of the difference between "no response" and "incorrect" responses suggests that the zero scores are, in fact, children who did not understand the story or the question, as opposed to children who just did not participate in the assessment by staying silent.

Children who reported speaking either Tagalog or Filipino at home had slightly higher scores (76% and 77% respectively), but the same was true for children who reported speaking Ilokano at home, so this is hard to explain, but other co-factors may have been at work here that were independent of language. Children who reported speaking Cebuano, Bisaya, or Tausug also had much lower scores—38%, 59%, and 51% respectively.

Other factors correlated with higher or lower scores were consistent with those mentioned earlier—gender, region, age, socioeconomic status, teaching behaviors such as reading aloud, etc. These factors are not intuitively linked to listening comprehension, so a more likely explanation is that a group of students was struggling overall (possibily because of compounding disadvantages) with school, causing them to perform poorly on all aspects of the assessment.

Dictation

Will be completed for an updated version of this report.

Summary of Filipino reading performance

To summarize, Filipino reading abilities at the end of grade 3 were strong for many children, but 65% of children are not meeting the fluency and accuracy benchmarks. The extremes that affected the overall averages were the high-performing children in Metro Manila and low-performing children in ARMM and Mindanao. Socioeconomic status also was a factor—probably also linked to regional differences (i.e., urban vs. rural; children in Manila also had higher SES). While this survey was not meant to evaluate classroom practices directly, there were some indications that Filipino was not being consistently used in the classroom to the extent intended, and the regional variations may also be explained by a combination of

community, home, and school language usage that undermines Filipino competency. Other more in-depth studies that measure the actual language use in the classroom may be of interest to DepEd moving forward.[22]

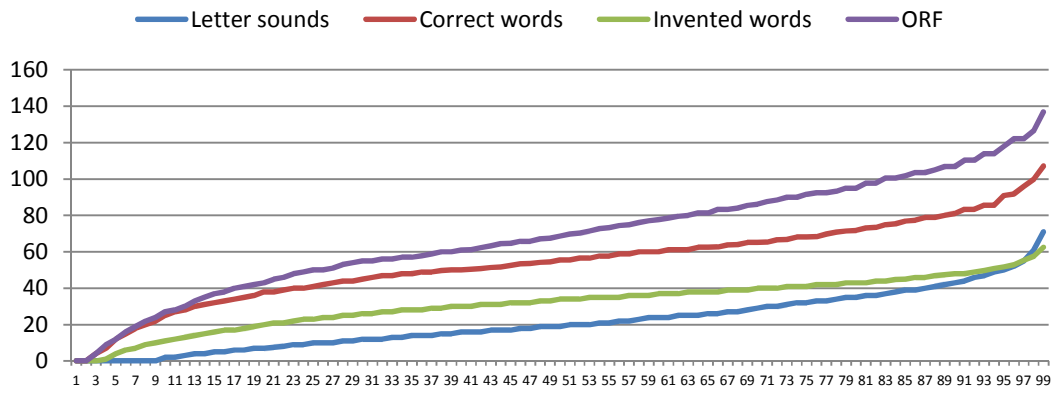
What this EGRA study does show is how prereading skills contribute to later oral reading fluency. All of the skills measured by EGRA are also teachable skills, so it will be important to ensure that children have a strong foundation in phonemic awareness and letter-sound correspondence well before grade 3—and that they are distinguishing them correctly in Filipino, English, and other local languages, all of which share the same alphabet but may have different letter-sound correspondences and points of articulation. However, for future grade 3 studies, it probably would be more informative to drop the phonemic awareness and letter-sound tasks and instead focus on higher-order skills such as comprehension and dictation. We were able to establish that children were decoding, and therefore had implicit letter-sound knowledge; and that low-performing children most likely had other barriers to learning other than lack of these basic skills. The key question remains whether or not the children's levels of reading fluency, as measured by this study, will be sufficient for understanding and learning independently from a variety of academic texts. A future EGRA might consider two reading and comprehension texts—one narrative and one expository—and a comprehension method based on silent reading.

The final graphic for this section, ***Figure 8,*** presents the percentile distribution of all of the fluency subtests, which may be useful for DepEd in reviewing benchmarks and standards, or eventually for teachers who conduct similar classroom-based measures and can see where their children fall relative to the larger national population. These *percentiles* should not be confused with *percentage* correct scores, nor do the units represent the same child's scores on the different subtests. The percentile distribution shows the number of correct items per minute associated with each percentile ranking, with examples of where the 50th percentile falls for each subtest. For example, the 50th percentile is defined by an ORF of 68 cwpm, 55 familiar words per minute, 34 invented words per minute, and 19 letter sounds per minute.

[22] See, for example, results from Kenya and Uganda of a study using classroom observations of language and instructional use: Piper, B. (2011). *Kenya Early Grade Reading Assessment findings report.* Prepared for the William and Flora Hewlett Foundation under the program Monitoring of Learning Outcomes in Sub-Saharan Africa. Research Triangle Park, North Carolina: RTI International and East African Development Consultants. Retrieved from https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=275

Piper, B. (2010). *Uganda Early Grade Reading Assessment findings report: Literacy acquisition and mother tongue.* Prepared for the William and Flora Hewlett Foundation under the program Monitoring of Learning Outcomes in Sub-Saharan Africa. Research Triangle Park, North Carolina: RTI International and Makerere University Institute for Social Research. Retrieved from https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=293

Figure 8: Filipino percentile distributions on all subtests



English Results

Children's English reading ability closely matched Filipino reading ability, with the notable exception that many fewer children were able to respond correctly to reading and listening comprehension questions in English. Results from the nonword reading and initial review of dictation subtests (details to follow in a supplement to this report) also reconfirmed that children were decoding and encoding using letter-sound correspondences. However, the simple transcription of the English phrase often did not correspond to the correct syntax or orthography, suggesting low comprehension/semantic knowledge and self-monitoring ability on the part of the children, stemming from a lack of exposure to English and poor oral language ability. These results are discussed in more detail in the sections that follow.

Table 12 summarizes the English results on all subtests.

Table 12: English results on all subtests disaggregated by gender

Subtest	Overall Mean	Female mean	95% CI – lower bound	95% CI – upper bound	% zero scores
Initial-sound segmentation (10 items - % correct)	55%	57%	51%	58%	17%
Letter-sound identification (100 items, correct per minute)	22	24	21	24	6%
Familiar word reading (50 items, correct per minute)	51	55	50	53	2%
Nonword reading (50 items, correct per minute)	26	29	25	28	6%
Oral reading fluency (59 items, correct per minute)	67	75	65	69	1%
Reading comprehension (5 questions – % correct)	32%	39%	30%	34%	37%
Listening comprehension (3 questions – % correct)	33%	37%	30%	36%	45%

Similar to the results of the Filipino language assessment, girls did better on all subtests, and this was statistically significant. There were also low percentages of zero scores on the prereading skills tests, with the exception of initial-sound segmentation; and very high zero

scores on reading and listening comprehension. Almost half of the students answered all three questions incorrectly (or provided no answer at all) on the listening comprehension subtest. It is important to recall that each child participated in both the English assessment and the Filipino assessment. The order in which the subtests were presented was alternated, so there should be no measurement bias caused by one subtest always being the first or last. The assessors were trained to be very careful about specifying which language each subtest was in. For example, the script for the assessors contained the read-aloud instruction "Here is a page full of words *in English,*" or "Please tell me the sound this letter makes *in English."*

Each of these subtests is reviewed in more detail below. As described in Section II Outcomes Summary, many factors were consistently associated with lower scores, including gender, region of origin, socioeconomic status, language background, and classroom characteristics. The English results were consistent with this pattern; therefore, it can be assumed that unless otherwise noted, the same patterns were present within each subtest. When the real scores varied dramatically from the mean, or when findings were otherwise unusual or notable—such as relationships to language background—they are pointed out below. However, absence of specific reference to the factors associated with performance on each subtest does not mean they were not present.

Finally, the purpose and procedure of each subtest were described in the Filipino Results section above, and are not repeated here. For further details about the EGRA test and administration procedures, see the USAID EdData II project website: www.eddataglobal.org.

Initial-sound segmentation

This English subtest followed the same procedure as in Filipino: The child was asked to repeat the first sound in a given word. The words of the subtest are provided in ***Table 13,*** with the frequency of incorrect answers.

Table 13: Items and percent of incorrect responses on English initial-sound segmentation (phonemic awareness) subtest

at	so	chalk	very	blue	for	man	clap	ear	go
50%	33%	57%	41%	51%	25%	31%	48%	27%	41%

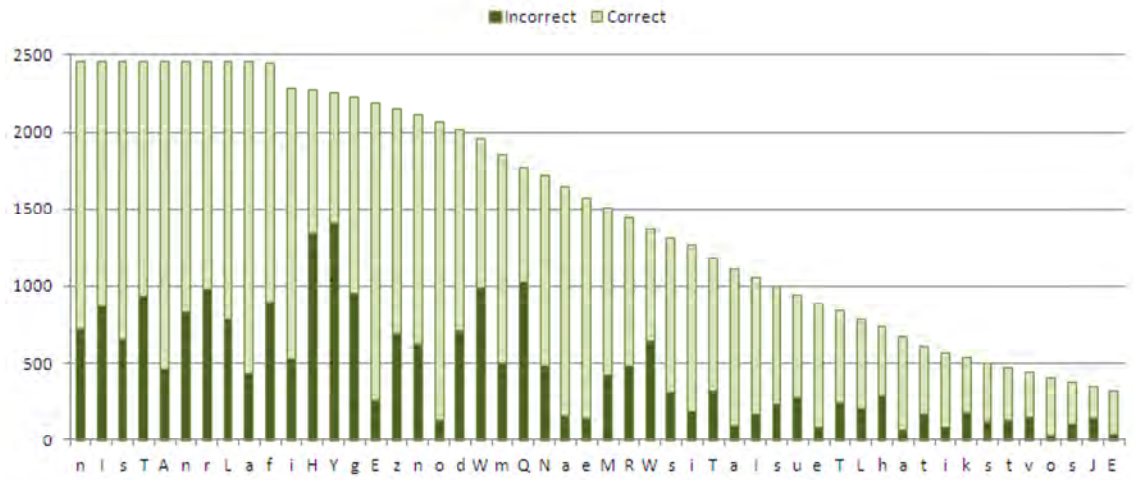
The average number of correct responses was 5.5 out of 10, compared to the Filipino mean of 6.5 out of 10—reflecting the higher percentage of zero scores. All children attempted the first five items, but per the testing protocol, children who answered all of the first five questions incorrectly discontinued the subtest. The reasons children had difficulty with this subtest may have been related to unfamiliarity with the concept of phonemes as distinct from their co-articulation with other vowels or blending with other consonants—for example, the first sound in "blue" is not /bl/ but /b/ and the first sound in "man" is not /ma/ but "m." Another reason may have been the inability to correctly reproduce the sound; although assessors were trained to be tolerant of regional accents, a child who answered that the first sound in the word "chalk" is /sh/ would have been judged incorrect even if the "ch" sound was one that was not present in the local language. Other known mother-tongue interferences were "v" being pronounced as "w" and "f" being pronounced as "v" and vice versa.

Girls did slightly better than boys on this subtest (4.5 percentage points); over-age children scored almost 1 correct answer less on average. Interestingly, children from ARMM and Mindanao—who scored lower on all of the Filipino subtests—did as well on this subtest as children from Metro Manila, and children in Visayas had 1.5 more correct items than counterparts in Manila. Although the actual number of Waray speakers was very low (N = 20), children who reported speaking Waray at home performed much better than the norm at 78% correct responses, as did children in Waray-speaking communities regardless of home language. Children in communities for which Tagalog was listed as the community language also scored much lower (46%). All of these findings most likely reflect local language interference more than instructional deficiencies, except that as in other areas, children who did not have homework assigned by their teacher also did worse on this subtest (43% correct responses).

