USAID/PHILIPPINES BASA PILIPINAS PROGRAM

EVALUATION REPORT FOR SCHOOL YEARS 2013/14 AND 2014/15

MAY 2015

This publication was produced for review by the United States Agency for International Development. It was prepared by Education Development Center, Inc. (EDC).

USAID/PHILIPPINES BASA PILIPINAS PROGRAM

EVALUATION REPORT FOR SCHOOL YEARS 2013/14 AND 2014/15

PROGRAM TITLE:	USAID/PHILIPPINES BASA PILIPINAS PROGRAM
SPONSORING USAID OFFICE:	USAID/PHILIPPINES - OFFICE OF EDUCATION
CONTRACT NUMBER:	AID-492-C-13-00004
CONTRACTOR:	EDUCATION DEVELOPMENT CENTER, INC. (EDC)
DATE OF PUBLICATION:	MAY 1, 2015
AUTHOR:	EDUCATION DEVELOPMENT CENTER, INC. (EDC)

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

EXECUTIVE SUMMARY

Basa Pilipinas is a four year early grade reading project that supports the Department of Education's national reading program. Basa is helping the Department of Education (DepEd) implement transformative literacy practices in selected divisions of Regions 1 and 7 by providing teacher and student materials, training teachers and school heads, and providing post-training support.

The project evaluation follows a cross-sectional design. In school year 2013/14, prior to the full implementation of the project, the Basa project team collected data on student achievement in literacy, teachers' classroom practices, teacher beliefs on literacy instruction, and data on school environment, to serve as a comparison cohort for the Basa intervention. Data were collected from 40 schools in Cebu and La Union. In school year 2014/15, after the rollout of the full Basa intervention, data collection was expanded to 80 intervention schools in Bohol, Cebu, Ilocos Norte, Ilocos Sur, and La Union. Key findings from the evaluation are outlined below. This evaluation report contains baseline and endline early grade reading and teacher practices results for the intervention and comparison cohorts.

Student Literacy Findings. As part of the study, in school year 2013/14, 469 second grade students were randomly selected from 40 schools in Cebu and La Union. These students, selected prior to the full implementation of the Basa intervention, form the comparison group. Subsequently, in school year 2014/15, a random sample of 1,216 students from 80 schools in Cebu, Bohol, Ilocos Norte/Sur and La Union were selected to be tested and form the intervention group. For both the comparison and intervention groups, student data were collected at two points in time (baseline and endline). The early grade reading assessment (EGRA) was administered to the comparison and intervention cohort students in Filipino by Basa-trained assessors who were native speakers. In addition to Filipino, the intervention cohort students were also tested in English literacy.

Students from the intervention cohort outperformed ($p<.001$) their comparison counterparts in all Filipino subtests. Since the comparison of means at the baseline showed that the tested cohorts were not perfectly identical (the intervention cohort students scored higher on all but one subtests), a statistical correction was conducted to estimate the amount of gains *after* the baseline differences were incorporated into the analyses and controlled for.

The results of the analysis are presented in Table E-1. The amounts of gains due to the Basa intervention are shown in the last column as the "intervention effect". As the table shows, the Basa intervention helped increase the fluency of second grade students by additional nine words per minute, above and beyond the natural increases expected to occur over the course of the school

year. Additional gains in dictation as well as listening and reading comprehension were also found to be substantial. The smallest additional gains were found in letter sounds subtest, and reading of familiar and nonsense word subtests.

Table E-1. Basa Intervention Effect on Filipino EGRA Results, by Subtest, Controlling for the Baseline

Filipino EGRA Subtests	Endline Difference of Means[1]*	Effect Size of Cohort Difference at Endline	Intervention Effect**
Initial Sound Identification (pct correct)	20.8	0.65	17.1%
Letter Sounds(percent correct)	10.4	0.64	5.6%
Letter Correct (per min)	10.2	0.63	5.5 lcpm
Familiar Words(pct correct)	5.1	0.17	2.7%
Familiar Words Correct(per min)	4.4	0.21	2.6 wcpm
Nonsense Words (pct correct)	5.2	0.20	2.9%
Nonsense Words Correct (per min)	2.7	0.20	1.5 wcpm
Oral Passage Reading (pct correct)	19.0	0.63	16.1%
Words Correct in a Text (per min)	11.1	0.45	9.1 wcpm
Prosody score	0.9	0.99	.50
Reading Comprehension (pct correct)	19.1	0.66	23.5%
Listening Comprehension (pct correct)	14.7	0.38	13.4%
Dictation Composite (pct correct)	16.3	0.61	13.7%

* All differences between the means are statistically significant at the *p*<.001 level.
** Intervention effect is computed using a linear regression model controlling for the baseline differences between the cohorts.

In both intervention and comparison cohorts, students had significantly fewer zero scores in all of the tested areas at the end of the school year than at the beginning of the school year. The two subtests with the highest percent of students scoring zero were the reading and listening comprehension subtests. This was particularly true for the English EGRA.

Several key trends also emerged for both the intervention and comparison groups from the data analysis:

- Students achieved higher in oral reading and lower in reading comprehension
- In both English and Filipino, students performed best on the initial sound identification and oral reading passage subtests.
- The subtests with the lowest scores were letter sounds for Filipino EGRA and listening comprehension and dictation on the English EGRA.

Overall, the majority (92%) of students performed better on the Filipino EGRA than on the English EGRA, which is not surprising given that English is introduced later in the curriculum than Filipino. An exception to this trend was found in the results of the oral reading fluency: on average, students read faster in English than they did in Filipino. Similarly, on the letter sound identification subtest, students performed better on the English test than in Filipino. These differences are statistically significant at the *p*<.001 level.

Girls on average demonstrate better EGRA results than boys. This pattern was consistent at the baseline and endline testing, in both intervention and comparison cohorts. The difference is statistically significant at *p*<.001 level for all subtests. This finding corroborates with a survey on

[1] Computed as follows: intervention cohort mean at baseline (weighted) – comparison cohort mean at baseline (weighted).

teacher beliefs and instructional practices (BIPI) collected in July/August 2014 which revealed that teachers thought it was harder for boys to learn to read than for girls. Further study is needed to see whether teacher gender bias in literacy expectations may have resulted in boys' lower literacy performance on the EGRA tests.

Quality of Language and Literacy Instruction Findings. The literacy practices of second grade teachers from a sample of schools were observed in November/December 2013 and again in December/January 2014/5, using the *Standard Classroom Observation Protocol in Education – Literacy (SCOPE Literacy).*[2] *SCOPE Literacy* is a structured observation tool designed to provide an overall assessment of classroom instruction in the areas of language, reading, and writing. The overall scores at the first observation were found to be very low, ranging between "deficient" and "inadequate" in November/December 2013. By the second observation in December/January 2014/5, improvements were seen in all teaching practices observed; however, the largest gains were seen in the Language and Literacy Instruction domain. In the Classroom Structure domain, teachers showed the largest improvements in ensuring participation of all learners, ensuring accessible classroom materials and effective management of reading and writing instruction. In fact, by December/January 2014/5, nearly half of observed teachers were scored as "strong" in classroom materials and "management of reading and writing instruction." For the Language and Literacy Instruction domain, teachers largely improved literacy instruction in terms of oral language development, developing reading fluency, and developing comprehension. These improvements are associated with the Basa intervention.

[2]*SCOPE* was developed by EDC for utilization with literacy projects and has been tested used in a number of countries at present.

Contents

LIST OF FIGURES 5
LIST OF TABLES 6
ACRONYMS 7
INTRODUCTION 8
EVALUATION DESIGN 10
Sample 10
STUDY PARTICIPANTS: STUDENTS 12
Student Context Interview 13
School and Teacher Environment 13
Socio-Economic Factors 14
Home Environment 15
SUMMARY EGRA FINDINGS 17
Overall Filipino EGRA Results 17
Effect Size 18
Overall English EGRA Results 20
Zero Scores 21
Results by Sex 23
Results by Division 24
Benchmarks 25
FILIPINO EGRA SUBTEST RESULTS 27
Initial Sound Identification, Letter Sounds, Familiar and Nonsense Words 27
Oral Reading Fluency and Comprehension in Filipino 29
Listening Comprehension and Dictation 34
ENGLISH EGRA SUBTEST RESULTS 37
Initial Sound Identification, Letter Sounds, Familiar and Nonsense Words 37
Oral Reading Fluency and Comprehension 39
Listening Comprehension and Dictation 42
IMPACT OF CONTEXT FACTORS ON ACHIEVEMENT 45
Impact of Parental Literacy 45
Impact of Home Context Factors 45
Impact of Home Language 46
School-Level Results in Reading 47
CHANGES IN TEACHER PRACTICE 49

CONCLUSION AND RECOMMENDATIONS 52
APPENDICES 1

LIST OF FIGURES

Figure 1: Basa Theory of Change 8
Figure 2. Sample Participants, by Sex and Group 13
Figure 3. Kindergarten Attendance Before Grade 1 14
Figure 4. Household Assets 15
Figure 5. Parental Literacy and Help with Reading at Home 16
Figure 6. Availability of Books at Home 16
Figure 7. Average Percent of Correct Answers for Filipino EGRA Subtests in All Provinces, by group. 18
Figure 8. Average Percent Correct on English EGRA Subtests 21
Figure 9. Change in Overall Zero Scores in Filipino EGRA Subtests, Across Baseline and Endline 22
Figure 10.Percent of Tested Students Scoring Zero on English EGRA Subtests 22
Figure 11. Average Filipino and English EGRA Subtest Percent Correct, by Sex 23
Figure 12.Percent of Filipino Text Read Correctly at the Beginning and the End of the School Year .. 24
Figure 13. Filipino Oral Passage Reading Subtest Results at Endline, by Group 25
Figure 14. Intervention Cohort Filipino Oral Passage Reading Subtest Results at Endline, by Sex 26
Figure 15. Filipino Initial Sound Identification – Endline Percent Correct, by Group 27
Figure 16. Filipino Correct Letter Sound Subtest Endline Results, by Group 28
Figure 17. Filipino Familiar Word Identification Endline Subtest Results, by Group 28
Figure 18. Filipino Nonsense Words Endline Subtest Results, by Group 29
Figure 19. Filipino Oral Passage Reading Endline Results, Percent of Words Read Correctly, by Group 29
Figure 20. Filipino Oral Passage Reading Average Percent Correct, by Group and Sex 30
Figure 21. Filipino Oral Reading Fluency (wcpm), by Group, at Endline 30
Figure 22. Filipino Fluency Measure (wcpm) at Endline, by Sex 31
Figure 23. Filipino Oral Passage Reading - Endline Prosody Score, by Cohort 31
Figure 24. Reading Comprehension Endline Results, Filipino EGRA, by Group 33
Figure 26. Comparison of Intervention Cohort Endline Comprehension Results Using Timed and Untimed Reading 34
Figure 27. Percent of Students Reading and Understanding Grade Level Text, at Endline, by Group. 34
Figure 28. Filipino Listening Comprehension Questions – Number Correct at Endline 35
Figure 29. Filipino Dictation Zero Scores at Endline, by Group 35
Figure 30. Filipino Dictation Breakdown, by Group 36
Figure 31. English Initial Sound Identification – Percent Correct 37
Figure 32. English Correct Letter Sound Subtest Results 38
Figure 33. English Familiar Word Identification Subtest Results 38
Figure 34. English EGRA - Nonsense Words Subtest Results 39
Figure 35. English Oral Passage Reading Subtest Results, Percent of Students Reading Correctly 39
Figure 36. English Oral Passage Reading (wcpm) 40
Figure 37. English Oral Passage Reading - Prosody Score 41
Figure 38. English EGRA Reading Comprehension Endline Results 41

Figure 39. Comparison of Intervention Cohort Endline Comprehension Results Using Timed and Untimed Reading 42
Figure 40. English Listening Comprehension Questions – Number Correct at Endline 43
Figure 41. English Dictation Subtest Results 43
Figure 42. English Dictation Breakdown (Out of 17 possible points) 44
Figure 43. EGRA Results at Endline for Intervention Students Disaggregated by Language Spoken at Home 46
Figure 44. Average percent correct among students on Filipino Oral reading fluency test and comprehension questions in study schools, by grade 47
Figure 45. Average percent correct among students on English Oral reading fluency test and comprehension questions in study schools, by grade 48
Figure 46. Language spoken at home among students in the lowest and in the highest 20 percent of study schools 48
Figure 47. Change in Average Scores for SCOPE, Change from 2013 to 2014 51
Figure 48. Correlation between Two Components of the SCOPE in 2014 52

LIST OF TABLES

Table 1. Timeline of Project Implementation and Data Collection 10
Table 2. EGRA Evaluation Sample 11
Table 3. Sample of Schools and Students by Province and Group 12
Table 4. What language do you speak at home and at school? 14
Table 5. Where do your parents work? 15
Table 6. Filipino EGRA Results, by Subtest and Cohort 17
Table 7. Basa Intervention Effect on Filipino EGRA Results, by Subtest, Controlling for the Baseline . 19
Table 8. English EGRA Results, Intervention Cohort Sample 20
Table 9. Basa Intervention Effect on Average Filipino EGRA Results, by Sex 24
Table 10. English Oral Fluency Measure at Endline, by Sex 40
Table 11. Effect of parental literacy on reading fluency* 45
Table 12. Effect of home context factors on reading fluency 45

ACRONYMS

Basa	Basa Pilipinas
BIPI	Beliefs and Instructional Practices Inventory (Survey)
DepEd	Department of Education
DNK	Do not know
EGRA	Early Grade Reading Assessment
ICC	Inter-class correlation
IRR	Inter-rater reliability
LAC	Learning Action Cell
M&E	Monitoring and Evaluation
PTA	Parent Teacher Association
RTI	Research Triangle Institute
wcpm	Words correct per minute

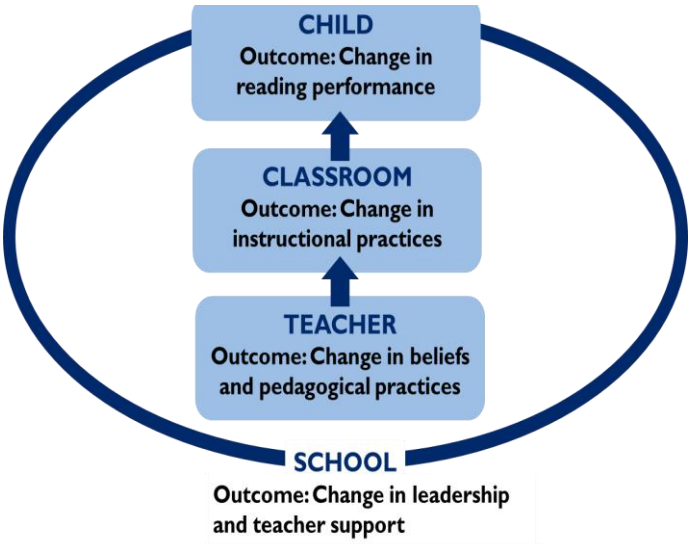
INTRODUCTION

The Basa Pilipinas (Basa) Project is a four-year project in support of the Philippine Government's literacy component of the K to 12 curriculum, and is implemented in close coordination with the Department of Education (DepEd) and other key education stakeholders in selected divisions and regions nationwide. The project is aligned within the framework of USAID's Global Education Strategy, USAID-Philippines' emerging Country Development and Cooperation Strategy, and the Philippine Government's priorities for basic education. Basa was established in 2013 to improve the reading skills for one million children in the early grades in English, Filipino and selected mother tongues. In addition, Basa works closely with DepEd to support and strengthen the literacy component of its K–12 Integrated Language Arts Curriculum for Grades 1 – 3.

In 2012, DepED adopted a K-12 Curriculum to be phased in over a six year span. The new Grade 1 curriculum was implemented nationwide in 2012/13, prior to the start of the Basa project. Grade 2 was implemented in 2013/14, and Grade 3 was rolled out nationwide in 2014/15. Basa's training focuses on enhancing the literacy component of the curriculum, with a particular emphasis on training teachers in techniques for bridging across languages. As per the DepEd policy of Mother Tongue-based Multilingual Education (MTB-MLE), Mother Tongue is the language of instruction in school for Grades 1 to 3 while Filipino and English are introduced gradually. The Basa project works in Regions 1 and 7, where the Mother Tongues of instruction are primarily Ilocano and Cebuano, respectively.

In 2013, Basa provided initial training in enhanced literacy instruction techniques to schools in Cebu (Region 7) and La Union (Region 1). In 2014, the project expanded to include the divisions of Bohol and Mandaue City in Region 7 as well as Ilocos Norte and Ilocos Sur in Region 1. All schools in these divisions were provided with a complete set of Grades 1 and 2 literacy teaching and learning materials, including teacher guides, read alouds and leveled readers, as well as training for school heads and teachers. In 2015, the project will also add the city divisions of Tagilaran and San Fernando, and will provide teaching and learning materials and training for Grade 3 teachers in all Basa divisions.

Figure 1: Basa Theory of Change



As part of its scheduled monitoring and evaluation (M&E) activities,[3] Basa is conducting outcome evaluations to measure changes at the school level with Grade 2 and 3 teachers and students, as well as principals, participating in the Basa intervention. The results of the evaluation activities will be used to not only inform the project technical and management teams (program management)

[3] For the evaluation design and details see the project's M&E (Performance) Plan with Contract Monitoring Plan, April 20th 2013.

but also, to measure outcomes (change in learner and teacher performance as well as principal viewpoints).

Data that are reported in this document were collected in Bohol, Cebu, Ilocos Norte/Sur and La Union and Mandaue City[4]—the focal divisions of the Basa intervention in 2014/15. The evaluation study was conducted to measure changes in student achievement between the comparison cohort (SY 13/14) and the Basa intervention cohort (SY 14/15), as well as to assess the growth in student literacy skills from the beginning of the school year to the end of the school year. Comparison cohort students were tested in Filipino, and intervention cohort students were tested in both Filipino and English by trained data collectors. Data on the quality of classroom instruction were collected using EDC-developed *Standards-Based Classroom Observation Protocol for Educations – Literacy* (SCOPE-Literacy).

This report presents findings from student assessment and teacher observations for both a comparison and Basa intervention cohorts. Principal surveys and teacher beliefs and practices results collected in 2014 can be found in the Annex. All survey and observation data were collected in English. Interpreting the results of student assessment, it is important to recognize that grade 2 students have had different amount of formal instruction in each language.

Filipino and English are introduced as subjects during the second and third quarters of Grade 1, with a focus on oral language development. Each language lesson is allotted 30 minutes daily. During Grade 2, students begin reading and writing in Filipino in the first quarter and in English in the third quarter. Each language lesson is allotted 50 minutes daily. In Grade 3, all three languages are taught as subjects; Mother Tongue remains the main language of instruction for all other subjects. Later, in Grade 4, Mother Tongue is phased out as both a subject and language of instruction and students are taught all subjects in Filipino, except math and science which are taught in English.

Therefore, by the end of the school year, second graders will have had seven quarters of instruction in the development of oral language in Filipino and four quarters in reading and writing Filipino. They will have received less instruction in English; six quarters in developing oral language skills in English and just two quarters of instruction in reading and writing English.

Introduction of Filipino and English by Grade Level and Quarter

<table>
<tr><th rowspan="2"></th><th colspan="4">Grade 1</th><th colspan="4">Grade 2</th><th colspan="4">Grade 3</th></tr>
<tr><th>Q1</th><th>Q2</th><th>Q3</th><th>Q4</th><th>Q1</th><th>Q2</th><th>Q3</th><th>Q4</th><th>Q1</th><th>Q2</th><th>Q3</th><th>Q4</th></tr>
<tr><td>Mother Tongue</td><td colspan="12">As language and medium of instruction for all other subjects</td></tr>
<tr><td>Filipino language</td><td></td><td colspan="3">oral</td><td colspan="8">reading/writing</td></tr>
<tr><td>English language</td><td></td><td></td><td colspan="4">oral</td><td colspan="6">reading/writing</td></tr>
</table>

[4] Manduae City data was removed from the longitudinal analysis because only endline data were collected there.

EVALUATION DESIGN

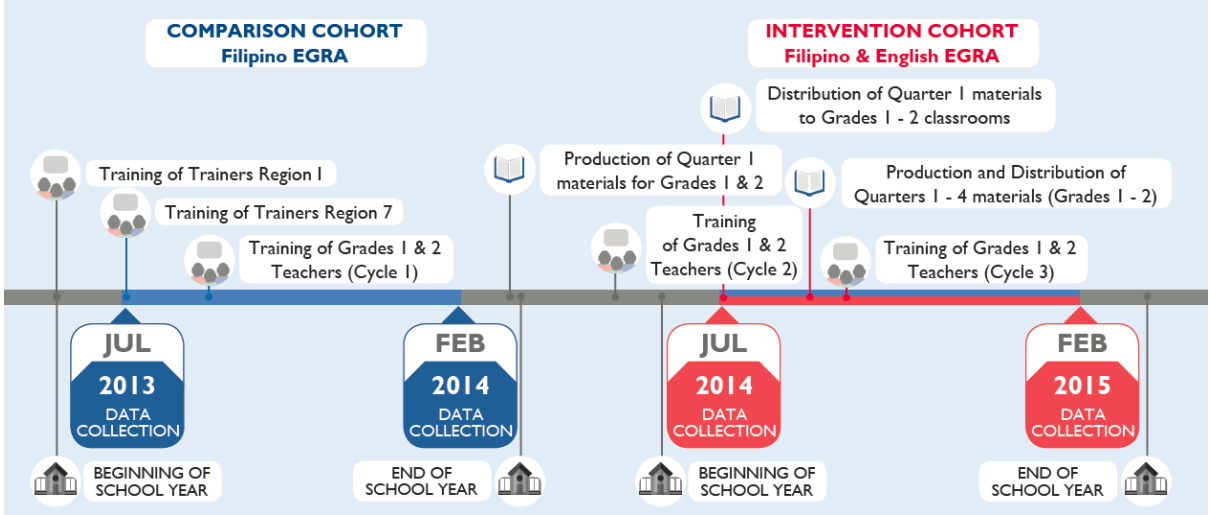
The purpose of this evaluation is to measure changes in Grade 2 student achievement as well as teacher practice associated with the Basa intervention. Specifically, the evaluation study was designed to answer the following questions:

1. Do students, both male and female, demonstrate improved reading and comprehension skills in Filipino at the end of Grade three per DepEd Standards?
2. Have teachers' instructional practices in reading (in the Filipino language) improved in target schools?

To answer question one on student progress, the evaluation followed the cross-sectional design. In 2013/14 before the full intervention began, a comparison cohort in a sample of schools in Cebu and La Union was assessed in reading (Filipino) to measure the natural growth in literacy skills over the course of a school year. The subsequent year (SY 2014/15), a random sample of grade 2 students in the same schools were tested in reading (Filipino) at the end of the school year.

To provide a more complete picture of the Basa outcomes, during school year 2014/15 the evaluation was expanded to additional schools in Cebu and La Union, as well as to schools in Ilocos Norte, Ilocos Sur and Bohol. Student assessment results are compared with comparison cohort scores established in 2013/14.

Table 1. Timeline of Project Implementation and Data Collection



The *SCOPE* data collected from 40 schools in Cebu and La Union were used to analyze changes in instructional practices among teachers trained by Basa. SCOPE data are used to answer evaluation question #2 which follows a longitudinal design with the same teachers being observed annually. This report will provide findings from November/December 2013 and December 2014.

Sample

The student sample for the comparison cohort (SY 2013/14) was drawn from Cebu and La Union. In SY 2014/15 the sample was drawn from Bohol, Cebu, Ilocos Norte/Sur, and La Union. The table below details the sample used in this report.

Table 2. EGRA Evaluation Sample

	Evaluation Sample			
	Grade	**Divisions**	**# of schools**	**# of students**
Comparison cohort	Grade 2	Cebu & La Union	40	469
Intervention cohort	Grade 2	Bohol, Cebu, Ilocos Norte, Ilocos Sur & La Union	80	1216

To enable the computation of estimates of literacy skills among students in *all* schools affected by Basa intervention based on the data from the tested sample, design weights were applied to the analyses of EGRA data. Consequently, the sample numbers (*n*) are only reported in the section describing study participants; the Findings section only shows the results for weighted data.

Appendices include a detailed description of methods and data collection tools, results of the BIPI and SSME surveys from the baseline of SY 2014/2015, and detailed assessment results.

STUDY PARTICIPANTS: STUDENTS

The evaluation study included 1,685 students:

- 1,216 intervention cohort students in all 5 provinces, and
- 469 the comparison cohort students in Cebu and La Union

Table 3 shows the number of schools in each province from which the samples were drawn.

Table 3. Sample of Schools and Students by Province and Group

	Intervention		Comparison	
PROVINCE	Number of Schools	Number of Students Matched at Endline	Number of Schools	Number of Students Matched at Endline
Bohol	20	298	-	-
Cebu	20	311	20	230
Ilocos Norte/Sur	19	291	-	-
La Union	20	316	17	239
Total	**79**	**1,216**	**37**	**469**

All participating students were randomly selected from second grade classrooms and tested twice: at the beginning of the school year, and at the end of the school year[5]. Design weights were applied to compensate for differences in provincial sampling and to ensure an appropriate representation of learners in all provinces in the sample. Weights were used in all analyses of EGRA data. Actual *n*'s are only reported in this section; in subsequent sections *n*'s will not be reported and weighted data will be used.

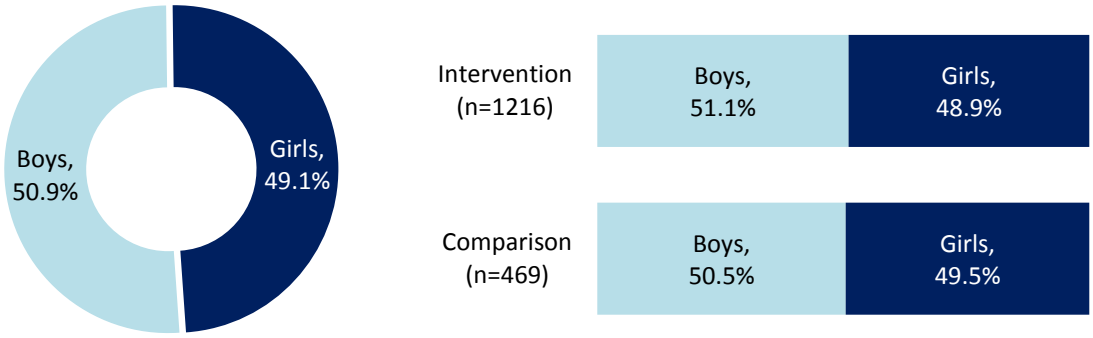
Most of the demographic data collected among comparison and intervention cohort students were very similar. In the instances where the results differed, they were reported separately for the two study groups. In the instances where the results were similar, they were presented in aggregate.

Sex and Age of Learners. A near gender parity was achieved in the total sample, with 49.1% female and 50.9% male student participants; this is consistent across intervention and comparison cohorts. Although all students were selected from the second grade, students

[5] Intervention cohort tested students were tracked using their names and classrooms where they were sampled from. The study team could not locate 5.1% of intervention cohort sample students at the endline. These students were replaced with students from the same classrooms with a similar demographic profile. For the comparison cohort, sample students were matched using propensity score matching with classrooms, gender and age as propensity variables.

ranged in age from 6 to 10 years in the intervention cohort sample and 5 to 12 in the comparison cohort sample. On average, mean ages of students in the intervention and comparison cohort samples were nearly the same with a mean age of 7 in the intervention cohort students and 7.3 in the comparison cohort students. Participating girls were on average slightly younger than boys.

Figure 2. Sample Participants, by Sex and Group



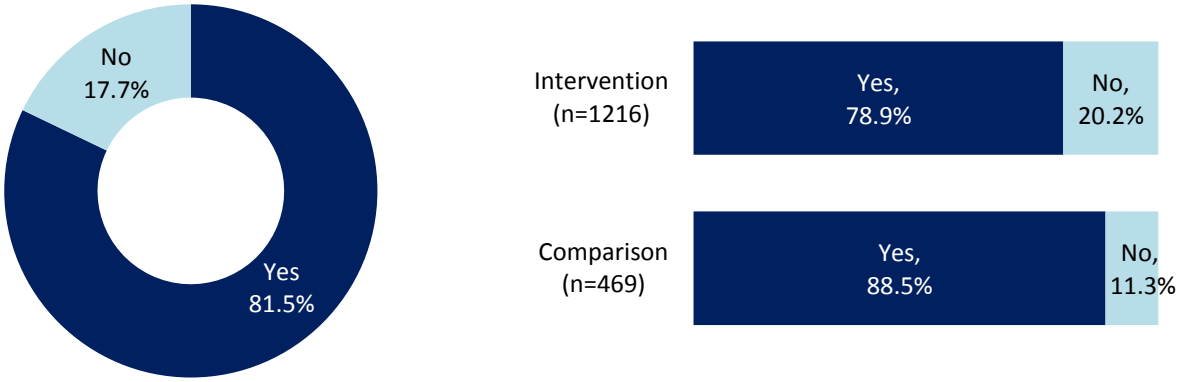
Student Context Interview

It is widely recognized in the field of education that contextual factors, such as supportive home environments, adequate nutrition, and early exposure to literacy, play prominent roles in helping children succeed academically. Additionally, school factors such as teachers assigning homework or teachers reading to children have also been found to be associated with improved performance. To assess these contextual factors, students were asked a series of questions about their home environment, student/teacher practices and their socioeconomic status. Below are results from the student context interview at baseline for both the intervention and comparison groups.

School and Teacher Environment

Nearly all students reported that they've been attending their school since the beginning of the school year. Most students reported that they walked to school. More than three in four students (81.5%) said they attended kindergarten before starting Grade 1. Slightly more comparison cohort students reported attending kindergarten than the intervention cohort students (Figure 3).

Figure 3. Kindergarten Attendance Before Grade 1



Students reported using a variety of different languages at home and at school.[6] While over half said they speak English at school, less than 5% said they speak English at home. Two-thirds of students (67.2%) reported speaking Filipino/Tagalog at school, while fewer than 10% said they speak Filipino/Tagalog at home. English (62.5%), Filipino/Tagalog (67.2%), Bisaya/Cebuano (58.8%), and Ilokano (20.1%) were reported to be most widely spoken languages in school. Cebuano/Bisaya was mostly widely spoken at home in Bohol (97%), and Cebu (91%) while Ilokano was more widely used in Ilocos Norte (82%), Ilocos Sur (84%) and La Union (74f%). Note that Filipino, which is originally based on the Tagalog dialect, is the national language of the Philippines and is introduced gradually beginning in the second quarter of Grade 1, with students' mother tongues forming the basis of instruction up through grade 3.

Table 4. What language do you speak at home and at school?

	Intervention		Comparison	
Language	**School**	**Home**	**School**	**Home**
Filipino/Tagalog	67.1%	8.0%	67.4%	15.1%
English	64.6%	2.1%	57.2%	4.2%
Bisaya/Cebuano	64.2%	70.8%	44.6%	62.9%
Hiligaynon	0.0%	0.0%	0.0%	0.0%
Tausug	0.1%	0.0%	0.0%	0.0%
Ilokano	21.3%	21.4%	17.1%	18.7%
Other languages	0.1%	0.1%	0.0%	0.0%

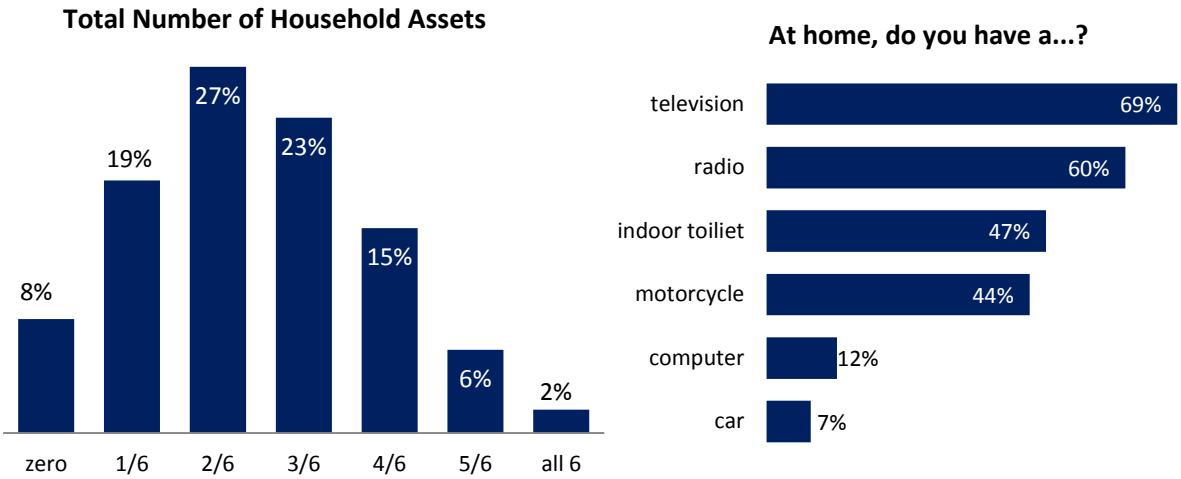
Socio-Economic Factors

Reports of common higher priced household items are commonly used as a proxy for household income as well as overall socio-economic status. On average, students said that their families had between two or three household possessions out of the six possessions listed in the survey. A

[6] Note respondents were allowed to report multiple responses, so the table below does not add up to 100%.

television, a radio, an indoor toilet and a motorcycle were the most commonly reported household possessions, with nearly half or more of participating students reporting having these items in their household.

Figure 4. Household Assets



Provincial comparisons of socio-economic factors show some variation across provinces. Students in Ilocos Norte and Ilocos Sur reported having the most household items with the majority of students reporting having between three to four of the six household possessions while the other provinces ranged between two and three.

Home Environment

Nearly two-thirds of surveyed students reported that their mothers were not engaged in formal employment, and 30% of mothers make a living through menial labor or in the informal economy. Three quarters of fathers were reportedly engaged in menial labor or the informal economy but only 10.4% were reported as unemployed. These results must be interpreted with caution since it is likely that children are not always aware of the occupation of their parents.

Table 5. Where do your parents work?

