USAID
FROM THE AMERICAN PEOPLE

PAQUED: DRC

Projet d'Amélioration de la Qualité de l'Education (PAQUED)

Midterm Report, Early Grade Reading Assessment (EGRA)

December 2012
This publication was produced for review by the United States Agency for International Development and by Education Development Center. It was prepared by RTI International.

PAQUED: DRC

Projet d'Amélioration de la Qualité de l'Education (PAQUED)
Midterm Report, Early Grade Reading Assessment (EGRA)

Contract: AID-623-A-09-00010
EDC Project: 11224
Date: December 2012

Prepared for
USAID/DRC and EDC

Prepared by
Jennae Bulat, Penelope Collins,
Timothy Slade, Christine Davies, and Michael Costello
RTI International 3040 Cornwallis Road
Post Office Box 12194 Research Triangle Park, NC 27709-2194

Table of Contents

List of Tables

List of Figures

Acronyms

CELTA	Centre de Linguistique Théorique et Appliquée
clpm	correct letters per minute
clspm	correct letter sounds per minute
cnonwpm	correct non-words per minute (invented words)
cwpm	correct words per minute
DRC	Democratic Republic of Congo
EDC	Education Development Center
EGRA	Early Grade Reading Assessment
IRR	inter-rater reliability
MEPSP	Ministry of Primary, Secondary, and Professional Education
orf	oral reading fluency
PAQUED	Projet d'Amélioration de la Qualité de l'Education
PCA	principal components analysis
RTI	RTI International
SES	socio-economic indicators
USAID	U.S. Agency for International Development

Executive Summary

Funded by the U.S. Agency for International Development (USAID) and led by the Education Development Center (EDC), the Projet d'Amélioration de la Qualité de l'Education (PAQUED) project is a five-year initiative that aims to raise student learning through improved teaching and school environments in the Democratic Republic of the Congo (DRC). As a partner in PAQUED, RTI International (RTI) is responsible for designing and implementing impact assessments of this program.

In September and October of 2010, RTI collaborated with the DRC's Ministry of Primary, Secondary, and Professional Education (MEPSP) to conduct a baseline assessment of the reading skills of students enrolled in Grades 2, 4, and 6 in three provinces participating in this project: Bandundu, Equateur, and Orientale. The instrument used in the DRC was RTI's Early Grade Reading Assessment (EGRA). This instrument is designed to collect information on the level of competency in foundational reading skills areas. All of the competencies that it measures have been shown through research to be highly predictive of later reading ability and susceptible to improvement through teaching.[1] The EGRA instrument contains a series of individually administered protocols designed to assess performance on the following discrete skills that constitute key building blocks of reading: vocabulary, phonemic awareness (in particular, initial sound identification), letter sound knowledge, sight word reading, decoding, oral comprehension, and reading comprehension.

In May 2012, RTI again collaborated with the MEPSP to administer a midterm assessment designed to measure student skills in early reading. Assessments were carried out in treatment (PAQUED-supported) schools and comparison (non-PAQUED-supported) schools in the three provinces. A total of 2,453 students from Grade 2 and Grade 4 in 95 schools were assessed. The goal of the midterm assessment was to identify gains in reading skills attributable to PAQUED interventions beyond the baseline assessment. Important to note, however, are substantive sampling and implementation issues. Numerous technological and logistical problems delayed and impeded implementation of planned interventions, with the result that very few schools had received the full complement of services and support as planned. In addition, the project encouraged shifting the assessment's sampling frame to focus on those schools which, by virtue of being ~~"a~~ccessible" (defined as being within approximately 20 km of an urban center), might be more likely to have received the intervention as planned and might, therefore, present a more accurate portrait of the intervention's potential impact. As a result, schools could no longer be tracked longitudinally from 2010 through 2012, as originally intended. The project therefore transitioned to a cross-section design. Additionally, despite its focus on accessible schools, the extent to which the assessment

[1] For a discussion of the EGRA research basis, see Gove, A. & Wetterberg, A. (Eds.) (2011). *The Early Grade Reading Assessment: Applications and interventions to improve basic literacy.* Research Triangle Park, NC: RTI International.

accurately reflects the impact of the intervention is unknown, as reliable information regarding fidelity of implementation is unavailable.

To be interpreted within the aforementioned constraints, analyses suggest the following.

- As may be expected, overall Grade 4 students tended to produce higher scores on EGRA subtasks than did Grade 2 students, although these differences were not tested for statistical significance.
- Results in either Grade 2 or Grade 4 did not consistently favor treatment-group students, although where statistically significant differences did emerge, the overall trend was for students in treatment schools to outperform students in comparison schools at the highest levels of performance within a subtask (e.g., treatment students more likely than comparison students to score 5 out of 5, with comparison students more likely to produce lower scores).
- While not a consistent trend, participation in the treatment group at times seems to have eliminated differences between males and females.
- No consistent patterns emerged between or within provinces.

A summary of each subtask's results follows.

Vocabulary

- Grade 2: Treatment-group students appear to have outperformed their comparison peers, for both males and females.
- Grade 4: Treatment-group students appear to have outperformed their comparison peers, for both males and females. Furthermore, while comparison males tended to outperform comparison females, no gender differences were observed within the treatment group.
- No meaningful differences were observed between or within provinces in Grade 2. In Grade 4, Orientale students seem to have outperformed those in other provinces, with Orientale treatment students outperforming their comparison peers.

Initial Sound Identification

- Grade 2: No meaningful trends were observed among males or females.
- Grade 4: Overall, treatment students outperformed comparison students on two of the ten stimuli, although when comparing across genders female treatment students outperformed female comparison students on five stimuli, with male treatment students outperforming male comparison students on one.
- Across provinces, in both Grade 2 and Grade 4, students in Equateur tended to generate higher scores where significant differences were observed. Within each

grade, only one significant difference between treatment and comparison groups emerged, in each instance favoring the treatment group.

Letter Sound Knowledge

- Grade 2: No significant percentage differences emerged across treatment and comparison groups. Significant percentage differences favoring males appeared within the comparison group, while no gender differences were observed within the treatment group. Sounds identified per minute, however, showed a treatment effect for females and a leveling of gender effect within the treatment group.
- Grade 4: No significant percentage differences emerged across treatment and comparison groups. Within each treatment group, two significant percentage differences were observed, with the trend favoring males. Looking at sounds identified per minute, while there was no gender effect within the treatment group, comparison males outperformed females.
- Looking at either percentages or letter sounds identified per minute, no significant treatment effects emerged within any of the three provinces for Grade 2 or Grade 4.

Familiar Word Reading

- Grade 4: Considering both percentage and fluency scores, significant differences favoring the comparison group emerged on this subtask; in addition, within the comparison group, a trend emerged favoring males over females, while no such trend was observed within the treatment group.
- Within each of two provinces a significant difference emerged that favored the treatment group; no other significant difference across treatment conditions surfaced.

Invented Word Reading

- Grade 4: Looking at both percentage and fluency scores, there were no significant differences across treatment conditions although, within both groups, gender differences favoring males emerged.
- No significant differences across treatment conditions were observed for any of the three provinces.

Oral Reading Fluency

- Grade 4: Looking at both percentage and fluency scores, there were no significant differences across treatment conditions, although, within the comparison group, gender differences favoring males emerged.
- No significant differences across treatment conditions were observed for any of the three provinces.

Reading Comprehension

- Grade 4: No significant differences were observed across or within treatment conditions.
- Looking at results by province, on two questions treatment students outperformed comparison peers: one for Bandundu students and one for Orientale students.

Listening Comprehension

- Grade 2: Treatment students outperformed their comparison peers on three questions. No gender differences emerged within either group.
- Grade 4: Treatment females outperformed comparison females on one question; while within the comparison group males outperformed females on two of the five questions, this occurred on only one question within the treatment group.
- Within Orientale, treatment students outperformed comparison students on three questions. No differences between treatment and comparison groups emerged in the other two provinces.

Comparisons of the reduced 2010 sample baseline with the 2010 midterm suggest that students in treatment schools had significantly higher scores on each of the Grade 2 EGRA measures at midterm than at baseline. Thus, treatment schools yielded higher student performance at midterm in vocabulary knowledge, phonemic awareness, letter sound knowledge, and listening comprehension than did comparable schools at baseline. Comparison schools yielded midterm scores that exceeded those of baseline schools in the two tasks more closely related to beginning reading achievement, initial sound identification, and letter sound knowledge; however, the gains shown by comparison schools were not as large as those shown by treatment schools. Within Grade 4, compared to baseline, students in treatment schools had higher scores on the Grade 4 EGRA measures that assessed oral language skills in French (vocabulary knowledge, initial sound identification, and listening comprehension) and in letter sound identification. In contrast, comparison schools did not show gains in midterm scores on any of the EGRA tasks, with the exception of letter sound identification. Students in both treatment and comparison schools had lower dictation scores than at baseline.