Letter-sound identification

Performance on letter-sound identification in English was very similar to the results for Filipino: The average fluency rate was 22 correct items per minute, and the number of children who attempted each item, as depicted in ***Figure 9***, below, declined at a similar rate for both languages. Although the data were not analyzed at the student level on a comparative basis (i.e., does a given child miss both the Filipino letter and the English letter?), it would be possible, and the similarity in performance between the two languages suggests children may not have been distinguishing at all between letter sounds in English and Filipino.

Figure 9: Performance on the first 50 English letter sounds



Like Filipino, English vowel sounds had the lowest rate of error. H, Y W, and Q had some of the highest rates of error. These are all less common letters, but also upper-case letters; however, it is unlikely that this signals a problem with recognition of upper-case letters, since other instances like “I” and “N” did not cause significant difficulty.

Region of origin was not associated with differences in scores on this subtest, and other factors such as gender, age, and home/school characteristics had only minor effects (i.e., three correct letters per minute).

Familiar word reading

The familiar word reading subtest was constructed based on an analysis of grade 3 texts from different subject areas, as well as existing guidelines for English word frequencies.

The average rate of fluency was 51 cwpm, compared to 54 in Filipino. Girls read 54 cwpm on average, and children in Metro Manila 64. About two thirds (67%) of the children attempted 50 words in the minute given for the subtest, with an average of 43 words attempted), and the average accuracy (number of items correct out of number of items attempted, regardless of time) was 79%. The first 10 words from the English familiar word reading subtest are provided in ***Table 14,*** with the frequency of incorrect answers.

Table 14: English familiar word reading – items and percent of incorrect responses (first 10 words)

when	stop	help	boy	many	are	the	old	was	two
12%	11%	9%	5%	25%	4%	2%	6%	4%	10%

The word with the highest frequency of error was "could," with 69% of children providing an incorrect answer. Words missed by more than 25% of the children were "are," "please," "thank," and "every." As mentioned elsewhere, assessors were trained to accept certain regional or familiar accents in pronunciation, and to treat this as a reading assessment, not an oral pronunciation assessment. However, it can be hard to distinguish between a reading error and a pronunciation error, and therefore in case of doubt, answers were marked wrong. It would be necessary to check with a sample of assessors to understand why there were so many errors on a word such as "many," and if other errors were more often a case of struggling to decode, total silence, or common mispronunciations due to language interference (i.e., "thank" being pronounced as "tank"). Considering that the overall fluency and accuracy were high, it probably was not a case of poor decoding or word recognition, and a large proportion of the percentages shown in Table 14 above can be accounted for by the general population of children in difficulty because of region and socioeconomic status.

Consistent with findings in Filipino, children in all regions outside of Metro Manila scored below the national average in English, with ARMM being the lowest at 44 cwpm. Students in Tausug-speaking communities again had very low scores on this subtest—39 cwpm—but language spoken at home did not result in large significant differences. Over-age children also scored 44 cwpm on average, and children in multigrade classrooms, 42 cwpm.

The following familiar words also reappeared in the English oral reading fluency passage; the rate of error is in parentheses after the word: "are" (5%), "going" (15%), "the" (3%), "play" (3%), "where" (11%), "but" (14%), "two" (11%), "they" (3%).

Nonword reading

Nonword reading in languages with transparent orthographies usually measures decoding through direct letter-sound correspondences. English is not a transparent language, and this English subtest also measured knowledge of orthography rules such as the "silent e" that changes a vowel sound from short to long. Therefore it is not surprising that nonword fluency was lower and the number of zero scores higher in English than Filipino.

The average number of words *attempted* was 37, and more than half of the sample read 41 out of 50 words. A third (32%) of the children attempted all 50 words, while in 6% of the cases, the assessor discontinued the subtest because the child read no items correctly out of the first five words. The average rate of fluency (number of *correct items* read per minute) was 26—slightly more for females (29). Accuracy was 49% (number of correct items out of total attempted). Therefore it is interesting that although the mean fluency scores for English and Filipino were similar—26 and 31, respectively—in English, children were actually reading faster, but with more errors. The first 10 words from the nonword subtest are provided in ***Table 15,*** with the frequency of incorrect answers.

Table 15: English nonword reading – items and frequency of incorrect responses (first 10 words)

nale	oad	nat	murst	brish	zom	quith	niph	jaf	fing
26%	34%	16%	44%	37%	36%	66%	48%	17%	17%

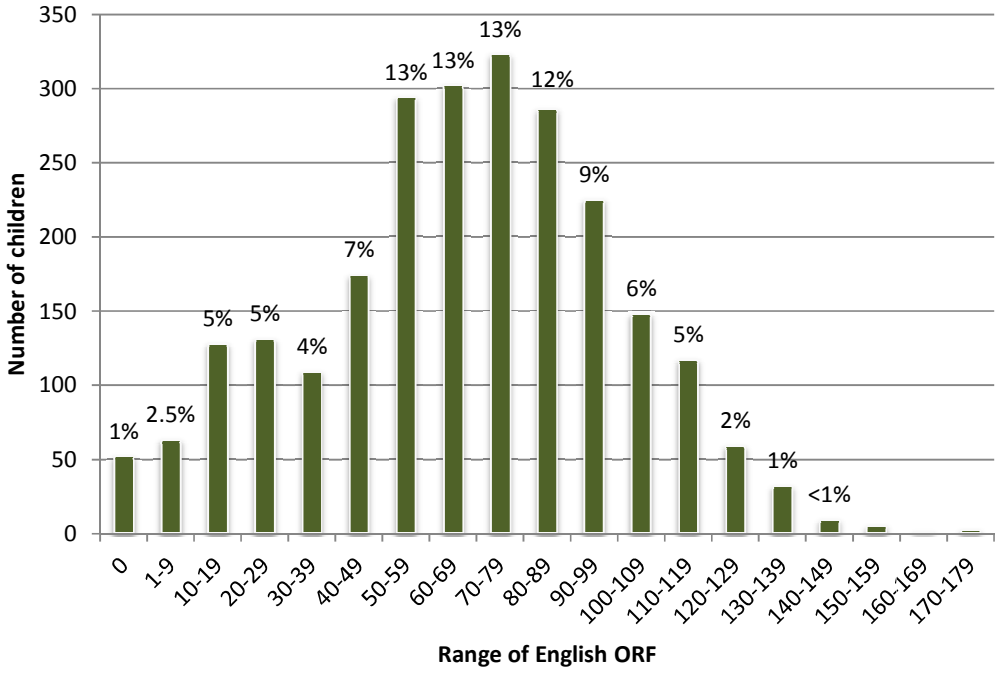
In addition to "quith," "grike," "onny," "fute," and "kise" had an error rate of more than 50%, but all of these latter words appeared after the 25th word in the subtest, when about 20% of children had already discontinued.

Regional differences were not large, but where Tausug was recorded as the language of the community, children scored much lower—15 cwpm on average. Note that only four children out of the entire sample of nearly 2,500 children reported speaking English at home. These children scored better on this subtest—39 cwpm.

Oral reading fluency

Filipino third graders were reading English at an average of 67 cwpm; girls scored higher at 75 cwpm. As was the case with Filipino reading, the English EGRA yielded a wide range of skills, from those who did not read a single word correctly, to a few who read at a rate above 150 cwpm. ***Figure 10*** shows the distribution of scores by range of English ORF. Slightly more than half of the students were situated around the mean, or between 50 and 90 cwpm, while the other half were split nearly evenly between those who scored more than 90 cwpm (23%) and those who scored less than 50 cwpm (26%). The percentages shown above each bar in the graph are the weighted percentages, while the number shown in the vertical axis is the number sampled.

Figure 10: English ORF – distribution of scores by range of fluency



The short story that the children were asked to read is copied in ***Figure 11.*** It consisted of 59 words total. The story was created, along with two others, during the instrument adaptation process involving English language specialists and DepEd staff. This passage was chosen from among the three that were pilot tested based on Rasch analysis. The words highlighted in yellow were missed by at least 20% of the children, and the words highlighted in blue were missed by more than 40% of the children.

Additionally, Figure 11 shows the required distance in the English text that the child must have reached in order to attempt the comprehension question corresponding to that text. The words underlined are those that were part of the response. Words in bold served as a reminder of frequent words that appeared in the familiar word reading subtest.

Figure 11: English ORF text

Dana and her sister **are** walking. They **are going** to **the** park [Q1] to **play** with friends. Suddenly it begins to rain. [Q2]

"**Where** is your umbrella?" asks Dana's sister. Dana opens her bag, [Q3] **but** she finds no umbrella.

"I'm sorry. [Q4] It's not in my bag." **The two** run back home so wet and sad. [Q5] Now **they** have to stay home.

Words that children missed frequently were often words ending in "s," which is a known difficulty in Philippine English speech—children and adults often do not pronounce the final "s." The extent to which this pronunciation changes the meaning for the speaker and has an effect on comprehension is uncertain. The words that were critical to understanding either the question or the response were not those missed most frequently, however; as the reading and listening comprehension sections show, comprehension in English is very low.

Reading prosody

Readers in English were also given a score of prosody based on how they signaled punctuation and meaning in the rate and inflection of reading. In particular, the direct-speech sentences were the opportunity for the child to show some degree of prosody, for example, by raising the voice to signal a question, changing the voice to represent a character of the story, or clearly marking a pause where there was a comma. ***Table 16*** summarizes the results.

Table 16: English prosody – distribution of scores

Level	Definition	*N*	%	Mean ORF
1	Reads primarily word-by-word.	409	17%	31
2	Word groupings may seem awkward and unrelated to larger context [i.e., punctuation].	852	35%	61
3	Little or no expressive interpretation is present.	814	34%	82
4	Some or most of the story is read with expressive interpretation.	339	14%	97

Note: Based on unweighted data.

Less than half of the children were in categories 3 and 4, where some expressive interpretation should be present, or at minimum, reflect the meaning of punctuation in the rate of reading. More analysis will be presented on this measure in a final version of the report, including the extent to which the prosody score may be related to comprehension.

Reading comprehension

Children were asked up to five questions related to the short story that they read (Figure 11 above), and the number of questions answered correctly was divided by five to get a score as a percentage of correct responses out of the total possible. The average score was 32%. Because children were reading, on average, at a rate of more than 60 cwpm, the majority were asked all five questions. Of those asked all five questions, however, only 13% answered them all correctly. This is equivalent to 3.5% of the entire sample when we factor in zero scores. This is in sharp contrast to the 41% of children who answered all five questions correctly in Filipino. 15% answered four or five questions correctly in English.

The low average reflects the high proportion of children who did not answer the questions correctly, either because they did not say anything at all or because they gave an incorrect response. Analysis of response types indicates that, compared to Filipino, a much larger proportion of children provided no answer to all of the questions they were asked—in many cases, even all five questions. This suggests that many students were reading with enough fluency to be asked all five questions, but were inhibited from answering for some reason, maybe because they did not understand the question that was asked. (Children who actually said "I don't know" were scored as "incorrect," not "no response.")

Question 5 ("Why were Dana and her sister sad?") was the hardest, which was expected because it was somewhat inferential, and several possible responses could have been correct. Question 3 ("What was Dana looking for in her bag?) was the easiest, with the largest share of children answering this question correctly.

Regardless of the number of questions a child attempted, the largest proportion of children answered none correctly, with the exception of children who were asked five questions. The largest proportion of children (27%) answered only one question, followed by children who answered two out of five questions correctly (21%). This is the opposite of what we expected to find, which is that as children acquire more fluency (and thus read to the end of the passage and are asked more questions), they are also able to understand them better and answer the questions correctly. This indicates a notable problem with either accuracy or oral language comprehension; in other words, children are decoding correctly or recognizing words by sight, but they are not registering the meaning of those words.