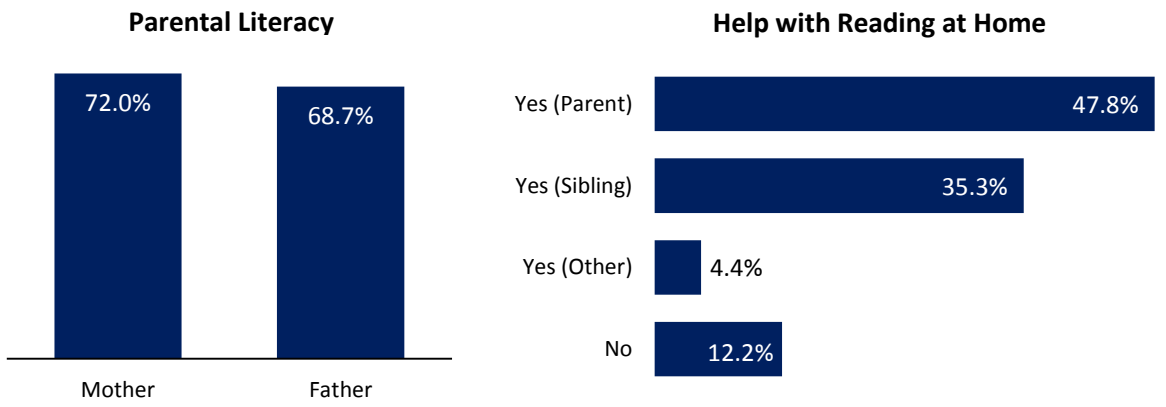
Parental Occupation	Mother	Father
Overseas Foreign Worker	2.2%	2.1%
Professional	3.2%	3.4%
Informal/Menial/Self	30.0%	73.3%
Unemployed	61.1%	10.4%
Do Not Know / No Response	3.5%	10.9%

Parental involvement is a key predictor in the early literacy success as well as future academic achievement of children. As such, the student context interview also aimed to find whether students receive any help with reading at home. The majority of students reported that both their parents were literate though slightly more mothers than fathers were reported as literate. When analyzed by province, Cebu and La Union reported lower literacy rates among parents with roughly half of students reporting that their parents were literate. In the other provinces, parental literacy

was nearly universal with more the 90% of students reporting that their mother and father could read.

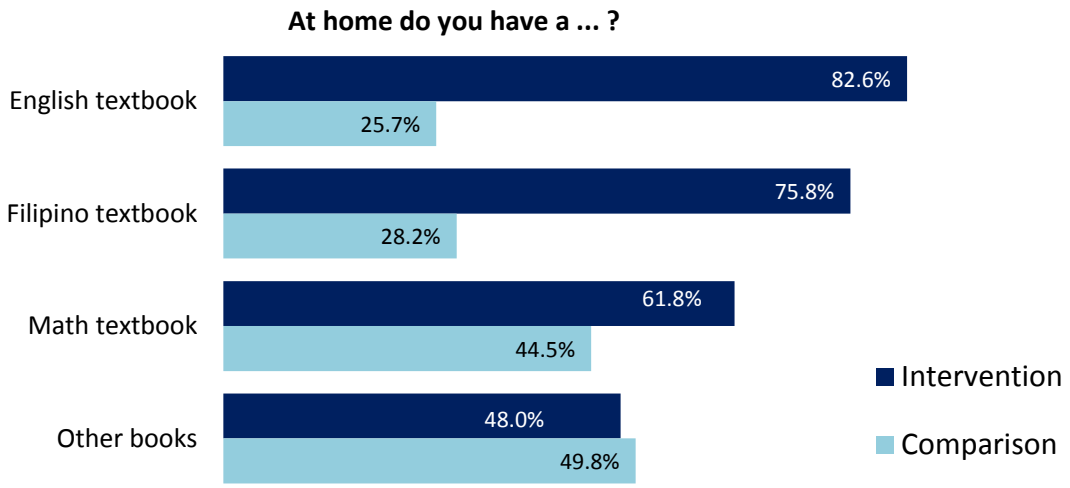
The majority of students said they receive help at home with reading, either from a parent or from a sibling. Just over 10% of surveyed students said they do not receive help at home with reading. No statistically significant differences in help with reading at home was found between the comparison and intervention cohort students.

Figure 5. Parental Literacy and Help with Reading at Home



Students were also asked the types of books[7] they have at home. Intervention cohort students largely reported having books at home. The most common books to have at home for intervention cohort students were in English, followed by Filipino. Math and other books were slightly less common to have at home. The majority of the comparison cohort students reported not having books at home; only a quarter of students had English or Filipino books at home.

Figure 6. Availability of Books at Home



One possible factor for the difference is that 2013 was the first year of Grade 2 implementation under the new K to 12 reform, and the textbooks and materials were not distributed to the schools until the end of the school year.

[7] This student context interview question is carried over from the instrument developed by RTI.

SUMMARY EGRA FINDINGS

The EGRA study presented in this report produced a considerable amount of data. To condense the data into a digestible report on the current status of reading skills in Filipino and English in five provinces of the Philippines, this section will present the summary findings from each subtest for both languages assessed for students sampled from both the intervention and comparison cohorts. It should be noted that comparison cohort students were not tested in English. As a result, comparisons across cohorts cannot be made for English EGRA.

Overall Filipino EGRA Results

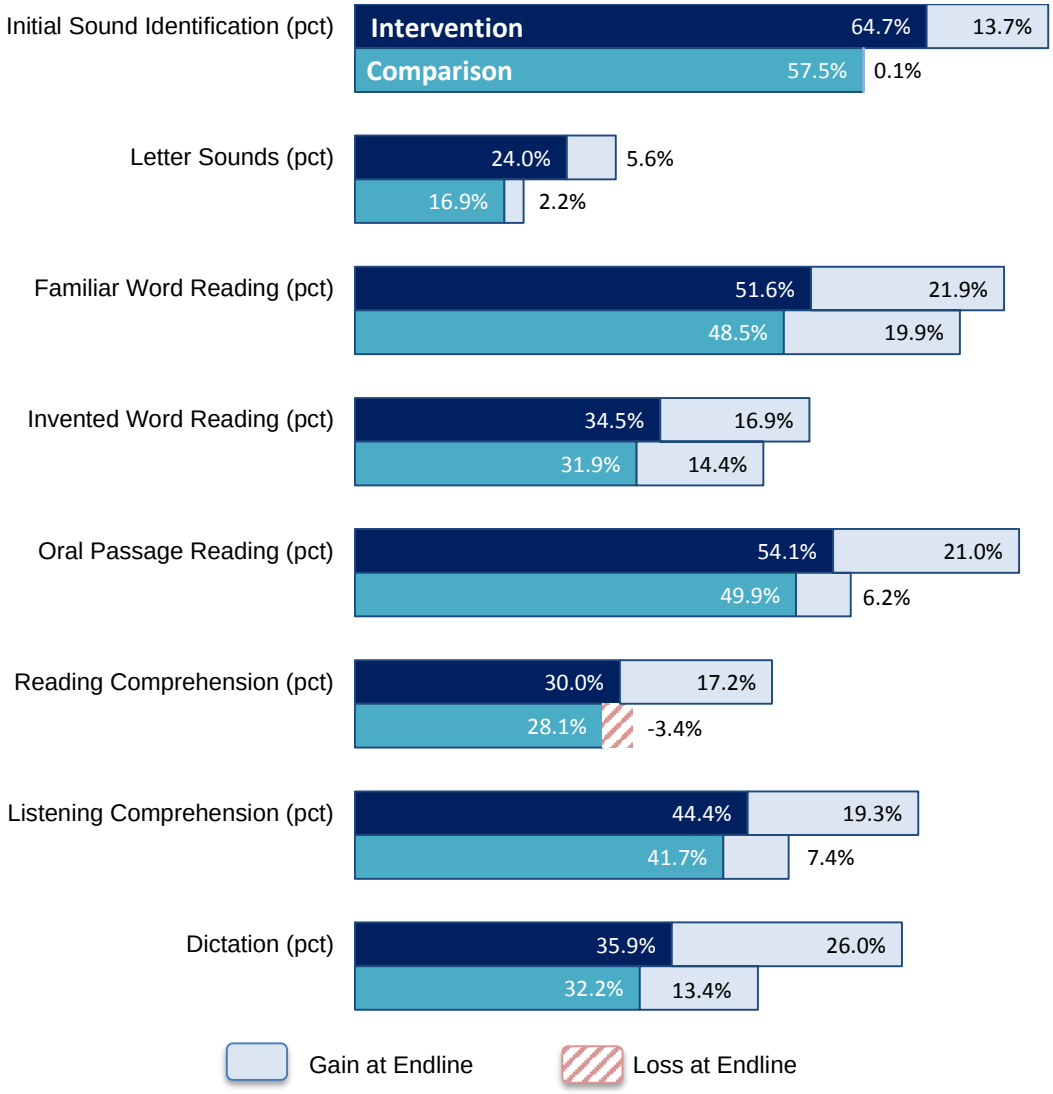
The analysis of data showed that students tended to perform best on the initial sound identification, reading familiar words, and oral reading passage subtests. The subtests with lowest scores were letter sounds for Filipino EGRA and listening comprehension and dictation on the English EGRA.

Table 6. Filipino EGRA Results, by Subtest and Cohort

	Intervention			Comparison		
Filipino EGRA Subtests	**Baseline**	**Endline**	**Gain**	**Baseline**	**Endline**	**Gain**
Initial Sound Identification (pct correct)	64.7%	78.4%	13.7%	57.5%	57.6%	0.1%
Letter Sounds(percent correct)	24.0%	29.5%	5.6%	16.9%	19.1%	2.2%
Letter Correct (per min)	24.0	29.5	5.5	17.1	19.2	2.1
Familiar Words(pct correct)	51.6%	73.4%	21.9%	48.5%	68.4%	19.9%
Familiar Words Correct(per min)	26.9	41.9	15	25.0	37.5	12.5
Nonsense Words (pct correct)	34.5%	51.4%	16.9%	31.9%	46.2%	14.4%
Nonsense Words Correct (per min)	17.3	26.0	8.7	16.1	23.2	7.1
Oral Passage Reading (pct correct)	54.1%	75.1%	21.0%	49.9%	56.1%	6.2%
Words Correct in a Text (per min)	31.3	48.1	16.8	29.3	37.0	7.7
Prosody score	2.5	3.0	0.5	1.6	2.1	0.5
Reading Comprehension (pct correct)	30.0%	47.2%	22.9%	31.4%	28.1%	-3.4%
Listening Comprehension (pct correct)	44.4%	63.7%	19.3%	41.7%	49.1%	7.4%
Dictation Composite (pct correct)	35.9%	61.9%	26.0%	32.2%	45.6%	13.4%

Intervention cohort students performed better than the comparison cohort students in all Filipino subtests. Figure 7 shows visually the differences in gains between the comparison cohort students and intervention cohort students.

Figure 7. Average Percent of Correct Answers for Filipino EGRA Subtests in All Provinces, by group



Effect Size

The difference in gains between intervention and comparison cohort students was statistically significant at $p<.001$ level for all subtests. The effect size[8] ranged between .17 for the familiar word reading subtest to .99 for the prosody subtest, with seven out of thirteen measures showing the effect size of .6 or higher. The effect size calculations are reported in Table 7.

[8] Effect size is a statistical measure that is used to estimate the magnitude of difference between two measures. It is computed by dividing the differences between the means of the two groups by the pooled standard deviation. Effect sizes are interpreted as follows, according to Cohen (1998): "small, d = .2," "medium, d = .5," and "large, d = .8". (reference: Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.)

Since the comparison of means at the baseline showed that the tested cohorts were not perfectly identical (the intervention cohort students scored higher on all but one subtests), a statistical correction is required to assess the true intervention effect[9]. A regression analysis was conducted to estimate the amount of gains *after* the baseline differences were incorporated into the analyses and controlled for.

Basa intervention helped increase the fluency of 2nd grade students by additional 9.1 words per minute over the expected scores.

The results of the analysis are presented in Table 7. The amounts of gains due to the Basa intervention are shown in the last column as the "intervention effect". As the table shows, the Basa intervention helped increase the fluency of second grade students by additional nine words per minute, above and beyond the natural increases expected to occur over the course of the school year. Additional gains in dictation as well as listening and reading comprehension were also found to be substantial. The smallest additional gains were found in letter sounds subtest, and reading of familiar and nonsense word subtests.

Table 7. Basa Intervention Effect on Filipino EGRA Results, by Subtest, Controlling for the Baseline

Filipino EGRA Subtests	Baseline Difference of Means[10]*	Effect Size[11] of Cohort Difference at Baseline	Endline Difference of Means[12]*	Effect Size of Cohort Difference at Endline	Intervention Effect**
Initial Sound Identification (pct correct)	7.2	0.20	20.8	0.65	17.1%
Letter Sounds(percent correct)	7.1	0.47	10.4	0.64	5.6%
Letter Correct (per min)	6.8	0.45	10.2	0.63	5.5 lcpm
Familiar Words(pct correct)	3.0	0.09	5.1	0.17	2.7%
Familiar Words Correct(per min)	1.9	0.10	4.4	0.21	2.6 wcpm
Nonsense Words (pct correct)	2.7	0.11	5.2	0.20	2.9%
Nonsense Words Correct (per min)	1.2	0.10	2.7	0.20	1.5 wcpm
Oral Passage Reading (pct correct)	4.2	0.13	19.0	0.63	16.1%
Words Correct in a Text (per min)	2.0	0.10	11.1	0.45	9.1 wcpm
Prosody score	0.8	0.80	0.9	0.99	.50
Reading Comprehension (pct correct)	-1.4	-0.05	19.1	0.66	23.5%
Listening Comprehension (pct correct)	2.8	0.07	14.7	0.38	13.4%
Dictation Composite (pct correct)	3.7	0.15	16.3	0.61	13.7%

* All differences between the means are statistically significant at the $p<.001$ level.

** Intervention effect is computed using a linear regression model controlling for the baseline differences between the cohorts.

[9] *Procedures and Standards Handbook, v.3.0*. What Works Clearinghouse, Institute of Education Sciences, US Department of Education, 2015, p. 14.

[10] Computed as follows: intervention cohort mean at baseline (weighted) – comparison cohort mean at baseline (weighted).

[11] Cohen's d is calculated using the mean and SD of endline subtest scores for both the intervention and comparison group.

[12] Computed as follows: intervention cohort mean at baseline (weighted) – comparison cohort mean at baseline (weighted).

Overall English EGRA Results

An English EGRA assessment was conducted with a sample of students in Basa intervention schools during 2014/15, the first year the project provided the full intervention.. The assessment was not administered to comparison cohort students. As a result, the report provides results for the intervention cohort students only.

The analysis of baseline and endline English EGRA data showed higher achievement in some EGRA subtests than others. Overall, students tended to perform best on initial sound identification and oral reading passage subtests. The subtests with lowest scores at endline were reading comprehension, listening comprehension and dictation. The table below shows the results.

Table 8. English EGRA Results, Intervention Cohort Sample

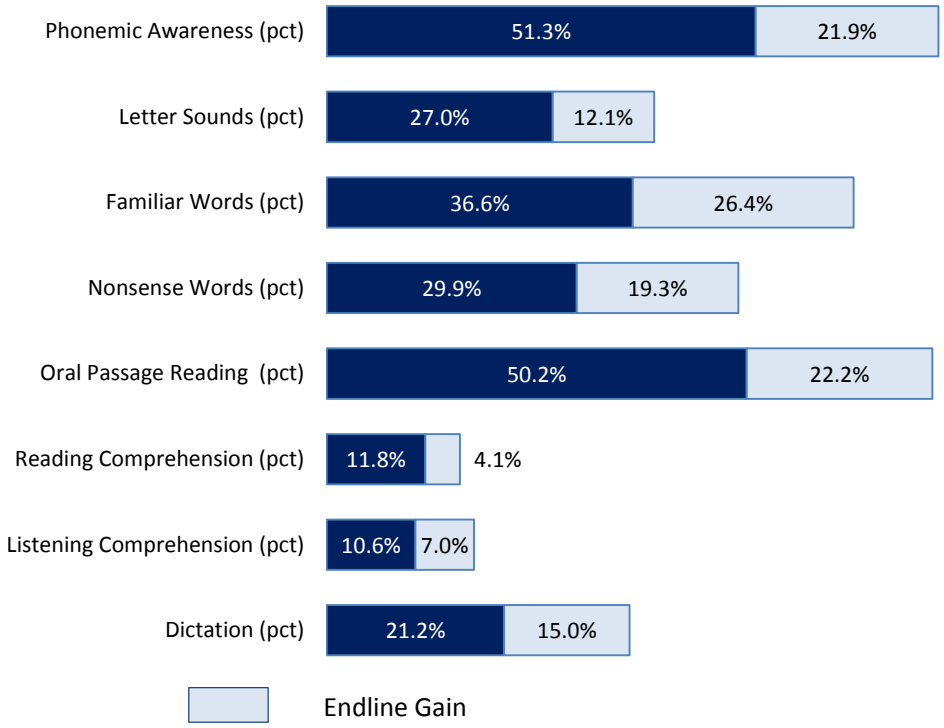
	Intervention		
English EGRA Subtests*	**Baseline**	**Endline**	**Gain****
Initial Sound Identification (percent correct)	51.3%	73.1%	21.9%
Letter Sounds(percent correct)	27.0%	39.1%	12.1%
Letter Correct (per min)	27.0	39.1	12.1
Familiar Words(percent correct)	36.6%	63.0%	26.4%
Familiar Words Correct(per min)	20.2	39.5	19.3
Nonsense Words (percent correct)	29.9%	49.2%	19.3%
Nonsense Words Correct (per min)	15.1	25.9	10.8
Oral Passage Reading (percent correct)	50.2%	72.4%	22.2%
Words Correct in a Text (per min)	34.9	58.1	23.2
Prosody score	2.2	2.8	0.6
Reading Comprehension (percent correct)	11.8%	15.9%	4.1%
Listening Comprehension (percent correct)	10.6%	17.6%	7.0%
Dictation Composite (percent correct)	21.2%	36.2%	15.0%

* The comparison cohort students were not tested in English.

** All gains are statistically significant at the p<.001 level

An analysis of changes in student scores from the beginning to the end of the school year show statistically significant gains (p<.001) for boys and girls, across all subtests. For most subtests the assessment found students performing better in Filipino than in English.

Figure 8. Average Percent Correct on English EGRA Subtests

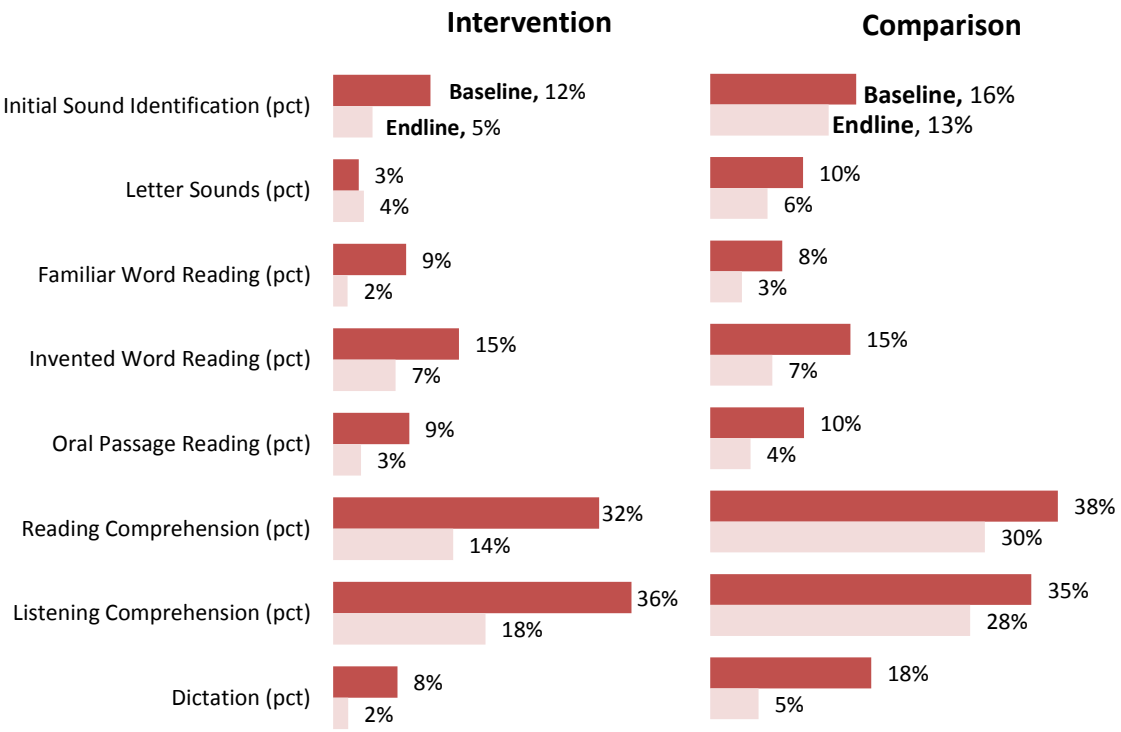


Zero Scores

The analysis of assessment results found a substantial proportion of tested students having zero scores on EGRA subtests at baseline which were reduced significantly at endline. On both the English and Filipino EGRA tests, the reading comprehension and listening comprehension subtests had the largest percentage of student with zero scores in both comparison and intervention cohorts at baseline. The assessment results showed a substantial decrease in zero scores for these subtests at endline, particularly for the intervention group. By the end of the school year in all but two subtests (Letter Sounds and Invented Words) the intervention cohort students had significantly fewer students with zero scores than the comparison cohort students ($p<.01$). This can be seen particularly in the reading comprehension and listening comprehension subtests where percent of students with zero scores in the intervention cohort students nearly halved from baseline to endline, while the comparison cohort students only saw 7-8% decreases in zero scores.

For every subtest, with the exception of letter sounds and oral passage reading, percentages of zero scores were much higher for the English test.

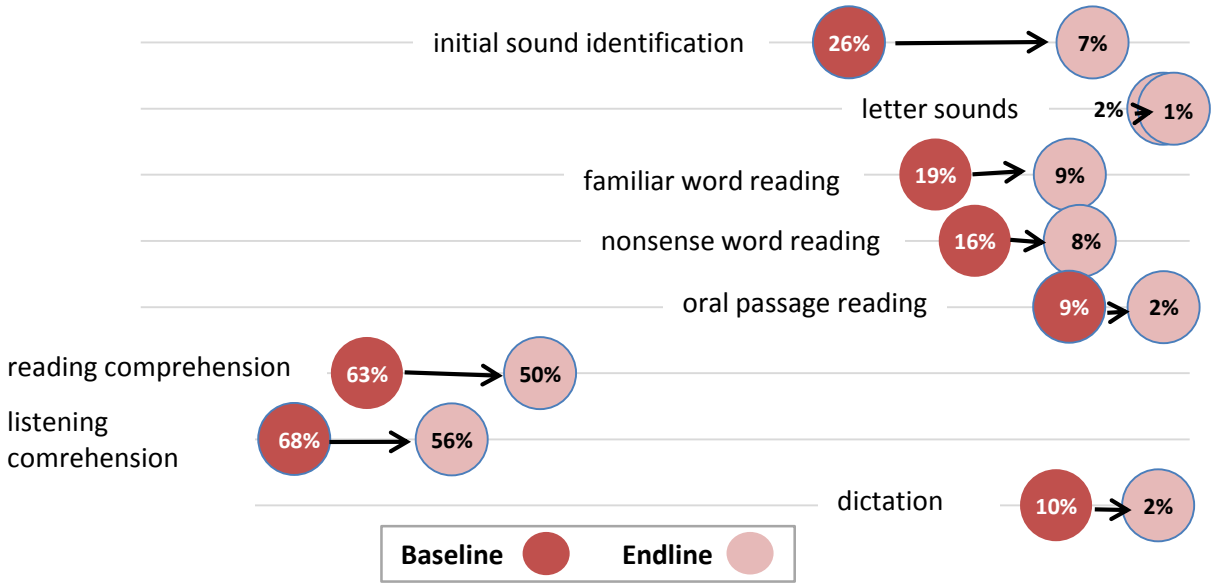
Figure 9. Change in Overall Zero Scores in Filipino EGRA Subtests, Across Baseline and Endline



The details of the statistical analyses are found in the Annex 5, which shows the proportion of students with zero scores on each subtest for each study cohort, at baseline and endline.

The assessment found a substantial proportion of tested students having zero scores on English EGRA subtests. As the graph below shows, the subtests with the largest percentages of students scoring zero was listening and reading comprehension. Zero scores significantly decreased from the beginning to the end of the school year for all subtests (p<.001). However, roughly half of tested students were still unable to answer a single reading or listening comprehension at endline.

Figure 10.Percent of Tested Students Scoring Zero on English EGRA Subtests

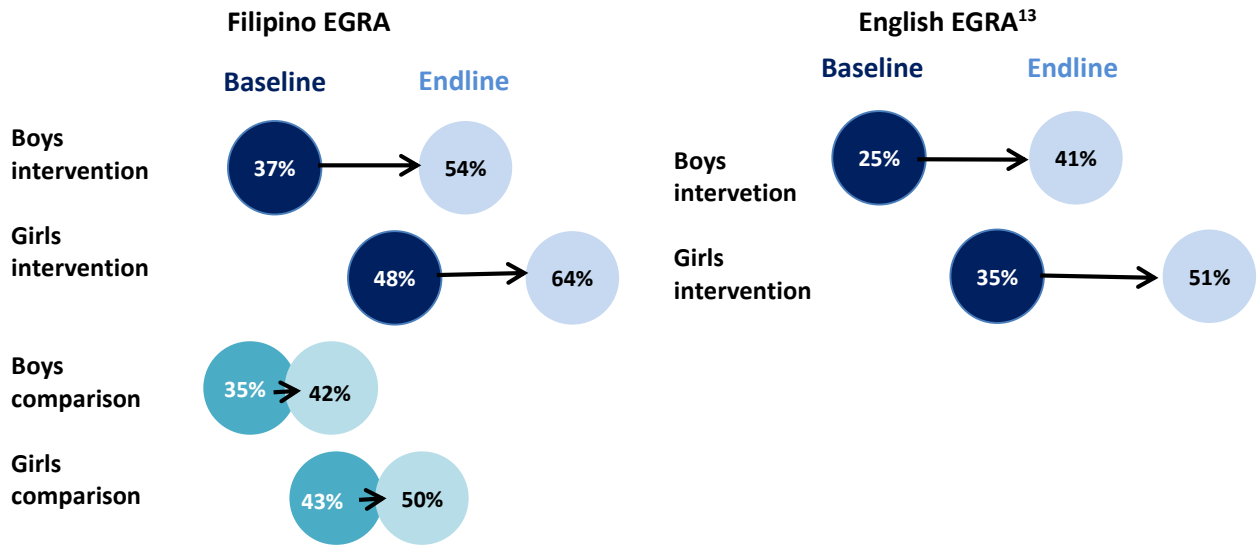


Results by Sex

Data analysis found that girls, on average, demonstrate far better EGRA results than boys. This pattern was consistent from baseline to endline and across group. Girls, on average, outperform boys by about five to fifteen percentage points (see Table A-18 for sex-disaggregated Filipino EGRA results, and Table A-22 for sex-disaggregated English EGRA results in Annex 5). The difference is statistically significant at $p<.001$ level for all subtests.

This finding is consistent with BIPI data collected and reported at the beginning of the 2014/2015 school year (summarized in the Annex 2 of the present report). The BIPI survey of teachers found that the majority of teachers thought it is harder for boys to learn to read than for girls. Further study is needed to find out whether teacher gender bias in literacy expectations may impact boys' reading performance.

Figure 11. Average Filipino and English EGRA Subtest Percent Correct, by Sex



As seen in the figure above, the average percent correct across subtests for both Filipino and English by sex. As seen in the figure, the intervention cohort boys and girls gained, on average, 17 and 16 percentage points on the Filipino EGRA over the course of the school year, respectively, while the comparison cohort boys and girls only gained 7 and 8 percentage points over the course of the

[13] Students in the comparison cohort were not tested in English EGRA.

school year. For English EGRA, the intervention boys and girls gained 15 and 17 percentage points over the course of the school year, respectively.

To establish the comparability of the intervention and comparison cohort samples, the comparison of baseline means was conducted. It showed that the intervention cohort students, particularly girls, had higher average Filipino EGRA scores. To estimate the intervention effect while taking into consideration this difference in baseline scores, a regression analysis was conducted. The results are showed in the table below. The amounts of gains among boys and girls due to the Basa intervention are shown in the last column as the "intervention effect".

Table 9. Basa Intervention Effect on Average Filipino EGRA Results, by Sex

GENDER	Intervention Cohort		Comparison Cohort		Baseline Difference of Cohort Means[14]*	Effect Size of Cohort Difference at Baseline	Endline Difference of Cohort Means[15]*	Effect Size of Cohort Difference at Endline	Intervention Effect**
	B	E	B	E					
Boys	37.1%	54.3%	34.9%	42.2%	2.21%	0.10	12.07%	0.56	10.28%
Girls	47.9%	63.9%	42.6%	50.4%	5.26%	0.24	13.54%	0.66	9.48%

B = baseline; E = endline

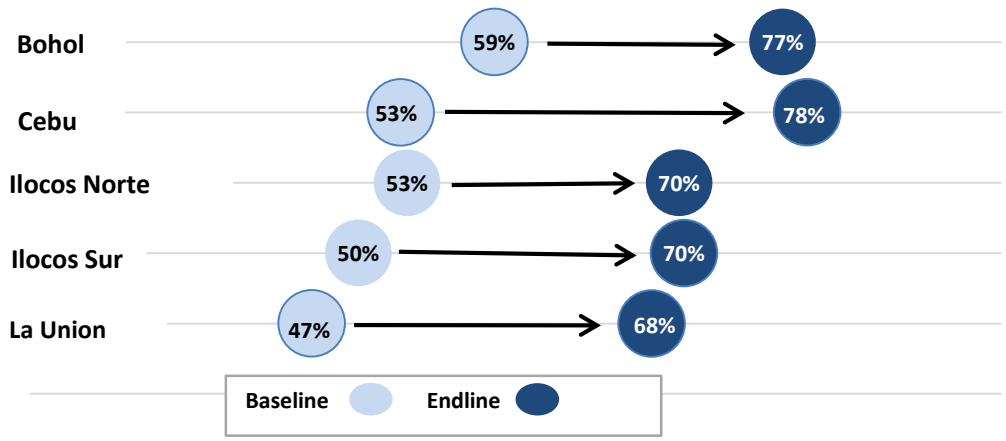
* All differences between the means are statistically significant at the $p<.001$ level.

** Intervention effect is computed using a linear regression model controlling for the baseline differences between the cohorts.

Results by Division

Comparisons of the results by division are problematic due to the sample size that might not be large enough at a division level to guarantee accuracy. While complete results of data analyses by division are provided in Annex 5, a graph below shows the results of oral reading passage subtest, as an illustration of division-level results. The graph shows the average percent of the text read correctly by students in each division at the beginning and at the end of the school year.

Figure 12.Percent of Filipino Text Read Correctly at the Beginning and the End of the School Year



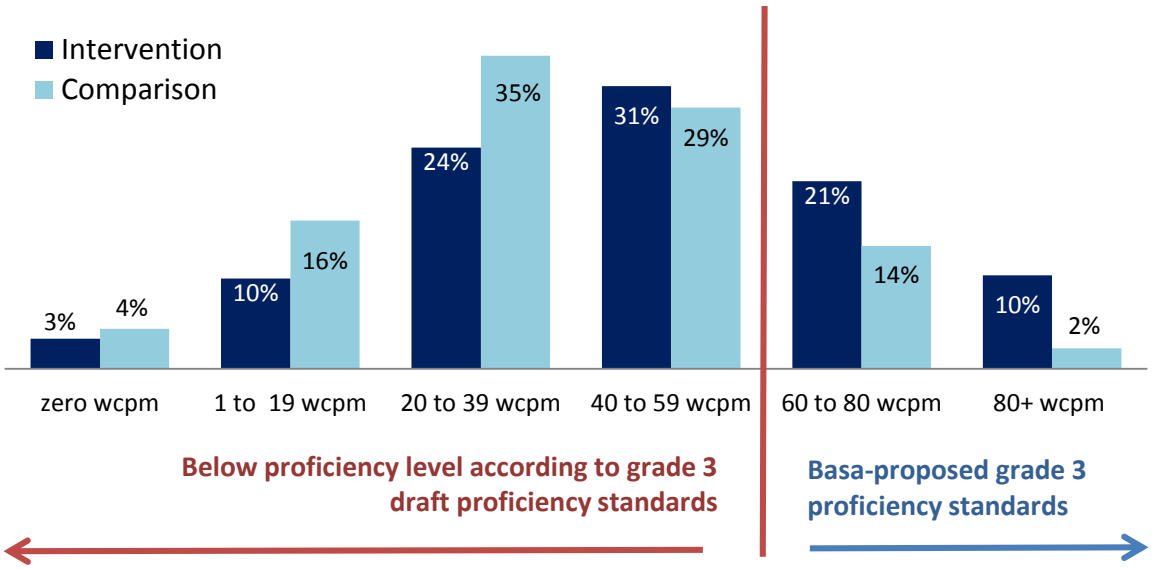
[14] Computed as follows: intervention cohort mean at baseline (weighted) – comparison cohort mean at baseline (weighted).

[15] Computed as follows: intervention cohort mean at endline (weighted) – comparison cohort mean at endline (weighted).

Benchmarks

Basa will be working with DepED to set fluency and comprehension benchmarks for grade 2, and finalize proposed a grade 3 Filipino fluency benchmark of 60 words correct per minute. The results of the assessments show that, as expected, at the end of a school year the majority of grade 2 students from both cohorts were reading below the Basa-proposed grade 3 proficiency standard of 60 wcpm[16] (Figure 10). Nonetheless grade 2 students, particularly those in the intervention cohort, appear to be on track to attain the grade 3 proposed benchmarks. On average, at the end of the school year intervention cohort students read 48 words correct per minute, compared to 37 words correct per minute for comparison cohort students. Twenty-one percent of second graders in the intervention cohort have already reached the grade 3 benchmark of 60 wcpm, as compared to 14% of students in the comparison cohort. Very few students read 80 words correct per minute or more, with 10% of intervention and 2% of comparison students reading at this level.