1. Introduction

Funded by the U.S. Agency for International Development (USAID) and led by the Education Development Center (EDC), the Projet d'Amélioration de la Qualité de l'Education (PAQUED) project is a five-year initiative that aims to raise student learning through improved teaching and school environments in the Democratic Republic of the Congo (DRC). As a partner in PAQUED, RTI International (RTI) is responsible for designing and implementing impact assessments of this program.

In September and October of 2010, RTI collaborated with the DRC's Ministry of Primary, Secondary, and Professional Education (MEPSP) to conduct a baseline assessment of the reading skills of students enrolled in Grades 2, 4, and 6 in three provinces participating in this project: Bandundu, Equateur, and Orientale. The instrument used in the baseline assessment was RTI's Early Grade Reading Assessment (EGRA). This instrument is designed to collect information on the level of competency in foundational reading skills areas. All of the competencies that it measures have been shown through research to be highly predictive of later reading ability and susceptible to improvement through teaching. The EGRA contains a series of individually administered protocols designed to assess performance on the following discrete skills that constitute key building blocks of reading: vocabulary, phonemic awareness (specifically, initial sound identification), letter/sound knowledge, familiar and invented word reading, passage reading and comprehension, oral comprehension, and dictation.

The 2010 baseline assessment established that students in comparison and treatment schools were similar, with no statistically significant differences in student performance across the various EGRA subtasks. Overall, students had poor reading skills, regardless of treatment or comparison status, province, or gender. Student performance on measures of simpler reading skills (such as knowledge of common vocabulary) was higher than on more difficult reading skills (such as oral reading fluency and comprehension). Grade 6 students performed better than Grades 2 or 4 students, but across all grades the typical student was unable to read and comprehend connected text, which is the ultimate goal of reading. Students, in general, also came from low-socioeconomic (SES) backgrounds and had low levels of preschool attendance.

Three themes, however, emerged from the baseline assessment: (1) low levels of mastery of oral and aural French, which hinder early French reading acquisition; (2) consistent regional differences in reading performance, with students in Orientale outperforming students in Equateur, who, in turn, outperformed students in Bandundu; and (3) differences in performance between genders, with male students outperforming female students on most subtasks.

In May 2012, RTI again collaborated with the MEPSP to administer the EGRA at the project's midterm in order to identify gains in reading skills attributable to PAQUED

interventions. Assessments were carried out in treatment (PAQUED-supported) schools and comparison (non-PAQUED-supported) schools in the three provinces. A total of 2,453 students from Grade 2 and Grade 4 in 95 schools were assessed.

Report Structure and Constraints

The initial intent for this report was to compare student performance from the 2010 baseline with that from the 2012 midterm. However, given modifications made to the sampling design (as described in the Sample Design subsection of this report) and the resulting reduction of the revised 2010 baseline sample (to 13 treatment and 3 comparison schools, with a total of 339 treatment and 78 comparison students, —educed 2010 baseline"), EDC and RTI opted to constrain the scope of this report primarily to comparisons between treatment and comparison schools within the 2012 sample, with a summary of high-level comparisons between the reduced 2010 sample and the 2012 sample.

Interpretation of these midterm results must take into consideration key logistic and sampling constraints. Numerous technological and logistical problems delayed and impeded implementation of planned interventions, with the result that very few schools received the planned complement of services and support. In addition, the assessment's sampling frame was modified to focus on those schools that, by virtue of being —accessible" (defined as being within approximately 20 km of an urban center), might be more likely to have received the intervention as planned and might, therefore, present a more accurate portrait of the intervention's potential impact. As a result, schools could no longer be tracked longitudinally from 2010 through 2012, and the sampling methodology transitioned from a longitudinal comparison-led study to a cross-section sampling design.

This report includes the following main sections: Research Design, Sample Characteristics, PAQUED Midterm Findings, Comparison of Baseline and Midterm Assessments, and Conclusions and Recommendations. The Research Design section describes the EGRA subtasks, sampling methodology, test administration, data collection, and analytical methodology. The Descriptive Statistics section describes the characteristics of the reduced 2010 and 2012 samples. The PAQUED Midterm Findings section describes students' performance on the EGRA subtasks. The Comparison of Baseline and Midterm Assessments summarizes analyses of student performance between the reduced 2010 sample and the 2012 midterm sample. Finally, the Conclusions and Recommendations section contextualizes the results to the DRC and offers suggestions to improve early grade reading among students as the program continues.

2. Research Design

The original PAQUED EGRA design was intended to be a longitudinal comparison study meant to isolate the reading gains attributable to the PAQUED program. Due to sampling constraints discussed earlier, it was converted to a cross-sectional design at midterm. The baseline data collection conducted in 2010 examined the competency levels of students in both treatment schools (those schools participating in the PAQUED program) and comparison schools (those schools with comparable demographics but not participating in the PAQUED program) prior to the start of the intervention. The current report examines the results of the data collection conducted in 2012, with the goal of identifying changes in indicators of reading performance and determining whether they are attributable to PAQUED program intervention.

Both the baseline and current studies included a sufficiently large sample to be representative of the three PAQUED provinces[2] and to allow for comparisons across these provinces. The ability to allow interprovincial comparisons is important because of systematic differences between the provinces PAQUED serves, including recent history with conflict, geography, economy, and languages spoken.

To determine impact attributable to the PAQUED program, the sample included both treatment schools and comparison schools that share similar characteristics to the program schools but are not participating in PAQUED. At the outset of the project, it was intended that both the treatment and comparison schools be tested over the five years of the program; this would allow any learning gains in PAQUED schools to be compared with those in comparison schools. The decision was made to modify the sampling frame for the 2012 data collection; however, due to obstacles encountered while implementing the program (discussed further under Limitations to Program Impact), the decision was made to focus on schools that were considered to be accessible, as they would have had more opportunity to receive program inputs. As a result, new samples of treatment and comparison schools were drawn (see the Sampling Methodology subsection of this report for a fuller discussion).

EGRA Instruments

The EGRA instrument is a 15- to 20-minute orally administered set of subtasks designed to assess foundational literacy skills crucial to becoming a fluent reader. The EGRA assessment is designed to be a method-independent approach to assessment, in that the instrument does not reflect a particular method of reading instruction (i.e., ―whole language" or ―phonics-based" approach"). Rather, EGRA measures basic skills that a child must have to eventually be able to read fluently and with comprehension—the ultimate goal of reading. The EGRA subtasks are based on research regarding a

2 Note that this is different from stating that the provinces sampled are representative of the DRC as a whole.

comprehensive approach to reading acquisition across languages, including five essential components: phonemic awareness, phonics, reading fluency, vocabulary, and comprehension.[3]

EGRA assessments are adapted to the language(s) and locality where they are administered. The French EGRA contains various subtasks, with items selected on the basis of the properties of French. Specific measures assessed in the DRC included vocabulary, phonemic awareness, letter/sound knowledge, familiar and invented word reading, passage reading and comprehension, oral comprehension, and dictation. The DRC EGRA was designed to pinpoint where students were having difficulty in reading and provide the MEPSP and PAQUED with information as to which basic skills they were lacking. A powerful diagnostic tool, EGRA is able to assess competencies that cannot be measured using a traditional pen-and-pencil examination.

Adaptation of the French EGRA instrument for the DRC took place in December 2009. DRC ministry officials and Centre de Linguistique Théorique et Appliquée (CELTA) language experts participated in the instrument adaptation process, which was led by Dr. Souhila Messaoud-Galusi of RTI.[4] The resulting EGRA instrument included the following subtasks.[5]

The **vocabulary** subtask assessed students' basic oral vocabulary and their understanding of basic commands. In this untimed subtask, students were asked to point to body parts (such as "nose") and objects in the environment (such as "eraser") as well as to follow simple commands (such as "put the pencil under the paper"). After two practice items, students were asked to identify eight body parts, identify six objects in the environment, and put a pencil in six different locations. The final score was the number of words and simple commands that students successfully identified, with the maximum possible overall score being 20 (8 possible points for parts of the body, 6 possible points for objects in the environment, and 6 possible points for spatial terms).

The **initial sound identification** subtask assessed students' 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 (such as "sack") and identify the first sound in that word (in this case, /s/). After two practice items, students were given 10 test items. The final score was the number of words of which students successfully identified the initial sound, with the maximum possible score being 10.