Either case is probably due to lack of sufficient exposure to English in a variety of contexts (spoken, written, heard), and possibly also a lack of experience with this type of activity. Unfortunately, the subtest does not allow us to know whether the problem was listening comprehension or reading comprehension; in other words, did the child understand the question (asked orally by the assessor) but not know the response—a reading comprehension problem; or did the child not understand the question and therefore could not respond—a listening comprehension problem?

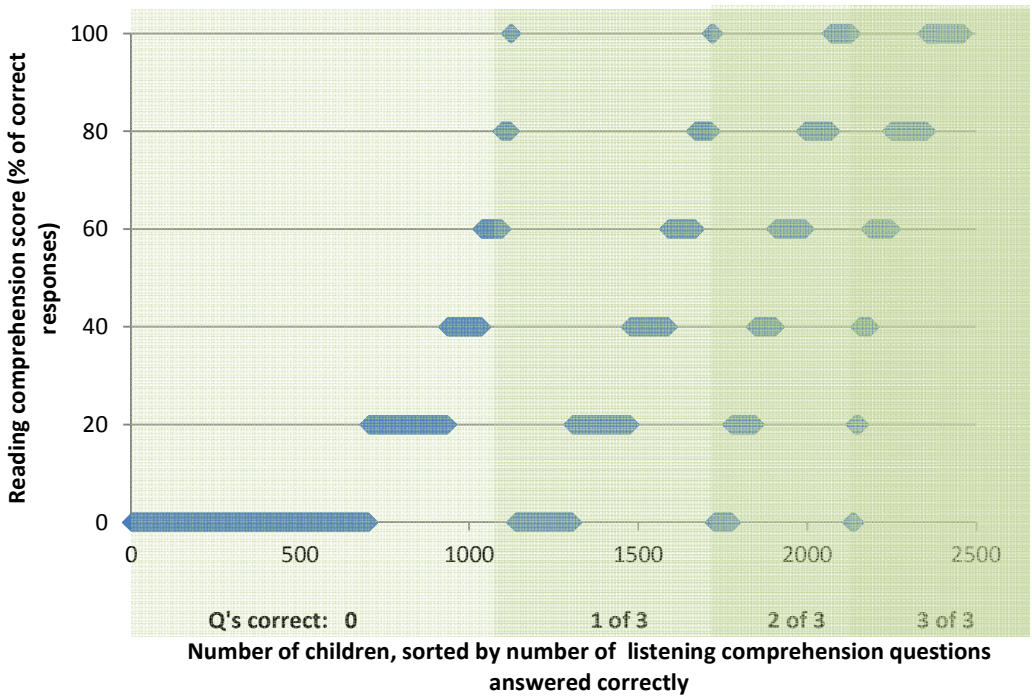
Consistent with other findings, English reading comprehension scores were lowest in ARMM and Mindanao (25% and 28%, respectively), and higher than average in Visayas and Metro Manila (38% and 40%, respectively), but even the highest average score for Metro Manila was equivalent to only three out of five questions answered correctly.

The next section describes the listening comprehension subtest.

Listening comprehension

This exercise involved a short story read aloud by an assessor, followed by three comprehension questions about the story. The story was developed by English-language experts from the Philippines and pilot tested prior to being selected as having the most appropriate level of difficulty for the range of abilities detected by the pilot sample. The average on this subtest was 33%, or one out of three questions. Almost half (45%) of children gave no correct response, resulting in zero scores; and only 13% answered all three questions correctly. Therefore we can confirm that a lack of oral language capacity also had an effect on the low reading comprehension scores, discussed above. ***Figure 12*** shows the relationship between English reading and listening comprehension. The graph plots the score of children according to their listening comprehension score and their reading comprehension score.

Figure 12: Relationship between English reading and listening comprehension



From this graphic we can see that children with zero scores on listening comprehension also, most often, had a zero score in reading comprehension. However, moving to the right on the horizontal axis, children who demonstrated listening comprehension ability in English (because they answered two or three out of three questions correctly) had a tendency to score higher on reading comprehension as well. The correlation between listening and reading comprehension was stronger for English ($r = 0.65$), and ***Table 17*** below indicates how large the difference was between the average reading comprehension scores of children with poor as opposed to good listening comprehension.

Table 17: Comparison of average reading and listening comprehension scores for English

	Listening comprehension score (no. correct out of 3 questions)			
	0/3	**1/3**	**2/3**	**3/3**
Average reading comprehension percent score	13%	30%	52%	74%

Similar to reading comprehension, children in ARMM and Mindanao scored lower on listening comprehension than the national average (25% and 27%, respectively) while Visayas scored higher than the average, with 44%, and higher than Metro Manila, at 36% correct responses. The real value of these differences is small, but reflects fewer children with zero scores. Children in Waray-speaking communities had an average of 48%, and children in Tausug-speaking communities only 9%.

Dictation

In general, the EGRA dictation subtest measures a combination of skills. The child must listen to a short sentence read by the assessor, and then write it on a piece of paper. "Reverse"

decoding skills are required to transcribe the written sounds on paper, but listening comprehension is necessary in order to choose the right word forms based on the meaning of the entire sentence. Automatic word recall is also required for common words like "the," "to," "we," and "school," and the ability of the child to write all 13 words in the sentence reflects a type of writing fluency, since children are given a limited amount of time in which to write the sentence.

As with Filipino, the results of this subtest will be described in more detail in an addendum to this report.

Summary of English reading performance

To summarize, English reading performance was nearly equivalent to Filipino reading performance in terms of fluency, decoding, and accuracy of reading words in isolation. However, children appear to have been only decoding or recalling words from memory, but without registering their meaning. Similar to Filipino, there were important inequalities among children, with some able to perform well above benchmarks, but others unable to demonstrate even pre-reading skills by the end of third grade. The factors of socioeconomic status, gender, region, and linguistic background all were consistently correlated with reading achievement, as were certain school-level characteristics.

This assessment did not seek detailed information on teacher qualifications or teaching strategies beyond a few self-reported questions from the children about whether teachers engaged them in reading and speaking activities. It is important to remember that the correlations reported here are not exhaustive, and there may be other school or teacher-level factors with even more influence on reading outcomes than SES and language background. According to a 2013 study by the National Educational Testing and Research Center (NETRC) [23], teachers' English proficiency was rated low to moderate on measures of "structure," "written expression," and "reading comprehension." Only in Visayas, Ilocos, NCR, and Cordillera Administrative Region (CAR) was teachers' mean performance on the Test of English Proficiency for Teachers (TEPT) rated moderate; these are the regions where children in this study performed better than others. Visayas and Caraga regions also surpassed the 75% minimum criterion in the NAT English reading comprehension test. As a result of the data from the NAT and TEPT, and as reconfirmed here, DepEd recognizes the need to strengthen teachers' abilities to use English proficiently and to teach using the language.

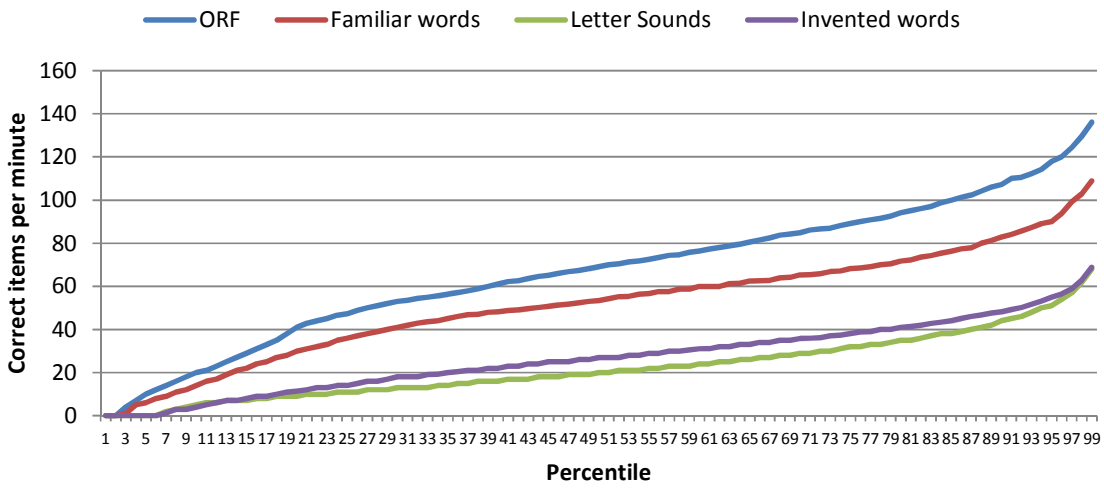
Based on the skills measured in this English EGRA, and the level of the actual items (frequent words, simple narrative story, simple short sentence to write), it is unlikely that these children will be prepared for independently learning higher-level academic content in English without the continued scaffolding appropriate for second-language learners (or third-language, in this case). Research by Thomas and Collier in the United States with English-language learners found that students who had completed at least two to three years of schooling in their L1 required five to seven years to reach the 50th percentile in English as L2 when taught in L2; students who completed no schooling in L1 required at least seven to ten years, and students taught bilingually in L1 and L2 usually needed 4 to 7 years to catch up

[23] Data from the national studies reference in this paragraph come from a presentation, "Multi-Literacy and Numeracy Program" (DRAFT). Personal Correspondence. April, 2013.

with native English-speaking peers.[24] Considering children in the Philippines have been immersed in academic English for only 3 years, and few use English outside the home, it is commendable that so many are actually meeting the fluency benchmarks, but if not accompanied by comprehension, then these gains are of little use.

Figure 13 shows the percentile distributions for English measures of fluency. It tells us, for example, that the 50th percentile is defined by an ORF of 69 cwpm, 53 familiar words per minute, 27 nonwords per minute, and 20 letter sounds per minute. Recall from the Filipino results section that these *percentiles* should not be confused with *percentage* correct scores, nor do the units represent the same child's scores on the different subtests. The percentile distribution shows the number of correct items per minute associated with each percentile ranking, with examples of where the 50th percentile falls for each subtest.

Figure 13: English percentile distributions on all subtests



Ilokano Results

The Ilokano assessment differed from the English and Filipino in several key areas:

- The sampling was purposeful, starting from selection of a subset of schools that had reported to regional authorities that they started implementing Ilokano as the mother tongue in the 2012 or 2011 school year. Within that sample, school and student selection was random.
- Grade 1 children were assessed rather than grade 3.
- The initial sound segmentation (phonemic awareness) subtest was a different format, and there were no dictation or prosody measurements.

Otherwise, the methodology was the same (see Section I, About the EGRA Tool, for more details) and the sample of 500 children in 50 schools was large enough to draw inferences at

[24] Please see the following: (1) Thomas, W. P., & Collier, V. P. (1997). *School effectiveness for language minority students.* Washington, DC: National Clearinghouse for English Language Acquisition. Retrieved from www.ncela.gwu.edu/ncbepubs/resource/effectiveness/index.html. (2) Thomas, W. P., & Collier, V. P. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement.* Santa Cruz, California, and Washington, DC: Center for Research on Education, Diversity & Excellence. Retrieved from www.crede.ucsc.edu/research/llaa/l.l_final.html

the regional level for schools that had been implementing Ilokano as language of instruction and teaching reading in Ilokano.

The results show that average reading ability was still far from the established benchmark, and there was a wide range of ability—from zero to +60 cwpm. Nearly half the children were reading less than 10 cwpm on the oral reading passage, and the average was 18. As in the E/F results described in the previous sections, girls achieved higher average scores than boys across all subtests, and girls outnumbered boys in the highest percentiles, whereas boys outnumbered girls in the lowest percentiles. ***Table 18*** summarizes the Ilokano results across subtests, and the sections that follow provide more details on each subtest.