Figure 13. Filipino Oral Passage Reading Subtest Results at Endline, by Group

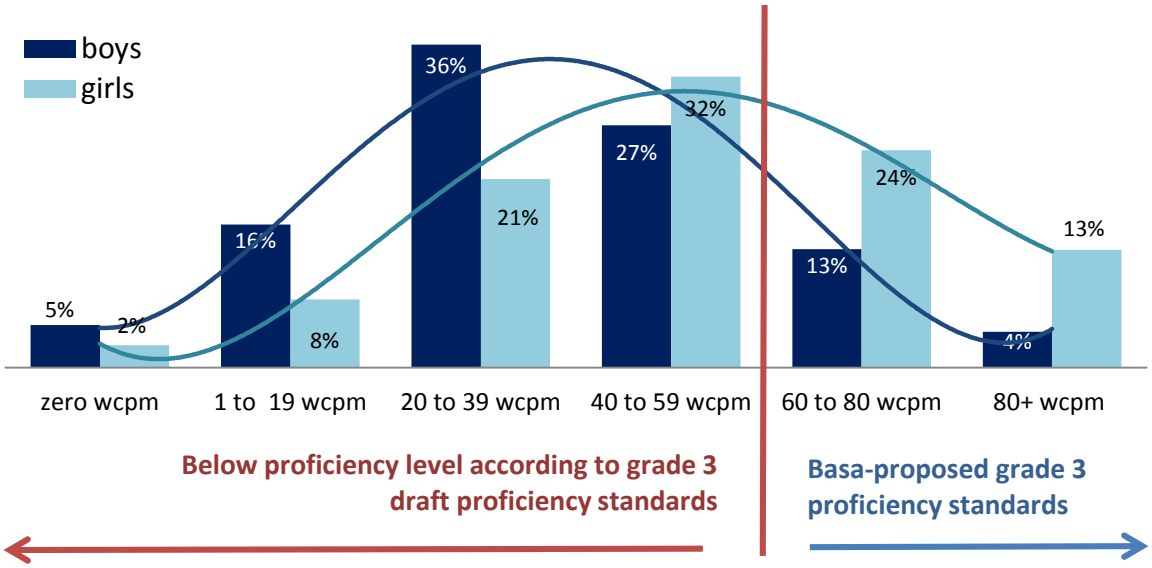


Girls on average demonstrate better results than boys both in the beginning and in the end of a school year, in both cohorts. The figure below shows the average percent correct for EGRA subtests for boys and girls at baseline and endline. Endline Filipino EGRA results show that girls outperform boys on all subtests. The difference between boys and girls is statistically significant at the $p<.001$ level for all subtests except for the listening comprehension subtest for the comparison cohort (*See Annex 5 for detailed results by subtest*). Across seven subtests, girls sampled from the intervention cohort demonstrated, on average, 10% more correct answers than boys, while girls sampled from the comparison cohort

[16] Basa is working with the Ministry to establish grade 2 proficiency standards.

averaged 8% more correct answers than boys. Within their cohort (intervention and comparison), boys and girls showed a similar level of gains between the baseline and endline. The gains were statistically significant for all subtests. As can be seen in Figure 14, twice as many intervention cohort girls meet the benchmark at the end of the second grade (35% of girls versus 17% of boys).

Figure 14. Intervention Cohort Filipino Oral Passage Reading Subtest Results at Endline, by Sex



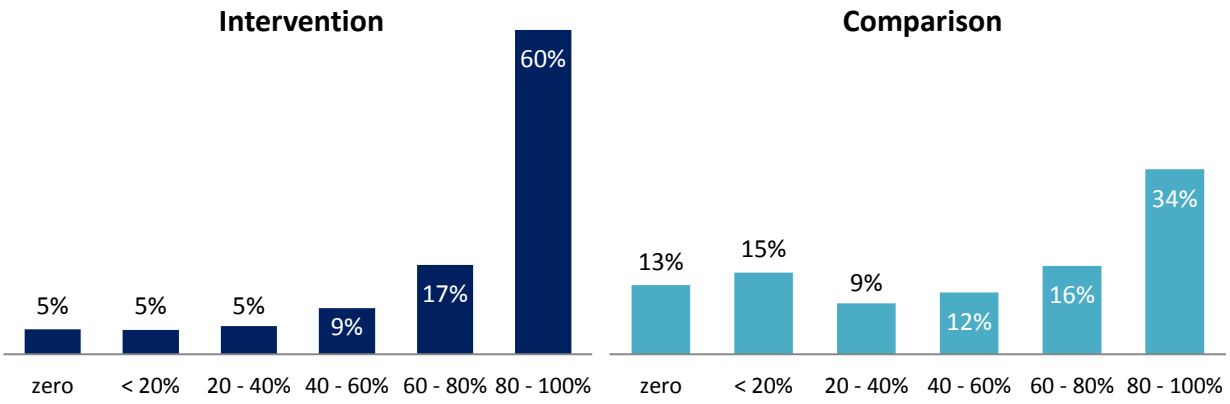
To help better understand the patterns of student achievement by subtest, the next sections of the report present results for each subtest, as well as disaggregations by sex.

FILIPINO EGRA SUBTEST RESULTS

Initial Sound Identification, Letter Sounds, Familiar and Nonsense Words

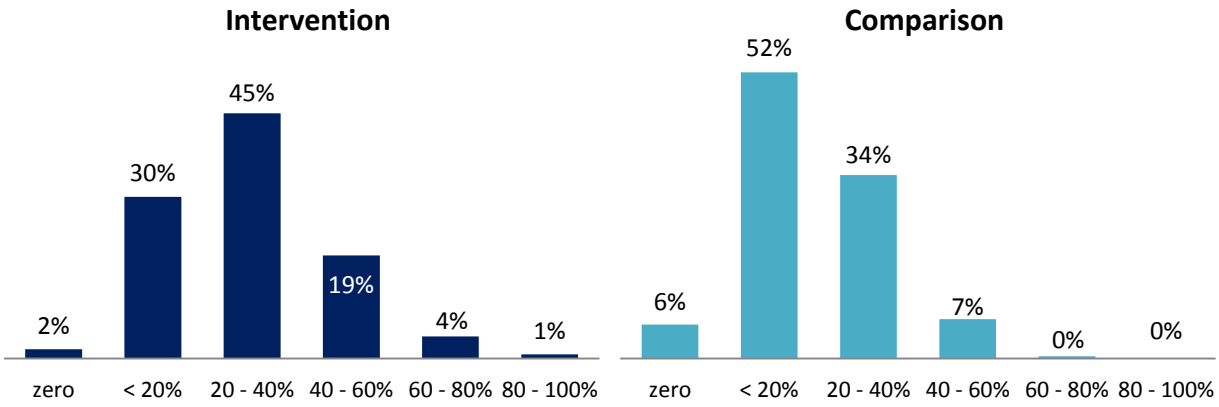
On the **initial sound identification** subtest (total possible ten letters) that measures phonemic awareness of students, responses ranged between zero to ten sounds correct, with a mean of 7.8 initial sounds at endline for the intervention cohort students, and 5.8 initial sounds for the comparison cohort students. The graph below shows differences in the distribution of student's initial sound identification across cohorts. For both cohorts, the Filipino EGRA initial sound identification results skews to the left with nearly three quarters of students in the intervention cohort scoring 60% or higher and roughly half of students in the comparison cohort students. Only 5% of students in the intervention cohort and 13% of students in the comparison cohort scored zero. Comparisons by sex showed that girls and boys in the comparison cohort had very similar distributions. In the intervention cohort, a more distinguished difference between boys and girls was seen, with 68% of girls scoring between 80 and 100%, while less than half (47.1%) of boys scored similarly highly.

Figure 15. Filipino Initial Sound Identification – Endline Percent Correct, by Group



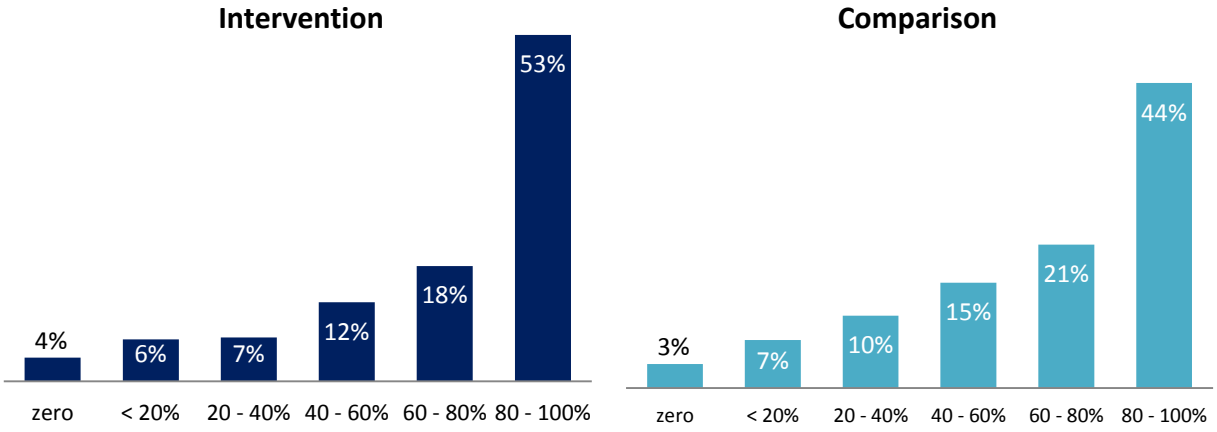
On the **letter sounds subtest** (total 100 letters) at the end of a school year, number of letters correctly pronounced by students ranged from zero to 100 letters (mean 29.5) in the intervention cohort and from zero to 84 letters (mean 19.1) in the comparison cohort. As the graphs below show, the majority of the students in both groups correctly named fewer than 40% of the letter sounds with only 1% of students from the entire sample able to identify between 80 to 100% of letter sounds. Students were timed on the responses. The number of sounds pronounced correctly was divided by the seconds it took to answer and then multiplied by 60 seconds to find the correct letter sounds per minute. The results suggest that students may not be used to being asked to name letter sounds in an assessment context. Distributions by sex were similar for boys and girls in both groups with girls naming slightly more letter sounds than boys.

Figure 16. Filipino Correct Letter Sound Subtest Endline Results, by Group



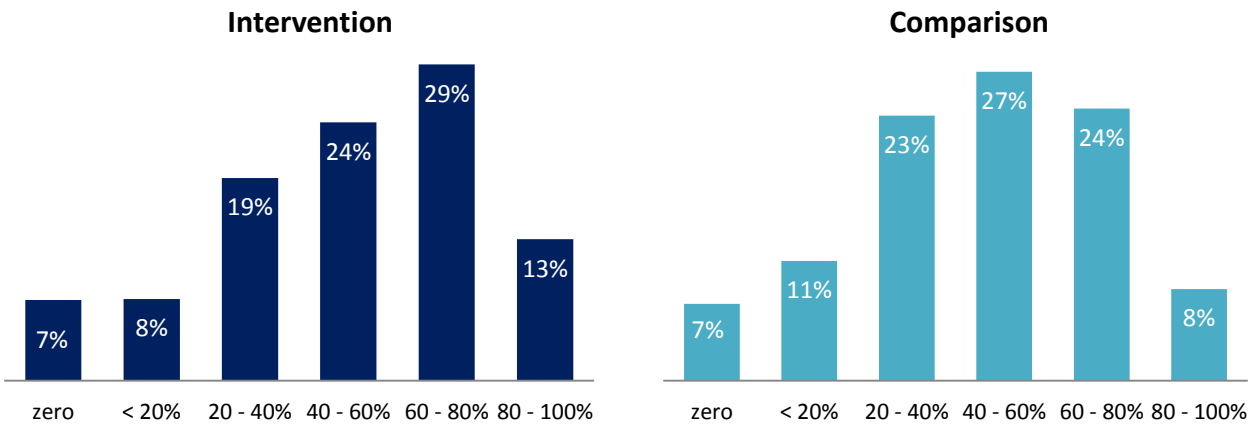
On the **familiar word identification** (total possible 50 words), students were timed on the responses. Responses ranged from zero to 50 familiar words read correctly, with a mean of 36.7 (73.4%) for the intervention cohort students and a mean of 34.2 (68.4%) for the comparison cohort students. The graph below shows that the distribution of Filipino EGRA scores is skewed to the left, with more than half of tested students reading between 60 and 100% of the words correctly. Analysis by sex showed that girls in both cohorts outperform boys.

Figure 17. Filipino Familiar Word Identification Endline Subtest Results, by Group



On the **simple nonsense word (non-word) decoding** (total possible 50 words), students were asked to read invented words in order to measure their ability to decode words rather than recognize them. Therefore, it is expected that this test is more difficult than oral passage reading or familiar word reading. Students were timed on the responses. The number of correct responses ranged from zero to 50 nonsense words, with a mean of 25.7 (51.4%) at endline for students in the intervention cohort and a mean of 23.1 (46.2%) for students in the comparison cohort. The distribution is close to normal for the Filipino test for both groups, with nearly a half of students reading correctly between 20 and 60%. Girls generally outperformed boys in both cohorts. The distribution of girls' scores in this subtest is also fairly normal for both tests. Conversely, for boys, the distribution of scores is skewed to the left, showing a large percentage of boys reading between zero and 40% of nonsense words correctly.

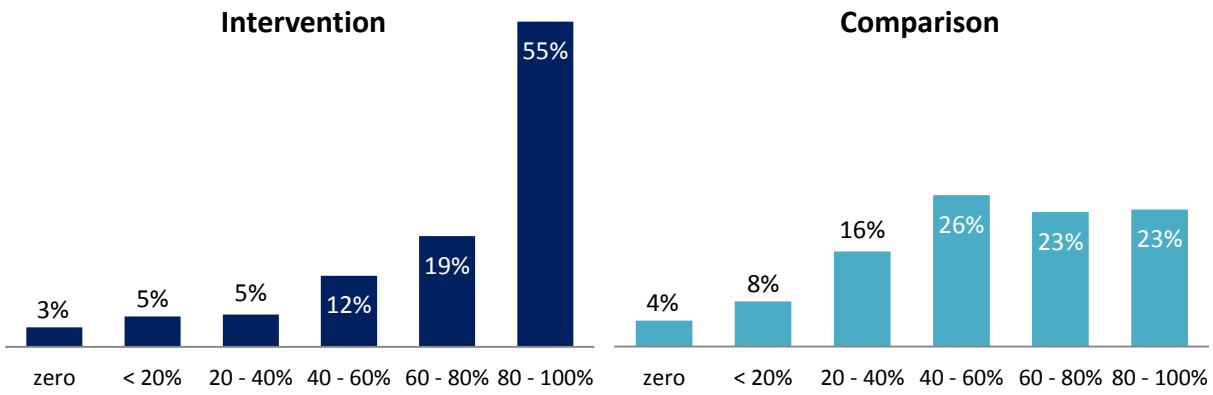
Figure 18. Filipino Nonsense Words Endline Subtest Results, by Group



Oral Reading Fluency and Comprehension in Filipino

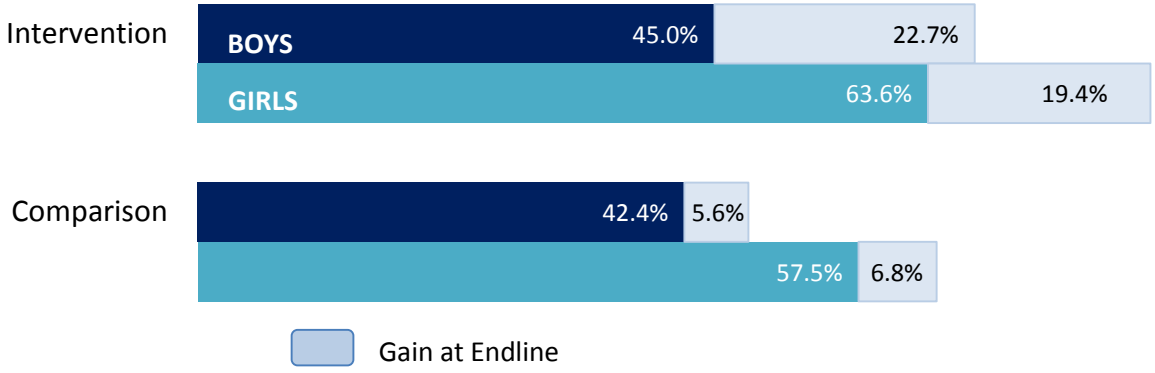
On the **passage reading and comprehension** subtest, students were scored on the number of words they read correctly in the passage (total possible 55 words for intervention cohort students and 64 words for the comparison cohort students), reading comprehension (total possible 5), and prosody (total possible 4). The number of words read correctly in Filipino by intervention students ranged from zero to 55 a mean of 41.33 words (75.1%); for the comparison cohort students, results ranged from zero to 64 with a mean of 35.9 (56.1%). The results presented in the graph below show for both groups results were largely skewed to the left, with the majority of students reading more than 40% of the text. Analysis by group shows that the intervention cohort students largely outperformed the comparison cohort students, with more than half of students in the intervention cohort students able to read 80-100% of the text at endline. Only a quarter of students in the comparison cohort showed such results.

Figure 19. Filipino Oral Passage Reading Endline Results, Percent of Words Read Correctly, by Group



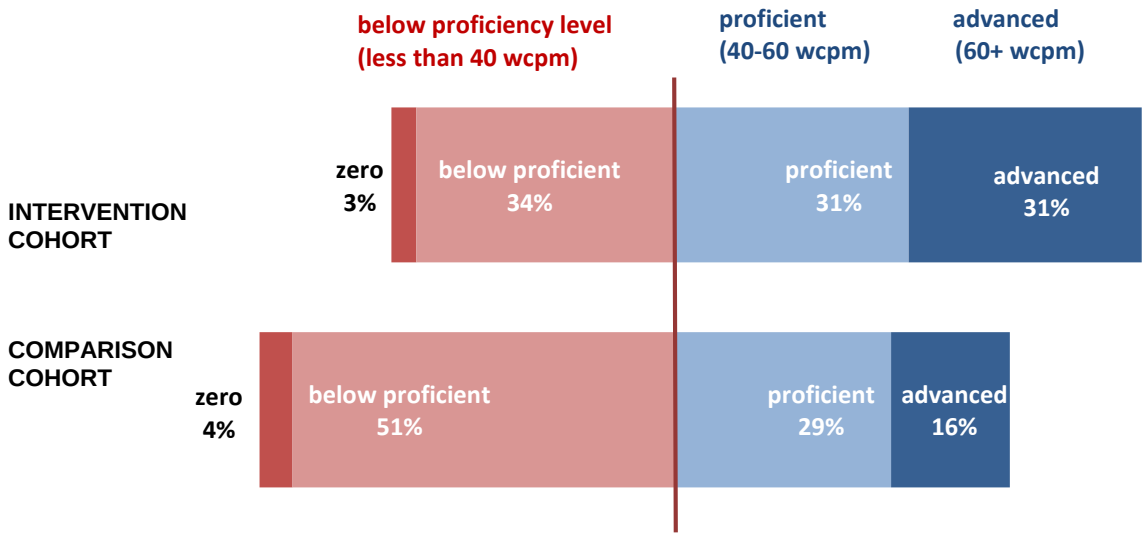
Similar to other subtests, girls in both groups significantly (p<.001) outperformed boys in the oral reading subtest. At the end of a school year, girls in the intervention cohort were able to read 83.0% of the text compared to boys who read 67.6% of the text. Similarly, in the comparison cohort, girls read 64.3% of the text at endline compared to 48.0% for boys.

Figure 20. Filipino Oral Passage Reading Average Percent Correct, by Group and Sex



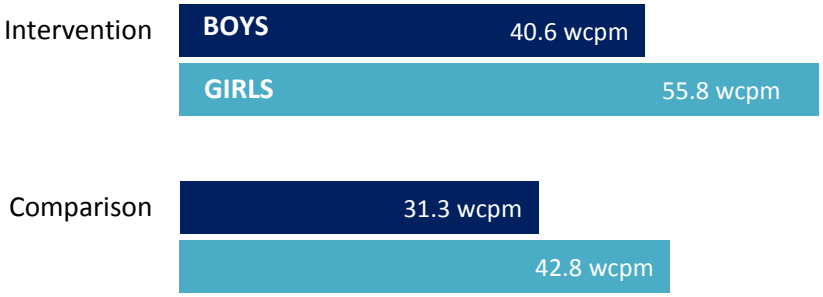
Words correctly read per minute. Students were timed on reading the text, with the limit of 60 seconds. The number of words read correctly was divided by the seconds it took to read and then multiplied by 60 seconds to find the number of correct words per minute, which is the standard fluency measure used to measure USAID reading interventions. Intervention cohort students on average read with the speed of 48.1 words correct per minute, compared to 37.0 words correct per minute for the comparison cohort students. The graphs below show the distribution of fluency scores by group, at baseline and at endline. As expected, the majority of students in both groups fell below the grade 3 proposed proficiency level of 60 wcmp, between 1 to 59 wcpm, at both baseline and endline. However, as can be seen in the graphs below, there is a shift in both groups toward grade 3 proficient levels from baseline to endline. For instance, in the intervention cohort students at baseline only 8% of students were proficient or above (60+ wcpm). By endline, 31% of students were reading at a proficient grade 3 level or above. As seen below, larger improvements were found among the intervention cohort students, with 31% of students who were proficient or above at endline, compared to 16% for students in the comparison cohort.

Figure 21. Filipino Oral Reading Fluency (wcpm), by Group, at Endline



There were significant differences (p<.001) in fluency rates between boys and girls. The range in girls' scores was wider than boys. On average, girls read at a faster rate than boys in both cohorts.

Figure 22. Filipino Fluency Measure (wcpm) at Endline, by Sex

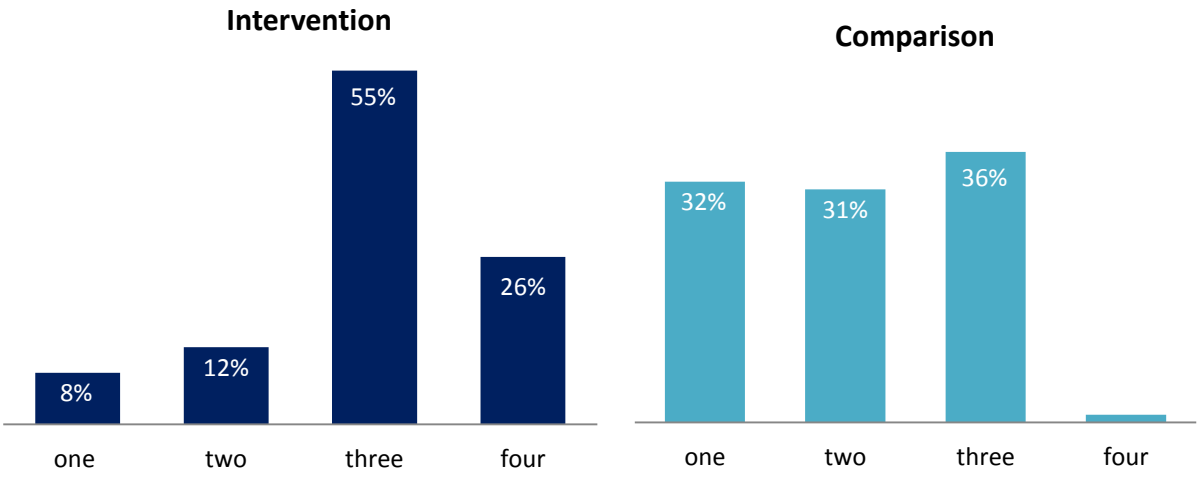


The **prosody scores** measured the degree to which students were able to read with intonation and expression. When students have difficulty reading, their prosody will often be slow and arduous whereas those who have ease in reading can focus on adding expression and tone to their reading. Prosody is just as linked to comprehension as automaticity because reading with good prosody resembles speech, which is embedded with underlying clues about the meaning of a text (Rasinki, 2011). Prosody measures were divided into four categories and scored accordingly:

1 = word-by-word, slow, laborious;
2 = small chunks, awkward;
3 = fluent, but does not mark punctuation; incorrect phrase groups, no expression;
4 = fluent, with expression to mark punctuation and/or direct speech.

Students reading the Filipino passage ranged from one to four, with a mean of 3.0 in the intervention cohort students at endline, and 2.1 among the comparison cohort students. The graphs below show the distribution in students' prosody on both tests. The distribution shows that in the intervention cohort more than three-quarters of students read fluently (scored 3 or 4); only 8% read laboriously. For the comparison cohort students, about a third of students read the passage at a 3, fluently without expression; nearly a third of students read laboriously.

Figure 23. Filipino Oral Passage Reading - Endline Prosody Score, by Cohort



Disaggregation by sex showed that girls from both groups read the Filipino passage with greater prosody than boys, which is consistent with the overall pattern of girls demonstrating higher reading proficiency.

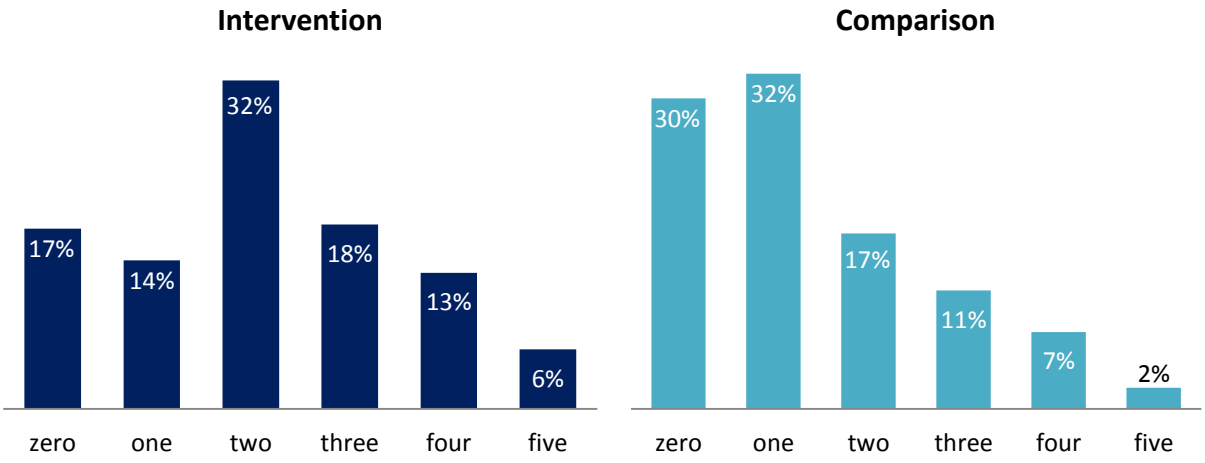
Reading Comprehension. After reading each passage, students were asked five questions about the text they read to check for comprehension. They were not allowed to look back at the text to help them answer questions. At the beginning of a school year, students in the intervention cohort answered on average 1.5 (30.0%) of reading comprehension questions, while the comparison cohort students answered 1.6 (31.5%) reading comprehension questions correctly.

At the end of a school year, the total number correct ranged from zero to five, with a mean of 2.1 (47.2%) for the intervention cohort students, and 1.4 (28.1%) for the comparison cohort students. The intervention cohort students saw an improvement in reading comprehension; however, conversely, the comparison cohort students saw a slight decrease in reading comprehension scores from baseline to endline. One possible explanation for why reading comprehension score was lower at endline for the comparison cohort students had to do with the type of comprehension questions asked. At baseline, of five comprehension questions four were locators and one was inferential, while at endline only two questions that the comparison cohort students were asked were locators and three were inferential. It is usually easier for students to answer locator questions than inferential questions.[17]

The figures below show the distribution of endline reading comprehension results by cohort. Students in the intervention cohort were able to answer more reading comprehension questions at endline correctly than students in the comparison cohort. Intervention scores approximate a normal curve, with the majority of students answering two questions correctly. Approximately 17% of students were unable to answer a single reading comprehension question. Only 19% were able to answer 4 or 5 questions correctly. As for the comparison cohort students, nearly two-thirds of students being unable to answer more than one or less reading comprehension questions correctly.

[17] The EGRA tests used for this cohort were not developed by EDC, but by another project, as described in the methodology section. It appears that that the two versions of the readings were not fully equated by means or linear methods prior to implementation. For the second round of testing (cohort 2) tests have been fully tested and equated by EDC using means equating.

Figure 24. Reading Comprehension Endline Results, Filipino EGRA, by Group

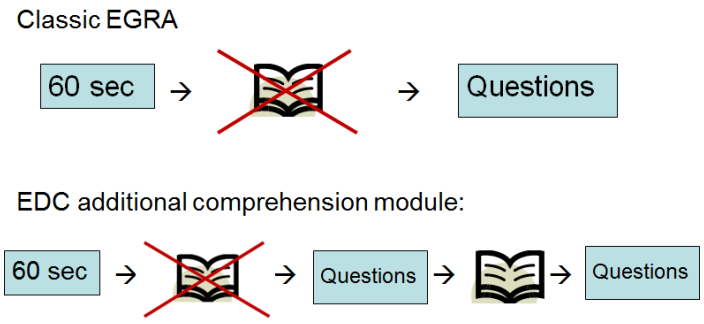


Comparisons by sex show girls significantly (p<.001) outperforming boys in reading comprehension at the end of a school year. This is evident in the higher proportion of zero scores for boys in comparison to girls in both cohorts. Detailed results by sex are found in Annex 5.

In addition to detailed results of statistical analysis by subtest and by sex, the annexes contain correlations between subtexts (Annex 1).

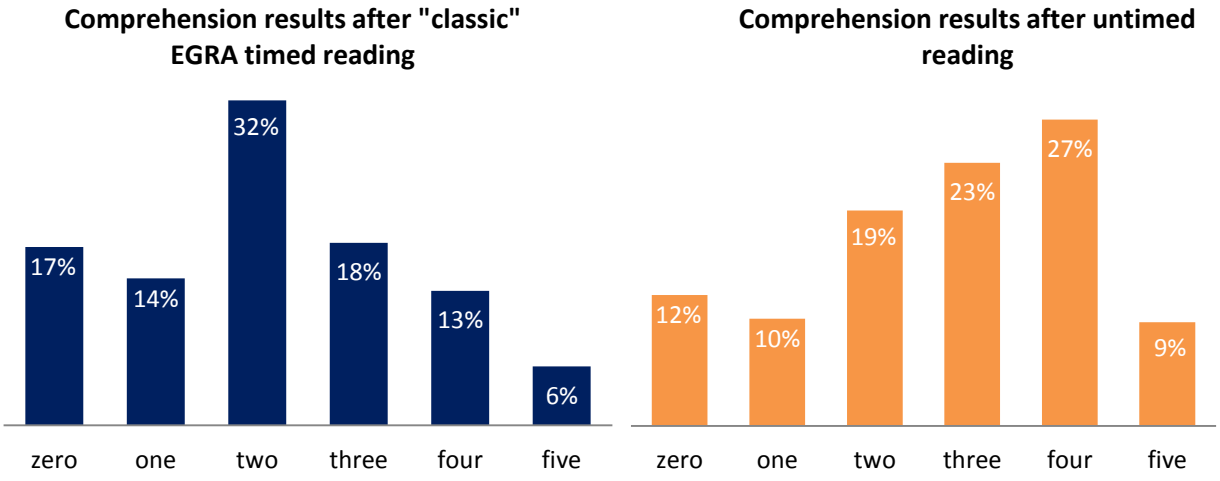
An examination of the testing process suggested to the Basa evaluation team that low comprehension results might be due to the test bias rather than to actual lack of comprehension abilities among students. To test this hypothesis, during the endline testing the evaluation team administered the Filipino reading comprehension questions in two rounds. The first round the administration followed the standard EGRA administration procedures. The second round immediately followed the first round. During the second round, assessors gave the text back to the students if allowed them to finish reading the passage (if they haven't done so already), and then asked them comprehension questions without taking the text away from the students. The figure below compares the two rounds:

Exhibit 1. Measuring Reading Comprehension



The results of the comparison between the two models of testing students' comprehension skills are presented in Exhibit1. As seen from Figure 26, the zero scores dropped from 17% to 12%, and the percent of students who were able to meet comprehension proficiency benchmark of 80% almost doubled (from 19% to 36%).

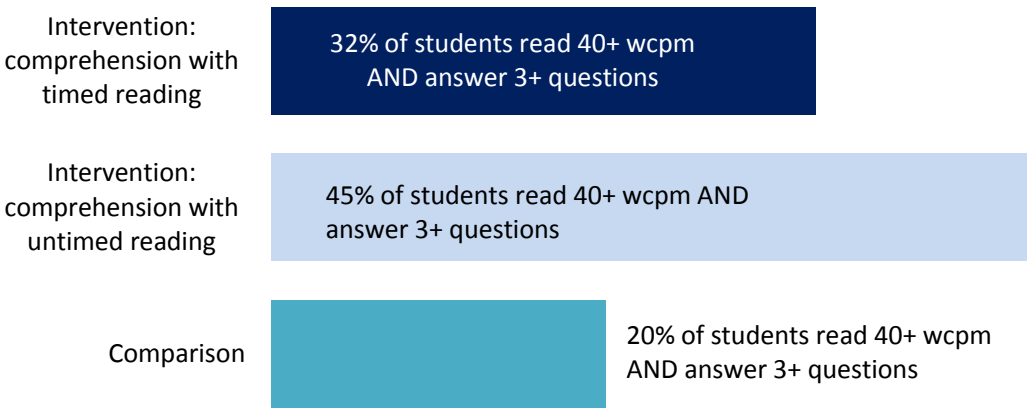
Figure 25. Comparison of Intervention Cohort Endline Comprehension Results Using Timed and Untimed Reading



These findings suggest that reading comprehension results from "classic" EGRA assessments must be interpreted with caution since scores might be as much impacted by the process of administration as by the abilities of the tested students.

To compute the percent of students who *both* can read with fluency *and* can answer comprehension questions, we combined results of the fluency assessment and the comprehension subtest. The graph below shows the results for both timed and untimed comprehension administration. As the graph demonstrates, the percent of students who "qualify" for the USAID standard indicator "percent of students who, by the end of two years of schooling, can read and understand grade-level text" is 32% at the end of the first year of Basa intervention, if measuring with timed reading. The percent increased from 20 to 45% if measuring comprehension with the untimed reading.