3 The definitions are adapted from Report of the National Reading Panel—Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction: Reports Of The Subgroups, Eunice Kennedy Shriver National Institute of Child Health and Human Development, 2000, NIH Publication No. 00-4754, available at http://www.nichd.nih.gov/publications/nrp/smallbook.cfm.

4 Details of the adaptation process are outlined in the pilot report, EGRA Pilot for PAQUED Baseline Assessment, DRC, prepared by RTI International for Education Development Center with funding from USAID, 2010.

5 Five of the subtasks that tested more advanced reading skills—familiar word reading, invented word reading, oral reading fluency, reading comprehension, and writing—were administered to students in Grade 4 only.

The **letter sound knowledge** subtask assessed students' knowledge of the letter-sound relationships critical for sounding out new words. In this timed subtask, students were shown a chart containing 10 rows, each containing 10 letters arranged in random order. Students were asked to tell the examiner the sounds of as many letters as possible within one minute, yielding a score of correct letter sounds per minute (clpm).[6]

The **familiar word reading** subtask assessed students' skill at reading high-frequency words (those words most frequently used in reading and writing by the target age in the target language). Recognizing familiar words is critical for developing reading fluency. In this timed subtask, students were presented a chart of 50 familiar words and asked to read as many words as they could within one minute, yielding a score of correct words per minute (cwpm).[7]

The **invented word decoding** subtask assessed students' skill at applying letter-sound correspondence rules to decode unfamiliar words. To ensure that students were sounding out the words rather than recognizing them by sight, administrators showed the students a chart of 50 pronounceable, but invented, words that followed legal spelling patterns of French. Students were asked to sound out as many invented words as they could within one minute, yielding a score of cwpm.

The **oral passage reading** subtask assessed students' fluency in reading aloud a passage of grade-level text and their ability to understand what they had read. There were two parts to this subtask that were scored separately:

- Oral reading fluency: The ability to read passages fluently is considered a necessary component of reading comprehension. In this subtask, students were given a 50-word story and asked to read it aloud in one minute. The oral reading fluency score was the number of cwpm.
- Reading comprehension: After the students finished the passage, or after the minute ended, the passage was removed. Students were orally asked five questions that required them to recall basic facts from the passage. The reading comprehension score was the number of correct answers, with a maximum possible score of 5.

Listening comprehension is considered to be a critical skill for reading comprehension, as it is the ability to make sense of oral language. In the listening comprehension subtask, the examiner read a 50-word passage to students. Students were then orally asked five questions about that passage. The listening comprehension score was the total correct answers, with a maximum possible score of 5.

6 Although 100 letters were presented, it was possible for students to obtain scores greater than 100 correct letter sounds per minute if they had completed the subtask in less than one minute. The score was calculated by dividing the total letter sounds correctly identified by the time taken to complete the task and multiplying the quotient by 60. For example, a child identifying 100 letter sounds correctly in 50 seconds would receive a score of (100clpm ÷ 50s) X 60s = 120 clpm.

7 Although only 50 familiar words were presented, it was possible for students to obtain scores greater than 50 cwpm if they had completed the task in less than 1 minute. The score was calculated by dividing the total words read correctly by the time taken to complete the task and multiplying the quotient by 60. For example, a child who read all 50 words correctly in 50 seconds would receive a score of (50cwpm ÷ 50s) X 60s = 60 cwpm.

The **dictation** subtask assessed students' skill at spelling. Students were allowed to spell from memory or by sounding out words and applying sound-spelling correspondences. In this subtask, the examiner read a short sentence to the students ("Mon ami s'appelle Bola."), and students attempted to write the sentence. The three target words that were scored were mon, ami, and appelle. The score was the number of words spelled correctly.

Administration of the EGRA includes a "stop" rule, which requires assessors to discontinue the administration of a subtask if a pupil is unable to respond correctly to any of the items in the first line (i.e., the first 10 letters, the first five words, or the first line of the oral reading fluency story). This rule was established to avoid frustrating pupils who do not understand the subtask or who lack the skills to respond. Before they administered the EGRA, administrators were required to read to students explicit information about the test and how it would be used. Pupils were asked to provide verbal assent to participate in the assessment.

Sampling Methodology

Sample Design. The 2012 EGRA data collection assessed Grade 2 and 4 students sampled from 95 schools across the provinces of Bandundu, Equateur, and Orientale. Of these schools, 60 students were drawn from treatment schools, and 35 were drawn from comparison schools (12 each in the Bandundu and Orientale provinces, and 11 in the province of Equateur). This approach was designed to allow for comparisons between (1) treatment schools versus comparison schools by grade, and (2) treatment schools in the different provinces across grades. This sampling method differed slightly from that used in 2010.

The 2010 sampling design was a three-stage random sample of subdivisions within the three provinces, schools within subdivisions, and students within grades. In 2010, the sample of treatment schools was drawn from approximately 3,000 schools that were receiving PAQUED support, with approximately 1,000 in each province; the final 2010 sample included 5,461 students from 144 schools in the provinces. The sampling criteria for that year were as follows.

- Safety: A school could not be located in a subdivision that would put teams of assessors at high security risk as they attempted to execute their school visits, and
- Accessibility: A school must (1) not be located in a flood zone, (2) be no more than three days' travel from the center of the subdivision, and (3) be primarily accessible by land.

Unlike the 2010 sampling plan, however, the 2012 sampling design was a two-stage simple random sample of schools within provinces, followed by students within grades. In addition, in 2012 the accessibility criterion was redefined to include only schools located within 20 km of an urban center based on available information, in order to focus assessment within schools that were geographically more likely to have received program inputs.

Sampling Frame. The first step in drawing the 2012 sample, therefore, involved building a list of schools that met the revised criterion for accessibility. The sampling design required that a minimum of six schools within a given subdivision be classified as accessible for reasons of feasibility[8]; therefore, a list was compiled of (1) all of the subdivisions that contained at least six ―accessible" schools and (2) all of the ―accessible" schools located within those subdivisions. The first stage of the sample (schools within provinces) was drawn by RTI, in collaboration with EDC, from that list. At each sampled school, a total of 13 students per grade were randomly selected to participate in the assessment (the second stage of sampling). Students were chosen randomly from within each classroom by the MEPSP team supervisor. Assessment took place in the morning, the time when students would normally be in class and likely be most alert.

Sample Selection. As shown in ***Table 1***, approximately 400 students were sampled for each grade within province, for a final 2012 sample of 2,453 students from 95 schools across the three provinces. This sample size follows best practice guidelines established from early grade assessments in countries in Africa, South America, and the Caribbean, from which it was possible to determine the most efficient sample size that would most likely show meaningful differences between groups of interest, such as those between grade levels or genders.[9]

Table 1. Sample Size by Province and Grade[10]

Province	Number of Schools	Number of Students		
		Grade 2	Grade 4	Total Students
Bandundu	32	406	411	817
Equateur	31	404	402	806
Orientale	32	416	414	830
Total	95	1,226	1,227	2,453

Probability of Selection and Design Weight. The probability of selection for schools within province was the total number of schools sampled in the region divided by the population number of schools in the region

$$poshi = \#(\text{schools sampled})h \ / \ \#(\text{population schools})h$$

where #(schools sampled) h was the number of schools sampled in the hth region and #(population schools)h, is the number of population schools in the hth region. The design

8 Any school drawn from a subdivision with only four ―accessible" schools, for example, would ultimately have needed to be replaced or resampled because they would have an insufficient number of ―similar" or ―peer" schools. The decision was, therefore, made to exclude those schools from the outset.

9 This number varies depending on standard deviation, but RTI has found that standard deviation for the purposes of the core measures of EGRA is reasonably constant across countries.

10 Throughout this report, sums reported in tables may not equal 100% due to rounding.

weight in a region was the inverse probability of selection in the region. That is, the design weight for the ith school in the hth region (dhi) was

dhi = 1 / poshi.

The probability of selection for students within school by grade and gender was the total number of students sampled by gender in the grade divided by the population number of students in that grade

posjlk = #(students sampled)jlk / #(population students)jl

where #(students sampled)jk was the number of students in the jth school for the kth gender and #(population students)j was the number of students in the jth school for the lth grade. The design weight at this level for a student in a school was the inverse probability of selection in the school by grade and gender

djlk = 1 / posjlk.

The final analysis weight used for all the analyses in this report was the product of the first-stage design weight and the second-stage design weight

WT= dhi*djlk.