Table 18: Ilokano results on all subtests disaggregated by gender

Subtest	Overall Mean	Female mean	95% CI – lower bound	95% CI – upper bound	% zero scores
Differentiating initial sounds (10 items – % correct)	41.9%	41.8%	38.9%	44.9%	7.8%
Letter-sound identification (100 items, correct per minute)	14.5	16.7	12.3	16.8	9.2%
Familiar word reading (50 items, correct per minute)	21.6	25.7	18.4	24.7	20.2%
Nonword reading (50 items, correct per minute)	18.3	21.7	15.2	21.4	32.0%
Oral reading fluency (37 items, correct per minute)	18.2	21.9	15.47	20.91	17.0%
Reading comprehension (5 questions – % correct)	22.9%	27.2%	18.7%	27.1%	45.5%
Listening comprehension (3 questions – % correct)	45.4%	45.9%	41.5%	49.4%	17.8%

A high proportion of children scored "zero" on tests of higher-order reading skills, such as decoding invented words, reading a short story, and comprehending what they read. The following sections look in detail at the items and factors that may have been associated with performance on each subtest, as well as relationships between subtests.

Differentiating initial sounds (phonemic awareness)

Recall that grade 3 children in E/F were asked to isolate the first sound they heard in a given word. Grade 1 children were given a different version of the subtest to measure phonemic awareness skills; they were asked to repeat the word that began with a different sound. That is, the assessor would say three common words: "Tasa, mita, mata—which word begins with a different sound?" The correct response would be "tasa." There were 10 total questions. A child who did not answer any of the first five questions correctly received a zero score and this subtest was discontinued. This was the case for fewer than 8% of the children.

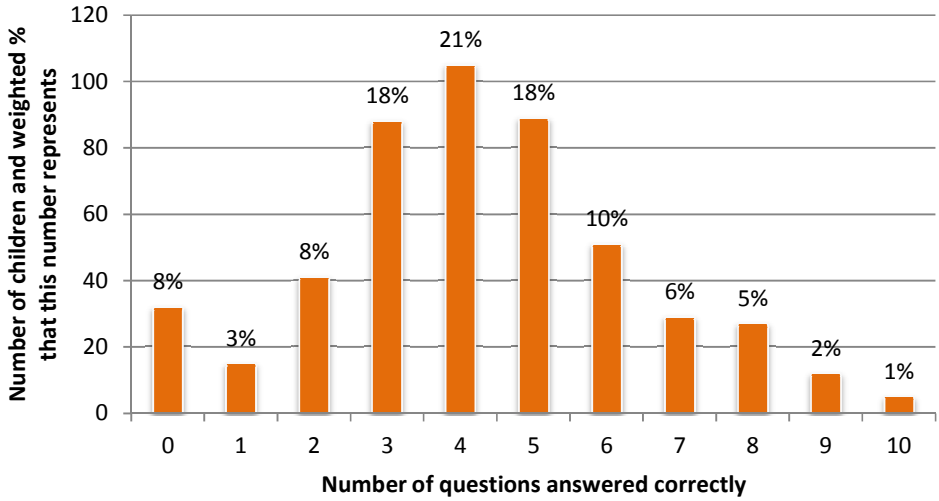
Table 19 provides an overview of the initial sound combinations for each of the 10 questions, and the frequency of incorrect responses for each question. For example, the first question included the words "baka," "bulsa," "kama."

Table 19: Items and frequency of incorrect responses on Ilokano differentiating initial sounds (phonemic awareness) subtest

b-b-k	m-a-m	p-l-p	d-t-t	t-t-n	s-s-d	b-p-p	g-s-s	k-k-w	l-t-l
48%	46%	68%	80%	34%	32%	85%	60%	35%	69%

Students demonstrated a wide range of ability on this subtest, with some items proving much harder than others. It is important to note that on a test with this format, it would be possible to get a score of 30% (3 of 10) just by, for example, always repeating the last word that the assessor said without considering the beginning sound at all. Therefore the mean of 45% was only slightly more than random guesswork. ***Figure 14*** shows the distribution of scores from 0 to 10, and the percentages of the total sample that each category represents. More than half of the children were in the range of three to five correct answers, with only about 10% of the children showing strong ability to differentiate between beginning sounds.

Figure 14: Distribution of Ilokano scores on differentiating initial sounds (phonemic awareness) subtest



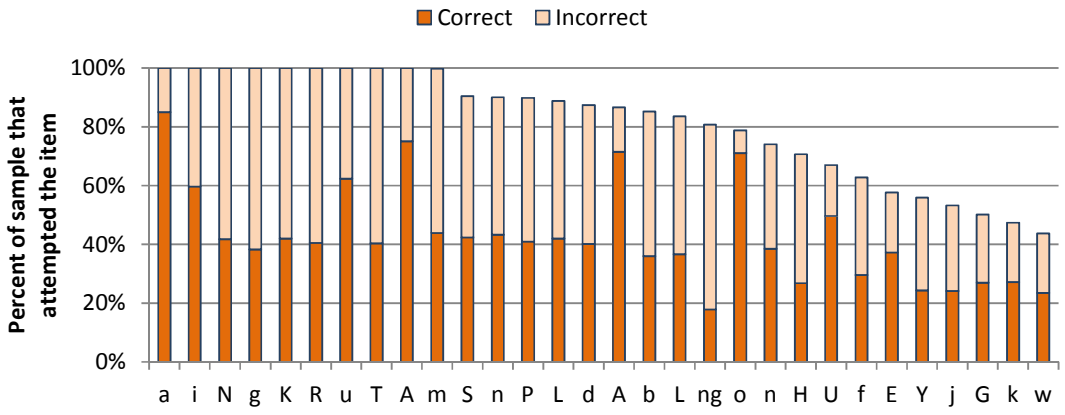
The next subtest, letter-sound identification, provided more information about the extent to which children were mastering letter sounds.

Letter-sound identification

According to the Ilokano K–12 competencies, children in grade 1 should be able to "Identify all of the letters of the alphabet, in both upper and lower case," and "Give the sounds of all letters of the alphabet." As described in earlier sections, this subtest measured this skill by asking children to give the sound of the letters presented in random order on a page, containing both upper- and lowercase letters.

The mean at the end of grade 1 was 14.5 correct letter sounds per minute, with girls slightly higher (16.7). On average, children attempted 29 letters, which means it took about two seconds for the child to provide the letter sound, but there was a high rate of error—about 46%. The first 30 letters on the page provided to the child included all of the letters of the alphabet, with the exception of c, q, v, and x, so according to the goals of the MTB-MLE curriculum, children should have known these by the end of grade 1. ***Figure 15*** shows the first 30 letters, and the proportion of correct and incorrect responses.

Figure 15: Performance on the first 30 Ilokano letter sounds

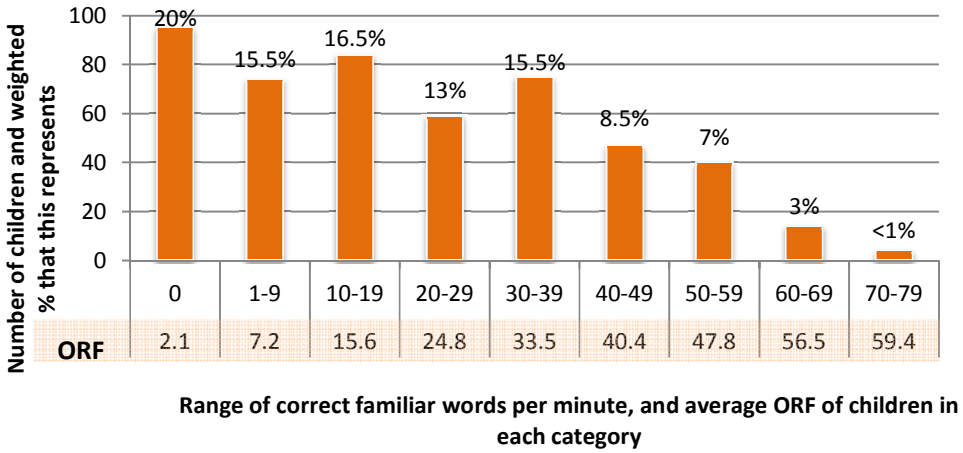


According to this item-level analysis, the majority of children succeeded at identifying only vowel sounds, but not consonants. This may be because of an instructional focus on vowel sounds and not consonants, or because children actually were giving the letter names and not the letter sounds (the names of vowels often overlap with the sound the letter makes, which is not the case with consonants).

Familiar word reading

The list of familiar words included in the Ilokano assessment were developed by language and teaching specialists from DepEd, based on frequent words in Ilokano reading materials used by teachers and produced by DepEd. The K–12 competencies established for mother-tongue instruction specify that at the end of grade 1, children should be able to "Read with automaticity 100 first grade high-frequency/sight words." Although "with automaticity" is not specifically defined, a good estimate is that it should take no more than one second to read a word and transition to the next; this would be a fluency of 60 words per minute with no errors, or 40–60 words with some errors. The average number of correct familiar words read per minute in this study was 21, but 20% of children couldn't read one word correctly out of the first five, resulting in a zero score. If we exclude the zero scores from analysis and consider only the children who did complete the subtest and had some measurable skills, the average increases to 27 correct words per minute. ***Figure 16*** shows the distribution of scores by category of correct response, with weighted percentages represented for each category.

Figure 16: Ilokano familiar words – distribution by range of correct words per minute



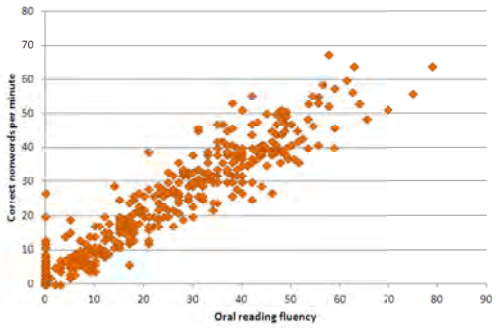
This chart also shows that as children were successful reading familiar words in isolation, they also were more likely to succeed in reading connected-text passages, as measured by the Ilokano ORF subtest (discussed below). The subtest revealed a broad range of skills, with a minority of children reading well above average, or as high as 77 words per minute. ***Table 20*** shows the first 10 of these familiar words and the rate of error for each word.

Table 20: Ilokano familiar word reading – items and frequency of incorrect responses (first 10 words)

idi	to	kayo	ket	da	ania	saka	no	isu	ulo
38%	22%	33%	42%	29%	61%	12%	10%	21%	16%

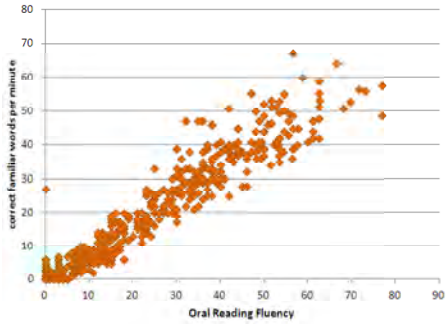
Of all of the words in the grid of 50 items, the ones missed by more than 30% who attempted to read it were "idi," "kayo," "ket," "ania," and "ngem." Familiar word fluency also was closely associated with the nonword reading and oral reading fluency results, but less so for letter-sound identification. This indicates that children may have been learning to read through syllables and whole words, but not learning to isolate the sounds of letters in words. ***Figure 17, Figure 18, Figure 19,*** and ***Figure 20*** below show the relationships among these different subtests.