Figure 26. Percent of Students Reading and Understanding Grade Level Text, at Endline, by Group

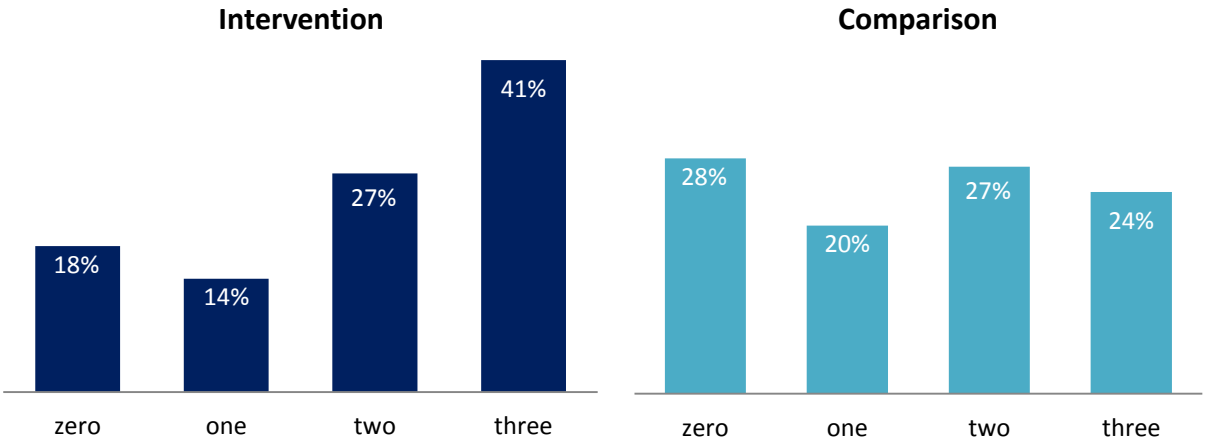


Listening Comprehension and Dictation

On the **listening comprehension subtest**, students were read a passage and asked comprehension questions. Three comprehension questions were asked after the Filipino passage. Results for Filipino listening comprehension ranged from zero to three. By the end of the school year, both comparison and intervention students showed gains in listening comprehension scores, with the intervention

cohort students answering on average 63.7% of listening comprehension questions correctly, compared to 49.1% in the comparison cohort students. Analysis of distribution results show that roughly two-thirds of the intervention cohort students answered two or three listening comprehension questions correctly; 18% were unable to answer any question. This is a decrease from baseline where 35.6% of students in the intervention cohort students scored zero on this subtest. For the comparison cohort students, scores were relatively evenly distributed. More than a quarter of students were unable to answer a single listening comprehension question; roughly half of students answered two or three questions.

Figure 27. Filipino Listening Comprehension Questions – Number Correct at Endline

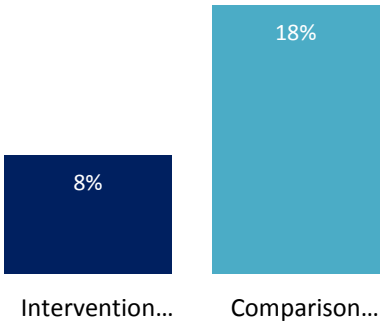


On the **dictation** (total possible 16 correct answers for the Filipino test), students were read a passage once, given a pencil and paper, and read a passage a second time with pauses, then read the entire passage a third time. Students wrote the words on the paper. Dictation scores were broke up into two subtests:

- Number of words spelled correctly (total possible 12)
- Other items relating to conventions of text in writing included spacing, text direction, capital letter, and using a period at the end of a sentence (total possible 4)

Number of correct answers for the dictation subtest at endline ranged from zero to 16, with a mean of 9.9 (61.9%) for the intervention cohort students and a mean of 7.3 (45.6%) for the comparison cohort students. Overall, distribution results were fairly similar across cohorts. The comparison cohort students had a larger percentage of students who scored zero on the dictation subtest, 18% compared to 8% for the intervention cohort students. Very few students scored between 80 and 100%: 5% and 6% for the intervention and comparison cohorts respectively.

Figure 28. Filipino Dictation Zero Scores at Endline, by Group

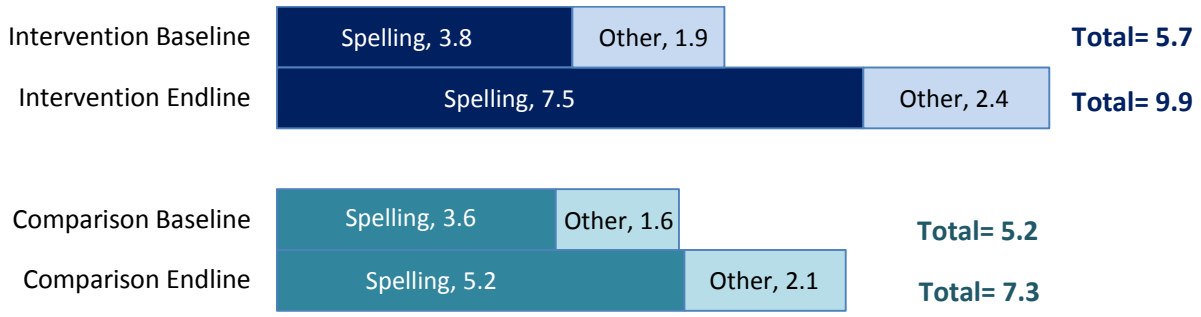


Breaking down the dictation composite at endline, scores for spelling ranged from zero to 12, with a mean of 7.5 for the intervention cohort students and 5.2 for the comparison cohort students at the end of a school year. With regard to the conventions of text, students in the intervention cohort students scored on average 2.4 out of possible 4, compared to 2.1 for the comparison cohort students. Most students used spacing and the direction of the text

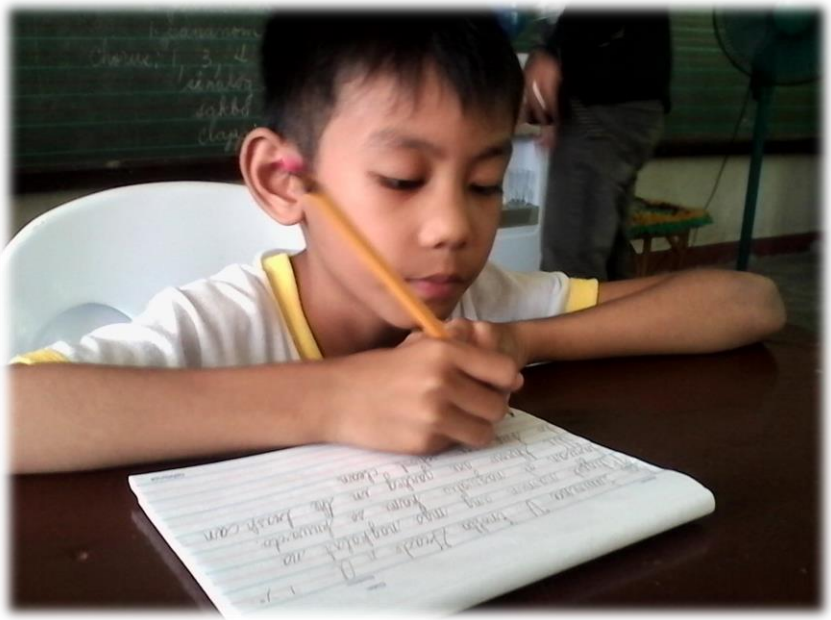
correctly. However, only one third of students capitalized correctly in both groups, and less than 20% of students used a period at the end of the sentence.

Both cohorts saw significant gains (p<.001) from the beginning to the end of a school year. However the intervention cohort students demonstrated larger gains over the course of the year. As seen in the figure below, the majority of gains for both groups were in spelling. Improvement in other conventions of text (spacing, direction of text, capitalization and use of a period) were relatively small.

Figure 29. Filipino Dictation Breakdown, by Group



Disaggregating by sex, girls significantly outperform boys in dictation. This is not surprising given girls have outperformed boys in all other subtests. Writing skills are strongly connected to reading skills in that it takes phonological awareness skills to translate sounds into individual graphemes and the written word.



ENGLISH EGRA SUBTEST RESULTS

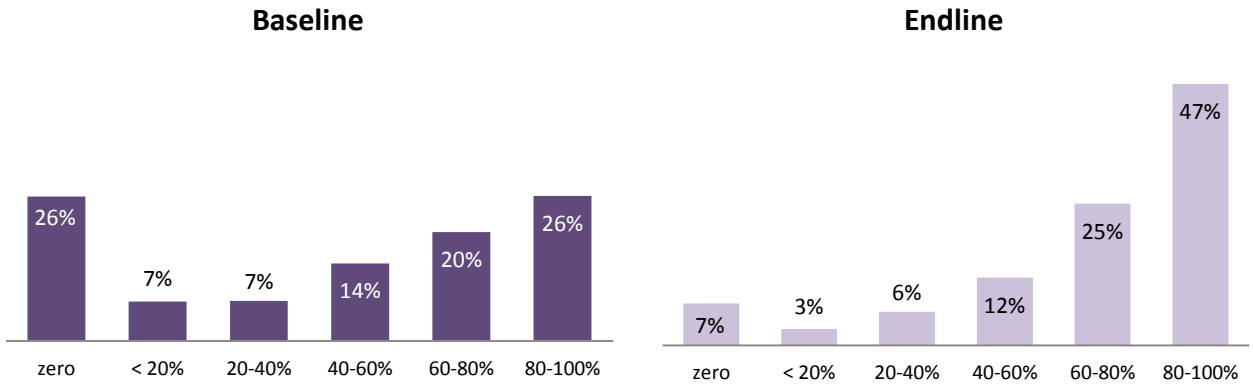
Initial Sound Identification, Letter Sounds, Familiar and Nonsense Words

On the **initial sound identification** subtest (total possible ten letters) that measures initial sound identification of students, responses ranged between zero to ten sounds correct, with a mean of 5.1 initial sounds at baseline and 7.3 initial sounds at endline. The graph below show differences in the distribution of student results at the beginning and at the end of a school year. At baseline the initial sound identification subtest results for the English EGRA show an overall U-shaped distribution, demonstrating a large proportion (26%) of students achieving 80 to 100% correct on this subtest, and an equal proportion having a zero score. However, by endline, the distribution showed a shift to the right, with the majority (72%) of students scoring between 60 and 100%. Additionally, the percentage of students with zero scores dropped from 26% to 7% at the end of the school year.

Comparisons by sex showed that girls outperform boys in this subtest. At endline, slightly more boys had zero scores (8.3% for boys and 6.5% for girls). Further, more than half of girls scored between 80-100%, compared to 40.8% of boys.

Results by division showed that results were fairly similar. Students from Bohol and Cebu performed the best, on average.

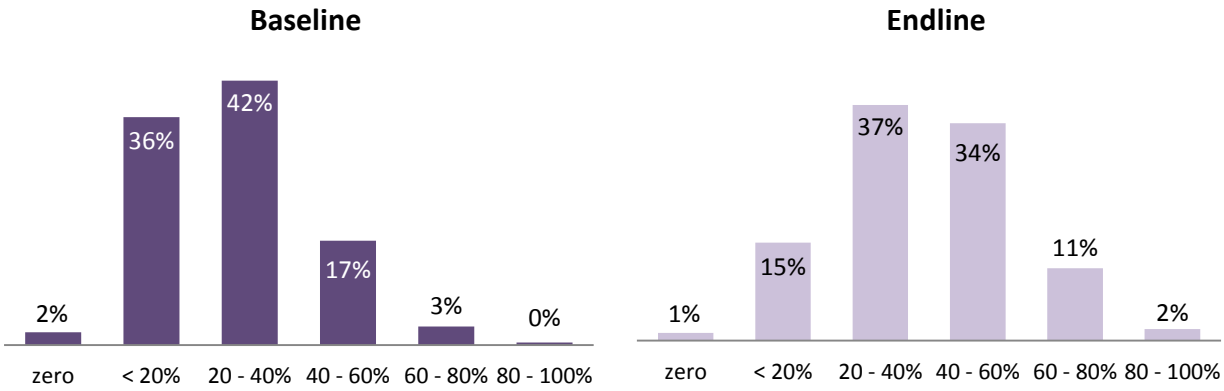
Figure 30. English Initial Sound Identification – Percent Correct



On the **letter sounds subtest** (total 100 letters, timed) number of correct letters ranged from zero to 85 letters (mean 27.0) at the beginning of a school year and from zero to 100 letters (mean 39.1) at the end of a school year. As the graph below shows, at baseline the majority of the students in both cohorts correctly named fewer than 40% of the letter sounds, with only 3% of students from the entire sample able to identify more than 60% of letter sounds. At the end of a school year, the majority of students scored between 20 and 60%.

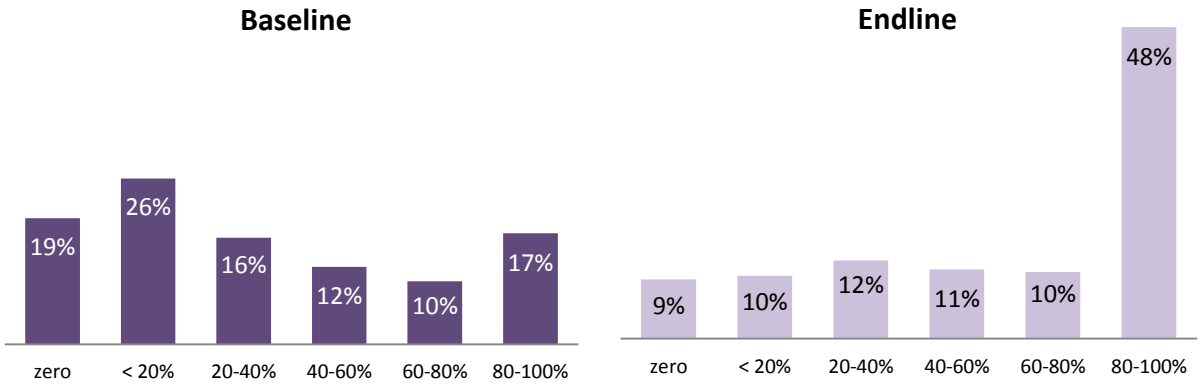
The mean letter sounds correct per minute improved from 27 letters sounds per minute at baseline to 39.1 sounds per minute at endline. Despite these improvements, results were still low at endline which may be due to the fact that many students were only recently introduced to English language. Analysis by sex showed that girls performed significantly better on this subtest, with girls being able to name on average 43.0% of letter sounds compared to 25.4% for boys (p<.001).

Figure 31. English Correct Letter Sound Subtest Results



On the **familiar word** reading (total possible 50 words, timed), responses ranged from zero to 50 familiar words identified correctly, with a mean of 18.3 (36.6%) at the beginning of a school year, and a mean of 31.5 (63.0%) at the end. The graph below shows that at endline the distribution of scores was skewed to the left, with nearly half of tested students reading between 80 and 100% of the words correctly. This is a large improvement from the beginning of the year, when only 17% of students were able to read 80-100% of the words. Analysis by sex showed a substantial difference between girls and boys with 71% of girls who scored between 60-100% compared to 46.2% for boys. Twice as many boys than girls had zero scores on this subtest, with 12.1% of boys failing to read a single word.

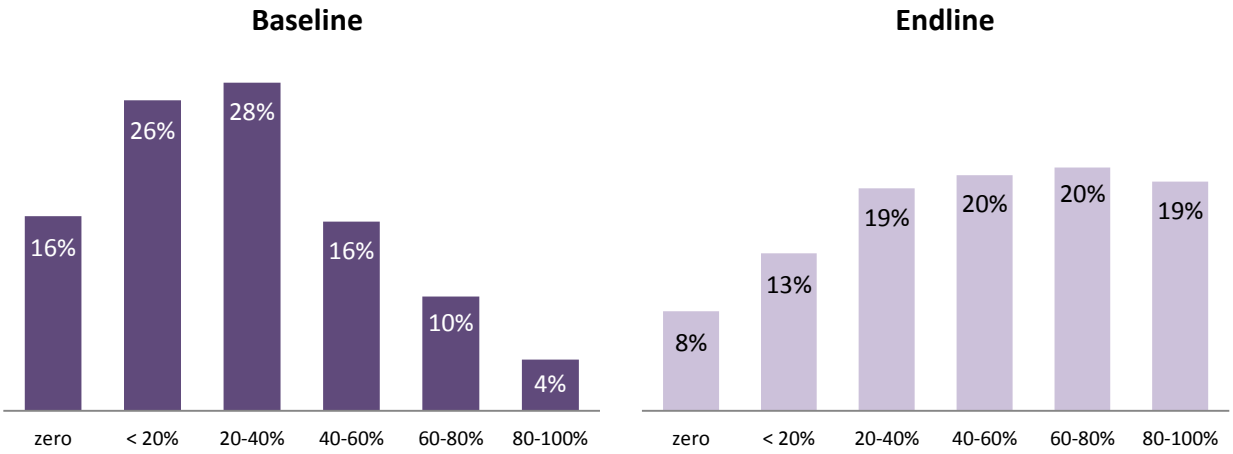
Figure 32. English Familiar Word Identification Subtest Results



On the simple **nonsense word (non-word)** decoding (total possible 50 words, timed), students were asked to read invented words to measure students' capacity to decode words rather than recognizing them or using a context to facilitate reading. Therefore, it is expected that this test is more difficult than oral passage reading or familiar word reading. The number of correct responses ranged from zero to 50 nonsense words correctly read, with a mean of 15.0 (29.9%) at the beginning of school year and a mean of 24.6 (49.2%) at the end of school year. While at the beginning of the school year the majority of students scored below 40%, by the end of the year the majority of students scored over 40%.

Similar to previous subtests girls generally outperformed. The distribution of boys' scores in this nonsense word reading is fairly normal at endline. Conversely, for girls, the distribution of scores is skewed to the right showing a large percentage (69.5%) of girls reading between 40% and 100% of nonsense words correctly.

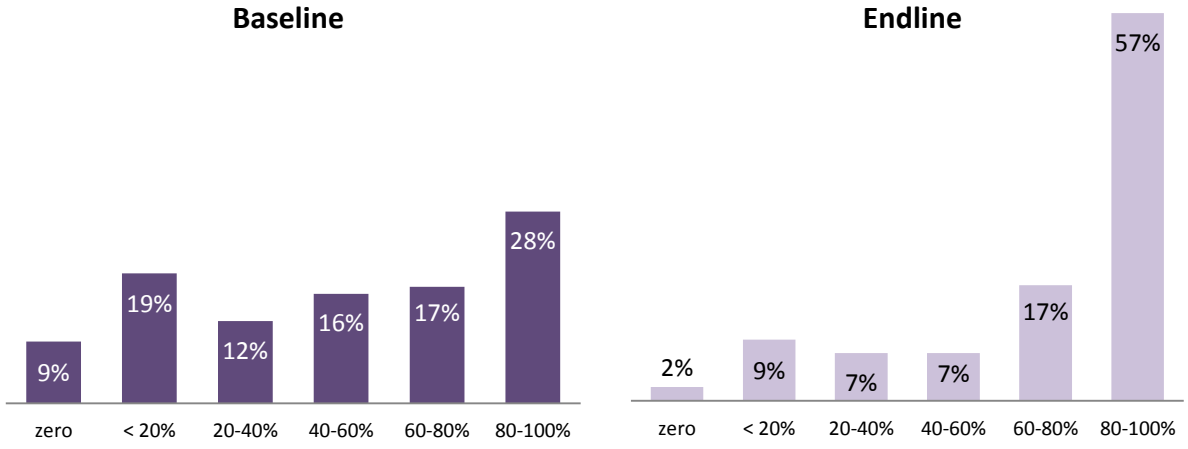
Figure 33. English EGRA - Nonsense Words Subtest Results



Oral Reading Fluency and Comprehension

On the **passage reading and comprehension** subtest, students were scored on the words they read correctly in the passage (total possible 60 words), reading comprehension (total possible 5), and prosody (total possible 4). The number of words read correctly for the English passage ranged from zero to 60 and a mean of 30.1 words (50.2%) at baseline and 43.5 (72.4%) at endline. The results presented in the graph below show that at endline the majority of students reading more than 60% of the text. Very few students (2%) were unable to read any words of the text at endline.

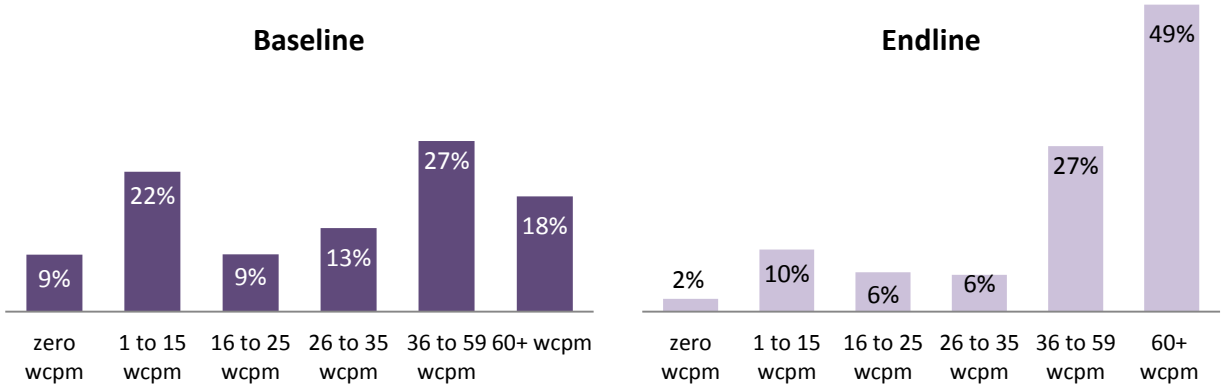
Figure 34. English Oral Passage Reading Subtest Results, Percent of Students Reading Correctly



Similar to other subtests, girls in both groups significantly ($p<.001$) outperformed boys in the oral reading subtest. At endline, girls in the intervention cohort students read on average 79.4% of the text compared to boys who read 65.8% of the text correctly.

Words correctly read per minute. Students were timed on reading the text, with the limit of 60 seconds. Students on average read 34.9 words correct per minute at baseline; by endline students were reading significantly faster with oral fluency results nearly doubled with students reading on average 58.1 words correct per minute. The graphs below show the distribution of fluency scores at baseline and endline. As seen in the graph below, by endline the distribution had shifted to the left, with nearly half of students reading 60 or more words correct per minute.

Figure 35. English Oral Passage Reading (wcpm)



There were significant differences between boys and girls (p<.001) in terms of words correctly read per minute. The range in girls' scores was wider than boys. On average, girls read at a faster rate than boys.

Table 10. English Oral Fluency Measure at Endline, by Sex

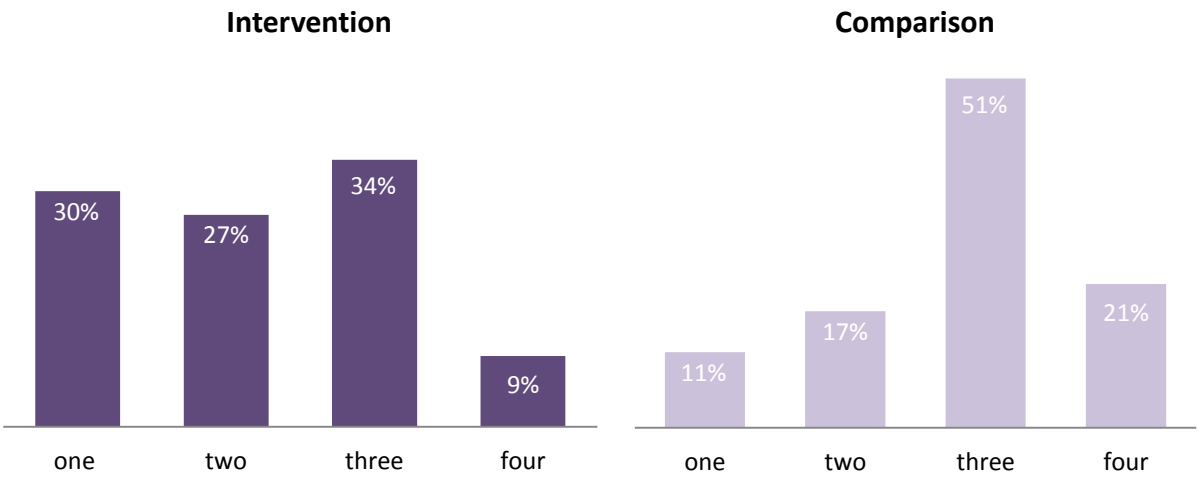
	Range	Mean wcpm	SD
Boys	0 to 150	48.87	31.0
Girls	0 to 171	67.76	33.4

Disaggregating by province for words correctly read per minute (WCPM) or fluency rate, students from Bohol and Cebu showed the highest fluency rates on average. In Ilocos Norte, Ilocos Sur and La Union, results were largely similar.

The **prosody scores** measured the degree to which students were able to read with intonation and expression. When students have difficulty reading, their prosody will often be slow and arduous whereas those who have ease in reading can focus on adding expression and tone to their reading.

Students reading the English passage ranged from one to four, with a mean of 2.8 at endline. The graph below shows the distribution in students' prosody at baseline and endline. The distribution shows that in the intervention cohort more than three-quarters of students read fluently (scored 3 or 4) at endline; only 11% read laboriously.

Figure 36. English Oral Passage Reading - Prosody Score

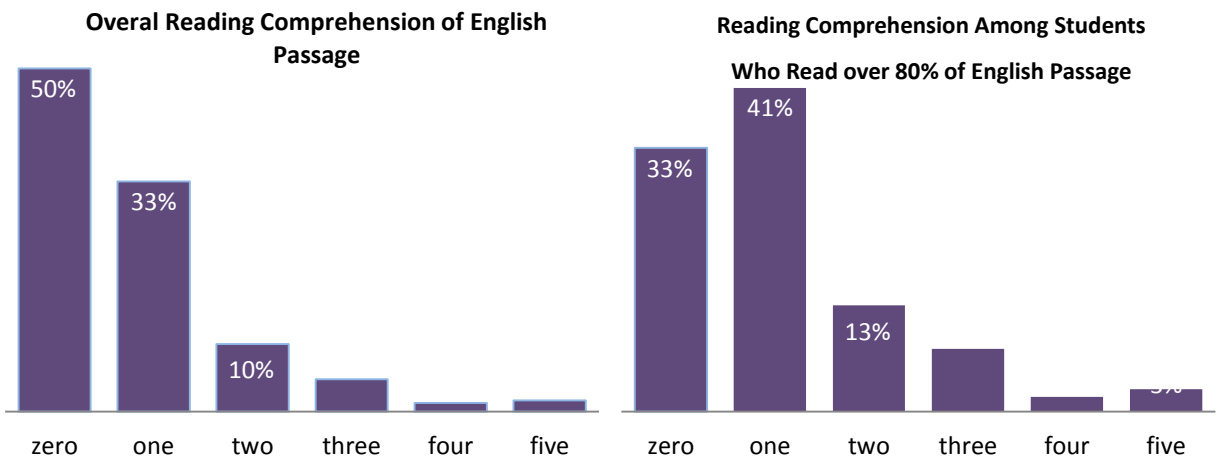


Disaggregation by sex showed that girls from both groups read the Filipino passage with greater prosody than boys, which is consistent with the overall pattern of girls demonstrating higher reading proficiency.

Reading Comprehension. After reading each passage, students were asked five questions they read to check for comprehension. They were not allowed to look back at the text to help them answer questions. At baseline, students group answered on average 0.59 (11.8%) of reading comprehension questions. At endline, minimal gains were seen in reading comprehension with a mean of 0.8 (15.9%). Comparisons by sex show girls performing slightly better than boys in reading comprehension at endline. Disaggregation by province was consistent across province.

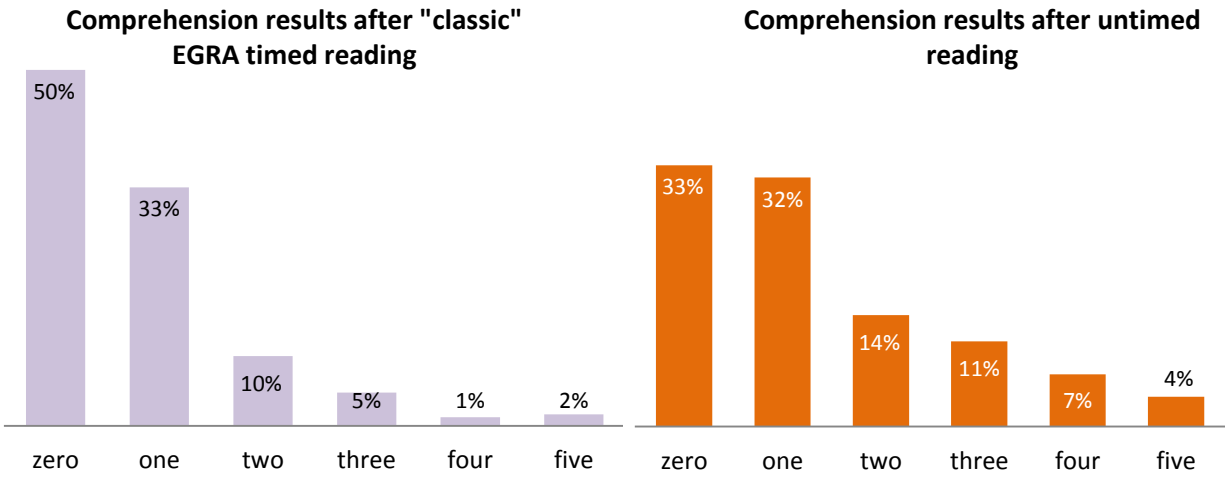
The figure below shows the distribution of endline reading comprehension results. Nearly half of students were unable to answer a single reading comprehension question in English. A third was able to answer one question correctly. Less than 20% of students were able to answer more than one comprehension question correctly. Even those students who were able to read over 80% of the English passage who comprised 57% of all tested students, the large majority only answered one comprehension question or less correctly.

Figure 37. English EGRA Reading Comprehension Endline Results



Similar to the Filipino EGRA, a hypothesis that the assessment process results in an underestimation of the reading comprehension capabilities of students was tested during the endline assessment. The evaluation team administered the English reading comprehension questions in two rounds. The first round the administration followed the standard EGRA administration procedures. The second round immediately followed the first round. During the second round, assessors gave the text back to the students if allowed them to finish reading the passage (if they haven't done so already), and then asked them comprehension questions without taking the text away from the students. The figure below compares results from the two rounds. As seen from the figure, the zero scores dropped from 50% to 33%, and the percent of students who were able to meet comprehension proficiency benchmark of 80% almost quadrupled (from 3% to 11%).

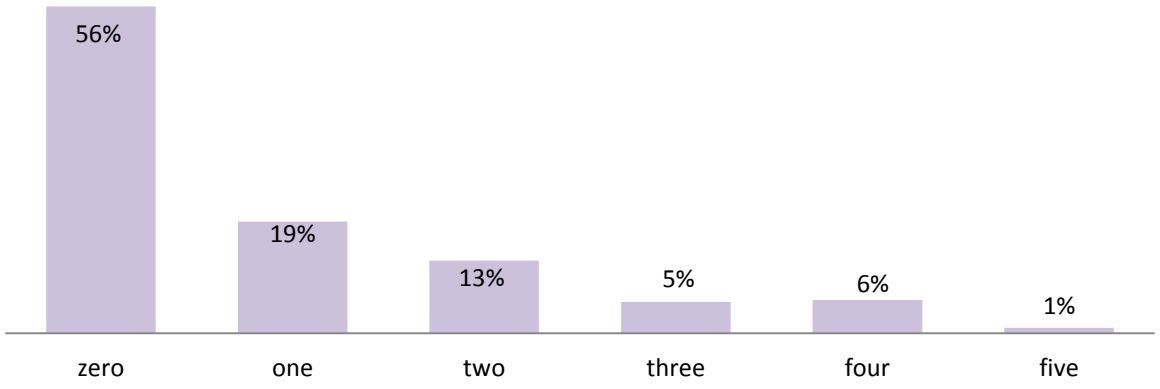
Figure 38. Comparison of Intervention Cohort Endline Comprehension Results Using Timed and Untimed Reading



Listening Comprehension and Dictation

On the **listening comprehension subtest**, students were read a passage and asked comprehension questions. Five comprehension questions were asked after the English passage. Results for English listening comprehension ranged from zero to five. By endline, very little improvement was seen in listening comprehension scores with students answering on average 17.6% of listening comprehension questions correctly at endline compared to 10.6% at baseline. Analysis of distribution results show that more than half of students were unable to answer a single listening comprehension questions. For students who were able to answer questions, the majority of students were only able to answer one (19%) or two (13%) questions. When disaggregating listening comprehension results by sex, girls did better than boys and were able on average to answer slightly more questions than boys.

Figure 39. English Listening Comprehension Questions – Number Correct at Endline

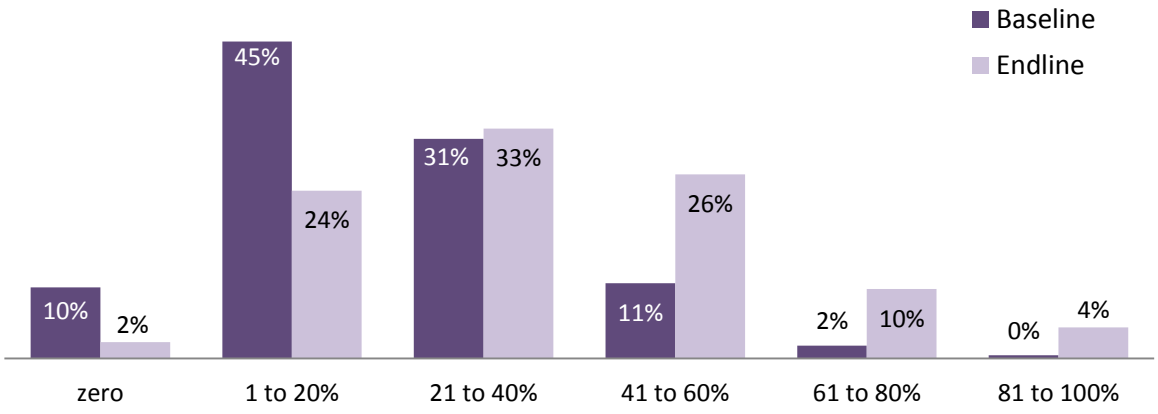


On the **dictation** (total possible 17 correct answers for the English test), students were read a passage once, given a pencil and paper, and read a passage a second time with pauses, then read the entire passage a third time. Students wrote the words on the paper. Dictation scores were broke up into two subtests:

- Number of words spelled correctly (total possible 13 for English EGRA)
- Other items relating to conventions of text in writing included spacing, text direction, capital letter, and using a period at the end of a sentence (total possible 4)

Number of correct answers for the dictation subtest ranged from zero to 17, with a mean of 3.6 (21.2%) at baseline and a mean of 6.2 (36.2%) at endline. The graph below shows that at baseline English dictation results were largely skewed to the right with the majority of students scoring 40% or below on this subtest. At endline there is a relatively normal distribution for dictation in English with the majority of students falling between 21 and 60%.