3. Sample Characteristics

Reduced 2010 Sample. The following two tables display descriptive statistics for the reduced 2010 sample: 417 students from 16 schools across the three provinces. Approximately four-fifths (81%, or 339) of the sample consisted of students in treatment schools, with 19% (78) of students sampled from comparison schools. Approximately the same numbers of students came from schools in the Equateur (183, or 44%) and Orientale (182, or 44%) provinces, with 52 students (13%) coming from Bandundu. Overall, the reduced 2010 sample was fairly evenly distributed across grade and gender. ***Table 2*** describes the general characteristics of the sample.

Table 2. Characteristics of the Reduced 2010 Sample (n = 417)

	Number of Students	Percentage
Province		
Bandundu	52	13%
Equateur	183	44%
Orientale	182	44%
Grade		
2	209	50%
4	208	50%
Gender		
Female*	214	51%
Male *	202	49%
Group		
Treatment	339	81%
Comparison	78	19%

*Due to data missing for gender, the sum of the female and male counts does not equal the total.

Table 3 displays the number of reduced 2010 sample students by province, gender, and grade within the treatment and comparison groups. The majority of reduced 2010 sample students came from treatment schools (percentages ranging from 71%, or 130 students, in Orientale to 100%, or 52 students, in Bandundu). While there were nearly equal proportions of students from Grades 2 and 4 (81% for each grade), the majority of students across grades came from treatment schools (81%, or 170, of Grade 2 students, and 81%, or 169, of Grade 4 students). Furthermore, while gender was relatively equally balanced, the majority of students across genders came from treatment schools (83%, or 177, females, and 80%, or 161, males). It should be noted, however, that since this sample reflects a reduction introduced after the baseline assessment was implemented, it does not reflect intended sample characteristics.

Table 3. Characteristics of the Reduced 2010 Sample, by Group (n = 417)

Variable	Treatment	Comparison	Total
Province			
Bandundu	52 (100%)	0 (0%)	52 (100%)
Equateur	157 (86%)	26 (14%)	183 (100%)
Orientale	130 (71%)	52 (29%)	182 (100%)
Grade level			
2	170 (81%)	39 (19%)	209 (100%)
4	169 (81%)	39 (19%)	208 (100%)
Gender			
Female	177 (83%)	37 (17%)	214 (100%)
Male	161 (80%)	41 (20%)	202 (100%)
Total*	**338 (81%)**	**78 (19%)**	**417 (100%)**

* Due to data missing for gender, the sum of the female and male counts does not equal the total.

2012 Sample. The final 2012 sample included 2,453 students sampled from 95 schools in three provinces. Approximately two-thirds (63%, or 1,548) of the sample consisted of students in treatment schools, with 37% (903) of students sampled from comparison schools and the student sample approximately even across province, grade level, and gender. ***Table 4*** describes the general characteristics of the sample.

Table 4. General Characteristics of the Overall Sample (n = 2,453)

Variable	Number	Percentage
Province		
Bandundu	817	33%
Equateur	806	33%
Orientale	830	34%
Grade level		
2	1,226	50%
4	1,227	50%
Gender *		
Female	1,212	49%
Male	1,239	51%
Group		
Treatment	1,548	63%
Comparison	903	37%

*Due to data missing for gender, the sum of the female and male counts does not equal the total.

Within the treatment and comparison groups, the number of students by province, gender, and grade was relatively equal, as displayed in ***Table 5***.

Table 5. Characteristics of the Student Sample, by Group (n = 2,453)

Variable	Treatment	Comparison	Total
Province			
Bandundu	507 (62%)	310 (38%)	817 (100%)
Equateur	520 (65%)	286 (35%)	806 (100%)
Orientale	521 (63%)	309 (37%)	830 (100%)
Grade level			
2	773 (63%)	453 (37%)	1,225 (100%)
4	775 (63%)	452 (37%)	1,228 (100%)
Gender*			
Female	788 (64%)	452 (36%)	1,241 (100%)
Male	758 (63%)	453 (37%)	1,210 (100%)
Total	1,546 (63%)	905 (37%)	2,451 (100%)

*Due to data missing for gender, the sum of the female and male counts does not equal the total.

Table 6 displays the proportion of schools in each group by school-type. As this table indicates, there were no differences in the proportion of private, unregistered (*~~–n~~on conventionnée*"), Catholic, and Protestant schools across comparison and treatment groups. In addition, a chi-square analysis showed no association between school type and comparison/treatment schools.

Table 6. School Type, by Group (n = 95)

School Type	Comparison	Treatment	Total
Private	4 (11%)	2 (3%)	6 (6%)
Unregistered	9 (26%)	18 (30%)	27 (28%)
Catholic	9 (26%)	16 (53%)	25 (27%)
Protestant	13 (37%)	24 (40%)	37 (39%)
Total	**35 (100%)**	**60 (100%)**	**95 (100%)**

Note: No differences across comparison and treatment groups were significant at the $p < 0.05$ level.

Table 7 shows how the students sampled in each school group responded to a series of questions targeting SES indicators regarding possessions in the home. As this table shows, with the exception of bicycle and canoe ownership, the students in treatment schools reported having more of the possessions listed in their homes, suggesting a higher average level of SES.[11]

Table 7. Student SES Indicators, by Group (n = 2,453)

SES Item	Percentages Comparison	Treatment	Total
Radio[b]	73.8	81.4	78.6
Telephone[b]	54.3	84.3	73.2
Electricity in the home[b]	8.5	36.9	26.4
Television[b]	11.0	41.9	30.5
Refrigerator[b]	2.2	13.7	9.5
Indoor toilets[b]	4.2	10.2	8.0
Bicycle[b]	70.2	65.3	67.1
Motorcycle[b]	21.4	31.0	27.5
Canoe[a]	12.9	9.6	10.8
Motor vehicle[b]	3.1	8.7	6.6

[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.

[11] Having several highly correlated independent variables (such as the SES-related questions shown in ***Table 7***) can produce unstable estimates; reducing those to a smaller number of principal components will yield a better test and/or more stable estimates of regression coefficients. Therefore, for the purposes of regression analyses incorporating SES, the students' answers to the questions regarding possessions in the home were collapsed into a single variable using principal components analysis (PCA). A comparison of these composite scores also showed treatment group students to have statistically significantly (at the 0.05 level) higher levels of ownership than comparison group students.

In order to account for the differences in SES between comparison and treatment groups, administrators designed the regression analyses reported later in this report to include SES in their models.

Test Administration and Data Collection

In each province, data collectors were hired through provincial institutions and universities. Two RTI EGRA specialists, in collaboration with local project staff, led intensive provincial-level training for data collectors. A majority of the data collectors brought prior experience administering EGRA through participation in this project's baseline assessment as well as in other EGRA projects in the province. Training included the following agenda.

- Welcome, introductions, EGRA overview, mock EGRA administration
- Reliability test, subtask practice administrations, review of test, guided practice, independent practice
- Additional assessment practice, reliability test, introduction to questionnaires
- Field testing in local schools, discussion of preliminary findings, additional practice
- Reliability tests, discussion of school-level data collection, team assignments

Inter-rater reliability (IRR) was evaluated throughout training. Data collectors averaged 93%, which is considered acceptable for this type of educational achievement testing and lends confidence to the findings.

The midterm assessments were administered by 21 teams of four data collectors, deployed over a period of approximately 3.5 weeks in May 2012, which is the final full month of the school year. A supervisor from the MEPSP led each team in the field, which tested students from a single school over the course of two consecutive days. Data entry followed the collection period and was carefully quality-comparison led through use of a software program designed to limit the possibility of error and to perform daily quality comparison checks for 10% of the data.

Analytical Methodology

The main analytical methods used to examine the EGRA data for this report included descriptive analyses and simple regression. All analyses utilized the final analysis weight (see earlier Probability of Selection and Design Weight subsection). Analyses included means comparisons, percentage comparisons, and regression analyses. Statistically significant differences are reported were applicable. Regression analysis was used to examine (and provide comparison for, as appropriate) the effects of key school characteristics (i.e., SES) on performance. Where the 2012 (midterm) data are compared to the 2010 (baseline) data, a subset of data that correspond with the reduced 2010 sample (see the Reporting Constraints subsection earlier in this report) was used. It is important to note that it was determined that treatment schools in the reduced 2010 sample were not

qualitatively different from the comparison schools in the reduced 2010 sample (given limited levels of program implementation within the treatment schools), and so the reduced 2010 sample comparison (3) and treatment (13) schools were collapsed into one group for greater statistical power.