Figure 17: Correct nonwords and Ilokano ORF



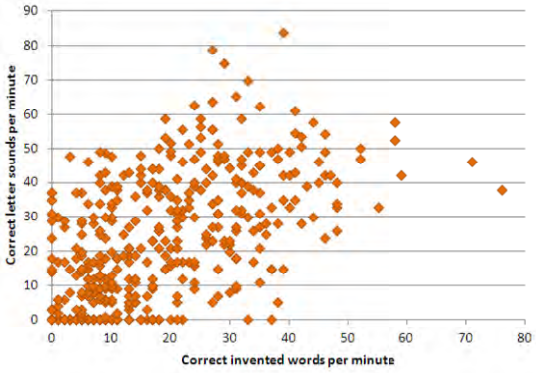
r^2=0.92

Figure 18: Correct familiar words and Ilokano ORF



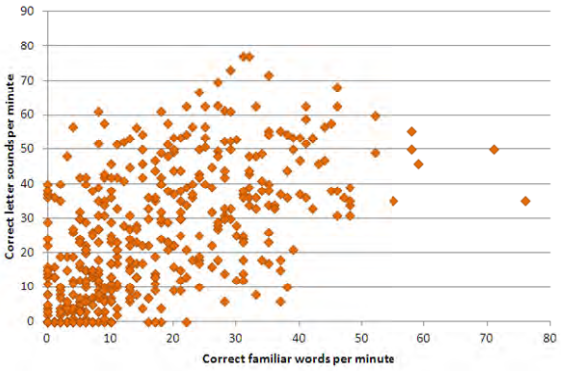
r^2=0.93

Figure 19: Ilokano letter sounds and nonwords



r^2=0.40

Figure 20: Ilokano letter sounds and familiar words

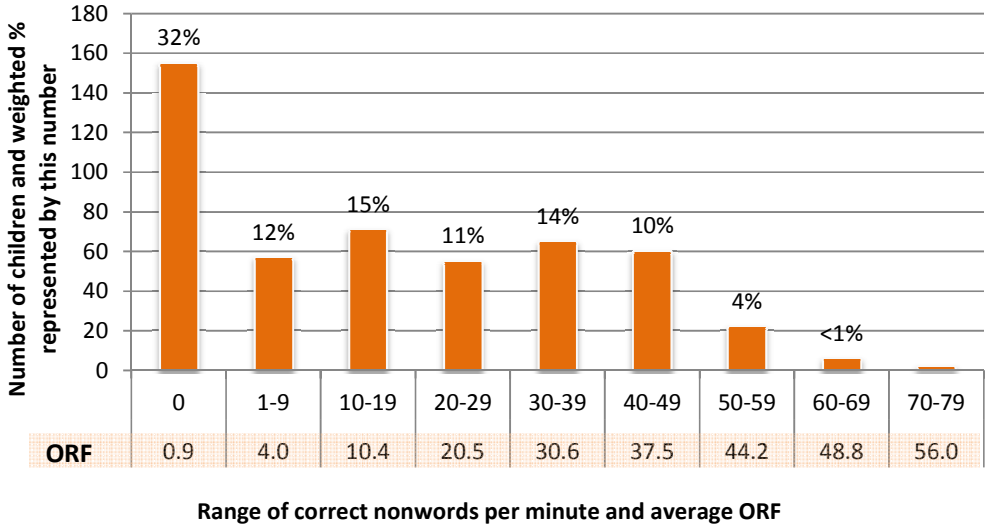


r^2=0.38

Nonword reading

Grade 1 children read, on average, 18 correct nonwords per minute, which was only slightly lower than for familiar words; girls read 22 correct nonwords per minute. There was also a very high rate of zero scores on this subtest, and the mean accuracy score (correct items out of items attempted) was 35%. ***Figure 21*** shows the distribution of scores across ranges of correct words per minute, including zero scores. Additionally, it shows the mean ORF for children in each range. For example, children who correctly read between 1 and 9 nonwords per minute had a score of 4 cwpm on the oral reading fluency measure to be discussed later. Once again, we find a range of scores, from little to no fluency at all, to children who read more than 40 cwpm. The reasons for the range of scores are explored in more detail in the next section—oral reading fluency.

Figure 21: Ilokano nonword reading – distribution of scores by range of correct words per minute



Oral reading fluency

The Ilokano reading passage followed the same format as the E/F ORF subtest, but it was shorter, with only 47 words. It is important to recognize that Ilokano is an agglutinative language, meaning words are formed by combining morphemes (small units of meaning) to change a word's meaning to signify tense, possession, plurality, etc. For this reason, many Ilokano words are very long. The average word length of this story was five letters, but some words were as long as 12 letters. ***Figure 22*** shows the reading passage that was used. Words in bold also appeared in the familiar word reading subtest. Words highlighted in yellow (e.g., "bimmangonda," "klase") were missed by more than half of those who attempted to read them, and words highlighted in blue were missed by 30% to 50% of those who attempted them. All children attempted the first eight words; the subtest was discontinued for any child who did not provide a correct answer among the first eight words.

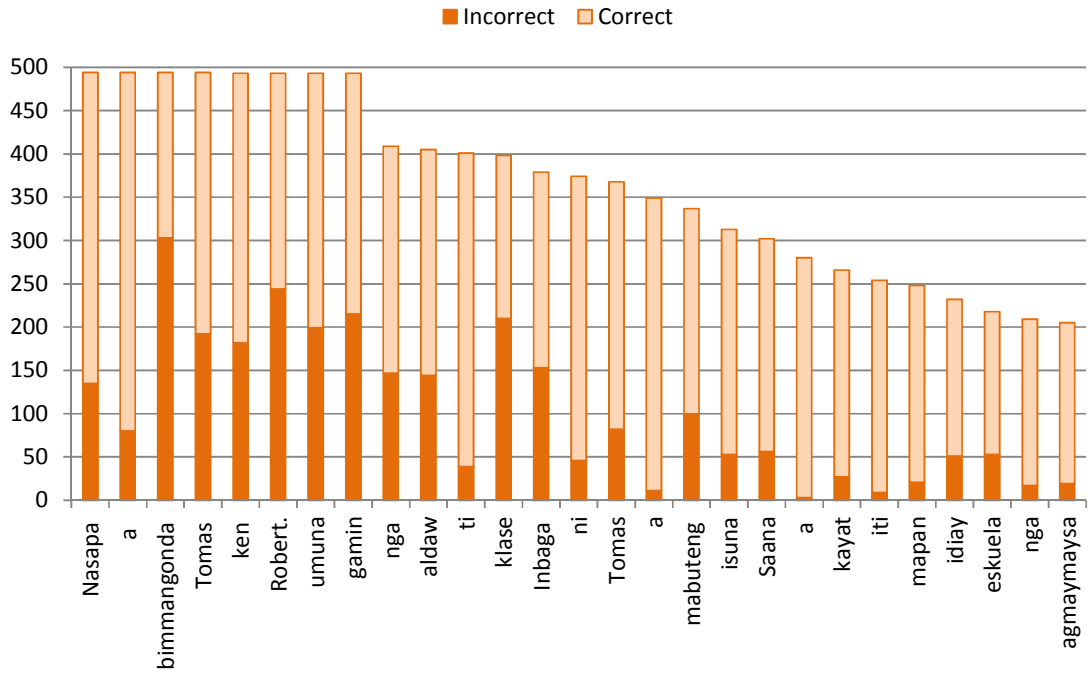
The average number of correct words per minute on this subtest was 18. The average number of words attempted was 24.5, while the average accuracy rate was 40%.

Figure 22: Ilokano ORF text

Nasapa a bimmangonda Tomas **ken** Robert. [Q1] Umuna gamin nga aldaw ti klase. [Q2] Inbaga **ni** Tomas a mabuteng **isuna**. Saanna a kayat **iti** mapan **idiay** eskuela nga agmaymaysa. [Q3] Intulod **ni** Robert ti adingna idiay eskuela. Nakita **ni** Tomas dagiti gagayyemna. [Q4] Kabigatanna, siraragsaken **ni** Tomas a napan idiay eskuela. [Q5]

Figure 23 further shows the rates of error on the first 30 words. While most children were successful reading short, one- or two-syllable words like "a," "ti," "iti," and "ni," words with multiple syllables proved much more difficult, and the words at the beginning of the passage had a high rate of error. This naturally affected comprehension, which we will see in the next section.

Figure 23: Grade 1 performance on Ilokano ORF subtest, first 30 words



The factors associated with low reading scores were similar to those found in the E/F data set, although this sample was not structured to capture differences based on region or subregion.

- Girls performed better on all subtests except differentiating initial sounds (phonemic awareness) and listening comprehension, and the difference was around 7 words per minute faster in ORF, familiar words, and nonwords.
- There were not as many over- or under-age children in the Ilokano classrooms, but where 5- or 9-year olds were present in grade 1, their scores also tended to be lower.
- Children in multigrade classrooms averaged only 9.5 correct words per minute.
- Children who missed two or three days of school the week prior to the assessment read 15 words per minute, compared to the average of 18.
- Children from wealthier homes, as measured by the wealth index (see note to Table 5 for explanation) or by the fact that the father was a professional or overseas foreign worker and not an informal laborer or unemployed, also achieved higher ORF scores.

There were also some important findings regarding the language environment of the children in the sample. Children who came from a Pangasinan-speaking household scored less than 8 words per minute. Also, children in classrooms where the teacher *never* read Ilokano stories aloud, nor asked the children to read Ilokano stories aloud, scored under 12 words per minute, compared to children in classrooms where the teacher *often* asked the children to read out loud—where the average was 20 correct words per minute. Fortunately, less than 10% of children reported "never" in these categories, but it is important to note that some classrooms were not providing adequate support for early reading.

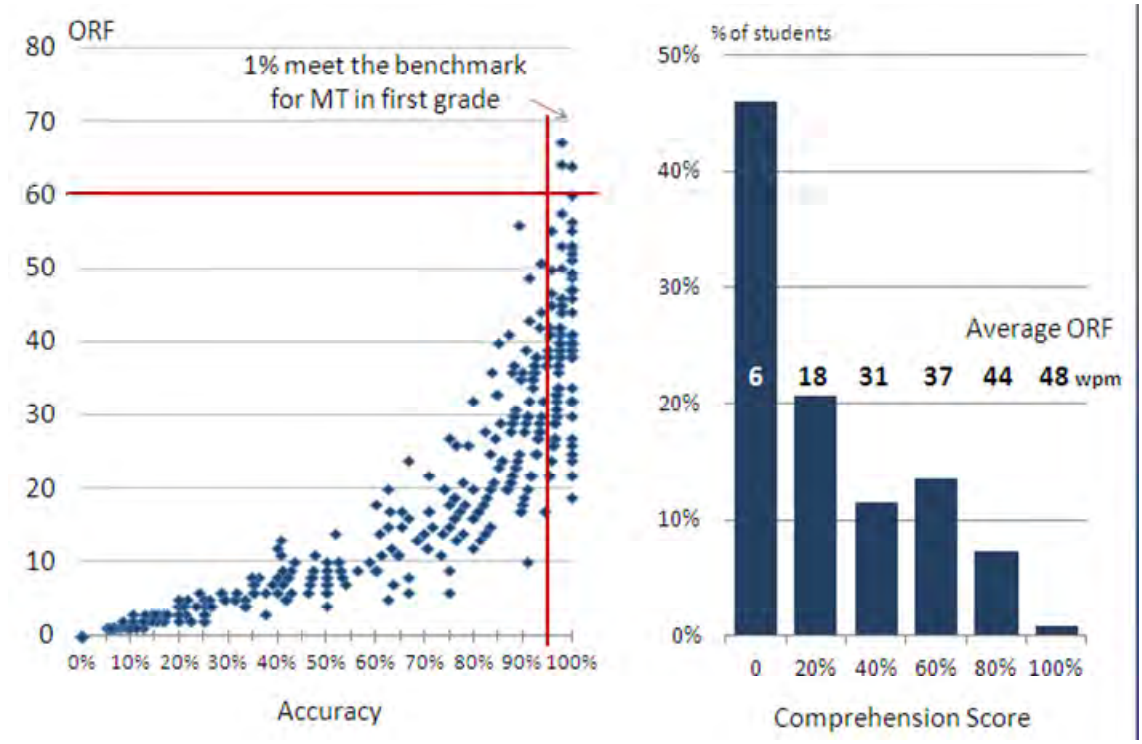
Additionally, as shown in Figure 17 through Figure 20 (scatterplots of relationships among the Ilokano subtests), prereading skills such as automatic word recognition and letter-sound correspondence were strongly correlated with reading fluency on a passage of connected text.