Figure 40. English Dictation Subtest Results



Disaggregating by sex, girls outperform boys in English dictation at endline and show larger gains from baseline to endline. This is not surprising given girls have outperformed boys in all other subtests. Writing skills are strongly connected to reading skills in that it takes phonological awareness skills to translate sounds into individual graphemes and the written word.

Breaking down the dictation composite, scores for spelling (total possible 19) for English ranged from zero to 13, with a mean of 3.9 at endline. With regard to the conventions of text (4 total possible points), at endline, students in the intervention cohort students scored on average 2.3 for other conventions of text. Most students used spacing and the direction of the text correctly. However, only one fifth of students capitalized the first word in the sentence correctly, and very few students (3.8%) used a period at the end of the sentence.

From baseline to endline, students saw significant gains ($p<.001$) in English dictation scores increasing from an average of 21.2% correct at baseline to 36.2% correct at endline. As seen in the figure below the majority of gains from baseline were in spelling. Improvement in other conventions of text (spacing, direction of text, capitalization and use of a period) were relatively small.

Figure 41. English Dictation Breakdown (Out of 17 possible points)

Overall, the results of EGRA in both Filipino and English showed that students are making significant progress as a result of Basa intervention. The next section of the report shows the changes in teacher practice.

IMPACT OF CONTEXT FACTORS ON ACHIEVEMENT

Impact of Parental Literacy

A factor that was found to have a very significant association with student performance on reading tests was parental literacy. As Table 11 shows, maternal literacy in particular is strongly associated with better results in both fluency and comprehension.

Table 11. Effect of parental literacy on reading fluency*

	Additional effect on fluency, in words correct per minute		Additional effect on comprehension, in percent correct	
	Filipino	English	Filipino	English
Maternal literacy	10.7 wcpm	13.5 wcpm	17.8%	1.5%
Paternal literacy	9.3 wcpm	2.5 wcpm	16.2%	2.7%

*The additional effect was computed using linear regression model, with end of the second grade results on fluency and comprehension as a dependent variable, and a dummy 0/1 parental literacy variables as predictors.

Impact of Home Context Factors

Other factors, such as attendance of the kindergarten, household assets, and availability of books at home were examined for association with key outcomes: fluency and comprehension. The data analysis revealed that attendance of preschool/kindergarten was not consistently associated with higher assessment results. Since 81.5% of students in the sample said they attended kindergarten, it is very likely that low variance in this variable was one of the reasons for why kindergarten was not found to be a strong predictor of achievement. Additionally, kindergartens vary in quality, and recent effort to strengthen kindergarten will undoubtedly influence results of next cohorts of students.

Having books at home, was found to have a positive association with higher EGRA scores in both Filipino and English. Household assets were found to have a weak association with higher assessment results

Table 12. Effect of home context factors on reading fluency[18]

	Pearson's correlation with fluency		Pearson's correlation with comprehension	
	Filipino	English	Filipino	English
Household assets	.049*	.074*	.135**	.120**
Books at home	.123**	.154**	.146**	.102**
Attended kindergarten	n/s	n/s	n/s	n/s

N/s = not significant
* Statically significant at the p<.05 level
** Statistically significant at the p<.01 level

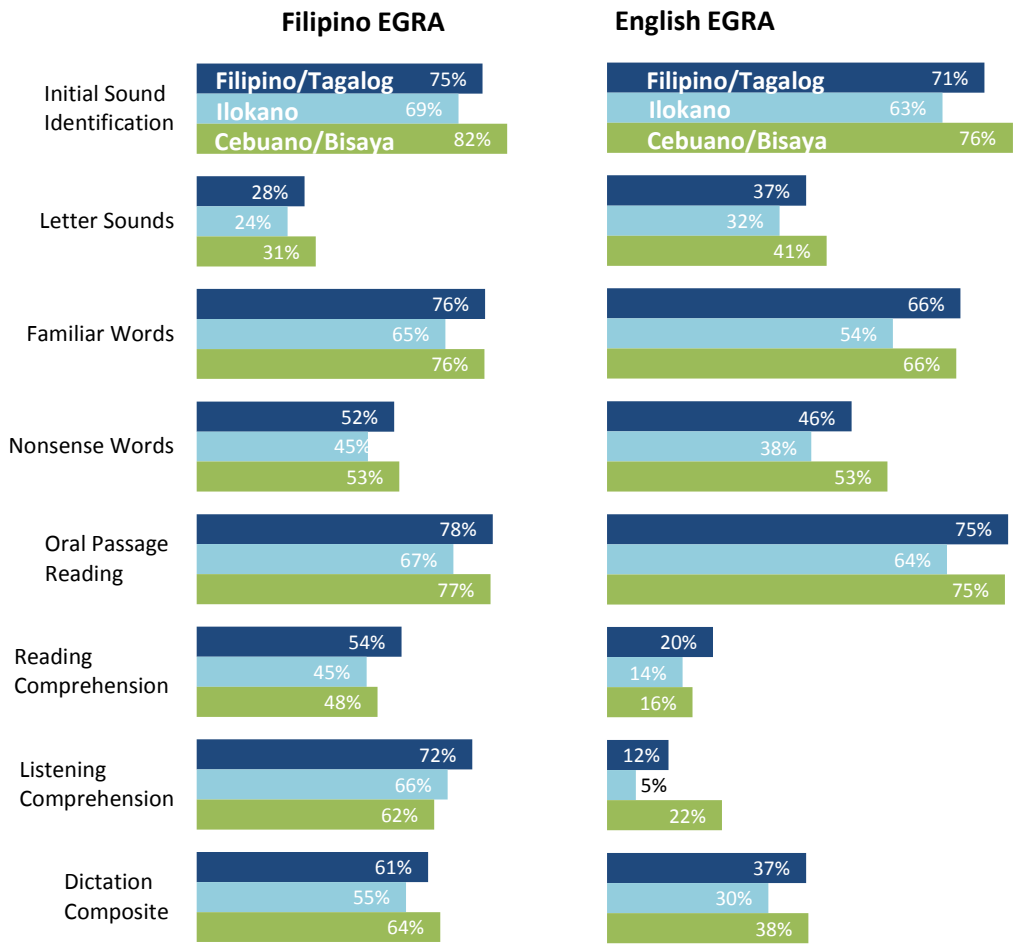
[18] For higher accuracy of the results, unweighted data were used in bivariate correlation analysis.

Impact of Home Language

When EGRA results were disaggregated by language spoken at home, students who spoke Cebuano/Bisaya or Filipino/Tagalog at home performed better on nearly all subtests on both the Filipino and English EGRA tests than students who reported speaking Ilokano at home. These differences were significant at the p<.001 level for both tests.

Students who reported speaking Filipino/Tagalog or Cebuano/Bisaya at home performed fairly similarly across sub-tests for both the Filipino and EGRA tests. For the Filipino EGRA, the largest differences between Filipino/Tagalog speakers and Cebuano/Bisaya speakers were in Listening Comprehension Initial Sound Identification and Reading Comprehension. For the English EGRA, Listening Comprehension had the largest difference in average percent correct between Filipino/Tagalog speakers and Cebuano/Bisaya speakers.

Figure 42. EGRA Results at Endline for Intervention Students Disaggregated by Language Spoken at Home



Ilokano's orthography has fewer similarities with Filipino than Cebuano. Therefore, Cebuano speakers may find it easier to transfer some aspects of phonics skills to Filipino and English than Ilokano speakers. This could explain the relatively higher scores for Cebuano speakers on initial sound identification, letter sounds and nonsense words. It may also explain how students are attaining high fluency scores despite the fact that they only started formal instruction in reading and writing Filipino at the start of grade two and English in the middle of second grade. However,

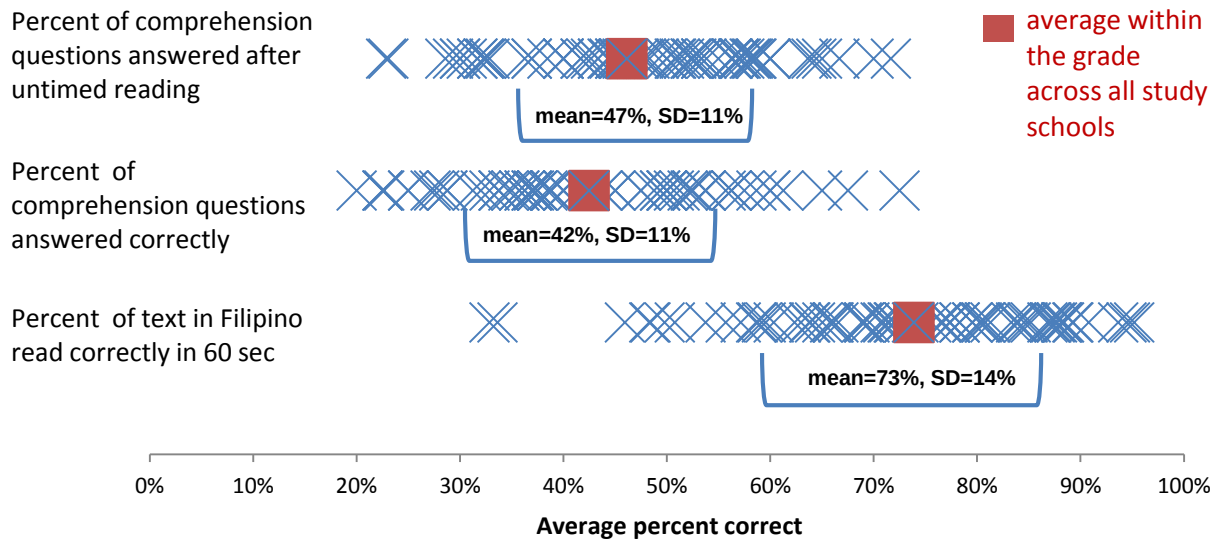
vocabulary knowledge, which is crucial to both listening and reading comprehension, are not transferable across languages especially to English. As a result, students need time to develop vocabulary in order to understand what they hear and read in Filipino and English, and they are not performing as well on comprehension tests.

School-Level Results in Reading

An analysis of school averages in fluency and timed and untimed comprehension[19] showed a remarkable diversity of results across schools, in each grade on both tests. Figure 44 shows a scatterplot of average percent of words in a grade-level text in Filipino read correctly by tested second grade students, in each study school. Each cross on the graph represents average results in a school.

The results show that in a majority of study schools most of the tested students read correctly between 60 and 90 percent of the text they were given to read. The difference in comprehension results between timed and untimed reading was not very large: the average scores only rose by about 5 percent. The distributions were found to be widely dispersed, with schools where most students were not found to perform very well as well as schools where students did well, on average. The average results for the comprehension subtest were found to be much lower than the average results for fluency of reading.

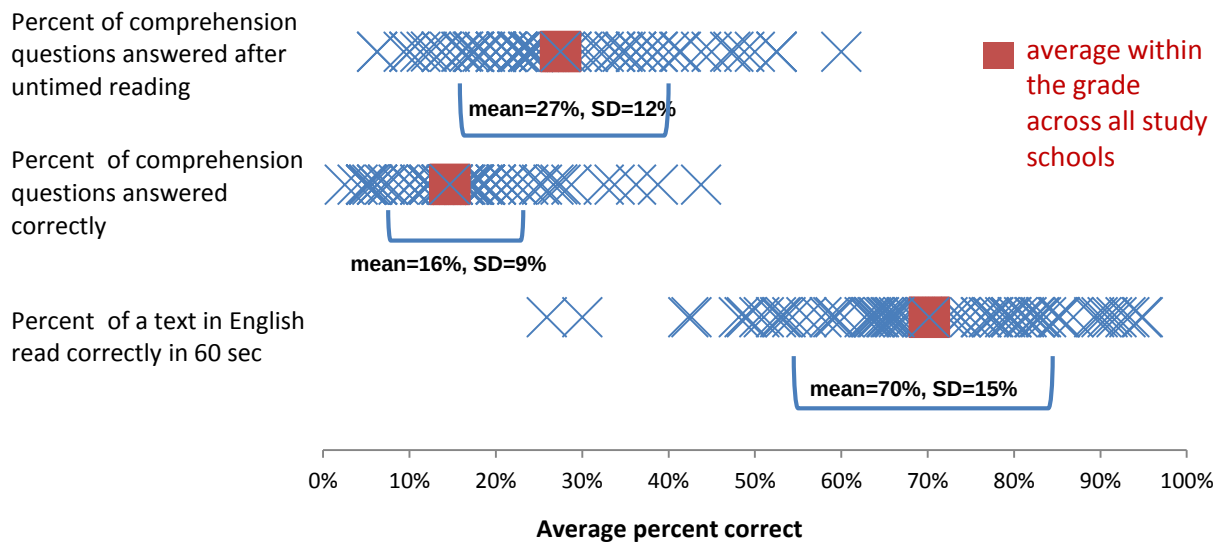
Figure 43. Average percent correct among students on Filipino Oral reading fluency test and comprehension questions in study schools, by grade



Similar analysis performed for the English oral reading passage and reading comprehension subtests across all schools in the sample. The results presented below show similar results for fluency subtest, but substantially lower results for comprehension subtest. The difference in average comprehension results in timed and untimed ways of administering the subtest was much more dramatic than in the Filipino test: the average results almost doubled, from 16 to 27 percent correct.

[19] The averages were computed using percent correct read by sampled students in each grade, for each school. For provincial comparisons, results across the three tested grades were averaged for each school.

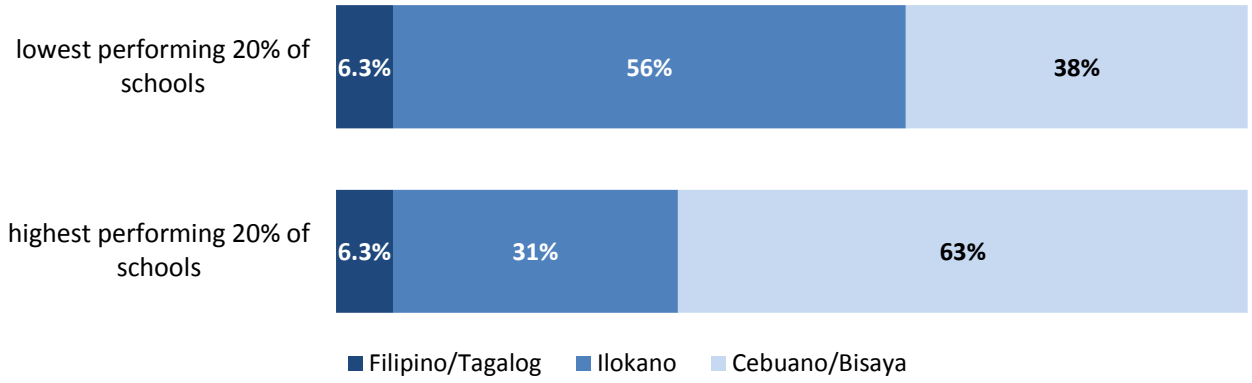
Figure 44. Average percent correct among students on English Oral reading fluency test and comprehension questions in study schools, by grade



Comparisons of school-level average results on oral reading fluency and comprehension tasks found that students in the same schools do below average on both subtests, and do above average on both subtests. Schools also have similar results among their students in both English and Filipino assessment: schools where students scored low on the Filipino EGRA, students also scored low on the English EGRA. In the vast majority of schools, however, the average student results varied and included both higher and lower performers.

Statistical comparisons of background characteristics of 20 percent of top performing and bottom performing schools showed that a significantly higher proportion of schools in the top 20 percent had students who spoke Cebuano at home, compared with the lowest performing 20 percent of schools. In the majority of the lowest performing schools students spoke Ilocano as a home language.

Figure 45. Language spoken at home among students in the lowest and in the highest 20 percent of study schools



A further investigation of why in some schools tested students averaged much higher than average would help document potential best practices in instruction, as well as shed light on structural deficiencies causing some schools to average significantly below their peers. Since the home

language difference appears to be an instrumental in explaining student reading ability in Filipino and English, students need to be tested in their home language, Cebuano or Ilocano, to better understand their true reading abilities.

CHANGES IN TEACHER PRACTICE

Standards-based Classroom Observation Protocol for Educators (SCOPE) was developed by EDC to measure the quality of classroom instruction across grades and subject matters. The observation protocol has subsequently been adapted to capture the quality of instructional practices with specific focus on literacy. The *SCOPE Literacy* tool has two major sections: *Classroom Structures* and *Language Literacy Instruction* and a total of 13 items describing specific literacy-related instructional practices.

Based on the observation of a class on literacy, each item is scored on a scale from 1 to 5:

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

Thirty-three grade two teachers from a sample of Basa schools in Cebu and La Union were observed and scored longitudinally using the *SCOPE Literacy* tool in November/December 2013 and again in December 2014 to measure changes in teaching practices as a result of the Basa intervention. As Figure 47 demonstrates, teachers started out with very low scores at baseline in 2013, ranging between "deficient" and "inadequate". By the end of 2014, teacher practices showed a broader range of scores with more teachers performing at "basic" level. This suggests some improvement from 2013 to 2014 which indicates that teachers are starting to apply new teaching practices. Ratings of 4 and 5, or "strong" and "exemplary," are not easy to attain. However, a "basic" rating of 3 is quite positive in the context of introducing new literacy instruction techniques.

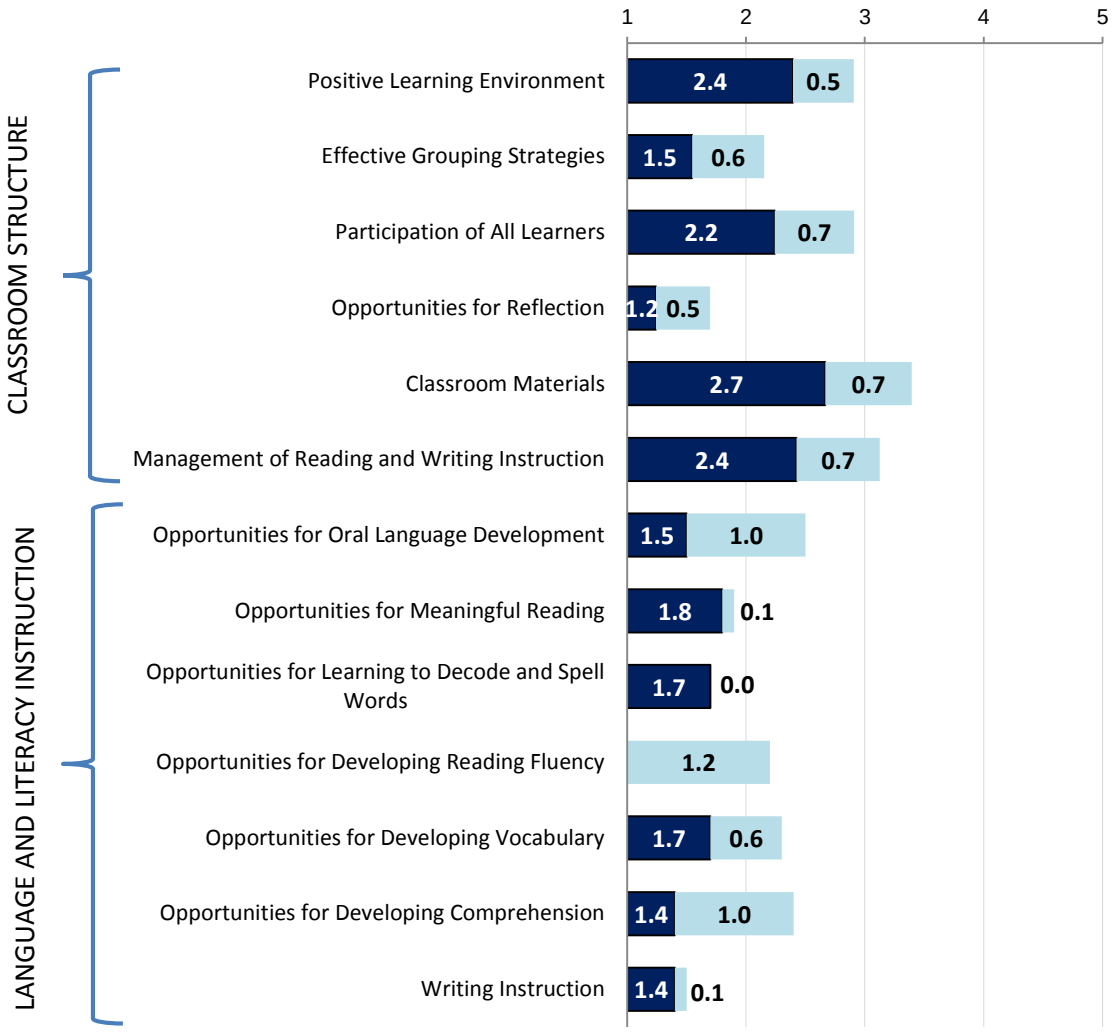
By December 2014, improvements were seen in all teaching practices observed; however, the largest gains were seen in the *Language and Literacy Instruction* domain. In the Classroom Structure domain, teachers saw the largest improvements in "ensuring participation of all learners", "ensuring accessible classroom materials" and "effective management of reading and writing instruction" items of the *SCOPE Literacy*. In fact, in 2014, nearly half of observed teachers were scored "strong" in "classroom materials" and "management of reading and writing instruction." This is important given that Basa has introduced a large number of new reading materials for teachers to manage in the classroom. Ensuring access to the materials is key for student learning, as is equitable participation in the classroom. Teachers didn't score as highly in the areas of "effective grouping strategies" and "opportunities for reflection," both items that require more advanced classroom management skills. Teachers who can effectively group students are better at providing differentiated learning opportunities as well as opportunities for reflection that can deepen students' understanding of text.

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

One general caveat, Basa teachers follow an instructional sequence in which all 14 domains of the K-12 curriculum are not taught everyday but over a period of 5 days for Filipino and 10 days for English. This is by design to allow adequate time for pupils to complete skills related tasks in their second and third language. Depending on the lesson plan for the day, it would not be expected that teachers teach all domains. In addition, domains such as phonics take on additional meaning in a language such as English which has an opaque orthography versus Filipino, a syllabic language that has a transparent orthography. Since the teachers were observed during their Filipino class, one possible reason for the no increase in the score for "opportunities for learning to decode and spell words" is because by Grade 2 quarter 3, majority of the students have already learned to decode and spell in Filipino.

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

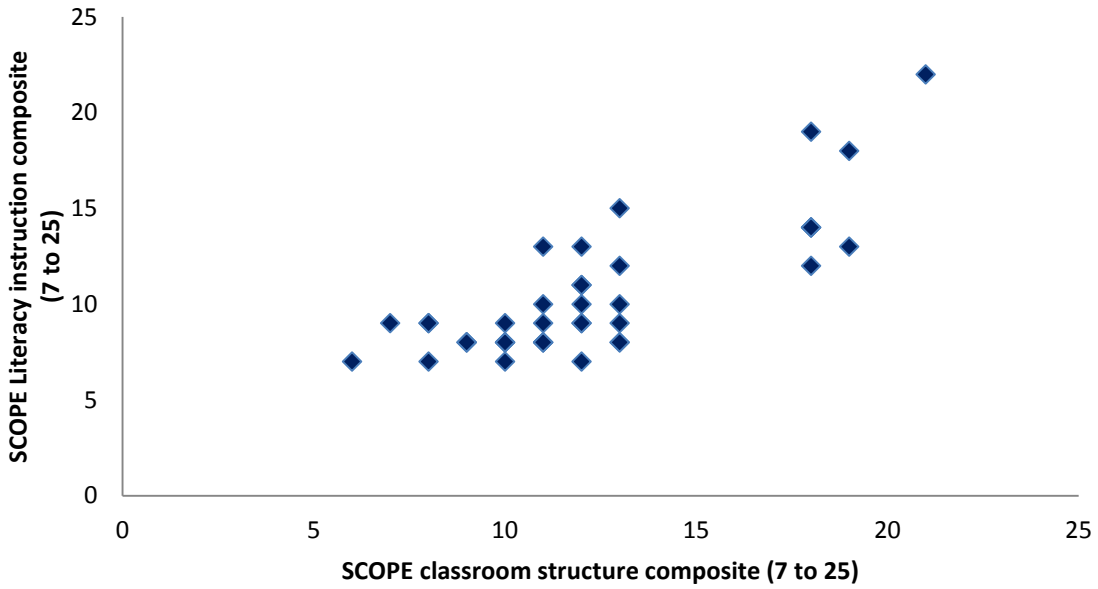
Figure 46. Change in Average Scores for SCOPE, Change from 2013 to 2014



A strong correlation between the two sections of the *SCOPE* results was found in 2014. The scatterplot below shows that the relationship between the two components of the SCOPE tool appears to be linear.[20]

[20] The coefficient of the correlation between the two sections of the *SCOPE* was high and significant (Pearson's r=.946; Kandall's tau = .820 and Spearman's rho = .905, all three significant at *p*<.001 level).

Figure 47. Correlation between Two Components of the SCOPE in 2014



These results confirm that teachers who have stronger practice in classroom structure, also have more advanced practices of teaching literacy.

CONCLUSION AND RECOMMENDATIONS

The main conclusion of the evaluation study is that students in the intervention cohort scored higher on all subtests than the comparison cohort students. This suggests that that Basa intervention did have a significant, positive impact on student achievement in reading. The study also found improvements in teacher performance, with the largest gains in the Language and Literacy Instruction domain. In the Classroom Structure domain, teachers showed the largest improvements in ensuring participation of all learners, ensuring accessible classroom materials and effective management of reading and writing instruction. In fact, by December/January 2014/5, nearly half of observed teachers were scored as "strong" in classroom materials and "management of reading and writing instruction." For the Language and Literacy Instruction domain, teachers largely improved literacy instruction in terms of oral language development, developing reading fluency, and developing comprehension. These improvements are associated with the Basa intervention.

Although 2012/13 was the first year of implementation of the new Grade 2 curriculum, teachers and students were just receiving new textbooks from DepED and were adjusting to a new curriculum. This may be one reason that comparison cohort students reported having fewer English and Filipino books at home, and may also contribute to the overall lower scores in the comparison cohort. In contrast, at the time of data collection for the intervention cohort, the new curriculum had been in place for a full year and Basa had provided additional read-alouds, leveled readers and accompanying teacher's guides as well as training in how to use the materials in class. This additional supported provided by Basa has undoubtedly contributed to improved student scores.

Since the comparison of means at the baseline showed that the tested cohorts were not perfectly identical (the intervention cohort students scored higher on all but one subtests), a statistical correction is required to assess the true intervention effect[21]. A.

The results of a regression analysis conducted to estimate the amount of gains *after* the baseline differences were incorporated into the analyses and controlled for, showed that the Basa intervention helped increase the fluency of second grade students by additional 9.1 words per minute, above and beyond the natural increases expected to occur over the course of the school year. Additional gains in dictation as well as listening and reading comprehension were also found to be substantial. The smallest additional gains were found in letter sounds subtest, and reading of familiar and nonsense word subtests. These results show that the intervention was associated with improvements in the two critical areas of literacy skills: fluency and comprehension.

Despite relatively similar baseline percentages of students with zero scores on EGRA subtests, by the end of the school year in all but two subtests (Letter Sounds and Invented Words) the intervention cohort had significantly fewer students with zero scores than the comparison cohort. This can be seen particularly in the reading comprehension and listening comprehension subtests where percent of students with zero scores in the intervention cohort nearly halved from baseline to endline, while the comparison cohort only saw a small decrease in the proportion of students with zero scores.

Thirty-one percent of second graders in the intervention cohort reached the proposed by Basa grade 3 oral reading benchmark of 60 wcpm in Filipino, and more than three quarters read fluently, scoring a 3 or 4 on prosody. This is despite the fact that fewer than 10% reported speaking Tagalog or Filipino at home, and that students have had only one year of formal instruction in reading and writing in Filipino. In English, an even higher percentage of students in the intervention group, 49%, were reading at more than 60 wcpm. Nearly three-quarters of the intervention cohort students scored a 3 or 4 on prosody, demonstrating fluent reading. This is surprisingly high, given that second graders have only had two quarters of formal instruction in reading and writing English, and that fewer than 3% reported speaking English at home. English words tend to be shorter than Filipino words which may be one reason why students were able to read faster in English.

While fluency scores in both languages are high, comprehension scores are relatively low. In Filipino, only 19% of the intervention cohort students were able to answer 4 or 5 out of 5 reading comprehension questions correctly. This may be due in part to the testing conditions, in which students are given asked to answer questions without being able to reference the reading passage. At endline in the intervention group, the study team also administered an untimed comprehension test and students were allowed to reference the text to help them answer the comprehension questions. In the untimed version, the results increased from 19 to 36% of students answering at least 4 out of 5 comprehension questions correctly in Filipino. Listening comprehension scores were higher, with 41% students in the intervention group answering all 3 questions correctly.

Both listening and reading comprehension in English were very low. Although zero scores significantly decreased from the beginning to the end of the school year, roughly half of tested

[21] *Procedures and Standards Handbook, v.3.0*. What Works Clearinghouse, Institute of Education Sciences, US Department of Education, 2015, p. 14.

students were still unable to answer a single reading or listening comprehension at endline. Very little improvement was seen in listening comprehension scores with students answering on average 17.6% of listening comprehension questions correctly at endline compared to 10.6% at baseline. Minimal gains were also seen in reading comprehension. At baseline, students answered on average 0.59 (11.8%) of reading comprehension questions while at endline they answered an average of 0.8 (15.9%). Even those students who were able to read over 80% of the English passage who comprised 57% of all tested students, the large majority only answered one comprehension question or less correctly. Allowing students unlimited time to read the passage and reference the text while answering the comprehension questions increased the percent of students answering at least 4 out of 5 comprehension questions correctly from 3 to 11%.

This highlights the need for the project to continue its focus on providing teachers with strategies for increasing comprehension. One way the project addresses this need is through the provision of leveled readers which are used to help students develop reading strategies such as in quickly identifying frequent words and decoding unfamiliar words. Leveled readers also build students' vocabulary acquisition skills using strategies such looking at pictures for clues to figure out the meaning of a word. Another way the project supports teachers is through training them on how to group students according to their reading ability. Grouping students allow teachers to provide more targeted instruction especially to the struggling readers. Classroom observations showed that teachers had improved literacy instruction in terms of oral language development, developing reading fluency, and developing comprehension, with the largest gain in developing reading fluency.

One important finding of the evaluation study is the importance of the language spoken at home. The evaluation found that students who spoke Cebuano/Bisaya or Filipino/Tagalog at home performed better on nearly all subtests on both the Filipino and English EGRA tests than students who reported speaking Ilokano at home. One of the reasons behind this difference can be higher print awareness among Cebuano/Bisaya students since there are reportedly more materials in Cebuano than Ilokano. Previous research has also found that parents who speak Ilokano are not that familiar with the language in its printed form.

Finally, similar with a number of previous studies and assessments, the evaluation found that girls consistently outperform boys, and the difference is significant. Further research is needed to explain this difference. Basa intervention strives to help address the gender gap through improved reading instruction and provision of appropriate reading materials. Additionally, more explicit attention to increasing boys' achievement is needed.

The next round of assessments will include testing a sample of students from grade 3 intervention cohort, to assess the cumulative impact of Basa support over the course of two years, within the context of the new curriculum implementation.

APPENDICES

ANNEX 1. METHODOLOGY

DESIGN

As part of its scheduled monitoring and evaluation (M&E) activities, Basa is conducting three phases of outcome evaluations to measure changes at different levels of the intervention, at the student level (reading), teacher and classroom level (beliefs, pedagogical and instructional practices) and school level (environment and management). The evaluation is designed to be carried out in three stages:

- SY 2013/ 2014: Comparison Cohort,
- SY 2014/ 2015: Intervention Cohort,
- End of SY 2016: follow-up with Comparison and Intervention Cohorts

In SY 2013/14 before the intervention began, a comparison cohort in a sample of schools in Cebu and La Union was assessed in reading (Filipino) to measure the natural growth before the Basa intervention. In school year 2014/15, after the rollout of the full Basa intervention, a random sample of students in grade 2 were tested in reading (Filipino) at the end of the school year within the same study schools. Additionally, during SY 2014/15 the evaluation was expanded to additional schools in Cebu and La Union, as well as schools in Ilocos Norte, Ilocos Sur and Bohol. Student assessment results are compared with baseline comparison scores established in SY 2013/14.

Student data will be analyzed longitudinally comparing data from the baseline (pretest) to data at the endline (posttests). The classroom observations will be analyzed cross-sectionally; the same classrooms will be tracked and measured at the endlines (posttests).

To justify the use of the comparison group in Cebu and La Union for the intervention group (all five provinces) baseline equivalence analysis was conducted. According to WWC guidelines, the effect size difference (Cohen's d) between intervention and comparison group means at baseline should fall between 0 and 0.05 to satisfy baseline equivalence or between 0.05 and 0.25 (requires statistical adjustment to satisfy baseline equivalence. For all but one subtest (Letter Sounds) the effect size difference falls between 0.05 and 0.25, and as such meets baseline equivalence. Given these results the comparison group was deemed a legitimate comparison group for comparison with intervention results in all five provinces and is referred to as the comparison group in the report.

Table A-1. Effect Size (ES) Difference between Intervention (All Provinces) and Comparison Cohort (Cebu/La Union) Means at Baseline

Subtest	Effect size of baseline mean differences
Phonemic Awareness	0.2008
Letter Sounds	0.4759
Familiar Word Reading	0.0949
Nonsense Word Reading	0.1103
Oral Passage Reading	0.13
Reading Comprehension	-0.0512
Listening Comprehension	0.0719
Dictation	0.1471

Specifically, the evaluation study was designed to answer the following questions:

Evaluation Question	Tools Used
1. Do students, both male and female, demonstrate improved reading and comprehension skills in Filipino at the end of Grade three per DepEd Standards?	Electronic Early Grade Reading Assessment (eEGRA) tests will be employed to detect changes in reading comprehension skills of primary schools students between the baseline (grade 2) and endline (end of grade 2). The tests will be administered in the Filipino and English languages.
2. Have teachers' instructional practices in reading (in the Filipino language) improved in target schools?	Teacher observation protocol (Standard Classroom Observation Protocol-SCOPE Literacy) to gauge changes pedagogical practices.
3. Have teacher beliefs in the area of teaching reading and writing changed?	A teacher survey (Beliefs and Instructional Practices Inventory-BIPI) to gauge changes in beliefs
4. What proportion of participating school heads, supervisors and teachers report that they have increased gender awareness in their management and teaching?	A principal's survey SSME will be administered with school principals at the sample schools to gather critical school-level demographics

SAMPLE

Sampling was conducted at three levels: 1) school, 2) classrooms, and 3) student. The **school sample** used for the Basa evaluation activities was drawn randomly from the project's five provinces: Cebu, La Union, Bohol and Ilocos Norte/Sur, excluding schools with multi-grade or combination classrooms. Only schools receiving direct interventions from the Basa project with activities in teachers training, curriculum development, provision of learning materials, and other targeted support were included in the sample. For Cohort 1 (2013/14) the comparison cohort of students 20 schools in both Cebu and La Union were randomly selected to be included in the evaluation. For Cohort 2 (2014/15) 20 intervention schools from each province, La Union, Cebu, Ilocos Norte/Sur and Bohol were randomly selected to be included in the evaluation.