4. PAQUED Midterm Findings

As described in the Sampling Methodology subsection of this report, EGRA was administered to both comparison and treatment students in Grades 2 and 4.[12] ***Table 8*** displays for all subtasks the percentages of students in treatment and comparison groups with zero scores (students unable to generate at least one correct response for a subtask), the percentages of items attempted, and mean scores. In each subsequent subsection of this report section, the results of descriptive analyses and significance testing are reported for each of the EGRA subtasks separately. Within the results of each subtask, means are reported by three key units: group (i.e., comparison or treatment), student gender, and province. Percent attempted percentages by gender and provinces are provided in the annex to this report.

Table 8. Percentages of Students with Zero Scores and Percent Attempted

Grade	Group	% Students with Zero Scores	Percent Attempted	Mean
Total Vocabulary				
2	Comparison	1.00%	36.4%	7.28
	Treatment	0.87%	41.8%	8.36
4	Comparison	0	50.2%	10.04
	Treatment	0	54.9%	10.98
Initial Sound Identification				
2	Comparison	68.7%	10.3%	1.03
	Treatment	68.8%	12.5%	1.25
4	Comparison	53.3%	21.1%	2.11
	Treatment	45.5	28.0%	2.80

[12] Familiar word reading, invented word reading, oral reading fluency, and reading comprehension subtasks were administered to Grade 4 students only.

Grade	Group	% Students with Zero Scores	Percent Attempted	Mean
Letter Sound Knowledge				
2	Comparison	23.4%	28.4%	6.55
	Treatment	18.9%	32.3%	7.65
4	Comparison	6.4%	59.8%	21.16
	Treatment	5.2%	60.1%	21.10
Familiar Word Reading				
4	Comparison	39.6%	31.6%	7.42
	Treatment	37.7%	32.4%	7.75
Invented Word Reading				
4	Comparison	46.4%	26.9%	6.23
	Treatment	50.7%	25.3%	6.39
Oral Word Reading				
4	Comparison	51.3%	28.3%	8.56
	Treatment	52.1%	28.0%	8.76
Reading Comprehension				
4	Comparison	80.8%	10.1%	0.33
	Treatment	82.6%	8.3%	0.31
Listening Comprehension				
2	Comparison	81.3%	6.6%	0.33
	Treatment	72.0%	11.3%	0.58
4	Comparison	59.2%	17.5%	0.87
	Treatment	49.0%	22.1%	1.10

Vocabulary

The Vocabulary subtask assesses students' basic French vocabulary across three categories of word knowledge: parts of the body, words of objects from the school environment, and spatial locations. Because this subtask taps three types of words or

concepts, student performance on each class of words is presented separately. Student performance on the aggregate vocabulary score—comprising all three categories—is also presented following the subtask results.

Table 9 presents results across the three vocabulary subtasks for students in Grade 2; statistically significant differences are bolded. ***Table 10*** presents results across the three vocabulary subtasks for students in Grade 4; again, statistically significant differences are bolded. Figures 1 through 12 display this information graphically to facilitate comparisons across grades and genders.

Table 9. Correctly Identified Vocabulary Words Among Grade 2 Students by Group and Gender: Percentages (Standard Errors)[13]

Number Correctly Answered	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
Parts of the Body						
0	18.7 (3.3)	13.8 (2.5)	15.7 (3.2)	21.7 (4.1)	10.8 (1.7)	16.8 (4.3)
1–2	43.5 (3.4)	42.2 (2.7)	45.7 (4.4)	41.2 (3.8)	44.1 (4.0)	40.4 (3.0)
3–4	30.2 (2.5)	32.7 (3.1)	32.2 (3.3)	28.2 (3.0)	34.9 (3.5)	30.6 (3.4)
5–6	7.4 (2.0)	9.9 (1.6)	6.1 (2.0)	8.8 (2.6)	9.5 (1.9)	10.2 (2.0)
7–8	**0.1[a] (.05)**	**1.2 (.50)**	.13 (.09)	nd (nd)	. 50 (.39)	1.8 (.96)
Words from the School Environment						
0	2.7 (.92)	1.4 (.55)	3.2 (1.3)	2.2 (1.0)	1.5 (.63)	1.4 (.71)
1–2	**10.2[a] (1.8)**	**5.0 (1.3)**	8.1 (1.8)	**12.5[b] (2.8)**	6.0 (1.9)	**3.9 (1.0)**
3–4	**55.4[a] (3.1)**	**43.9 (3.3)**	**59.4[b] (4.4)**	51.4 (3.6)	**40.7 (3.5)**	47.0 (4.1)
5–6	**31.4[b] (3.6)**	**49.6 (4.1)**	**29.2[b] (4.7)**	**33.7[a] (3.9)**	**51.6 (4.5)**	**47.6 (4.5)**

13 Throughout the remainder of the report, for tables in which weighted percentages are used, sample size is not provided.

Number Correctly Answered	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
Spatial Terms						
0	28.1 (3.3)	22.0 (1.7)	26.9 (4.0)	**29.3[a] (3.7)**	24.1 (2.8)	**19.8 (2.5)**
1–2	53.9 (3.1)	49.2 (2.0)	57.0 (4.4)	50.7 (3.4)	50.9 (2.8)	47.5 (3.0)
3–4	**16.1[b] (2.2)**	**24.6 (2.0)**	13.8 (2.7)	**18.4[b] (2.6)**	**20.0[c] (2.3)**	**29.1 (2.6)**
5–6	1.8 (.78)	4.2 (1.0)	2.2 (.83)	1.4 (1.1)	4.8 (1.5)	3.5 (1.1)

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED-supported schools and comparison schools are not PAQUED-supported schools.
NOTE: Percentages are defined as number of students who correctly identified the word divided by the number of students who attempted to identify the word.
[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

Table 10. Correctly Identified Vocabulary Words Among Grade 4 Students by Group and Gender: Percentages (Standard Errors)

Number Correctly Answered	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
Parts of the Body						
0	7.6 (2.3)	3.5 (1.0)	6.1 (2.5)	9.1 (2.6)	3.4 (.97)	3.6 (1.4)
1–2	**27.6[a] (2.8)**	**19.7 (2.3)**	23.3 (3.7)	**32.0[a] (3.5)**	19.4 (2.7)	**20.1 (2.9)**
3–4	**41.0[a] (3.2)**	**50.4 (2.0)**	**39.6[a] (4.5)**	42.5 (3.3)	**50.4 (2.9)**	50.3 (2.7)
5–6	21.9 (3.0)	23.7 (2.6)	**29.4[d](4.0)**	**14.3 (2.9)**	23.5 (3.2)	23.9 (3.1)
7–8	1.7 (.68)	2.5 (.96)	1.6 (.94)	1.9 (1.0)	3.1 (1.1)	1.9 (1.0)
Words from the School Environment						
0	.35 (.18)	0.5 (.27)	0.2 (.14)	0.5 (.34)	0.6 (.42)	0.5 (.34)
1–2	4.2 (1.3)	2.2 (.73)	5.9 (2.1)	2.6 (1.2)	2.2 (.81)	2.1 (.85)
3–4	**46.3[b] (3.1)**	**31.5 (3.7)**	**39.2[d](3.8)**	**53.4[b] (4.1)**	29.0 (4.7)	**34.2 (3.7)**
5–6	**49.1[b] (3.4)**	**65.7 (4.1)**	**54.6[a],[c](3.7)**	**43.5[b] (4.4)**	**68.2 (5.1)**	**63.0 (4.1)**

Number Correctly Answered	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
Spatial Terms						
0	13.2 (1.8)	11.3 (1.8)	14.1 (3.0)	12.3 (2.5)	11.6 (2.3)	11.0 (2.3)
1–2	37.4 (3.2)	31.6 (2.7)	**28.4 [d] (4.3)**	**46.5 [a] (4.0)**	27.9 (3.3)	**35.5 (3.3)**
3–4	34.1 (2.6)	35.0 (2.7)	37.5 (3.8)	30.7 (3.2)	37.9 (3.0)	32.0 (3.5)
5–6	15.1 (2.7)	21.9 (2.9)	**19.8 [d] (3.5)**	**10.4 [a] (2.6)**	22.4 (3.5)	**21.3 (3.6)**

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED-supported schools and comparison schools are not PAQUED-supported schools.
NOTE: Percentages are defined as number of students who correctly identified the word divided by the number of students who attempted to identify the word.
[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

Parts of the Body. Children's recognition of French words for parts of the body (such as foot, arm) is presented in ***Table 9*** and ***Figure 1*** for Grade 2 students and in ***Table 10*** and ***Figure 2*** for Grade 4 students. As indicated in ***Table 9*** and displayed graphically in ***Figure 1***, overall for Grade 2, students in the treatment group were more likely than comparison-group students to respond correctly to 7–8 (1.2% versus 0.1%, respectively, $p<0.05$) French vocabulary words; no other levels were statistically significant. Nearly 14% of Grade 2 treatment students were unable to correctly identify any vocabulary, whereas nearly 19% of comparison students were unable to.