Although this relationship was not quite as linear for letter sounds as it was for familiar words and nonwords, children who identified fewer than 20 correct letters per minute had an average ORF of 12 cwpm. However, children reading more than 20 correct letters per minute averaged 35 on Ilokano ORF. Similarly, children who did well on letter sounds and familiar words also did better on reading comprehension.

Reading comprehension

Given the low reading fluency and high rate of error described in the above section, it is not surprising that Ilokano comprehension levels were low, even if most children were reading in their native language. If children struggle to process each word in a sentence, they will not have sufficient cognitive resources for making meaning out of the text. The average reading comprehension score was 23%, reflecting the low average reading fluency (18 cwpm). Only four children in the entire sample (<1%) were asked all five questions and answered them all correctly. As shown in the ***Figure 24*** graph labeled "Comprehension Score," the ability of children to answer correctly all questions that they attempted increased as their ORF increased. This pattern was more evident in Ilokano than in either English or Filipino (see tables in Annex 3, particularly the column representing children who answered 100% of questions correctly, according to how many they answered).

Figure 24: Ilokano reading and comprehension data



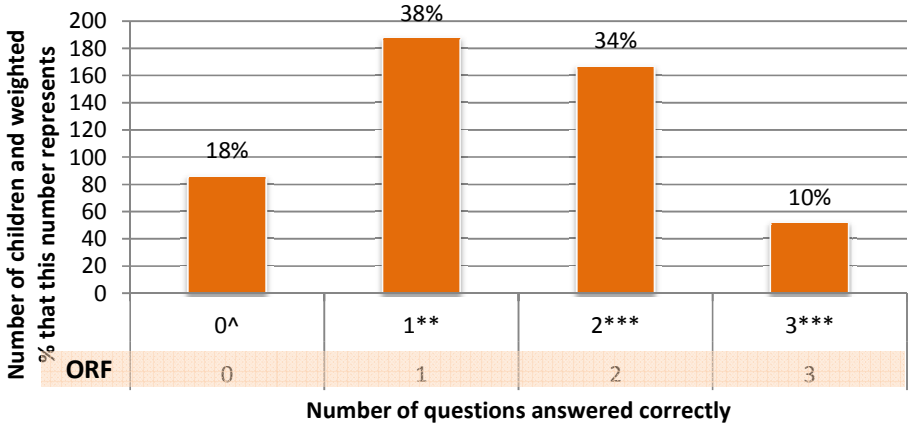
The children who answered either four or five questions correctly read on average 44 correct words per minute. Although speed is important, so is accuracy; therefore increased levels of fluency are necessary, but not sufficient, for comprehension. To clarify this point, 57% of children who read above 40 words per minute could not answer at least 4 out of 5 questions correctly. This may have been due to accuracy issues—i.e., they didn't understand one of the key words in the story or in the question asked by the assessor—or they may have had low

oral language ability in Ilokano. As mentioned above, some children in this sample said they did not speak Ilokano at home even though they had been placed in an Ilokano school. Or they may have spoken Ilokano at home but the specific background knowledge required for the story comprehension was not part of their habitual context. Additionally, all of the children were very young (average age 6.6 years), and may have been inhibited for a variety of reasons from answering questions, especially if this is something that is rarely asked of them in class. In some cases (Questions 1, 2, 5) the instances of "no answer" were greater than the instances of "incorrect" answers; however, there were also as many or more children responding with correct answers. Therefore it is possible that some children in grade 1 were acquiring and demonstrating reading and comprehension skills in Ilokano, but many other children also were being left behind, for a combination of the reasons outlined above.

Listening comprehension

The listening comprehension subtest is a way of confirming whether or not children are having difficulty with reading comprehension because of issues with oral language capacity. The short story read to the children consisted of 38 words, including a dialogue between two children. Three direct-recall questions were asked based on the story, so there were four possible scores a child could get: 0%, 33%, 67%, or 100%. The average on the Ilokano subtest was 45%, and there was no significant difference between boys and girls. The distribution of scores is presented in ***Figure 25,*** with the weighted percentages for each category.

Figure 25: Ilokano listening comprehension distribution of scores



Again, many children did not answer any questions correctly, resulting in a zero score. Although not presented in graphic format in this section, fewer children gave no answer than gave an incorrect answer for questions 1 and 2, but more children gave no answer for question 3. Therefore there is no reason to believe that incorrect answers were due to an unwillingness to respond rather than inability to do so.

Figure 25 also shows that Ilokano listening comprehension scores, like many other subtests, had a linear relationship to ORF—children who answered no questions correctly read only 11 cwpm on average while children who answered three listening comprehension questions correctly read 2 cwpm on average. This suggests several possible explanations:

- A category of children was in difficulty academically because of a combination of factors—SES, absences, classroom practices, etc.—and this affected them across all subtests of this assessment and probably across other subject areas, too.
- Some children may have had low oral language skills in Ilokano either because they did not speak Ilokano at home, or because of subregional differences in Ilokano pronunciation and vocabulary. There may also have been different styles of language use in the home versus in schools and lack of background knowledge and vocabulary required for answering listening comprehension questions.
- The Ilokano MTB-MLE program is not being implemented with the same degree of intensity and methods across schools, so some of these children may have been getting more academic exposure to the language and possibly more opportunities to engage in active listening and recall in the classroom.
- Implementation of the MTB-MLE program has been inadequately carried out in terms of teacher training, materials, and ongoing support to teachers. As a result, some teachers in these classrooms may be struggling to provide reading instruction in Ilokano or have given up trying and reverted to the old Filipino/English bilingual approach.

Summary of Ilokano findings

The findings indicate that on average, grade 1 children are not attaining the benchmarks or curricular goals outlined for the MTB-MLE program. Even basic skills such as letter-sound knowledge of all the letters of the alphabet have not been mastered by all children equally. The average scores mask large differences in ability, from children who demonstrated no reading skills at all, to children who could read up to 70 correct words per minute. The points listed above, explaining the low listening comprehension results, also provide a reasonable explanation of overall findings from the grade 1 study; individual factors, language factors, and implementation factors could have been responsible for these differences.

Individual factors. The schools in the sample demonstrated a range of achievement, as opposed to all the high achievers being in a few schools and all of the low achievers in a few schools. This indicates that achievement was dependent on factors not related to school, as there was as much of a range in scores within schools as between schools. Some of these non-school-related factors have been mentioned throughout this report—gender, SES, and attendance being the most prominent. For grade 1, self-reported absences were high, with 38% of children having been absent two or three days the week prior to data collection. Among children who scored above 30 cwpm, the percentage reporting being absent the previous week was much lower. While SES data are still being analyzed, initial analysis indicated that children from wealthier families were reading better than children from less wealthy families.

Language factors. Children who said they spoke Pangasinan at home, and heard Pangasinan spoken at home by others, were more likely to be in the low-achieving group. Our study did not investigate why these children were in schools teaching Ilokano, but these children and others speaking other indigenous languages at home may actually have had Ilokano as L2 and therefore had not had enough exposure to oral or written Ilokano to be able to succeed at these prereading tasks. Even within the Ilokano-speaking community in Region 1, variations in pronunciation and vocabulary could have given some children an advantage on this

assessment. More in-depth studies need to be carried out by language experts to understand to what extent this may be a problem in implementing standardized tests of Ilokano across the region. Additionally, how language is taught and how reading is taught should be reviewed to determine, for example, whether phonemic awareness is being taught in grade 1, including learning letter sounds, manipulating sounds in words, exploring rhymes, etc. These are all activities that can be done orally to support beginning reading as well as oral language development for Ilokano- and non-Ilokano-speaking children.

Implementation factors. The MTB-MLE program is new, and has made important progress in developing and distributing materials and teacher training. However, there may be challenges to implementation that this study was not designed to uncover that would explain why some children reached the targets and others did not. Schools have used and are using the MTB-MLE reading materials, yet 5% of children said they had never used the book in class. Many more were not using the MTB-MLE math materials (23%). More importantly, about 30% of students in the sample said their teacher mostly spoke a language other than Ilokano. The sample schools were drawn from those who reported at division level that they were implementing Ilokano, and this was again confirmed with them the day of the assessment. It appears that there may be wide variations among schools in the thoroughness with which they are implementing Ilokano instruction. Pioneering schools that had been in the program were not doing significantly better than nonpioneering schools; the mean scores were higher by two to four items on the fluency subtests for the pioneering schools, but the sample size did not allow us to detect statistical significance in this gain. So even in those schools where we are certain that the initiative and some training to implement MTB-MLE were in place, the average scores were still well below the benchmarks. This study only provided a few clues about what was actually happening in the classroom, and these were mainly reported by the students themselves, so it would be important to conduct follow-up activities to determine, for example, how much Ilokano is being used in the classroom, how much time in the school day is dedicated to activities that support reading, whether a range of effective activities is being used, whether teachers are teaching the prereading skills of letter sounds (as opposed to letter names), etc.

Children in a Vigan City school engaged in Ilokano reading instruction. Photos: S. Pouezevara.

IV. Conclusions

One of the key takeaways from this study is the existence of important inequities in achievement between boys and girls, and across regions. ***Figure 26*** and ***Figure 27*** below present again the same data that were presented previously, but in a format that shows the differences among regional populations. The challenges in these lowest-performing regions are well-known— instability, poverty, geography, lack of access to early childhood education, etc.—and many programs are in place to try to address these factors. However, understanding which factors are having the most influence and what specifically can be done most effectively and efficiently to address the achievement gaps will be critical going forward.