At the second level of sampling, one **classroom** will be selected randomly from the total number of classrooms at the school (i.e. one grade 2 from 5 total classrooms). The teacher for the selected classroom will complete the BIPI Survey and be observed using the SCOPE tool. Finally, a random sample of students will be selected from within the sample classroom. For cohort 1, a minimum of 10 students were selected from each of the 40 sample classrooms (Basa years 1/2), while for cohort 2 (Basa years 2/3), 13-15 students will be selected from each of the 80 sample classrooms to account for expected attrition. Equal numbers of male and female students will be selected as much as possible.

Reading performance of students will be measured at the start of the school year (grade 2) and at the end of the school year (grade 2) and longitudinal gains between these two data points in the same grade level will be calculated. All attempts to track unique children longitudinally were made to measure learning gains. However, due to attrition, students in Cohorts 1 and 2 were matched wherever possible longitudinally. For those students who could not be matched from baseline to endline, students were matched using rank order matching in which sex, province and school was taken into account in the rank matching process. The table below maps out the complete data collection schedule by the multiple data points, cohorts and grades.

Table A-2. Student Sample by Classroom

	CLASSROOM SAMPLE			STUDENT SAMPLE					
				SY 2013/14 Comparison Cohort		SY 2014/15 Intervention Cohort		SY 2015/16	
Cohort	**Grade**	**Regions**	**# of Classes & Measurement**	July 2013	July 2014	July 2014	Mar. 2015	July 2015	Mar 2016
1	Grade 2	Cebu & La Union	**40 classes** Filipino EGRA	Baseline (469)	Endline[1] **(469)**		Endline **(631)**		Post-test (400)
2	Grade 2	Bohol, Cebu Ilocos N/S & La Union	**80 classes** Filipino & English EGRA			Baseline (1,343)	Endline **(1,216)**		Post-test (1040)
3	Grade 3	Bohol, Cebu, Ilocos N/S & La Union	Same pupils as cohort 2						Post-test (1040)
Total (unique) units			**120 classes**		**469 pupils**		**1,616 pupils**		**1440 pupils**

[1] Endline sample represents matched sample. Students that could not be matched at endline are not included in this sample value.

The student sample was determined by the following sampling parameters: Independent t-test, two tail, effect size .25 (small), alpha = 0.025, Power = 80% and Attrition=15%.

DATA COLLECTION TOOLS

Four tools were used to collect data used in this study. The timeframe and number of administrators varied by the tool.

***Principal (School Snapshot) Survey*.** The school's environment and management is critical to understanding the teaching and learning that is taking place in the school. The Principal Survey[1] was designed to capture information on: 1) the physical infrastructure, 2) the teaching and learning demographic, 3) the school manager's background and characteristics, 4) school policies, practices, and monitoring, 5) the reading environment, 6) parent and community involvement, and 7) disabilities and gender policies and practices.

The survey was administered one-on-one with each of the 83 principals (or other school leaders) present at the time of the EGRA and BIPI data collection in July/August 2014 in intervention schools in Bohol, Cebu, Ilocos Norte/Sur and La Union. The survey data was collected on paper and processed in the SurveyToGo system, similar to the BIPI data. Detailed Principal Survey findings can be found in Annex 3.

***Beliefs and Instructional Practices Inventory (BIPI)*.** Teacher beliefs are known to impact classroom and instructional practices. The inventory was conducted to better understand what teachers in Basa intervention schools think about their students' abilities, the appropriateness of different instructional methods for teaching literacy to students, as well as their own classroom practice. The survey consisted of the following sections:

Section A. Teacher's demographic information

Section B. Questions about frequency of literacy-related instructional practices in the classroom

[1] Note that the Principal's Survey used was modeled on the Snapshot for School Effectiveness tools used in a number of different countries. RTI was not using a Principals' Survey in Philippines at the time of data collection, so Basa developed its own tool.

Section C. Series of statements about teaching literacy
Section D. Series of statements about students' abilities in relation to literacy

The survey was conducted at the same time as the baseline EGRA data collection (July 2014) for teachers in 82 out of the 84 intervention sample schools using paper forms by trained partner and Basa staff and processed using the SurveyToGo system. Detailed BIPI findings can be found in Annex 2 of this report.

Standard Classroom Observation Protocol in Education-Literacy (SCOPE-Literacy). The quality of literacy instruction was observed using *SCOPE-Literacy*, a tool developed by EDC and tested in several countries. The tool includes two sections with a total of 13 items; each item is rated on a scale from one to five. The first section focuses on **classroom structure** and includes the following six items:

1. Positive Learning Environment
2. Effective Grouping Strategies
3. Participation of All Learners
4. Opportunities for Reflection
5. Classroom Materials
6. Management of Reading and Writing Instruction

The second section focuses on **language and literacy instruction** and includes the following seven items:

7. Opportunities for Oral Language Development
8. Opportunities for Meaningful Reading
9. Opportunities for Learning to Decode and Spell Words
10. Opportunities for Developing Reading Fluency
11. Opportunities for Developing Vocabulary
12. Opportunities for Developing Comprehension
13. Writing Instruction

In Year 1, baseline data were collected in November 2013 to February 2014[2] by trained observers working in pairs. Subsequently in Year 2, the same teachers were observed in December 2014. In Year 1 while teachers had received some preliminary training on the DepEd curriculum through Basa, the full intervention[3] had not been delivered as of the time of observation. By Year 2, the full intervention had been rolled out.

During observation, a pair of data collectors was sent to observe teachers. Each pair observed sampled teachers' instruction separately and then agreed on a score for each of the items, following a discussion. The agreement score was used in the analysis. Individual observer scores were used for the inter-rater reliability (IRR) analysis, to measure how well individual assessors understood the

[2] Note that the scheduled observation was November 2013, but there were some delays experienced, notably scheduled school holidays and difficulty in getting to some of the school due to emergency relief situations involving the 2013 typhoon.
[3] The Basa intervention consists of a set of materials made up of Read-Aloud stories, leveled readers and teacher guides as well as trainings on how to use the materials.

observation criteria and to what extent they scored the same teacher practices similarly. Since the SCOPE rating data are categorical and not nominal, the interclass correlation coefficient (ICC) was used in the IRR analysis. The ICC assesses rating reliability by comparing the variability of different ratings of the same subject to the total variation across all ratings of all subjects, and results in a coefficient between 0 and 1 where 0 = complete lack of agreement across raters on how to score a particular item, and 1 = perfect agreement. It is considered to be desirable to achieve ICC of .8 or higher. The IRR analysis of observers showed high rates of agreement among observers, with an average ICC of .939 at baseline and 0.989 at endline.

A reliability analysis of the SCOPE Literacy found that the test reliability was high (Cronbach's alpha = .891 at baseline and .828 at endline). Items showing lower internal consistency to the overall *SCOPE-Literacy include* "writing instruction" and "opportunities for learning to decode & spell words."

Table A-3. SCOPE Literacy Reliability, Year 1 (SY 2013/14)

SCOPE Literacy Items	Item-Total Correlation	Cronbach's Alpha if Item Deleted
Positive Learning Environment	.595	.884
Effective Grouping Strategies	.396	.892
Participation of All Learners	.803	.871
Opportunities for Reflection	.754	.878
Classroom Materials	.720	.877
Management of Reading & Writing Instruction	.641	.880
Opportunities for Oral Language Development	.620	.882
Opportunities for Meaningful Reading	.624	.881
Opportunities for Learning to Decode & Spell Words	.476	.889
Opportunities for Developing Reading Fluency	.678	.880
Opportunities for Developing Vocabulary	.622	.883
Opportunities for Developing Comprehension	.548	.885
Writing Instruction	.113	.897

Table A-4. SCOPE Literacy Reliability, Year 2 (SY 2014/15)

SCOPE Literacy Items	Item-Total Correlation	Cronbach's Alpha if Item Deleted
Positive Learning Environment	.558	.810
Effective Grouping Strategies	.598	.807
Participation of All Learners	.583	.809
Opportunities for Reflection	.367	.824
Classroom Materials	.303	.827
Management of Reading & Writing Instruction	.602	.807
Opportunities for Oral Language Development	.704	.799
Opportunities for Meaningful Reading	.497	.815
Opportunities for Learning to Decode & Spell Words	.304	.830
Opportunities for Developing Reading Fluency	.576	.809
Opportunities for Developing Vocabulary	.398	.824
Opportunities for Developing Comprehension	.598	.806
Writing Instruction	.079	.844

***Early Grade Reading Assessment (EGRA*).** To assess **student reading proficiency**, the English language EGRA, as well as the EGRA[4] adapted for Filipino language by RTI[5] was used in the baseline. EGRA is a diagnostic instrument designed to assess the foundation skills for literacy acquisition for the early grades according to the following subtests:

1. *Initial sound identification* assessed student's phonemic awareness (the ability to explicitly identify and manipulate the sounds of language). Phonemic awareness has been found to be one of the most robust predictors of reading acquisition and is often used to identify students at risk for reading difficulties in the primary grades in developed countries. In this subtask, students were asked to listen to a word and identify the first sound in that word. After two practice items, students were given ten test items. This subtest was not timed.

2. *Letter sounds* assessed students' knowledge of the sounds that the letters of Filipino alphabet make. Students were presented with a random mix of 100 upper case and lower case letters of the alphabet, and asked to identify what sounds those letters make. Only letter sounds, not letter names, constituted correct answers. The test was timed at 60 seconds; the score was the number of correct letters per minute.

3. *Familiar word reading* assessed student's skill at reading high-frequency words. Recognizing familiar words is critical for developing reading fluency. In this timed subtask, students were presented a chart of 50 familiar words. Students were asked to read as many words as they could. The subtest was timed at 60 seconds and yielded a score of percent correct and correct words per minute.

4. *Simple non-word decoding* assessed student's skills in decoding words they could not have memorized. Tested students were asked to decode a list of 50 pronounceable nonsensical words that followed legal spelling patterns of Filipino for the Filipino EGRA test or English for the English EGRA test. Students were asked to decode as many invented words as they could within 60 seconds. The scores were percent correct and correct words per minute.

5. *Oral passage* reading assessed student's fluency in reading a simple connected text aloud and their ability to understand what they had read. The passage length varied, with 55 words in the Filipino version and 60 words in the English version. The subtest was timed at 60 seconds and yielded a score of correct words per minute. In addition to determining the fluency of reading, data collectors marked a prosody score for each student on a four-point scale, from "word by word, slow, laborious" (1) to "fluent, with expression to mark punctuation and/or direct speech" (4).

6. *Reading comprehension* indicates how well the students understood what they read. After the students finished reading the oral reading passage, or the minute ended, the passage was removed and students were asked five questions with varying difficulty about the passage they just read, but primarily locator.

7. *Listening comprehension* is considered to be an important skill for reading comprehension. In this subtask, the test administrator read a passage to students. Students were then asked questions about that passage. The number of questions asked varied by test; three questions were asked in the Filipino version and five questions were asked in the English version. This subtest was not timed.

[4] See Annex 3 for the summary of EGRA subtests.

8. *Dictation* was designed to assess student's skill at spelling and basic writing rules, such as capitalization, punctuation, text direction, and spacing between words. The data collector read a short sentence to the students and students attempted to write the sentence. The data collector scored the dictation results after the child was finished with the test. This subtest was not timed.

The *EGRA* was administered for the comparison group during July 2013 (baseline) and again in February 2014 (endline). Subsequently in July 2014 the EGRA was administered to the intervention group (baseline) and then again in February 2015 (endline). See Assessment schedule below.

Table A-5. Assessment Schedule

	Assessment	July 2013	Feb 2014	July 2014	Feb 2015
Comparison cohort	Filipino EGRA	Baseline →	Endline		
Intervention cohort	Filipino & English EGRA			Baseline →	Endline

Assessors, supervised by BASA staff, participated in data collection. All assessors attended training in data collection procedures, including random selection of boys and girls from the classrooms for the student assessment. No disruptions to the data collection process were reported. To measure how well individual administrators graded the sub-tests similarly, inter-rater reliability (IRR) exercises were conducted during the training. All administrators took part in IRR exercises. During the group role play, administrators scored the mock child respondent and the trainer noted the variances in the scores for each of the subtests. Administrators with consistent discrepancies were given additional training, monitoring and support. Items with larger discrepancies were furthered reviewed with the larger group during practice sessions. In addition, during the practice testing with actual children, two administrators were paired together to score the same child respondent. Each administrator scored the respondent separately. At the end of the testing, the administrators compared scoring data and discussed discrepancies with the oversight of the trainers. Those administrators that were not consistent in their scoring by the end of training were not allowed to participate in the baseline and endline testing.

In addition, IRR during the actual data collection was conducted on 30 students in the comparison cohort (Year 1 – SY 2013/14), who were tested by two assessors. The mean ICC score was .830, and the median was 1, which indicates very strong reliability. The analysis of paired assessments at endline showed similar results, with the median ICC of 1.

In Year 2 (SY 2014/15), IRR was conducted on 166 students at baseline. IRR was consistent for both EGRA tests, which suggests that assessors consistently scored the same behaviors similarly regardless of the language the test was administered in, English or Filipino. The mean IRR score was .993 for both English and Filipino, and the median was .998 for English and .999 for Filipino, which indicates very strong reliability. At endline, the median was .995 for Filipino and .992 for English which again indicates very strong reliability.

Student Assessment Reliability Analysis. A statistical analysis of test reliability is used to describe an internal consistency of the test, and is based on the correlations between different items

(subtests). Internal consistency of the test is measured with Cronbach's alpha which is the result of pairwise correlations between items. Cronbach's alpha ranges from zero to 1, where zero denotes an absence of any correlation across items on the test, and 1 denotes a perfect correlation across items. A typical and acceptable range for Cronbach's alpha is above .8. A good internal consistency of a literacy assessment means that a child who scores higher on some items would also score higher on other items in the test.

A test of internal consistency of *EGRA* found that the overall test reliability was relatively high (Cronbach's alpha = .855 for the Filipino test at baseline and .818 for the English test at baseline). At endline, the Cronbach's alpha was .891 for the Filipino test and .901 for the English test. On both tests, the item level analysis showed that listening comprehension did not correlate well with other items. If removed from the test, the Cronbach's alpha would go up to .922 and .909 for the Filipino and English tests respectively. The second least correlated item for both tests was the letter sounds subtest. On the English EGRA, phonemic awareness and reading comprehension also did not correlate well the other subtests (Cronbach's alpha = .654 and .424 respectively). Remaining items on the Filipino and English EGRA tests correlated very well with the rest of the test.

Overall, bivariate correlations between subtests for each test (English and Filipino) were found to be statistically significant. Familiar Word and Nonsense Word reading was found to be highly correlated; they were also highly correlated with oral passage reading. Tables 8 and 9 below show the results.

Table A-6. Filipino EGRA Reliability, Intervention Cohort (SY 2014/15)

	Baseline		Endline	
EGRA Subtests	**Item-Total Correlation**	**Cronbach's Alpha if Item Deleted**	**Item-Total Correlation**	**Cronbach's Alpha if Item Deleted**
1. Initial sound identification	.614	.813	.670	.877
2. Letter sounds	.549	.802	.625	.887
3. Familiar word reading	.888	.732	.857	.856
4. Nonsense word reading	.860	.745	.799	.865
5. Oral passage reading	.860	.745	.850	.858
6. Reading comprehension	.561	.829	.663	.877
7. Listening comprehension	.443	.831	.302	.922
8. Dictation	.671	.816	.811	.864

Table A-7. English EGRA Reliability, Intervention Cohort (SY 2014/15)

	Baseline		Endline	
EGRA Subtests	**Item-Total Correlation**	**Cronbach's Alpha if Item Deleted**	**Item-Total Correlation**	**Cronbach's Alpha if Item Deleted**
1. Initial sound identification	.696	.851	.654	.893
2. Letter sounds	.602	.844	.672	.893
3. Familiar word reading	.932	.787	.882	.869
4. Nonsense word reading	.904	.795	.844	.873
5. Oral passage reading	.914	.794	.864	.871
6. Reading comprehension	.727	.862	.473	.905
7. Listening comprehension	.330	.869	.424	.909
8. Dictation	.792	.846	.764	.884

Table A-8. Correlations of Filipino EGRA Subtest for Intervention Cohort (SY 2014/15) at Endline

Filipino EGRA Subtests	ISI	Letter Sounds	Familiar Words	Nonsense Words	Oral Reading	Reading Comp.	Listening Comp.	Dictation
ISI	1							
Letter sounds	.605**	1						
Familiar words	.613**	.552**	1					
Nonsense word reading	.564**	.539**	.904**	1				
Oral passage reading	.603**	.526**	.964**	.888**	1			
Reading comprehension	.497**	.464**	.560**	.493**	.571**	1		
Listening comprehension	.322**	.320**	.277**	.242**	.288**	.460**	1	
Dictation	.637**	.535**	.831**	.793**	.819**	.550**	.305**	1

** Correlation is significant at the p<.01 level

Table A-9. Correlations of English EGRA Subtest for Intervention Cohort (SY 2014/15) at Endline

English EGRA Subtests	ISI	Letter Sounds	Familiar Words	Nonsense Words	Oral Reading	Reading Comp.	Listening Comp	Dictation
ISI	1							
Letter sounds	.631**	1						
Familiar words	.616**	.608**	1					
Nonsense word reading	.569**	.611**	.887**	1				
Oral passage reading	.631**	.616**	.924**	.836**	1			
Reading comprehension	.311**	.308**	.428**	.412**	.414**	1		
Listening comprehension	.305**	.333**	.349**	.362**	.332**	.351**	1	
Dictation	.523**	.502**	.743**	.714**	.697**	.458**	.431**	1

** Correlation is significant at the p<.01 level

EGRA was programmed into tablets using SurveyToGo software, and sampled students were tested on a one-on-one basis by a trained assessor using a tablet.

Student Assessment Intra-Class Correlation (ICC) Analysis. While there are many uses for the intra-class correlation coefficient (ICC), we looked specifically at its use to determine the amount of clustering in a nested design. The present assessment drew students from classrooms, so we looked at students nested within classrooms model.

As students within classrooms tend to be similar, their test scores cannot always be considered as statistically independent from each other. This violates the assumption built into OLS regression models that cases are indeed independent. This leads to false precision in the estimated standard errors of the regression coefficients.

To properly model the data, in the presences of an obvious cluster effect, we turn to multi-level, or mixed models. To examine the extent to which clustering impacts the data (so we can determine which model to use) we calculate the intra-class correlation coefficient. The results of the EGRA assessment data analysis revealed the ICC for Philipino EGRA to be 0.1458, and for English EGRA to be 0.1762.

DATA ANALYSIS

All collected data were cleaned by EDC M&E staff and analyzed using standard statistical techniques such as univariate and bivariate statistics as needed for different analytical purposes. The results were disaggregated by sex, and province, as appropriate. Univariate and multivariate statistical analyses for were conducted. Central tendency analysis (e.g. mean, median) were conducted for continuous demographic variables. Comparison of means statistical tests (independent samples *t*-test) were conducted to estimate differences between groups such as province and sex, where appropriate. Bivariate statistical analyses (e.g., correlations) were conducted to examine the relationship between different variables.

LIMITATIONS

Since the evaluation design does not include random assignment of teachers and students into participant and non-participant groups to assess the true impact of the program, the attribution of the observed outcomes to the program will be limited since other factors may have contributed to the changes in the studied outcomes.

ANNEX 2. BIPI FINDINGS AND RESULTS

Teacher beliefs are known to impact their instructional practice. To better understand what teachers think about their students' abilities, the appropriateness of different instructional methods for teaching literacy to students, as well as their own classroom practice, Basa conducted a Beliefs and Instructional Practices Inventory (BIPI) survey with 82 teachers in the sample EGRA classrooms between July and August, 2014. The survey consisted of the following sections:

Section A. Teacher's demographic information
Section B. Frequency of use of literacy-related instructional practices in the classroom
Section C. Series of statements about teaching literacy
Section D. Series of statements about students' abilities in relation to literacy

This Annex presents the detailed results of the statistical analysis of the survey data for sections B, C and D.

AVAILABILITY OF RESOURCES AND TEACHERS' LITERACY PRACTICES

Access/Availability of Resources[6]

Curriculum and textbooks were reported to be accessible by the majority of teachers; however other types of reading materials were less accessible. While more than 90% of teachers reported having regular access to the K- 12 curriculum in their Mother Tongue, Filipino and English, only about 36% said they had access to a library or reading center. Analysis by income classification showed that there was a relationship between the income level of the district and access to libraries and reading centers. For instance, teachers working in high income areas were more likely to have access to a library or reading center, with 50% of teachers in high income areas reporting that they had access compared to 35% of teachers working in medium income areas, and 31% of teachers working at schools in low income areas.

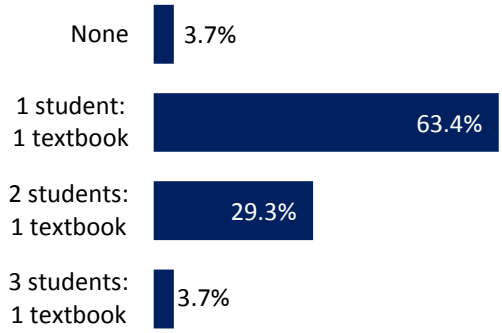
When analyzed by province, nearly all respondents in Cebu, Ilocos Norte, Ilocos Sur, La Union and Mandaue City said they had access to the K- 12 curriculum compared to approximately 70% in Bohol. In terms of access to a library/reading center, responses were largely consistent across province with roughly a third of teachers from each province reporting access to a library with the exception of Bohol, where 60% of teachers reported having access to a library in their community.

Similarly, nearly all teachers (92.7%) reported that they had enough reading learner manuals for every student (63.4%) or every two students (29.3%). Only around 40% reported they had sufficient reading materials (e.g. picture books or story cards) for every one to two students. Responses were consistent by province and income classification.

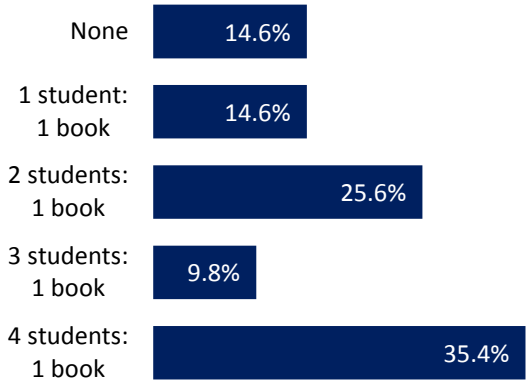
Figure A-1. Availability of Reading Learner Manuals **Figure A-2. Availability of Other Reading Materials**

[6] This section includes #23, 27, 28, 29

(n=82)



(n=82)

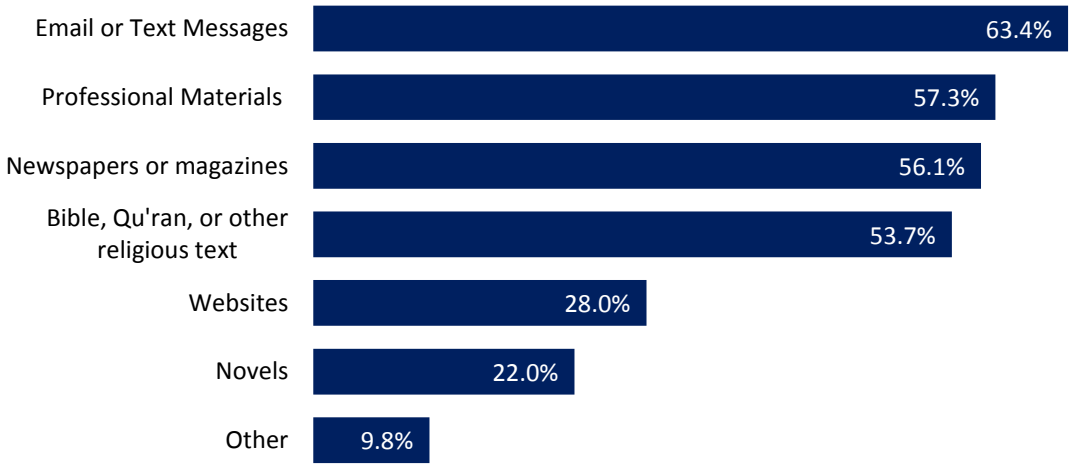


Teacher Literacy Practices[7]

TEACHER LITERACY PRACTICES OUTSIDE THE CLASSROOM

The vast majority of teachers surveyed had a radio (85.4%) or television (98.8%) in their home or neighborhood that they could access regularly. Most teachers said they read for pleasure outside the classroom (75.6%); the majority did so 1 or 2 times a week (67.2%). Across the seven different types of reading material presented in the survey (including "other"), teachers were most likely to read emails or text messages (63.4%), professional materials (57.3%) and newspapers/magazines (56.1%).

Figure A-3. Types of Reading Material Read for Pleasure (n=81, multiple response)

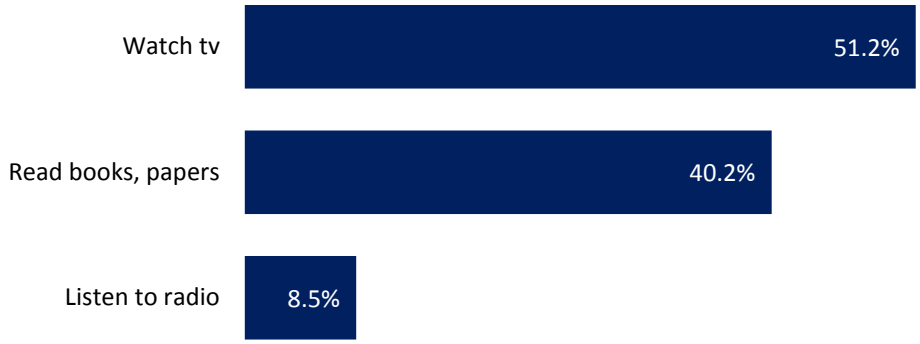


"Other" responses mostly centered on short stories/ story books, and education-related materials, such as encyclopedias, education/teacher journals, and textbooks. The number of reading materials read per respondents varied. The majority (64.6%) of teachers read one to three types of materials on a regular basis, though over a fifth (22.0%) reported reading four different types of materials.

[7] This section includes #18, 19, 20, 21, 22, 23, 24,25, 26, 30

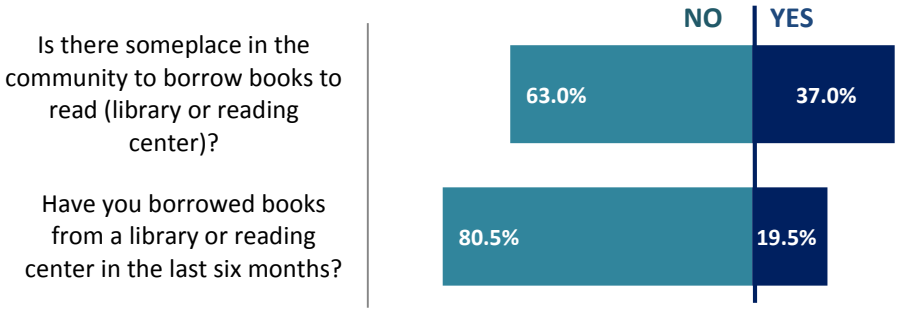
When asked to pick one type of activity they preferred to do (read, listen to radio, watch television), responses were fairly split between reading (40.2%) and watching television (51.2%) – few teachers preferred to listen to the radio (8.5%). Responses by teachers were fairly consistent across province, with the exception of Ilocos Norte where a large majority (83.3%) of teachers preferred to watch television. Conversely, teachers in Mandaue City reported that they preferred reading to watching television.

Figure A-4. Preference of Free Time Activity, by Province (n=82)



Although interest in reading seems to be high, only one third (37.0%) of respondents said they had access to a library or reading center from which to borrow books. Even fewer teachers (19.5%) reported borrowing books from a library or reading center. Respondents in Bohol were more likely to have access to a place to borrow books and more likely to have borrowed books than respondents from other provinces.

Figure A-5. Access to Reading Materials in the Community (n=82)



In terms of writing, teacher respondents wrote texts or emails more frequently than letters, reports, or documents. While half of respondents reported sending a text or email more than 20 times a week, the same number of respondents wrote one or less letters, reports or documents since the start of the school year.

TEACHER LITERACY PRACTICES INSIDE THE CLASSROOM

The large majority of teachers reported putting students into smaller groups based on reading level. Of those teachers who put students in groups, most reported doing so frequently; 87% stated they did so multiple times a month. Teachers were also asked how frequently they conducted literacy

instruction in smaller groups; the majority did so most of the time (39.1%) or some of the time (37.7%). Only 5.8% of teachers reported that they did not conduct literacy instruction very often.

Figure A-6. Respondent Puts Students Into Smaller Groups by Their Reading Level (n=81)

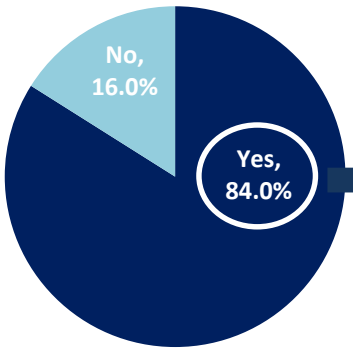
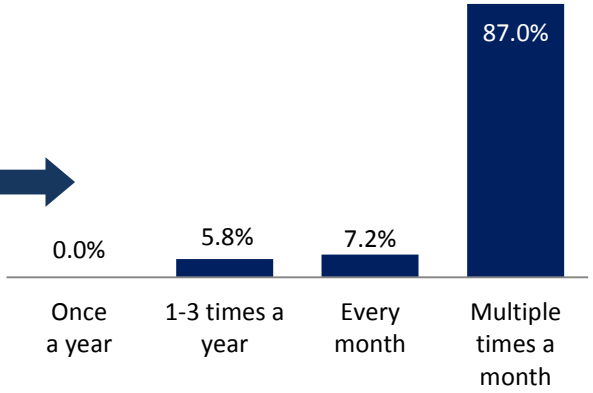
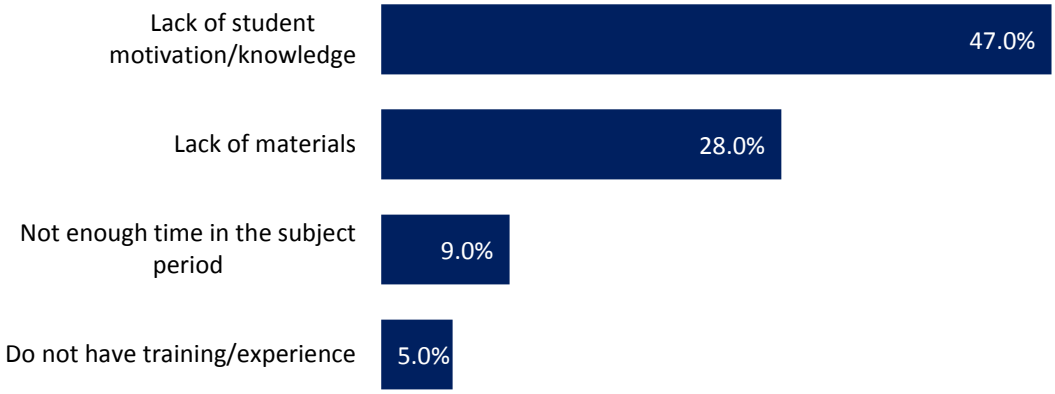


Figure A-7. Frequency of Small Reading Groups (n=69)



Teachers reported that teaching English was difficult primarily because of lack of student interest and lack of materials. Many students were not used to speaking English given that it wasn't their mother tongue. Teachers also added that lack of support and reinforcement of English language learning at home by parents also made it challenging to teach English.

Figure A-8. What makes it difficult to teach English?



BELIEFS ABOUT TEACHING LITERACY (SECTION C)

The beliefs sections of the BIPI survey (Sections C and D) explored how strongly respondents identified with research and evidence-based best practices around literacy, based on responses to a series of statements to which they could agree, disagree, or have no opinion. It is interesting to note that a widely used approach to teaching reading in the Philippines' schools emphasizes recitation and memorization, which is not the case in the evidence-based instruction. Part of the intention of using the BIPI was to understand where teachers' beliefs aligned with the evidence-based practices and where they did not.

Fourteen statements in Section C that are commonly supported in current research as key literacy teaching practices were selected to measure beliefs around teaching literacy. Given that these questions are based on the current research on literacy acquisition, a "correct" answer was

identified. During the data analysis, all answers were recoded into dichotomous variables 1/0 where 1 represented a correct answer, and 0 represented an incorrect answer, undecided or no answer. A composite score was created from the 14 statements selected.[8] This score was converted into a percent of correctly answered questions from the total number of questions (14) in the composite.

The following sections display cumulative findings for the five provinces (Bohol, Cebu, Ilocos Norte, Ilocos Sur and La Union) and Mandaue City. During analysis, data were disaggregated by province. Any major differences in findings between provinces are noted in the appropriate section.

Analysis showed that on average, teachers answered around eight out of 14 statements correctly (57.1%). However, further examination of Section C results showed that literacy practices are varied with scores ranging from 21.5% correct to 85.7% correct. Later sections will take a deeper look into responses based on subtopics of literacy.

Results were fairly consistent when analyzed by province.