As shown in ***Table 10*** and ***Figure 2***, Grade 4 comparison students were statistically significantly more likely than treatment students to correctly answer 1–2 words (27.6% versus 19.7%, respectively, $p<0.05$) and significantly less likely to correctly answer 3–4 words (41.0% versus 50.4%, respectively, $p<0.05$).

Figure 1. Grade 2 Percentages of Correct Identification of Parts of the Body (By Number of Words Identified Correctly)

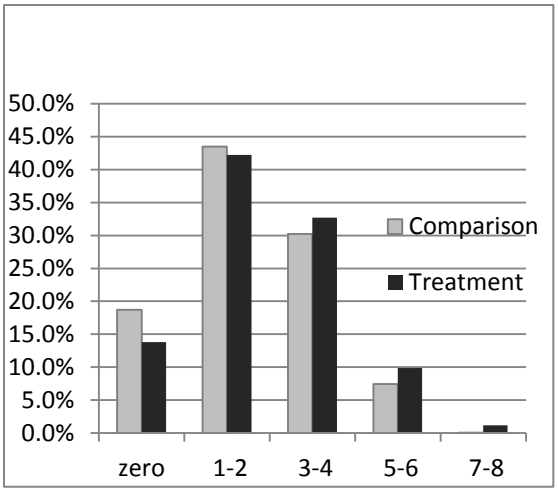
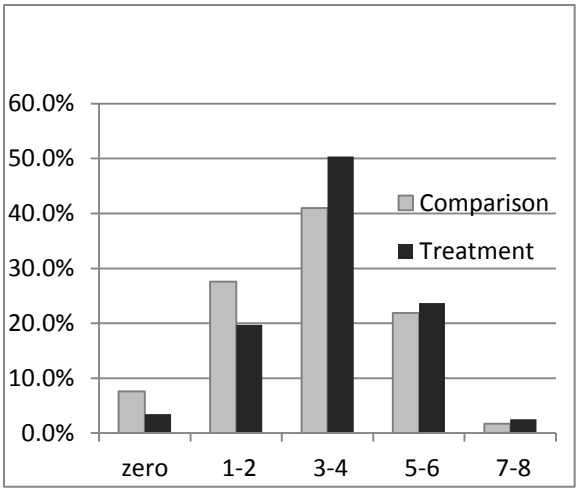
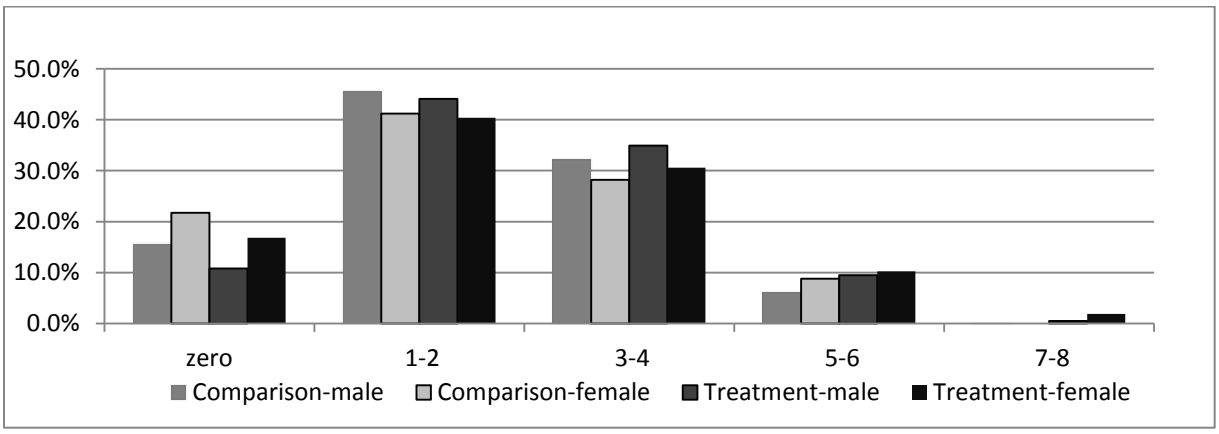


Figure 2. Grade 4 Percentages of Correct Identification of Parts of the Body (By Number of Words Identified Correctly)



Figures 3 and 4 display results by gender. As indicated in ***Table 9*** and graphically displayed in ***Figure 3***, patterns between Grade 2 comparison and treatment groups were similar across gender and similar within group; none of these differences is statistically significant at $p<0.05$.

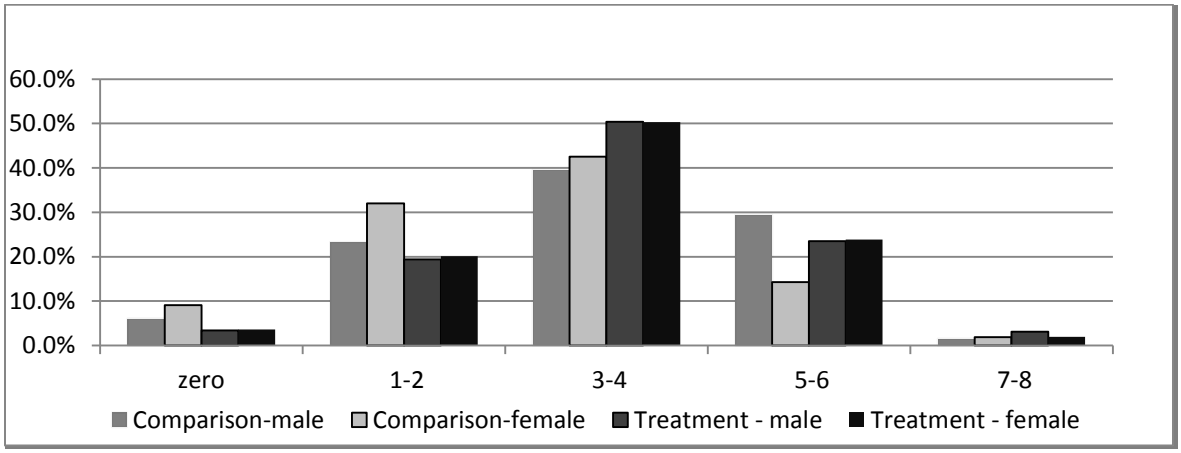
Figure 3. Grade 2 Percentages of Correct Identification of Parts of the Body by Group and Gender (By Number of Words Identified Correctly)



As shown in ***Table 10*** and ***Figure 4***, female Grade 4 treatment students performed on par with male Grade 4 treatment students, whereas female comparison students tended to underperform male comparison students. This difference between female and male comparison students is statistically significant at the 5–6 correct words level of

proficiency (14.3% versus 29.4%, respectively, p<0.01). Comparing across treatment conditions, comparison group females were more likely than treatment females to respond correctly to only 1–2 words (32.0% versus 20.1%, respectively, p<0.05), whereas treatment group males were more likely than comparison group males to respond correctly to 3–4 words (50.4% versus 39.6%, respectively, p<0.05).

Figure 4. Grade 4 Percentages of Correct Identification of Parts of the Body by Group and Gender (By Number of Words Identified Correctly)



Words from the School Environment. As indicated in ***Table 9*** and displayed graphically in ***Figure 5***, almost half the Grade 2 students in treatment schools (49.6%) and a third of students in comparison schools (31.4%) could identify at least five of the six words; difference between groups is statistically significant at p<0.05. Comparison students were more likely than treatment students to respond correctly to 1–2 words (10.2% versus 5.0%, respectively, p<0.05), and 3–4 words (55.4% versus 43.9%, p<0.01).

Table 10 and ***Figure 6*** show the same pattern for Grade 4, with almost two-thirds of the students in treatment schools (65.7%) and half the students in comparison schools (49.1%) identifying at least five of the six words (difference between groups statistically significant at p<0.01). Overall, few students were unable to identify any of the words from the environment; 2.7% of the Grade 2 students in comparison schools and 1.4% of the Grade 2 students in treatment schools obtained zero scores, while fewer than 1% of the Grade 4 students in both types of schools obtained zero scores.