Figure 26: Comparison of Filipino scores across regions

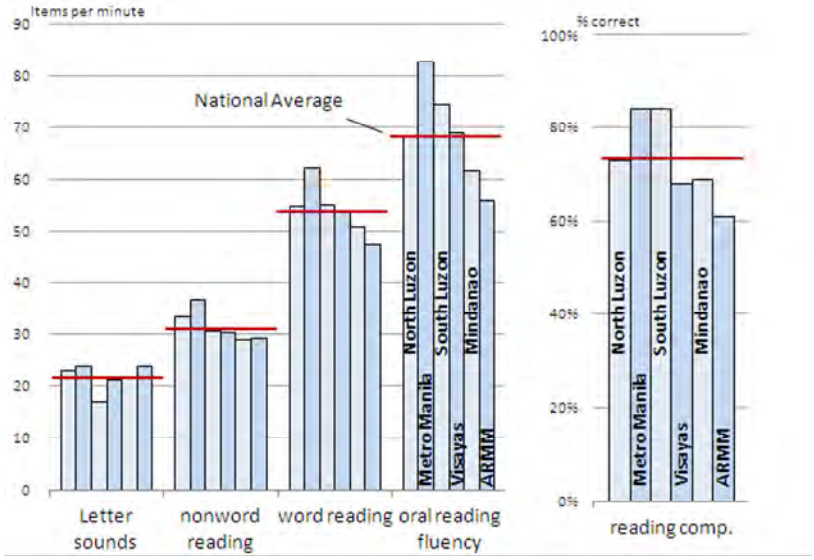
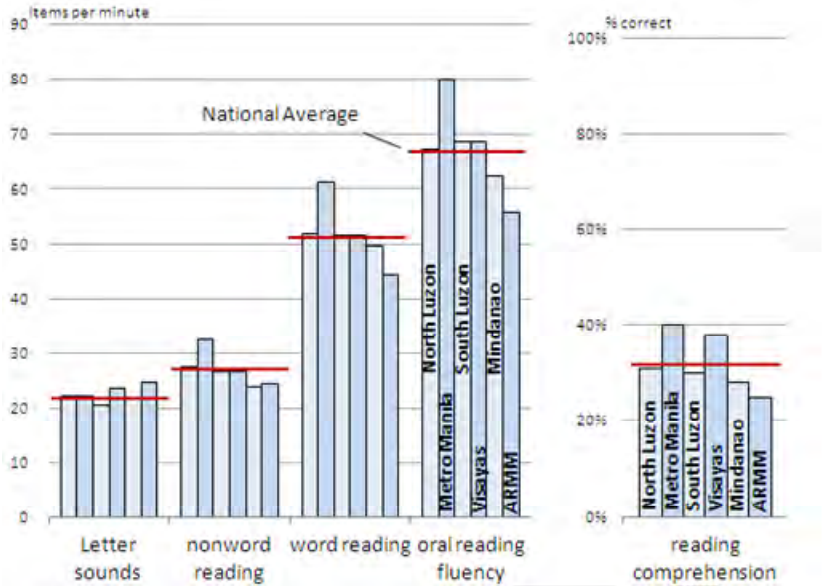


Figure 27: Comparison of English scores across regions



Additionally, the gap in achievement between boys and girls is remarkable, and specific strategies are needed to better engage boys in reading. However, equally important is ensuring that all children start Kinder at the appropriate age and continue their schooling consistently, grade by grade, with a low rate of absenteeism. Combined with common-sense teaching and learning inputs such as textbook availability and activities that promote reading both in school and out of school, policies to promote improvements in these areas would address many of the factors that were revealed to be associated with reading scores in this study.

While some factors cannot be controlled by policy makers, it is important to weigh the magnitude of the effect against the percentage of the population that is affected by the factor when making policy decisions. For example, although living more than an hour away from school was associated with the greatest negative change in ORF in English and Filipino in grade 3, policies to build schools in rural areas would affect only 2% of the population, and may not, therefore, be as cost-effective as interventions that target changes in larger populations. On the other hand, one third of the children in grade 3 classrooms sampled were over-age, indicating a wide prevalence of children who start late or policies that encourage holding children back to repeat a grade. It may be worthwhile to look into ways in which classrooms can be made more homogenous (including avoiding multigrade classrooms) in order to better address the needs of over-age students.

The reasons why children were over-age were not uncovered by this study, but the children likely started late or were required to repeat a grade or grades. However, the correlation between age in grade and reading fluency supports findings in other international literature and other EGRA studies. Children who are required to repeat a grade by definition are experiencing difficulty in school, possibly because of other characteristics such as poverty, high absenteeism, etc. Therefore, they likely were poor readers before being asked to repeat a grade. What is important to note, however, is that repeating a grade does not usually help to raise the level of achievement of the child. See, for example, research from Senegal by Glick and Sahn,[25] which recommended "alternative measures to improve the skills of lagging children" than repeating (p. 93); or the literature review by CfBT Education Trust on grade repetition in sub-Saharan Africa, which concluded that "The majority of the studies undertaken to explore the effects of repetition on learning suggest that the practice does more harm than good (p. 5)."[26]

The issue of age is just one factor that creates classrooms that are very diverse, both linguistically and demographically. This presents a significant challenge for teachers trying to match their instruction to learners' needs when those needs are necessarily so different. Yet research in bilingual education[27] shows that this is exactly what is needed. It has been found to be the case in other countries that teachers typically teach to those at the top of the class—teachers may always call on children they know have the ability to answer and as long as some children have understood and mastered the lesson, they will continue to move on, leaving others behind. Finding time and resources to intervene with remedial instruction for

[25] Glick, P., & Sahn, D. (2010). Early academic performance, grade repetition, and school attainment in Senegal: A panel data analysis. *The World Bank Economic Review*, *24*(1): 93–120. Washington, DC.: World Bank.

[26] Ndaruhutse, S. (2008). *Grade repetition in primary schools in sub-Saharan Africa: An evidence base for change*. Reading, Berkshire, UK: CfBT Education Trust. Retrieved from http://www.cfbt.com/evidenceforeducation/pdf/Grade%20Repetition_FINAL_8FEB08.pdf

[27] Krashen, S. D., & Terrell, T. D. (1983). *The natural approach: Language acquisition in the classroom* (1st ed). New York: Pergamon.

those who need it is a worldwide challenge, but it is one that the Philippines will need to address, given the large percentages of children who are being left behind, according to this study. In fact, a 1999 study published in the World Bank Economic Review found that out of four experiments to improve education quality in the Philippines, the one that was found to be the most promising was to provide teachers with learning materials, "which helped them to pace lessons according to students' differing abilities (p. 493)."[28]

A key question at the grade 3 level is whether, in either language, children are ready to read to learn in the more cognitively complex forms of the language used in subject areas like science and mathematics, and whether they are prepared to do so through independent study as well as classroom instruction. The results of this study raise important doubts about whether that is possible. What is clear is that multilingual education goals of the curriculum are very ambitious and will require a great deal of highly contextualized language input, in a language-rich environment, supported by trained teachers throughout the primary grades.

Areas for Further Research

Futher analysis of these data. The data collected by this study represent hundreds of individual variables for each language and thus many possible levels of analysis. In the interest of presenting key information to DepEd as soon as possible, this preliminary report is partially incomplete. Elements that will be completed and added to a final version include:

- More detailed analysis of dictation data.
- More detailed analysis of the relationship of prosody to comprehension.
- Updated data on socioeconomic status and its effect on achievement based on more precise levels of analysis.
- More cross-language analysis per child—i.e., did children who did well in Filipino also do well in English? Did children present the same error types across subtests?

Additionally, whereas this report presents a great deal of data and attempts to provide explanations and interpretations based on that data, the validity of this interpretation depends on participation of local experts—linguists, teachers, DepEd officials—who know more about how teachers are trained, what materials are in circulation, and what actually happens in classrooms. This dialogue will be taking place and will contribute to the final version.

Potential follow-on studies. Interpreting these results, and monitoring the implementation of MTB-MLE, will also benefit from more in-depth studies that measure the actual language use in the classroom. See, for example, studies done in Africa using classroom observations of instruction and language use.[29] These studies provide detailed information about how much time in the day was being spent—across curricular subjects—by teachers and students using

[28] Tan, J., Lane, L., & Lassibille, G. (1999). Student outcomes in Philippine elementary schools : An evaluation of four experiments. *The World Bank Economic Review, 13*(3): 493–508. Washington, DC.: World Bank.

[29] Studies were conducted by RTI in Mali, Senegal, Kenya, and Uganda with funding from the William and Flora Hewlett Foundation. See the two Piper reports cited earlier, as well as Varly, P. (2010). *The monitoring of learning outcomes in Mali: Language of instruction and teachers' methods in Mali grade 2 curriculum classrooms.* Prepared for the William and Flora Hewlett Foundation under the program Monitoring of Learning Outcomes in Sub-Saharan Africa. Research Triangle Park, North Carolina: RTI International. Retrieved from https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=348 and Varly, P. (2010). *The monitoring of learning outcomes in sub-Saharan Africa: Senegal languages of instruction and teachers' methods in senegal grade 3 classrooms.* Prepared for the William and Flora Hewlett Foundation under the program Monitoring of Learning Outcomes in Sub-Saharan Africa. Research Triangle Park, North Carolina: RTI International. Retrieved from https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=349

different languages, as well as what types of activities teachers and children were engaged in (i.e., reading silently, reading out loud, etc.). These results can also be associated with reading outcomes measured by EGRA and help determine how variations in instructional practice can influence reading outcomes. This will be especially important for the Ilokano schools, since such large differences were seen in achievement, and overall low results; before concluding that MTB education in the early grades is unsuccessful, one must first determine whether MTB education was actually being implemented in the schools in the way it was intended by the program. Similarly, such a study could also look into teacher competencies in the language of instruction. As discussed briefly above, national studies of teacher competency in English have shown low levels of proficiency; additionally, across the Philippines, different linguistic backgrounds affect how English is spoken and understood.

Theories of language acquisition, which inform different models of bilingual education, stress the importance of "comprehensible input",[30] which means not only that the speaker needs to be speaking clear and correct language, but also that whatever is said needs to be appropriate to the level of language that children are capable of understanding, with gradual additions to increase vocabulary and comprehension. This "input +1" model is similar to the concept of leveled readers for children learning to read: Providing texts that are too difficult for a child will result only in frustration and failure, but providing decodable texts reinforces the concepts the child has already learned and creates confidence and ability to attack increasingly difficult texts. To summarize, teachers' competency needs to be measured not only by their own language speaking and reading competency but also by the extent to which they know how to adjust their language use to match the level of the children in the classroom and provide comprehensible input that results in "comprehended intake."

The results presented in this report are the first of their kind on this scale for the Philippines, and represent a baseline against which progress can be measured in the future. It will also be interesting for the Ilokano group to determine intergrade progression in mother-tongue reading by assessing grade 2 learners as well as grade 1 learners. This could be done, for example, in January 2014 using the same instrument in both grade 1 and grade 2, then in 2015 in grade 2 and grade 3, again using the same instrument. In the future, it is recommended to limit this externally administered, sample-based EGRA to grade 2 students who are a bit more mature and may be less intimidated by the procedure.

Similarly, a study completed in grade 3 after the MTB-MLE implementation had been under way for several years would help detect whether children were able to acquire E/F competency on the foundation of their mother-tongue literacy as fast as, or faster than, in the previous structured bilingual immersion method. Again, before definitive conclusions are drawn about the effectiveness based only on student outcome measures, more detailed studies need to be carried out to determine whether mother-tongue literacy was actually taught and acquired by the children. Thus, for example, in 2015 the children who were tested in Ilokano will be completing grade 3. At this time, a measurement of Ilokano, English, and Filipino literacy could be conducted with each child to determine proficiency across languages.

Note that for repeated use of the EGRA instrument, the items will have to be changed and equivalency studies conducted to ensure that the measurements are the same, yet the items cannot be known by teachers in advance such that they can artificially prepare students for success on the assessment. More details will be provided on this and other considerations for

[30] Krashen, S. (1982). *Principles and Practice in Second Language Acquisition.* New York: Pergamon Press.

comparability and scale-up in a separate EGRA toolkit that is under development for DepEd Philippines.

This proposed list of follow-on studies is not exhaustive, and there are certainly many other qualitative and quantitative studies that could add important contextual information with which to improve the interpretation of these findings and to improve implementation of programs to address them.

Annex 1: Description of the Sample

Student Characteristics

As demonstrated above, descriptive statistics can be used to explain differences in reading ability, where differences exist either in characteristics of the sample or in reading ability. This type of analysis has some natural limitations. For example, results cannot be taken to mean that the relationships are necessarily causal or exhaustive—certain factors may simply vary along with learner achievement, and other intervening factors exist that were not measured in this study. Also most of the information is self-reported, and therefore is only indicative and not definitive, since unstated interpretations and biases may be introduced in respondents' answers. Nevertheless, the demographic and descriptive information gathered by the instruments allows us to generate a probable picture of Philippine schools today, and particularly grade 3 classrooms, but also among the grade 1 classrooms that are the first to implement MTB-MLE. ***Table 21*** provides an overview of key characteristics of students based on weighted percentages and averages.