TEACHER BELIEFS ABOUT ABILITIES OF THEIR STUDENTS (SECTION D)

Teacher responses to this section of the survey help better understand what expectations teachers set for their students, and what skills they view as essential. The following topics were covered in this section of the survey:

- Literacy beliefs (reading, writing, foundational skills and language, comprehension and higher-order thinking);
- Literacy practices ;
- Training, gender and disability

BIPI data also showed that many teachers possess a mixture of beliefs about literacy development and instruction—some of which are supported by the research literature and some that are not. Data analysis showed that students in the earlier grades (before Grade 1 or in Grade 1) were mostly viewed as able to perform basic language and reading skills, such as letter recognition, basic reading and answer simple oral questions about the text. For example, more than one third of respondents (37.8%) felt students were able to recognize letters and sounds before Grade 1 and around 59% felt students were able to do so in Grade 1.

Comprehension and other higher order thinking skills were, for the most part, seen as abilities that younger students did not yet possess. Most respondents did not feel students entering first grade or in the first three months of Grade 1 were able to write original pieces, hypothesize or predict a story, explain what they liked or didn't like about a story or text, infer or deduce meaning of new words by looking at how they are used in a sentence, or other skills that involved authentic thinking. Fairly equal numbers of teachers responded that these students were able to achieve these skills by the end of Grade 2, or in Grade 3.

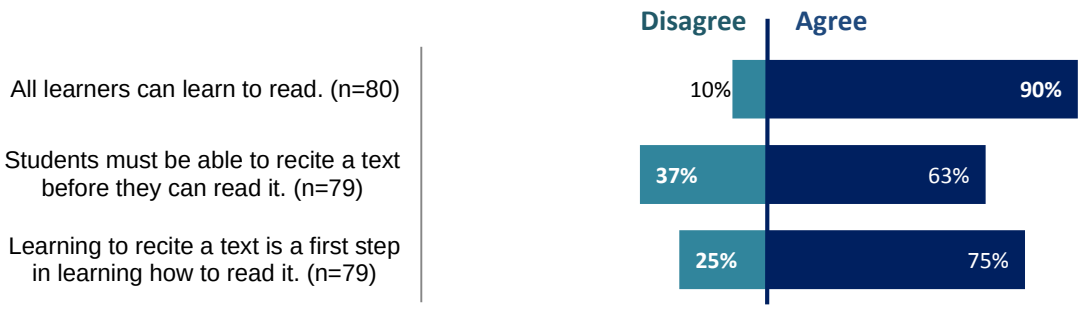
[8] Fourteen items from Section C (#33, 34, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, and 55) were selected for the composite, with correct answers coded as 1, incorrect answers coded as 0, and the total computed.

LITERACY BELIEFS

Reading[9]

A large percent of respondents (90%) felt that all learners could read, though more than half of respondents (52.4%) felt that most students found it difficult. Responses indicated that teachers were largely supportive of recitation practices for literacy – 63% of teachers agreed that a student needs to be able to recite a text before reading it and three-quarters felt reciting a text was the first step in learning how to read it (75.0%). Responses were fairly consistent across province, with the exception of respondents from Ilocos Sur and La Union who were substantially more supportive of recitation practices for literacy; over 80% of teachers from both provinces supported recitation to teach literacy.

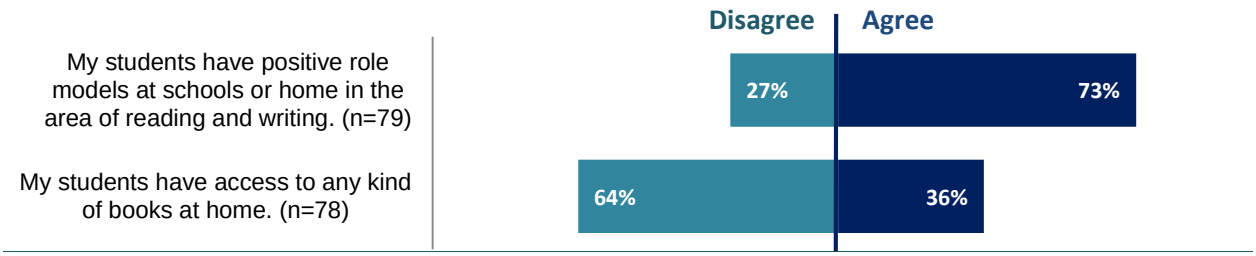
Figure A-9. Selected Beliefs in Learning How to Read



Most respondents felt that students were capable of reading out loud a simple text (two to three sentences) that they had never seen before in Grade 1 – either during the first three months (28.0%) or by the end (38%). Reading texts of the students' own choosing was seen as a skill for slightly older students – those at end of Grade 1 (39%) or end of Grade 2 (43%).

The majority of teachers (73%) reported that their students had a positive role model at school or at home in the areas of reading or writing. However, results show that access to books at home may be an issue, with only one third (36%) of respondents agreeing that students have access to books at home. Responses were consistent across province.

Figure A-10. Selected Beliefs in Home Literacy Environment



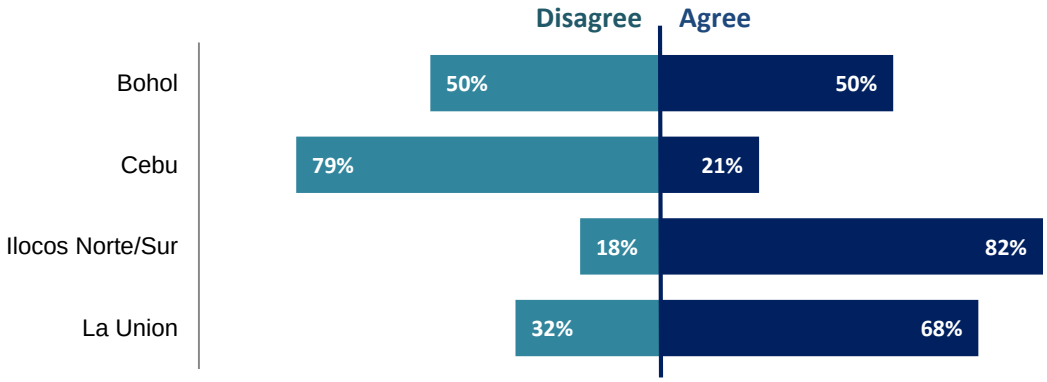
[9] This section includes #33, 36, 44, 53, 61, 62, 63, 73

Writing[10]

Teachers were also asked about students' writing abilities. Respondents largely felt that students could learn how to write, with only 27% of teachers who reported that they felt that students had difficulty learning how to write. Despite teachers' reported confidence in students' ability to learn to write, nearly two-thirds (63%) of teachers felt students needed to learn to read before learning to write. Similarly, only one half (51%) of respondents felt that all learners could learn to compose their own stories.

When analyzed by province, responses on whether all learners could learn to compose their own stories varied substantially. Respondents in Cebu were the least likely to agree that all students could learn to compose their own stories, with less than one quarter of respondents in Cebu who agreed with this statement. Conversely, the majority of respondents in Ilocos Norte/Sur felt that student could compose their own stories. Respondents in Mandaue City largely agreed that students could learn to compose their own stories as well, with 67% of respondents that agreed with that statement.

Figure A-11. All learners can learn to compose their own stories, by province



Many teachers believed that spelling correctly was extremely important. Only around a third (35.4%) agreed that spelling errors made when attempting to write for the first time were not a major concern. The vast majority (88%) felt correcting all errors in a student-produced sentence was important. That said, most teachers (78%) did not think that writing "well" meant perfect spelling and grammar.

Students in the early grades were seen as able to write original sentences. Only one third (32%) of respondents felt students couldn't do so until at least Grade 3 or 4. When asked in which grade students could write an original text of two or more sentences, responses were fairly evenly distributed between end of Grade 1, end of Grade 2, and Grade 3.

Foundational Skills and Language[11]

For the purpose of this analysis, foundational skills and language include the ability to:

[10] This section includes #34, 37, 43, 46, 47, 49, 55
[11] This section includes #65, 68, 69, 70, 74, 75

- Recognize all the letters of the alphabet and the sound each letter makes
- Spell common or frequently encountered words correctly
- Use common punctuation (period, question mark, exclamation mark) correctly in their original productions
- Infer or deduce the meaning of a new word by looking at how it is used in the sentence
- Decode new words without the teachers' help by making correct letter-associations
- Recognize and read common or frequently encountered words

Respondents were asked whether students were able to perform these skills before Grade 1, within the first three months of Grade 1, end of Grade 1, end of Grade 2, or Grade 3. There was also an option to state that the skill was not important.

By and large, respondents found each of the skills important. Recognition of letters was seen as a skill very young children should be able to do – a third (38%) of respondents felt students should be able to do so coming in to first grade and a quarter (26%) thought students should have it within the first three months.

For the remaining skills, most respondents did not think children were able to possess them before Grade 1 or within the first three months of Grade 1.

Comprehension and Higher-Order Thinking[12]

Comprehension and higher-order thinking skills include students being able to:

- Understand the meaning of texts they are reading
- Understand a text without memorizing it first
- Express their opinions on a text they have read
- Express their opinions about a text that the teacher has read to them
- Make a hypothesis or a prediction about what a text or story is about by looking at the title or the illustrations
- Explain what they liked or didn't like about a story or text they have read
- Explain what they liked or didn't like about a story or text they have had read to them
- Make predictions about what will happen next in a text or story
- Answer simple oral questions (where a text takes place, who are the main characters, when it takes place…) about a text they have read
- Answer simple oral questions (where a text takes place, who are the main characters, when it takes place…) about a text they have read to them

Very few teachers (22.0%) agreed that students must memorize a text before they could understand it and some felt that younger students (before Grade 1 or in first three months of Grade 1) could answer simple questions about texts or stories (read to them or read on own). However, respondents generally did not believe that younger children were able to have authentic and original thoughts such as understanding meaning of texts read, expressing opinions around texts, making hypothesis or predictions about a text or story by looking at the title or illustrations, explaining what they liked or didn't like about stories or texts, or making predictions.

[12] This section includes #52, 64, 71, 72, 76, 77, 78, 79, 80, 81

LITERACY PRACTICES[13]

Literacy practices delve into how teachers support the reading and writing environment in their classroom. Responses were mixed in this section. Many teachers expressed a belief in research-supported best practices, such as giving students time each day to read freely materials of their own choosing (81%), giving students time each day to write freely on topics of their own choosing (83%), discussing what students know about the subject addressed in a new text before reading it (95%), reading stories to students even after they can read on their own (76%).

However, around half (51%) still believed that teaching reading and writing as two separate subjects would cause less confusion for students and two-thirds (66%) of teachers believed silent reading should be avoided (because the teacher couldn't check if students were actually reading or reading correctly). Respondents in Bohol and Cebu were more likely to hold this belief than their counterparts in Ilocos Norte/Sur, La Union and Mandaue City.

Teachers supported having students review a classmate's text in order to help him/her correct spelling or grammar mistakes, though most felt this was a skill students were best able to perform in later grades (the end of grades 2 or 3).

TRAINING, DISABILITY, AND GENDER[14]

Half or less of respondents felt they had received adequate training on how to teach reading (50%) or writing (39.0%). However, nearly all teachers (90%) felt they often had opportunities to engage with colleagues about how to teach reading or writing.

Few respondents had received specific training in working with students with disabilities (26%). Particularly in Bohol and Cebu, only 15% and 20% respectively of teachers reported receiving training on teaching students with disabilities. More than half (52.4%) of surveyed teachers believed in the inclusion of students with disabilities in the regular classroom and literacy activities.

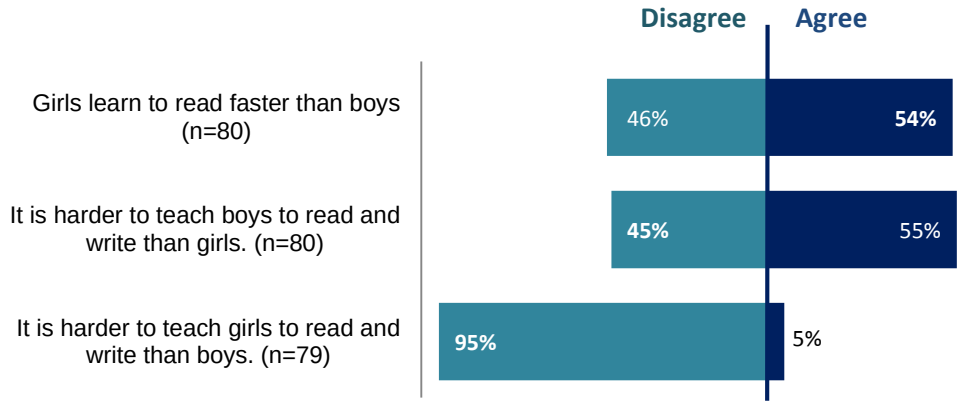
More respondents had gender training (48%) than disability training, albeit the majority of teachers reported not having gender training. Provincially, the majority (90%) of La Union teachers reported having received training on how to promote gender equity in their classroom, compared to Bohol and Cebu which had substantially less teachers trained (15% in Bohol and 30% in Cebu).

To assess gender bias in literacy beliefs and practices, teachers were asked to agree or disagree to a series of statements related to gendered beliefs around literacy and literacy instruction. In terms of literacy instruction, a large majority of teachers (89%) felt boys and girls should not be separated during reading activities. However, more than half of teachers (55%) felt it was harder to teach boys to read and write than girls. Conversely, almost no teachers (5%) felt that girls were harder to teach to read and write than boys. Responses were split as to whether girls learned to read faster than boys. This apparent gender bias in the expectations of teachers of students' literacy skills is an

[13] This section includes #42, 45, 48, 50, 51, 54, 67
[14] This section includes #35, 38, 39, 40, 41, 56, 57, 58, 59, 60

important concern given that research has shown that teacher expectations is an important predictor of later reading achievement.[15]

Figure A-12. Selected Beliefs on Literacy and Gender



Detailed statistical results of the BIPI survey are found in the original BIPI report produced as part of 2014/2015 cohort baseline assessment.

[15] Hinnant, B. J., O'Brien, M., Ghazarian, S.R. (August, 2009). The longitudinal relations of teacher expectations to achievement in the early school years. Journal of Educational Psychology, 101 (3) 662-670.

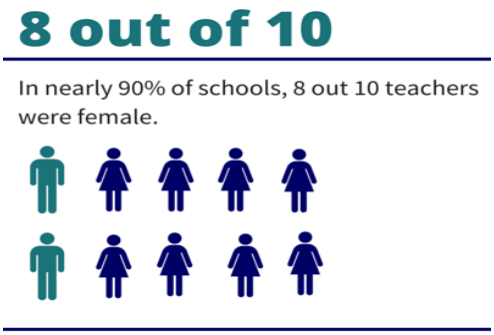
ANNEX 3. SCHOOL DEMOGRAPHICS

SCHOOL ENVIRONMENT

The school's environment and management is critical to understanding the teaching and learning that is taking place in the school. The Principal Survey[16] was designed to capture information on: 1) the physical infrastructure, 2) the teaching and learning demographic, 3) the school manager's background and characteristics, 4) school policies, practices, and monitoring, 5) the reading environment, 6) parent and community involvement, and 7) disabilities and gender policies and practices.

The survey was administered one-on-one with each of the 83 principals (or other school leaders) present at the time of the baseline EGRA and BIPI data collection in intervention schools in July/August 2014. Twenty principals were included from each province with the exception of Bohol where only 19 principals were surveyed. Four principals were also included from Mandaue City. Schools included in the sample were from a range of income classifications, with 23 schools each from high income and medium income areas; in addition 37 schools were included from low income areas.[17]

Figure A-13.Teaching Force, by Sex (n=83)



The majority of schools had between 5 and 20 teachers. Only five of the schools had more than 21 teachers. The teaching forces at the sample schools are largely female; in nearly 90% of schools, 8 out of 10 teachers were female. Only one school had more male teachers than female teachers.

Nearly three quarters of sample schools had zero school closures since the start of the current school year, but 25% reported one to two days of closure. The majority of schools reported that schools were closed due to natural disasters (typhoon) in which schools were affected.

A quarter of sampled schools reported having undergone some type of *EGRA* testing prior to the baseline, between the years 2010 and 2014.

In terms of teacher attendance, sixteen principals reported that on the day of the survey, one to two of their teachers had been absent. One principal indicated that 6 teachers had been absent. Six principals reported that one to two of their teachers had been late on the day of the survey.

[16] Note that the Principal's Survey used was modeled on the Snapshot for School Effectiveness tools used in a number of different countries. RTI was not using a Principals' Survey in Philippines at the time of data collection, so Basa developed its own tool.

[17] High income areas are defined as municipalities with an average annual income of 55₱ million or more. Medium income areas are municipalities with average annual income between 35 and 45₱ million. Low income areas are municipalities with average annual income between 15 and 35 ₱ million.

READING ENVIRONMENT AND MATERIALS

Adequate levels and timeliness of receipt of learning materials remains a challenge for schools, with only 12% of schools that reported that they had the appropriate number of textbooks for all students at the beginning of the year, according to the current Department of Education (DepEd) policy. When analyzed by income category, schools from high income areas reported that they were more likely to receive the appropriate number of textbooks for all students (26%) compared to schools from medium income (4%) and low income (8%) areas.

The majority of schools (72%) reported that their school received the appropriate number of textbooks for one or more grades, but not for all grades. Of the 73 schools that reported that they did not have enough materials, 71% had yet to receive the books three months after the school year started. It should be noted that the principals were asked about whether they received enough books at the beginning of the school year (timeliness of receipt), while the teachers (BIPI survey) were asked for the student to book ratio (sufficient texts).

Less than half of schools (42%) reported having a school library. Of those schools that did have a library, the majority of students were allowed to read library books at school, largely in the classroom or library. Fewer schools allowed students to read library books at home.

Figure A-14. Where can students read library books? (n=35, multiple response)

DESCRIPTION OF THE SCHOOL POLICIES, PRACTICES, AND MONITORING

The principals reported using official forms for collecting data in their schools, which suggests that monitoring procedures are clearly defined and followed, especially for teacher and student attendance.

Table A-10. Use of an Official Form (n=83)

Type of Attendance Data	% of schools
Student attendance	100%
Teacher attendance	98%
Student enrolment	86%
School census	76%

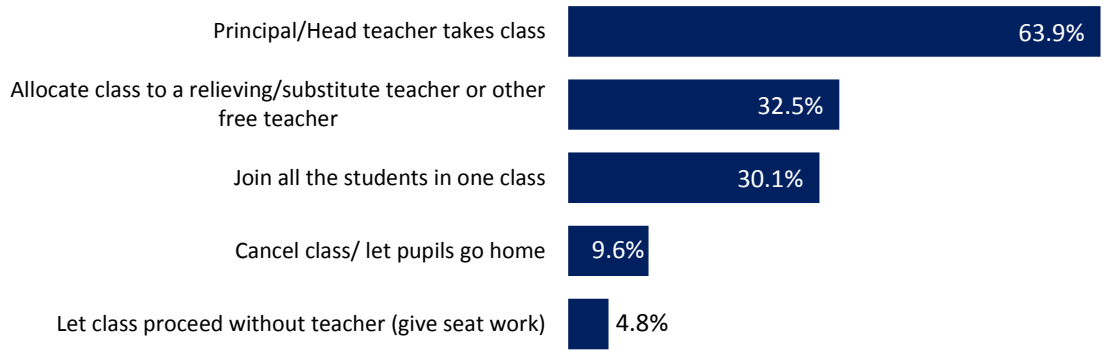
Roughly two thirds of schools reported that either the principal or an assigned teacher was responsible for tracking teacher attendance. The remaining schools reported that teachers were responsible for either recording their own attendance or signing a logbook. One school reported that the guidance counselor kept track of attendance. In nearly every sampled school, the teacher attendance records were easily available for the interviewer to review.

On average the principals reported the following practices for collecting and processing data:

- **Record teacher attendance:** completed daily
- **Compilation of student attendance data:** 1-2 times a month
- **Verify teacher lesson plans:** at least once every week
- **Observation of classrooms:** at least once every once every week (21% of principals reported visiting classrooms daily)

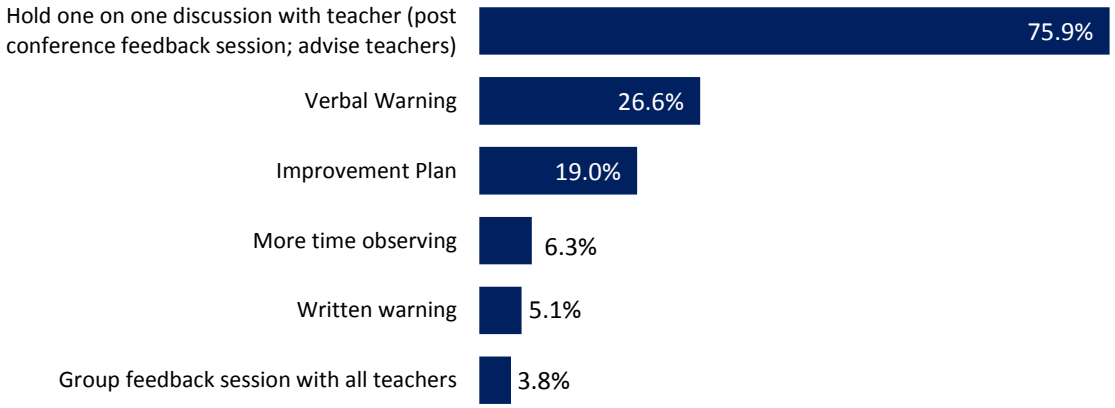
When a teacher is absent, the majority of principals reported that they themselves assume the absent teacher's class or that they assign the class to a relieving/substitute teacher. Several principals also reported that they combine classes. Very few reported that they allow pupils to go home or proceed without a teacher.

Figure A-15. What do you do with a class whose teacher is absent? (n=83, multiple response)



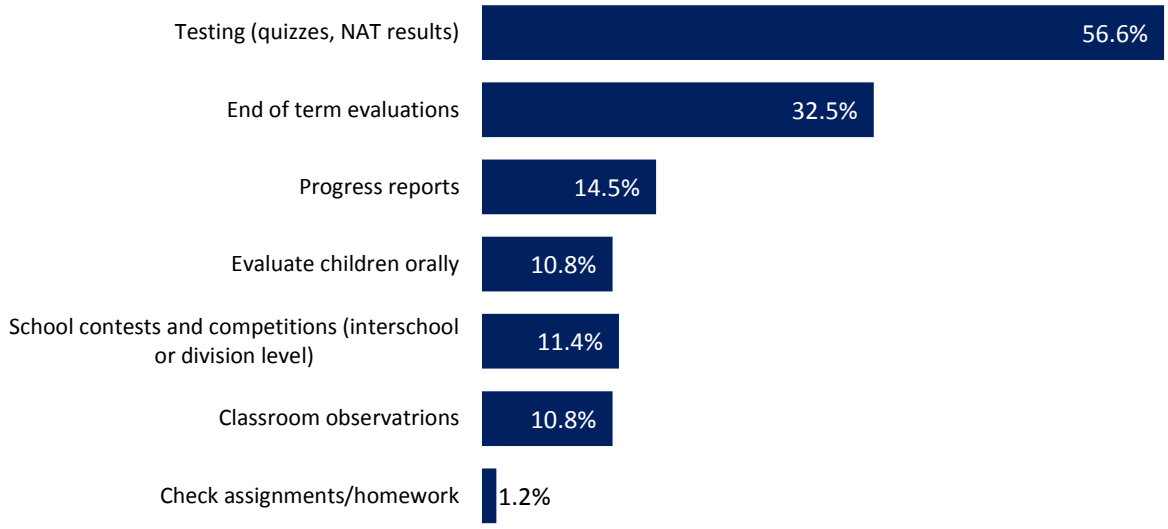
Only four principals reported ever having had to sanction a teacher, a process that ranged from two weeks or less to two to three months. In the cases where principals had grievances with their teachers they reported using the following courses of action:

Figure A-16. If you are VERY dissatisfied with a teacher's performance in the classroom, what are the immediate actions you take? (n=83, multiple response)



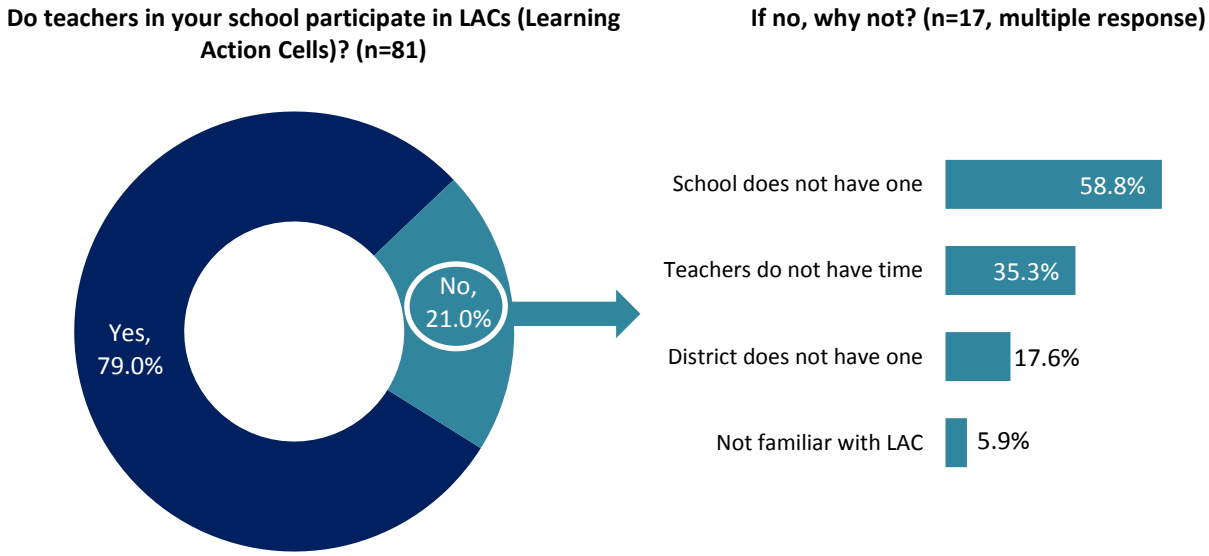
The principals said they monitored student progress through various methods, the main method being testing and end of term evaluations. Note that testing includes end of lesson quizzes, as well as high stakes testing, such as the National Achievement Test (NAT).

Figure A-17. How do you know whether your students are progressing academically? (n=83, multiple response)



The majority of principals reported that teachers in their schools participate in Learning Action Cells (LACs). Seventeen schools reported that they did not participate in LACs, largely due to the fact that schools did not have LACs or because teachers did not have time to participate.

Figure A-18. Learning Action Cells



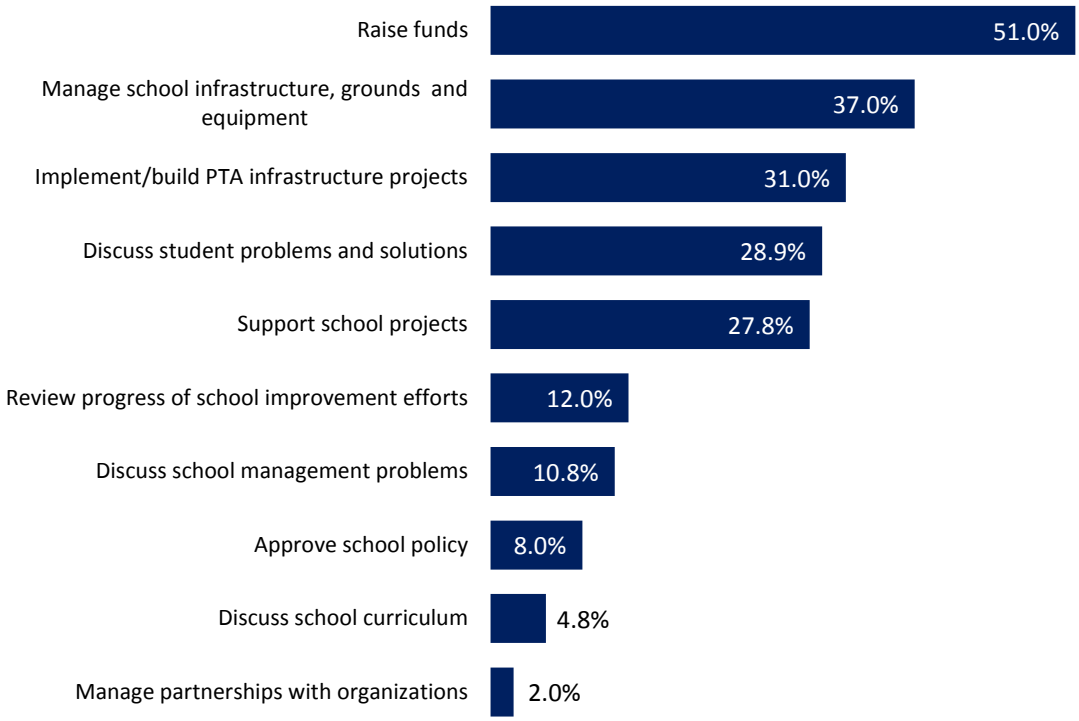
For those schools who participate in LACs, principals stated that they assumed various roles, ranging from lead instructor, to overseer/manager, to facilitator, to no role at all. LACs convene one to two times a month in the majority of surveyed schools.

PARENTAL AND COMMUNITY INVOLVEMENT

Nearly all schools (98.8%) had a parent teacher association (PTA) that met at various intervals during the year. On average, principals reported that PTAs met five times in the past school year. The major roles of the PTA, from the perspective of principals, were raising funds, managing school infrastructure and school improvement, troubleshooting problems and supporting school projects.

The following figure details various tasks and the percentage of principals that agreed that this was one of the PTA's roles.

Figure A-19. What are the roles of the PTA of your school? (n=83, multiple response)

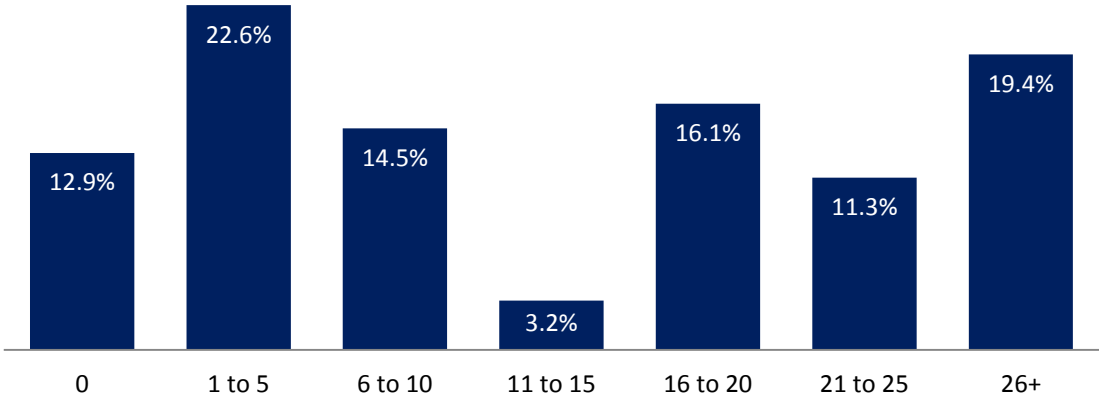


In regards to the overall support provided by the PTAs, nearly all principals sampled were either satisfied or very satisfied with the support provided to their schools.

In the last school term, 63% of the principals reported that they had met with the majority (76-100% of parents) of their students' parents or guardians at some point during the previous school year. Overall, principals were satisfied with the level of parental involvement (PTA or non PTA) in their children's schoolwork.

In addition to PTA and general parental involvement, schools were asked if they kept log books and if official visitors had visited their school. Three-quarters of the schools kept visitor logbooks. The number of visits by official visitors varied substantially by school, ranging from zero to 100 visits in the last school year. On average, schools reported frequent visits with approximately 18 visits from official visitors. Schools in high income areas on average received more official visits than schools in medium and low income areas. Visits by DepEd officials were namely district supervisors, other supervisors, and medical professionals.

Figure A-20. Number of School Visits (n=62)



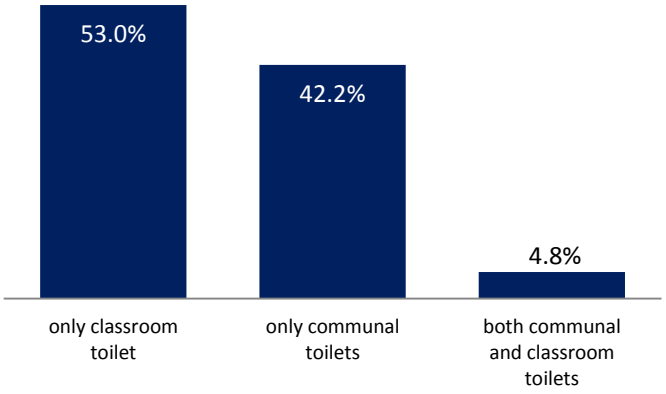
The major reasons for the visits were to:

- Check school infrastructure (water, toilets, etc.);
- Observe a class in session;
- Check lesson plans and school financial records;
- Check recent student assessment tests and evaluation process.

ISSUES RELATED TO GENDER AND DISABILITIES

Only 25% of the schools had a written gender policy in place. More than three-quarters of schools with gender policies also had guidelines on how to implement the policy.

Figure A-21. School Infrastructure (n=83)

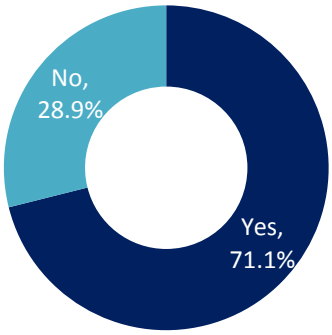


In regards to whether the physical infrastructure was both boy and girl friendly, principals were asked whether their schools had classroom or communal toilets. Slightly more principals responded that their schools had classroom toilets instead of communal toilets. Of the schools who had communal toilets, 80% of schools had separate communal toilets for boys and girls; the remaining schools reported that girls and boys had to share the same toilets.

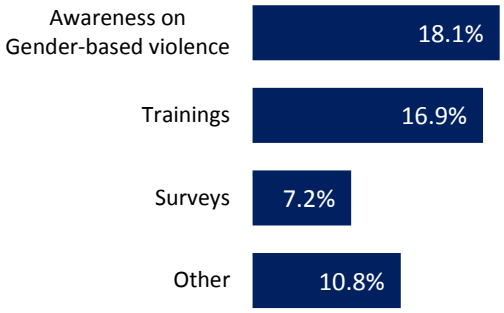
Nearly three-quarters of the principals surveyed reported having attended some form of gender awareness training; the majority of the trainings had been facilitated by DepEd.