Figure 5. Grade 2 Percentages of Correct Identification of Words from the School Environment (By Number of Words Identified Correctly)

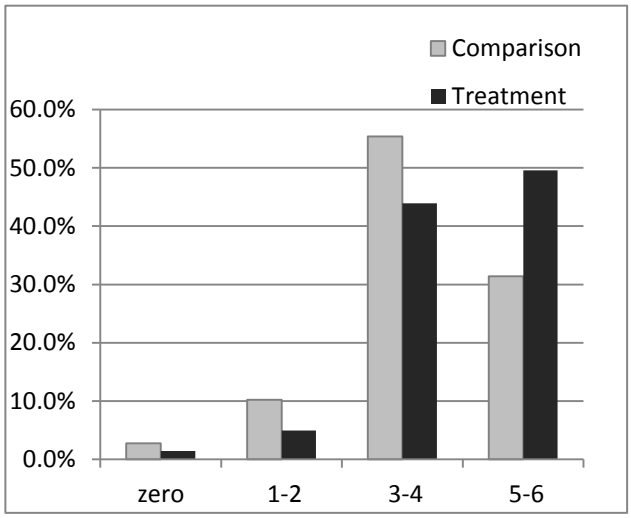
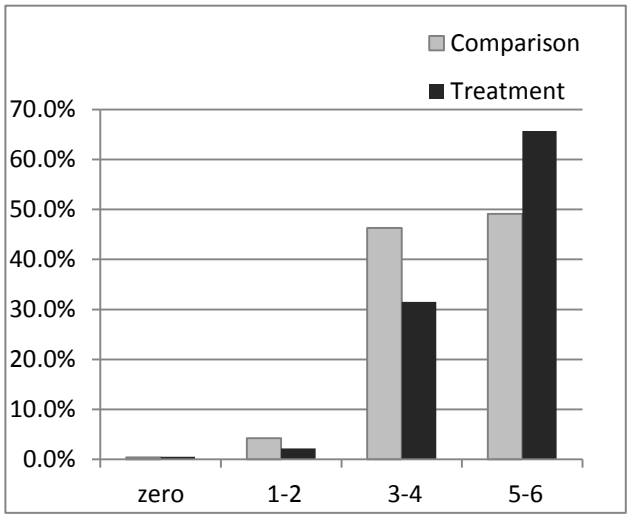
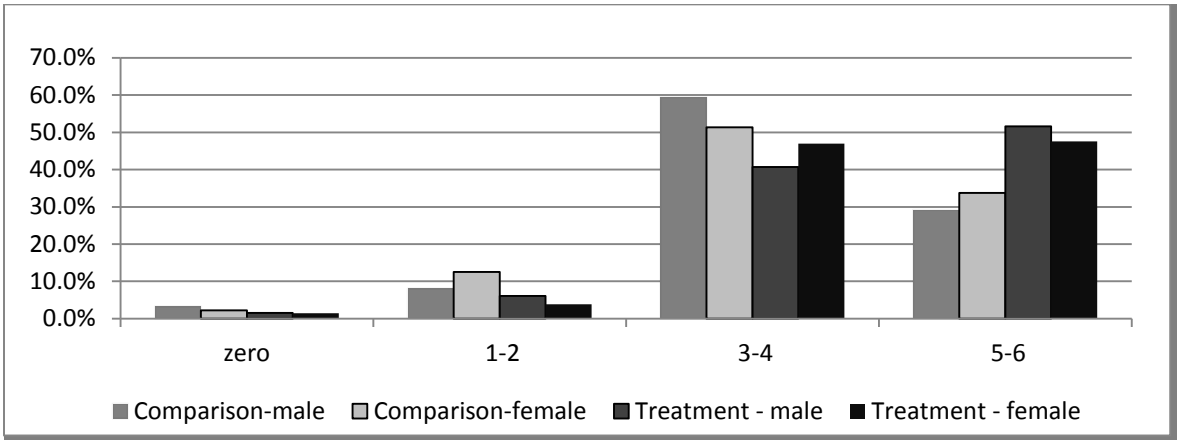


Figure 6. Grade 4 Percentages of Correct Identification of Words from the School Environment (By Number of Words Identified Correctly)



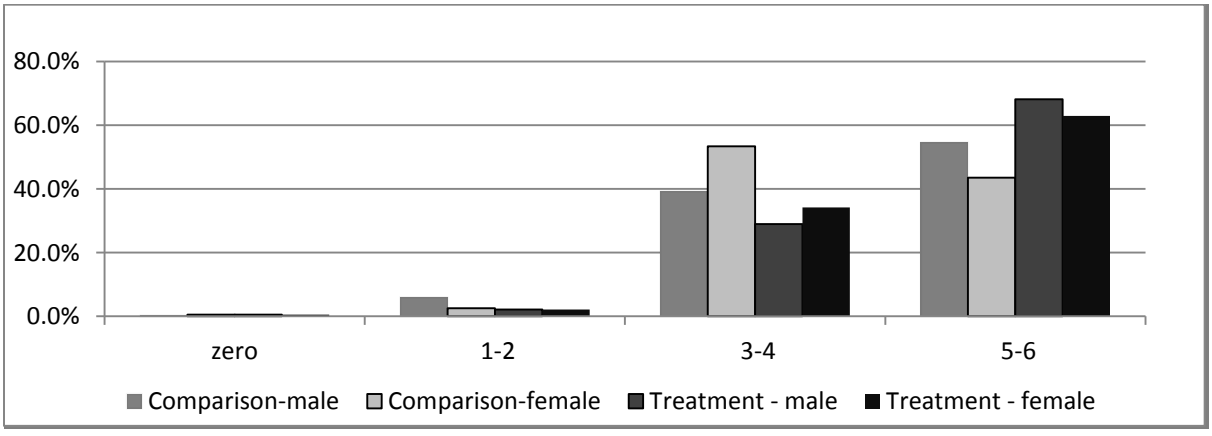
Comparing across gender, as indicated in ***Table 9*** and graphically displayed in ***Figure 7***, Grade 2 treatment group females were statistically significantly more likely than their comparison counterparts to respond correctly to 5–6 words (47.6% versus 33.7%, respectively, $p<0.05$), as were treatment group males (51.6% for treatment versus 29.2% for comparison, $p<0.01$). Comparison group males and females were more likely than treatment group males and females to respond correctly to fewer than five words, although only two of these differences were statistically significant (3–4 words correct: 59.4% for comparison males versus 40.7% for treatment males, $p<0.01$; and 1–2 words correct: 12.5% for comparison females versus 3.9% for treatment females, $p<0.01$).

Figure 7. Grade 2 Percentages of Correct Identification of Words from the School Environment by Groups and Gender (By Number of Words Identified Correctly)



As shown in ***Table 10*** and ***Figure 8***, Grade 4 treatment males were more likely than comparison group males to correctly identify 5–6 words (68.2% versus 54.6%, respectively, $p<0.05$), as were treatment females compared with comparison females (63.0% versus 43.5%, respectively, $p<0.01$). Comparison males were more likely than comparison females to respond correctly to that many words (54.6% versus 43.5%, respectively, $p<0.05$). Conversely, more comparison females than treatment females were likely to respond correctly to 3–4 words (53.4% versus 34.2%, respectively, $p<0.01$), and comparison females more likely than comparison males to respond correctly to 3–4 words (53.4% versus 39.2%, respectively, $p<0.01$).

Figure 8. Grade 4 Percentages of Correct Identification of Words from the School Environment by Group and Gender (By Number of Words Identified Correctly)



Spatial Terms. Unlike the words mastered in the school environment, it appears that Grade 2 students had not yet mastered more abstract, spatial terms (such as behind you, in front of you). As shown in ***Table 9*** and ***Figure 9***, only 4.2% of the students in treatment schools and 1.8% of students in comparison schools could follow directions to show their understanding of at least five of the six words (difference between groups not

statistically significant); 22.0% of students in treatment schools and 28.1% of students in comparison schools could not correctly follow even one direction (difference between groups not statistically significant). The only statistically significant difference between treatment and comparison groups was for the 3–4 words correct category, in which treatment students outperformed comparison students (24.6% versus 16.1%, $p<0.01$).

As indicated in ***Table 10*** and ***Figure 10***, 21.9% of the Grade 4 students in treatment schools and 15.1% of students in comparison schools were able to follow directions to show their understanding of at least five of the six words (difference between groups not statistically significant); only 11.3% of the Grade 4 students in treatment schools and 13.2% of students in comparison schools obtained zero scores (difference between groups not statistically significant).