Table 21: Descriptive statistics for children

Characteristic	Grade 3 (E/F)	Grade 1 (Ilokano)
Average age, years	9.3	6.6
Is over-age for the grade level	31%	11%
Child is female (% of sample)	50%	45%
Child is in a multigrade classroom	11%	7.8%
Child was not absent last week	50%	41%
Child was absent 2–3 days last week	21.5%	38%
Child is repeating the grade	6.3%	6.8%
Child attended Kindergarten before grade 1	81%	88%
Child mainly speaks the language of the EGRA at home	English – 2.5% Filipino/Tagalog – 28%	74%
Has other reading materials in the home	47%	76%
Takes more than an hour to get to school	1.5%	(not asked)

School Characteristics

The same questionnaire included information about practices in the classroom, and data gathered from school records provided details about class size. The results are summarized in ***Table 22.***

Table 22: Descriptive statistics about schools

Characteristic	Filipino	Ilokano
Average class size	19.9	29
Teacher mainly uses the language of the EGRA in class	English – 77% Filipino/Tagalog – 90%	71.5%
Teacher never assigns reading homework	3%	n/a
Teacher assigns English reading homework	79%	n/a
Teacher assigns Filipino reading homework	79%	n/a
Child has English reading textbook	79%	n/a
Child has Filipino reading textbook	76%	n/a
Teacher *often* asks the child to read aloud in E, F, or I	27% , 27%	6%
Child has math textbook	53%	n/a

School heads were asked if they were participating in any particular reading improvement programs. The cumulative answers for the grade 3 sample are presented in ***Table 23.*** Regression analysis was done to see if participating in any of these programs was correlated with reading ability. The results indicate that students in schools that self-reported implementing ECARP had a higher average oral reading fluency in both English and Filipino, and this difference was statistically significant. The intensity of implementation or the actual activities engaged in were not specifically defined or measured by this question; it was a yes-or-no question answered by head teachers, and the findings need to be interpreted with this caveat in mind. However, reporting participation in the program would indicate an awareness of the program, its objective ("every child a reader"), and thus a commitment to the national goals of improving reading. Schools who self-reported participating in more than one program were also associated with larger ORF scores.

Table 23: Participation in reading programs (grade 3)

Reading program	No. of schools that participate in this program (children)	Weighted % of schools participating in this program (children)	Average ORF	*p*-value
Read-a-thon	186 (1909)	80% (77%)	E: 68 F: 69	0.181 0.224
ECARP	119 (1201)	51% (49%)	E: 70 F: 72	0.018* 0.000***
Festival of Talents	121 (1228)	49% (50%)	E: 69 F: 70	0.149 0.138
Read more, be more	43 (437)	15% (18%)	E: 68 F: 68	0.834 0.993
Other	(1043)	(42%)	E: 65 F: 67	0.305 0.339

*p <0.05, *** p < 0.001

Annex 2: Details on Sampling and Weighting

Sampling

Grade 3 sample

Selection of the 2013 Philippine EGRA sample involved a **stratified three-stage sample** of divisions, schools, and grade 3 students. ***Table 24*** provides a quick summary of the sampling methodology, followed by further explanations in the text.

Table 24: Summary of sample stage selection for grade 3

Stage no.	Stage item	Stratified by	Selection probability	Total no. of items sampled
1	Divisions	Super-regions	Probability proportional to grade 1 enrollment	56
2	Schools	<None>	Probability proportional to grade 1 enrollment	241
3	Grade 3 students	Gender	Equal probability selection	2,410

The 2010-2011 Basic Education Information System (BEIS) Elementary School census data were used as the list frame for sampling divisions and schools. The initial list of schools contained 38,351 public elementary schools. After removal of 196 schools that had zero grade 1 enrollments,[31] a total of 38,155 schools remained in the final sample frame.

Stage 1 – Selection of divisions: The 2010–2011 BEIS data were aggregated to the division level, then stratified by the six super-regions. With the exception of the National Capital Region, divisions were sorted by region, then 10 divisions were selected proportional to total grade 1 enrollment using systematic random sampling. Because the NCR Grouped region contains only 16 divisions, and encompasses a small geographical area, all divisions were automatically selected.32

Stage 2 – Selection of schools: Schools from each selected division were selected with probability proportional to the grade 1 enrollment. About 40 schools were selected for each super-region;[33] in total, 241 schools were selected. For each selected school, two replacement schools were selected in the event the original sampled school was not able to be assessed (for example, the school was permanently closed, or did not contain any grade 3 students).

Stage 3 – Selection of students: Grade 3 students were selected the day the assessment team arrived at the school. All grade 3 students within the selected school were stratified by gender. Five males and five females were then selected with equal probability. If the school had fewer than 10 students, all of the students were automatically selected.

[31] Because the BEIS sample frame was two years old, a zero enrollment of grade 1 students two years ago meant that there would currently be a zero enrollment of grade 3 students.

[32] Another reason all divisions in the NCR were automatically selected in the sample: Selecting the divisions proportional to grade 1 enrollment using the systematic sampling process meant that 7 of the 10 divisions were automatically selected.

[33] Forty-one schools were selected from the super-region: Visayas (regions R6, R7, R8).

Grade 1 sample

The sample selection of the 2013 Philippine grade 1 Ilokano EGRA sample involved a **stratified two-stage sample** of schools and grade 1 students ***(Table 25).***

Table 25: Summary of sample stage selection for grade 1

Stage no.	Stage item	Stratified by	Selection probability	Total no. of items sampled
1	Schools	Ilokano pilot school	Probability proportional to grade 1 enrollment	50
2	Grade 1 students	Gender	Equal probability selection	500

The 2010–2011 BEIS Elementary School census data were used as the list frame for sampling schools. After exclusion of schools that (1) were known to not teach in Ilokano (n = 549), (2) had participated in the Ilokano EGRA pilot study (n = 4), (3) participated in the grade 3 national EGRA study (n = 9), and (4) had zero grade 1 enrollments, a total of 1,786 public elementary schools in the Ilokano Region remained in the list frame.

Stage 1 – Selection of schools: Ten schools that had participated in the Ilokano mother-tongue pioneering study were randomly selected and 40 schools that did not participate in the pioneering study were randomly selected. All schools were selected proportional to the grade 1 enrollment using systematic random sampling.

Stage 2 – Selection of students: Grade 1 students were selected the day the assessment team arrived at the school. All grade 1 students within the selected school were stratified by gender. Five males and five females were then selected with equal probability. If the school had fewer than 10 first graders, all of the first graders were automatically selected.

Weighting

Sample weights were created by the inverse of the probability of selections. The school weights were scaled to the total number of schools in the strata. Student weights were scaled to the total number of enrolled boys and girls at the Regional level.

The final sample of schools and students, with the weighted percentages, is provided in ***Table 26.***

Table 26: Final sample for grades 1 and 3: Divisions, districts, and schools

Grade 1 Sample

Division	Schools			Students		
	N	Weighted value	Weighted percent	*N*	Weighted value	Weighted percent
Alaminos City	2	35	2.0	22	1822	2.6
Candon City	1	6	0.3	10	449	0.6
Ilocos Norte	7	333	18.6	72	10325	14.8
Ilocos Sur	10	554	31.0	80	13462	19.3
La Union	9	150	8.4	92	12642	18.1
Laoag City	2	63	3.5	20	2335	3.3
Pangasinan I:Lingayen	5	174	9.8	50	9243	13.2

Pangasinan II:Binalonan	12	429	24.0	108	16919	24.2
San Fernando City	1	1	0.1	10	283	0.4
Udaneta City	1	17	0.9	10	621	0.9
Urdaneta city: r1	1	1	0.1	10	340	0.5
Vigan City	1	24	1.4	10	1462	2.1

Grade 3 Sample

Super-region	Region	Schools			Students		
		N	Weighted value	Weighted percent	*N*	Weighted value	Weighted percent
	I - Ilocos Region	10	2366	6.2	100	83103	3.56
	II - Cagayan Valley	6	2174	5.7	61	95679	4.1
	III - Central Luzon	20	2948	7.73	200	214889	9.22
North Luzon	CAR	4	1455	3.81	38	85942	3.69
Metro Manila	NCR	40	510	1.34	400	151040	6.48
	IV-A - CALABARZON	20	2708	7.1	206	215720	9.25
	IV-B - MIMAROPA	8	1814	4.75	81	123756	5.31
South Luzon	V - Bicol Region	12	3137	8.22	122	148988	6.39
	VI - Western Visayas	16	3369	8.83	160	193267	8.29
	VII - Central Visayas	16	2901	7.6	164	231372	9.92
	VIII - Eastern Visayas	9	3582	9.39	89	131637	5.65
Visayas	IX - Zamboanga Peninsula	6	2076	5.44	70	129011	5.53
	X - Northern Mindanao	12	2075	5.44	126	119508	5.13
	XI - Davao Region	9	1636	4.29	96	74135	3.18
	XII - SOCCSKSARGEN	9	1673	4.38	94	95828	4.11
Mindanao	XIII – Caraga	4	1609	4.22	42	109736	4.71
ARMM	ARMM	39	2122	5.56	414	128025	5.49
Total:		240	38155	100	2463	2331636	100.01

Annex 3: Reading Comprehension Distribution Scores

The following tables provide additional details on the distribution of reading comprehension scores. Unlike the values provided in the main body of this report, these tables exclude 'zero' scores and shows the distribution of scores according to how many questions a child attempted rather than the total number of questions possible (5). According the standard rules of EGRA administration, a child is only asked questions that correspond to text he or she has actually read. Thus, a child who only attempted two questions (because they read too slowly to advance in the text far enough to be asked two questions) has three possible scores—0% (0 of 2 question answered correctly), 50% (1 of 2 questions answered correctly) or 100% (2 of 2 questions answered correctly). Table 27 shows that the distribution of scores for Filipino reading comprehension, for children who only attempted 2 questions, is 41% who answered 0 of 2 questions correctly, 39% who answered 1 of 2 questions correctly, and 20% who answered 2 of 2 questions correctly. This allows us to see in what category the majority of students fall according to what they actually attempted to do.

Table 27: Distribution of Filipino reading comprehension scores by number of questions attempted

Q's asked	*N*	Percentage of questions answered correctly										
		0	20	25	33	40	50	60	67	75	80	100
1	39	**74%**										26%
2	71	**41%**					39%					20%
3	127	20%			27%				**33%**			20%
4	241	10%		9%			21%			29%		**30%**
5	1,929	<1%	1%			6%		13%			27%	**53%**

* Based on unweighted scores.
** Of children who got all questions correct.

Table 28: Distribution of English reading comprehension scores by number of questions attempted

Q's asked	*N*	Percentage of questions answered correctly*										
		0	20	25	33	40	50	60	67	75	80	100
1	62	**95%**										5%
2	124	**90%**					8%					2%
3	99	**81%**			18%				0%			1%
4	161	**65%**		24%			8%			2%		1%
5	1,951	9%	**27%**			21%		15%			15%	13%

* Based on unweighted scores.
* Average ORF of children who got all questions correct in each category (weighted values).

Table 29: Distribution of Ilokano reading comprehension scores by number of questions attempted

Q's asked	N	Percentage of questions answered correctly*										
		0	20	25	33	40	50	60	67	75	80	100
1	18	**100**%										0%
2	85	**71**%					26%					2%
3	205	32%			**36%**				16%			16%
4	62	5%		5%			23%			32%		**35%**
5	45	2%	7%			18%		**33%**			31%	9%

* Based on unweighted scores.

What is particularly striking from these tables is that no matter how far they read in the text, and subsequently, how many questions they were asked, most children could not answer even one question correctly in English. In Filipino and Ilokano, as children were asked more questions, they were also increasingly likely to answer most of them correctly.