Figure A-22. Gender Awareness and Initiative

Have you attended any training on gender awareness (i.e. classroom equity, etc.)? (n=83)



What kind of gender-related awareness activities have you initiated? (n=83)



A small percentage (36%) of principals had experience organizing gender awareness activities such as trainings, surveys or polls on gender issues, or awareness related to gender based violence. Additional activities reported by principals include, establishing a committee on child protection, integration of gender awareness into lessons and scouting, and establishing policies on child protection and anti-bullying.

Principals were also asked whether they track students with disabilities or special needs at their school. The majority of schools reported that they do track students with special needs or disabilities. Approximately, eight out of ten schools reported that they had at least one student with a disability or special needs. The large majority of sampled schools had between one and five students with a disability or special needs.

Figure A-23. Schools that Track Students with Disabilities or Special Needs (n=83)

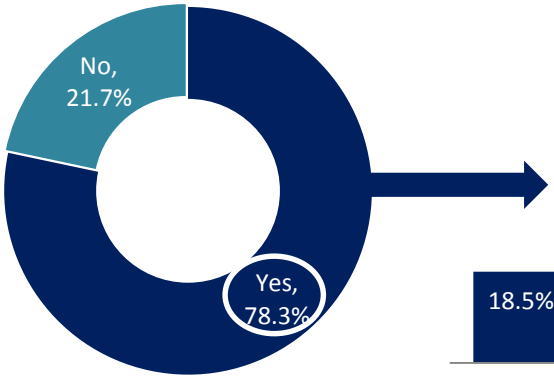
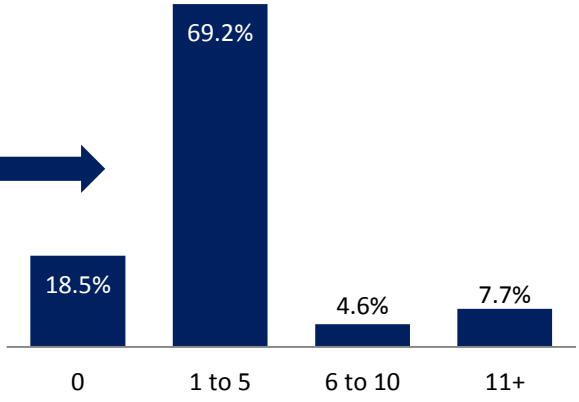
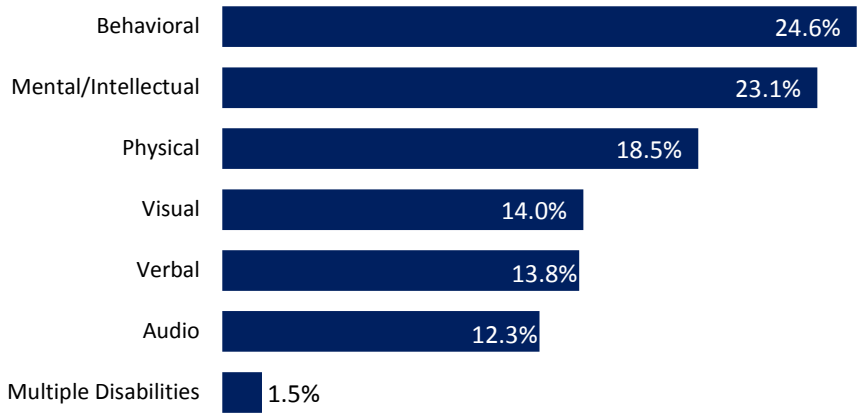


Figure A-24. Number of Students with Disabilities or Special Needs (n=83)



Principals reported that the types of disabilities that students in their schools had varied greatly. The most common disabilities were behavioral (ADD/ADHD), mental intellectual (autism, Down Syndrome), and physical disabilities (cerebral palsy, spinal cord injury, etc.).

Figure A-25. What types of disabilities do the students have? (n=65, multiple response)



When asked what the two main things they look for when they observe classrooms, the majority of principals responded that they observe instructional practice and classroom management. No principal reported that they observed for attention to gender and disability in the classroom.

ANNEX 4. EGRA SUBTESTS

Table A-11. Filipino EGRA Subtests

#	Description (Instrument)	Tasks/Max. Pts.	Reported %/wcpm	Timed
1	Initial sound identification	10 letters/sounds	percent: value / 10 * 100%	No
2	Letter Sound Knowledge	100 letters	percent: value * 1% lcpm: value / timer * 60	Yes (60 sec.)
3	Familiar Word Identification	50 words	percent: value / 50 * 100% wcpm: value / timer * 60	Yes (60 sec.)
4	Simple Non-word decoding	50 words	percent: value / 50 * 100% wcpm: value / timer * 60	Yes (60 sec.)
5A	Oral Passage Reading	56 words (baseline comparison) 55 words (baseline intervention) 64 words (endline comparison) 55 words (endline intervention)	percent: value / # of words in the text * 100% wcpm: value / timer * 60	Yes (60 sec.)
5B	Oral Reading Comprehension	5 questions	percent: value / 5 * 100%	No
6	Listening Comprehension	3 questions (Filipino)	percent: value / 5 * 100%	No
7A	Dictation (spelling)	12 words (Filipino)	percent: value / 13 * 100%	No
7B	Dictation (conventions of text)	4		No

Table A-12. English EGRA Subtests

#	Description (Instrument)	Tasks/Max. Pts.	Reported %/wcpm	Timed
1	Initial sound identification	10 letters/sounds	percent: value / 10 * 100%	No
2	Letter Sound Knowledge	100 letters	percent: value * 1% lcpm: value / timer * 60	Yes (60 sec.)
3	Familiar Word Identification	50 words	percent: value / 50 * 100% wcpm: value / timer * 60	Yes (60 sec.)
4	Simple Non-word decoding	50 words	percent: value / 50 * 100% wcpm: value / timer * 60	Yes (60 sec.)
5A	Passage Reading	60 words (baseline and endline intervention)	percent: value / 60 * 100% wcpm: value / timer * 60	Yes (60 sec.)
5B	Oral Reading Comprehension	5 questions	percent: value / 5 * 100%	No
6	Listening Comprehension	5 questions	percent: value / 5 * 100%	No
7A	Dictation (spelling)	13 words	percent: value / 13 * 100%	No
7B	Dictation (conventions of text)	4		No

ANNEX 5. SUMMARY EGRA RESULTS

In the summary table below, *mean* refers to the percentage of items answered correctly. *Mean for non-zero* refers to the percentage of items answered correctly among students who scored above zero. The final percentage (grey column) is the percent of students that had zero scores on that sub-test. Note the unit of analysis for the first two (means) is *items correct*, while the last column's unit of analysis is *students*. Note that design weights were applied to calculate EGRA results presented in this Annex in order to compensate for oversampling/under sampling at the provincial level and to ensure that results are representative of all provinces in the sample.

Filipino EGRA Results

Table A-13. Overall EGRA Results in Intervention Schools (SY 14/15)

	Baseline		Endline		Gain Score
Subtest	**Mean (SE)**	**95% Confidence Interval**	**Mean (SE)**	**95% Confidence Interval**	**Mean**
Initial Sound Identification (percent correct)	64.7% (0.36%)	(64%, 65.4%)	78.4% (0.3%)	(77.9%, 79%)	13.7%
Letter Sounds (percent correct)	24% (0.16%)	(23.7%, 24.3%)	29.5% (0.17%)	(29.2%, 29.9%)	5.6%
Letter Correct (per min)	24 (0.16)	(23.7, 24.3)	29.5 (0.17)	(29.1, 29.8)	5.5
Familiar Words (percent correct)	51.6% (0.33%)	(50.9%, 52.2%)	73.4% (0.31%)	(72.8%, 74.1%)	21.9%
Familiar Words Correct (per min)	26.9 (0.19)	(26.5, 27.3)	41.9 (0.22)	(41.4, 42.3)	15
Nonsense Words (percent correct)	34.5% (0.25%)	(34.1%, 35%)	51.4% (0.28%)	(50.9%, 52%)	16.9%
Nonsense Words Correct (per min)	17.3 (0.13)	(17.0, 17.5)	26.0 (0.14)	(25.7, 26.2)	8.7
Oral Passage Reading (percent correct)	54.1% (0.32%)	(53.5%, 54.7%)	75.1% (0.3%)	(74.6%, 75.7%)	21.0%
Words Correct in a Text (per min)	31.3 (0.21)	(30.9, 31.7)	48.1 (0.26)	(47.6, 48.6)	16.8
Prosody score	2.5 (0.01)	(2.9, 2.5)	3.0 (0.01)	(2.9, 3.0)	0.5
Reading Comprehension (percent correct)	30% (0.27%)	(29.4%, 30.5%)	47.2% (0.32%)	(46.6%, 47.8%)	22.9%
Listening Comprehension (percent correct)	44.4% (0.4%)	(43.7%, 45.2%)	63.7% (0.39%)	(63%, 64.5%)	19.3%
Dictation Composite (percent correct)	35.9% (0.25%)	(35.4%, 36.3%)	61.9% (0.26%)	(61.4%, 62.4%)	26.0%

Table A-14. EGRA Filipino Subtests Summary – Intervention Cohort (SY 14/15)

	Baseline			Endline		
Subtest	**Mean (SE)**	**Mean (SE) non-zero**	**% of students with zero scores**	**Mean (SE)**	**Mean (SE) non-zero**	**% of students with zero scores**
Initial Sound Identification (percent correct)	64.7% (0.36%)	73.2% (0.3%)	11.6%	78.4% (0.3%)	82.3% (0.2%)	4.7%
Letter Sounds(percent correct)	24% (0.16%)	24.7% (0.2%)	3.0%	29.5% (0.17%)	30.1% (0.2%)	1.7%
Letter Correct (per min)	24 (0.16)	24.7 (0.2)	3.0%	29.5 (0.17)	30.0 (0.2)	1.7%
Familiar Words(percent correct)	51.6% (0.33%)	56.5% (0.3%)	8.7%	73.4% (0.31%)	76.2% (0.3%)	3.6%
Familiar Words Correct(per min)	26.9 (0.19)	29.5 (0.2)	8.7%	41.9 (0.22)	43.5 (0.2)	3.6%
Nonsense Words (percent correct)	34.5% (0.25%)	40.7% (0.2%)	15.0%	51.4% (0.28%)	55.6% (0.2%)	7.5%
Nonsense Words Correct (per min)	17.3 (0.13)	20.3 (0.1)	15.0%	26.0 (0.14)	28.0 (0.1)	7.5%
Oral Passage Reading (percent correct)	54.1% (0.32%)	59.5% (0.3%)	9.1%	75.1% (0.3%)	77.7% (0.3%)	3.3%
Words Correct in a Text (per min)	31.3 (0.21)	34.4 (0.2)	9.1%	48.1 (0.26)	49.7 (0.3)	3.3%
Prosody score	2.5 (0.01)	2.5 (0)	-	3.0 (0.01)	3.0 (0)	-
Reading Comprehension (percent correct)	30% (0.27%)	43.9% (0.2%)	31.8%	47.2% (0.32%)	55.1% (0.3%)	14.3%
Listening Comprehension (percent correct)	44.4% (0.4%)	69% (0.3%)	35.6%	63.7% (0.39%)	77.9% (0.3%)	18.2%
Dictation Composite (percent correct)	35.9% (0.25%)	38.8% (0.2%)	7.7%	61.9% (0.26%)	63.0% (0.3%)	1.8%

Table A-15. Overall EGRA Results in Comparison Schools (SY 13/14)

	Baseline		Endline		Gain Score	Effect Size
Subtest	**Mean (SE)**	**95% Confidence Interval**	**Mean (SE)**	**95% Confidence Interval**	**Mean**	**Cohen's *d***
Initial Sound Identification (percent correct)	57.5% (0.61%)	(56.3%, 58.7%)	57.6% (0.6%)	(56.4%, 58.8%)	0.1%	0.0
Letter Sounds (percent correct)	16.9% (0.21%)	(16.5%, 17.4%)	19.1% (0.23%)	(18.7%, 19.5%)	2.2%	0.16
Letter Correct (per min)	17.1 (0.21%)	(16.7%, 17.5%)	19.2 (0.23%)	(18.8%, 19.7%)	2.1	0.16
Familiar Words (percent correct)	48.5% (0.54%)	(47.5%, 49.6%)	68.4% (0.51%)	(67.4%, 69.4%)	19.9%	0.63
Familiar Words Correct (per min)	25.0 (0.29%)	(24.5%, 25.6%)	37.5 (0.33%)	(36.9%, 38.1%)	12.5	0.67
Nonsense Words (percent correct)	31.9% (0.4%)	(31.1%, 32.6%)	46.2% (0.41%)	(45.4%, 47%)	14.4%	0.59
Nonsense Words Correct (per min)	16.1 (0.21%)	(15.7%, 16.5%)	23.2(0.21 %)	(22.8%, 23.7%)	7.1	0.57
Oral Passage Reading (percent correct)	49.9% (0.56%)	(48.8%, 51.0%)	56.1% (0.48%)	(55.2%, 57%)	6.2%	0.2
Words Correct in a Text (per min)	29.3 (0.35%)	(28.6%, 30.0%)	37(0.34%)	(36.4%, 37.7%)	7.7	0.37
Prosody score	1.6 (0.01%)	(3.9%, 1.6%)	2.1(0.01 %)	(2.0%, 2.1%)	0.5	0.55
Reading Comprehension (percent correct)	31.4% (0.53%)	(30.4%, 32.4%)	28.1% (0.44%)	(27.2%, 28.9%)	-3.4%	-0.11
Listening Comprehension (percent correct)	41.7% (0.61%)	(40.5%, 42.9%)	49.1% (0.63%)	(47.8%, 50.3%)	7.4%	0.2
Dictation Composite (percent correct)	32.2% (0.44%)	(31.4%, 33.1%)	45.6% (0.45%)	(44.7%, 46.5%)	13.4%	0.5

Table A-16. EGRA Filipino Subtests Summary – Comparison Cohort (SY 13/14)

	Baseline			Endline		
Subtest	**Mean (SE)**	**Mean (SE) non-zero**	**% of students with zero scores**	**Mean (SE)**	**Mean (SE) non-zero**	**% of students with zero scores**
Initial Sound Identification (percent correct)	57.5% (0.61%)	68.4% (0.5%)	15.9%	57.6% (0.6%)	66.1% (0.5%)	12.9%
Letter Sounds(percent correct)	16.9% (0.21%)	18.8% (0.2%)	10.1%	19.1% (0.23%)	20.4% (0.2%)	6.2%
Letter Correct (per min)	17.1 (0.21%)	19.0 (0.2)	10.1%	19.2(0.23%)	20.5 (0.2)	6.2%
Familiar Words(percent correct)	48.5% (0.54%)	52.7% (0.5%)	7.9%	68.4% (0.51%)	70.8% (0.5%)	3.4%
Familiar Words Correct(per min)	25.0 (0.29%)	27.2 (0.3)	7.9%	37.5 (0.33%)	38.8 (0.3)	3.4%
Nonsense Words (percent correct)	31.9% (0.4%)	37.6% (0.4%)	15.3%	46.2% (0.41%)	49.6% (0.4%)	6.8%
Nonsense Words Correct (per min)	16.1 (0.21%)	19.1 (0.2)	15.3%	23.2(0.21%)	24.9 (0.2)	6.8%
Oral Passage Reading (percent correct)	49.9% (0.56%)	55.6% (0.5%)	10.2%	56.1% (0.48%)	58.7% (0.5%)	4.4%
Words Correct in a Text (per min)	29.3 (0.35%)	32.7 (0.3)	10.2%	37(0.34%)	38.7 (0.3)	4.4%
Prosody score	1.6 (0.01%)	1.6 (0)	-	2.1 (0.01%)	2.1 (0)	-
Reading Comprehension (percent correct)	31.4% (0.53%)	50.6% (0.5%)	37.9%	28.1% (0.44%)	40% (0.5%)	29.9%
Listening Comprehension (percent correct)	41.7% (0.61%)	64.1% (0.5%)	35.0%	49.1% (0.63%)	68.5% (0.5%)	28.3%
Dictation Composite (percent correct)	32.2% (0.44%)	39.1% (0.4%)	17.5%	45.6% (0.45%)	48.1% (0.4%)	5.3%

Table A-17. EGRA Filipino Subtests Summary, by Province

Descriptive Statistics for Filipino Subtests, by Province					
Province	Subtest	Intervention (SY 14/15)		Comparison (SY 13/14)	
		Baseline Mean (SE)	Endline Mean (SE)	Baseline Mean (SE)	Endline Mean (SE)
BOHOL	Initial Sound Identification (percent correct)	70.1% (0.5%)	82.1% (0.4%)	-	-
	Letter Sounds (percent correct)	28.8% (0.3%)	35.2% (0.3%)	-	-
	Familiar Words (percent correct)	57.3% (0.5%)	75.3% (0.5%)	-	-
	Nonsense Words (percent correct)	38% (0.4%)	53.5% (0.5%)	-	-
	Oral Passage Reading (percent correct)	58.7% (0.5%)	76.5% (0.5%)	-	-
	Prosody	2.7% (0%)	2.9 (0)	-	-
	Reading Comprehension (percent correct)	32.5% (0.4%)	51.1% (0.5%)	-	-
	Listening Comprehension (percent correct)	48% (0.7%)	68.7% (0.6%)	-	-
	Dictation Composite (percent correct)	40.9% (0.4%)	63.7% (0.4%)	-	-
CEBU	Initial Sound Identification (percent correct)	65.7% (0.6%)	81% (0.5%)	63.4% (0.7%)	61.4% (0.7%)
	Letter Sounds (percent correct)	21.5% (0.2%)	27.4% (0.2%)	17.9% (0.2%)	19.4% (0.2%)
	Familiar Words (percent correct)	49.6% (0.5%)	76.1% (0.5%)	50.5% (0.6%)	70.6% (0.5%)
	Nonsense Words (percent correct)	33.7% (0.4%)	53.3% (0.4%)	33.4% (0.5%)	47.9% (0.5%)
	Oral Passage Reading (percent correct)	52.9% (0.5%)	78% (0.5%)	51.6% (0.6%)	58.1% (0.5%)
	Prosody	2.5% (0%)	3.0 (0)	1.6 (0)	2.1 (0)
	Reading Comprehension (percent correct)	28.9% (0.5%)	43.9 (0.5)	30.1% (0.6%)	27.7 (0.5)
	Listening Comprehension (percent correct)	36.4% (0.7%)	56.1% (0.7%)	36.1% (0.7%)	44.3% (0.7%)
	Dictation Composite (percent correct)	33.6% (0.4%)	64.4% (0.4%)	33.5% (0.5%)	48.2% (0.5%)
ILOCOS NORTE	Initial Sound Identification (percent correct)	52.5% (1.7%)	68.7% (1.5%)	-	-
	Letter Sounds (percent correct)	15.8% (0.6%)	21.3% (0.6%)	-	-
	Familiar Words (percent correct)	49.1% (1.6%)	68.7% (1.6%)	-	-
	Nonsense Words (percent correct)	34.2% (1.2%)	48.4% (1.3%)	-	-
	Oral Passage Reading (percent correct)	53.3% (1.5%)	70.1% (1.5%)	-	-
	Prosody	2.3% (0%)	3.0 (0)	-	-
	Reading Comprehension (percent correct)	27% (1.2%)	49.3 (1.4)	-	-
	Listening Comprehension (percent correct)	50.3% (1.7%)	71.1% (1.6%)	-	-
	Dictation Composite (percent correct)	34% (1.2%)	59.3% (1.3%)	-	-
ILOCOS SUR	Initial Sound Identification (percent correct)	54.5% (1.4%)	68.3% (1.2%)	-	-
	Letter Sounds (percent correct)	19.4% (0.5%)	23.1% (0.5%)	-	-
	Familiar Words (percent correct)	47.4% (1.2%)	68.7% (1.1%)	-	-
	Nonsense Words (percent correct)	31.8% (0.9%)	48% (0.9%)	-	-
	Oral Passage Reading (percent correct)	50.3% (1.2%)	70.4% (1.1%)	-	-
	Prosody	2.2 (0)	3.1 (0)	-	-
	Reading Comprehension (percent correct)	24.8% (0.9%)	43.6 (1.1)	-	-
	Listening Comprehension (percent correct)	45.2% (1.3%)	59.8% (1.4%)	-	-
	Dictation Composite (percent correct)	31.1% (0.9%)	56.9% (1%)	-	-

LA UNION	Initial Sound Identification (percent correct)	58.2% (1.1%)	71.8% (1%)	40.4% (1.2%)	46.6% (1.3%)
	Letter Sounds (percent correct)	23.4% (0.5%)	27.2% (0.5%)	14.2% (0.5%)	18.3% (0.5%)
	Familiar Words (percent correct)	45% (0.9%)	65.7% (1%)	42.9% (1.1%)	61.9% (1.2%)
	Nonsense Words (percent correct)	29.4% (0.7%)	44.1% (0.8%)	27.4% (0.8%)	41.3% (0.9%)
	Oral Passage Reading (percent correct)	47.4% (0.9%)	68.4% (0.9%)	45.1% (1.2%)	50.2% (1%)
	Prosody	2.1 (0)	3 (0)	1.7 (0)	1.8% (0%)
	Reading Comprehension (percent correct)	30.6% (0.8%)	46.9 (0.9)	35.2% (1.1%)	29.2 (1)
	Listening Comprehension (percent correct)	54.2% (1.1%)	71.3% (1%)	57.9% (1.2%)	62.8% (1.2%)
	Dictation Composite (percent correct)	32.1% (0.7%)	54% (0.8%)	28.7% (0.9%)	38.1% (0.9%)

Table A-18. Overall Endline Filipino EGRA Results in Sample BASA Schools, by Sex and Group

	Intervention		Comparison	
Subtest	**Boys Mean (SE)**	**Girls Mean (SE)**	**Boys Mean (SE)**	**Girls Mean (SE)**
Initial Sound Identification (percent correct)	74.3% (0.4%)	82.7% (0.4%)	55.7% (0.8%)	59.6% (0.8%)
Letter Sounds(percent correct)	26.4% (0.2%)	32.8% (0.2%)	17.2% (0.3%)	21.0% (0.3%)
Letter Correct (per min)	26.4 (0.2)	32.6 (0.2)	17.5 (0.3)	21.0 (0.3)
Familiar Words(percent correct)	66% (0.5%)	81.2% (0.4%)	61.4% (0.7%)	75.4% (0.6%)
Familiar Words Correct(per min)	36.2 (0.3)	47.8 (0.3)	33.1 (0.5)	42.0 (0.4)
Nonsense Words (percent correct)	44.4% (0.4%)	58.8% (0.4%)	40.5% (0.6%)	52% (0.6%)
Nonsense Words Correct (per min)	22.3 (0.2)	29.8 (0.2)	20.5 (0.3)	26.0 (0.3)
Oral Passage Reading (percent correct)	67.7% (0.4%)	82.9% (0.4%)	47.9% (0.6%)	64.3% (0.7%)
Words Correct in a Text (per min)	40.6 (0.3)	55.8 (0.4)	31.3 (0.4)	42.8 (0.5)
Prosody score	2.8 (0)	3.2 (0)	1.8 (0)	2.3 (0)
Reading Comprehension (percent correct)	44.7% (0.5%)	49.8% (0.4%)	24.9% (0.6%)	31.2% (0.7%)
Listening Comprehension (percent correct)	62.4% (0.5%)	65.1% (0.5%)	50% (0.9%)	48.1% (0.9%)
Dictation Composite (percent correct)	56% (0.4%)	68% (0.4%)	40% (0.6%)	51.3% (0.6%)

English EGRA Results

Table A-19. Overall English EGRA Results in Intervention Schools (SY 14/15)

	Baseline		Endline		Gain Score
Subtest	**Mean (SE)**	**95% Confidence Interval**	**Mean (SE)**	**95% Confidence Interval**	**Mean**
Initial Sound Identification (percent correct)	51.3% (0.39%)	(50.5%, 52%)	73.1% (0.3%)	(72.5%, 73.7%)	21.9%
Letter Sounds (percent correct)	27.0% (0.17%)	(26.7%, 27.3%)	39.1% (0.19%)	(38.8%, 39.5%)	12.1%
Letter Correct (per min)	27.0 (0.17)	(26.7, 27.3)	39.1 (0.19)	(38.7, 39.5)	12.1
Familiar Words (percent correct)	36.6% (0.34%)	(35.9%, 37.2%)	63.0% (0.36%)	(62.3%, 63.7%)	26.4%
Familiar Words Correct (per min)	20.2 (0.21)	(19.8, 20.6)	39.5 (0.29)	(38.9, 40.0)	19.3
Nonsense Words (percent correct)	29.9% (0.26%)	(29.4%, 30.4%)	49.2% (0.31%)	(48.6%, 49.8%)	19.3%
Nonsense Words Correct (per min)	15.1 (0.13)	(14.9, 15.4)	25.9 (0.18)	(25.5, 26.2)	10.8
Oral Passage Reading (percent correct)	50.2% (0.35%)	(49.5%, 50.9%)	72.4% (0.31%)	(71.8%, 73.1%)	22.2%
Words Correct in a Text (per min)	34.9 (0.28)	(34.4, 35.5)	58.1 (0.35)	(57.5, 58.8)	23.2
Prosody score	2.2 (0.01)	(2.9, 2.2)	2.8 (0.01)	(2.9, 2.8)	0.6
Reading Comprehension (percent correct)	11.8% (0.2%)	(11.4%, 12.2%)	15.9% (0.22%)	(15.5%, 16.3%)	4.1%
Listening Comprehension (percent correct)	10.6% (0.19%)	(10.2%, 11.0%)	17.6% (0.26%)	(17.1%, 18.1%)	7.0%
Dictation Composite (percent correct)	21.2% (0.16%)	(20.9%, 21.5%)	36.2% (0.22%)	(35.8%, 36.7%)	15.0%

Table A-20. English EGRA Subtests Summary – Intervention Cohort (SY 14/15)

	Baseline			Endline		
Subtest	**Mean (SE)**	**Mean (SE) non-zero**	**% of students with zero scores**	**Mean (SE)**	**Mean (SE) non-zero**	**% of students with zero scores**
Initial Sound Identification (percent correct)	51.3% (0.39%)	69.2% (0.3%)	26.0%	73.1% (0.3%)	79.0% (0.2%)	7.4%
Letter Sounds(percent correct)	27.0% (0.17%)	27.5% (0.2%)	2.0%	39.1% (0.19%)	39.6% (0.2%)	1.2%
Letter Correct (per min)	27.0 (0.17)	27.6 (0.2)	2.0%	39.1 (0.19)	39.5 (0.2)	1.2%
Familiar Words(percent correct)	36.6% (0.34%)	45.4% (0.4%)	19.4%	63.0% (0.36%)	69.4% (0.3%)	9.1%
Familiar Words Correct(per min)	20.2 (0.21)	25.0 (0.2)	19.4%	39.5 (0.29)	43.4 (0.3)	9.1%
Nonsense Words (percent correct)	29.9% (0.26%)	35.7% (0.3%)	16.4%	49.2% (0.31%)	53.7% (0.3%)	8.4%
Nonsense Words Correct (per min)	15.1 (0.13)	18.1 (0.1)	16.4%	25.9 (0.18)	28.2 (0.2)	8.4%
Oral Passage Reading (percent correct)	50.2% (0.35%)	55.2% (0.3%)	9.2%	72.4% (0.31%)	73.9% (0.3%)	2.0%
Words Correct in a Text (per min)	34.9 (0.28)	38.4 (0.3)	9.2%	58.1 (0.35)	59.3 (0.3)	2.0%
Prosody score	2.2 (0.01)	2.2 (0)	-	2.8 (0.01)	2.8 (0)	-
Reading Comprehension (percent correct)	11.8% (0.2%)	31.6% (0.3%)	62.7%	15.9% (0.22%)	31.6% (0.3%)	49.5%
Listening Comprehension (percent correct)	10.6% (0.19%)	33.4% (0.3%)	68.3%	17.6% (0.26%)	40.2% (0.4%)	56.3%
Dictation Composite (percent correct)	21.2% (0.16%)	23.6% (0.2%)	10.2%	36.2% (0.22%)	37.1% (0.2%)	2.4%

Table A-21. English EGRA Subtests Summary, by Province

	Descriptive Statistics for English Subtests, by Province (Intervention Cohort (SY 14/15)		
Province	**Subtest**	**All Students**	
		Baseline Mean (SE)	**Endline Mean (SE)**
BOHOL	Initial Sound Identification (percent correct)	58% (0.6%)	76.2% (0.4%)
	Letter Sounds (percent correct)	31.9% (0.3%)	45% (0.3%)
	Familiar Words (percent correct)	41% (0.6%)	64.7% (0.6%)
	Nonsense Words (percent correct)	33.3% (0.4%)	51.1% (0.5%)
	Oral Passage Reading (percent correct)	54.5% (0.6%)	73.3% (0.5%)
	Prosody	2.4 (0)	2.7 (0)
	Reading Comprehension (percent correct)	11.9% (0.3%)	15.8% (0.3%)
	Listening Comprehension (percent correct)	12.7% (0.3%)	21% (0.4%)
	Dictation Composite (percent correct)	24.3% (0.3%)	35.9% (0.3%)
CEBU	Initial Sound Identification (percent correct)	52.3% (0.7%)	76.4% (0.5%)
	Letter Sounds (percent correct)	24.6% (0.2%)	37.5% (0.3%)
	Familiar Words (percent correct)	34.9% (0.6%)	66.6% (0.6%)
	Nonsense Words (percent correct)	28.9% (0.4%)	54.2% (0.5%)
	Oral Passage Reading (percent correct)	49.1% (0.6%)	76.3% (0.5%)
	Prosody	2.2 (0)	2.8 (0)
	Reading Comprehension (percent correct)	12% (0.3%)	16.6% (0.4%)
	Listening Comprehension (percent correct)	9.6% (0.3%)	22.1% (0.5%)
	Dictation Composite (percent correct)	19.4% (0.3%)	40% (0.4%)
ILOCOS NORTE	Initial Sound Identification (percent correct)	41.3% (1.6%)	60.2% (1.5%)
	Letter Sounds (percent correct)	19.3% (0.6%)	27.3% (0.8%)
	Familiar Words (percent correct)	37.9% (1.7%)	57.6% (1.8%)
	Nonsense Words (percent correct)	29.6% (1.2%)	42.2% (1.4%)
	Oral Passage Reading (percent correct)	50% (1.8%)	66.7% (1.6%)
	Prosody	2.1 (0)	2.8 (0)
	Reading Comprehension (percent correct)	14.2% (1%)	16.4% (1.1%)
	Listening Comprehension (percent correct)	9.4% (0.8%)	6.4% (0.8%)
	Dictation Composite (percent correct)	20.8% (0.8%)	33.4% (1.1%)
ILOCOS SUR	Initial Sound Identification (percent correct)	38.9% (1.3%)	65.1% (1.1%)
	Letter Sounds (percent correct)	21.7% (0.5%)	32.4% (0.6%)
	Familiar Words (percent correct)	33.1% (1.1%)	56.1% (1.3%)
	Nonsense Words (percent correct)	25.3% (0.8%)	38.7% (1%)
	Oral Passage Reading (percent correct)	46.5% (1.2%)	63.8% (1.1%)
	Prosody	2.1 (0)	2.9 (0)
	Reading Comprehension (percent correct)	9.2% (0.6%)	13% (0.7%)
	Listening Comprehension (percent correct)	6.2% (0.5%)	5.4% (0.6%)
	Dictation Composite (percent correct)	17.8% (0.5%)	28.3% (0.7%)
LA UNION	Initial Sound Identification (percent correct)	41.5% (1.1%)	65.6% (1%)
	Letter Sounds (percent correct)	26.2% (0.5%)	36.5% (0.6%)

	Familiar Words (percent correct)	30.5% (0.9%)	55.2% (1%)
	Nonsense Words (percent correct)	26.4% (0.7%)	39.4% (0.8%)
	Oral Passage Reading (percent correct)	43.8% (1%)	67.2% (0.9%)
	Prosody	2 (0)	2.9 (0)
	Reading Comprehension (percent correct)	11.4% (0.5%)	16.3% (0.7%)
	Listening Comprehension (percent correct)	10.8% (0.6%)	7.7% (0.5%)
	Dictation Composite (percent correct)	19.8% (0.4%)	32.8% (0.7%)

Table A-22. English EGRA Results, by Sex

	Baseline		Endline	
Subtest	**Boys Mean (SE)**	**Girls Mean (SE)**	**Boys Mean (SE)**	**Girls Mean (SE)**
Initial Sound Identification (percent correct)	46.9% (0.5%)	55.8% (0.6%)	69.3% (0.4%)	77.1% (0.4%)
Letter Sounds(percent correct)	24.4% (0.2%)	29.7% (0.2%)	35.4% (0.3%)	43.0% (0.3%)
Letter Correct (per min)	24.4 (0.2)	29.7 (0.2)	35.4 (0.3)	42.9 (0.3)
Familiar Words(percent correct)	28.4% (0.4%)	45.1% (0.5%)	53.8% (0.5%)	72.7% (0.5%)
Familiar Words Correct(per min)	15.3 (0.3)	25.3 (0.3)	31.6 (0.4)	47.7 (0.4)
Nonsense Words (percent correct)	24.5% (0.3%)	35.5% (0.4%)	42.1% (0.4%)	56.6% (0.4%)
Nonsense Words Correct (per min)	12.3 (0.2)	18.1 (0.2)	21.7 (0.2)	30.2 (0.3)
Oral Passage Reading (percent correct)	41.5% (0.5%)	59.2% (0.5%)	65.8% (0.5%)	79.4% (0.4%)
Words Correct in a Text (per min)	27.6 (0.4)	42.5 (0.4)	48.9 (0.4)	67.8 (0.5)
Prosody score	2.0 (0)	2.5 (0)	2.6 (0)	3.0 (0)
Reading Comprehension (percent correct)	9.6% (0.3%)	14.0% (0.3%)	14.5% (0.3%)	17.4% (0.3%)
Listening Comprehension (percent correct)	9.0% (0.3%)	12.3% (0.3%)	14.5% (0.3%)	20.8% (0.4%)
Dictation Composite (percent correct)	18.0% (0.2%)	24.5% (0.2%)	30.3% (0.3%)	42.4% (0.3%)