Figure 9. Grade 2 Percentages of Correct Identification of Spatial Terms (By Number of Words Identified Correctly)

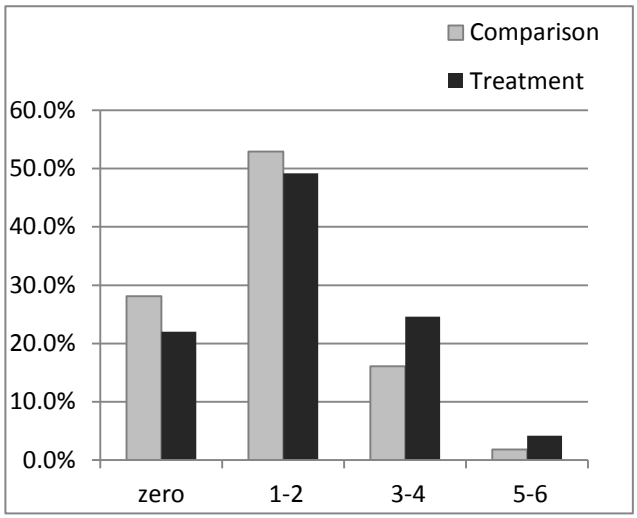
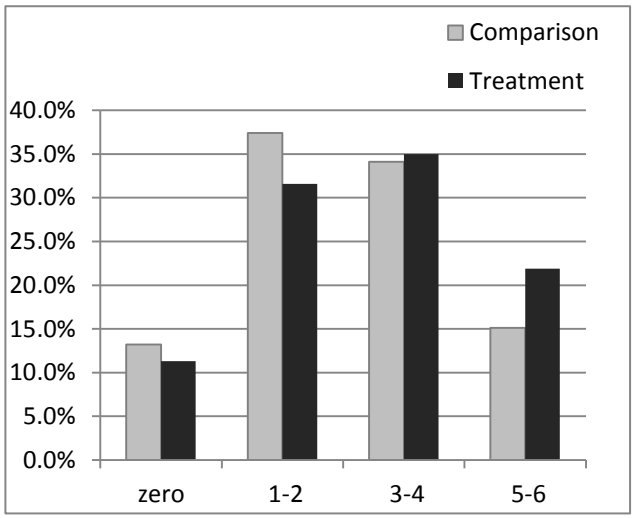
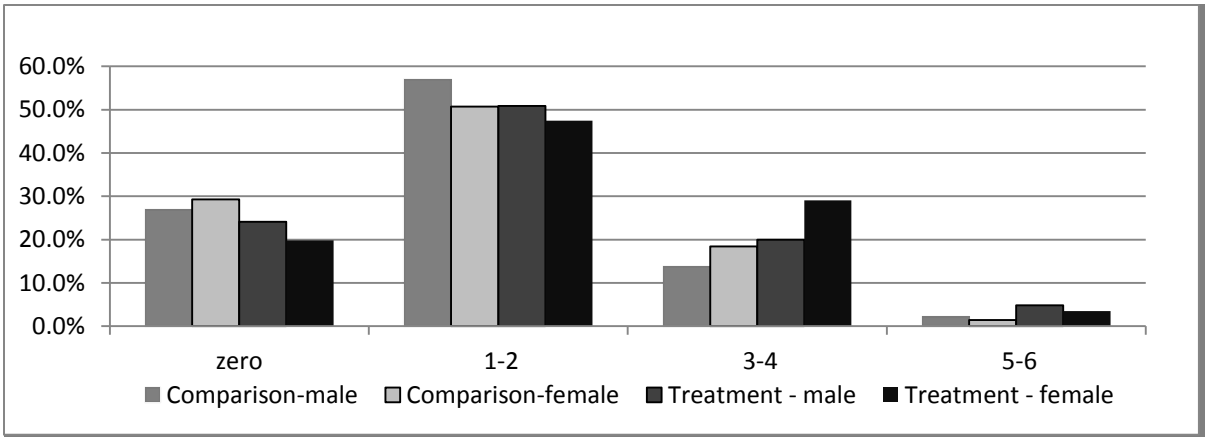


Figure 10. Grade 4 Percentages of Correct Identification of Spatial Terms (By Number of Words Identified Correctly)



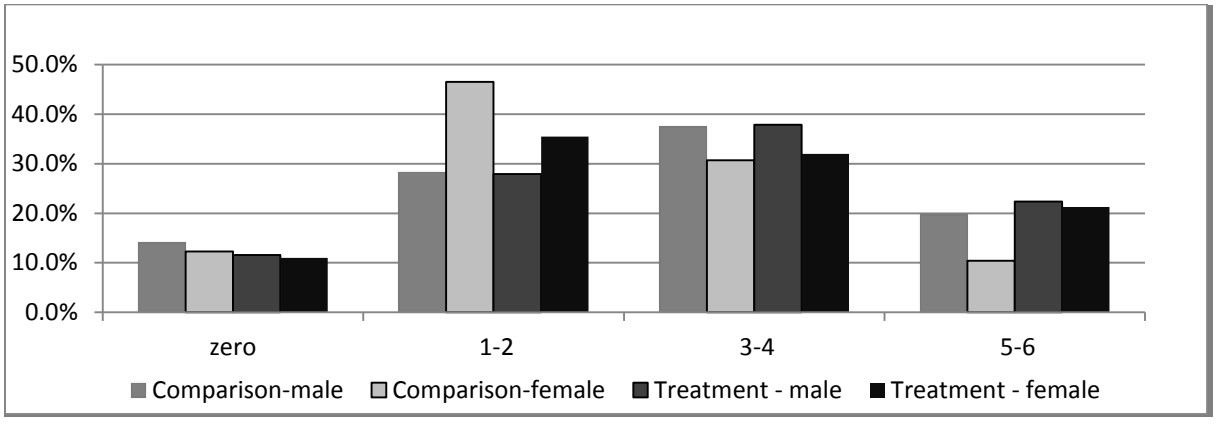
Comparing across gender, as indicated in ***Table 9*** and graphically displayed in ***Figure 11***, Grade 2 treatment group females were statistically significantly more likely than their comparison counterparts to respond correctly to 3–4 words (29.1% versus 18.4%, respectively, $p<0.01$), while 29.3% of comparison females were unable to identify any words compared with 19.8% of their treatment counterparts ($p<0.05$). No other differences across treatment conditions by gender were statistically significant. Within the treatment group, females were more likely than males to correctly identify 3–4 words (29.1% versus 20.0%, respectively, $p<0.05$).

Figure 11. Grade 2 Percentages of Correct Identification of Spatial Terms by Group and Gender Percentages
(By Number of Words Identified Correctly)



Within Grade 4, as indicated in ***Table 10*** and graphically displayed in ***Figure 11***, treatment group females were statistically significantly more likely than their comparison counterparts to respond correctly to 5–6 words (21.3% versus 10.4%, respectively, $p<0.05$), while 46.5% of comparison females were able to identify only 1–2 words compared with 35.5% of their treatment counterparts ($p<0.05$). No other differences across treatment conditions by gender were statistically significant. Within the comparison group, males were more likely than females to correctly identify 5–6 words (19.8% versus 10.4%, respectively, $p<0.01$), while females were more likely than males to identify only 1–2 words (46.5% versus 28.4%, respectively, $p<0.01$).

Figure 12. Grade 4 Percentages of Correct Identification of Spatial Terms by Group and Gender
(By Number of Words Identified Correctly)



Overall Vocabulary Performance. While the previous three subsections display results by type of vocabulary assessed, the analyses reported in this subsection combined all vocabulary words into a single overall vocabulary score.

Comparisons by Grade and Treatment Condition. ***Table 11*** presents overall vocabulary percentages and standard errors for Grade 2 students, by comparison and treatment group

and by gender, with statistically significant differences bolded. ***Table 12*** presents overall vocabulary scores for Grade 4 students, also by comparison and treatment group and by gender, and with statistically significant differences bolded.

Table 11. Correctly Identified Vocabulary Words among Grade 2 Students by Group and Gender: Percentages (Standard Errors)

Number Correctly Answered	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
Across 3 Subtasks						
0	1.0 (.73)	0.9 (.31)	1.5 (1.0)	0.6 (.51)	1.2 (.54)	0.5 (.27)
1–4	15.5 (2.5)	10.5 (2.4)	13.9 (2.7)	17.1 (3.6)	9.1 (2.0)	11.9 (3.1)
5–8	54.4[b] (2.9)	39.7 (2.7)	56.0[a] (3.9)	52.7[b] (3.7)	43.2[c] (3.2)	36.3 (3.1)
9–12	23.8[b] (2.6)	39.2 (3.3)	23.1[b] (3.8)	24.6[b] (3.4)	37.8 (3.6)	40.6 (4.0)
13–16	4.3 (1.7)	8.9 (1.8)	4.8 (2.1)	3.8[a] (1.8)	8.1 (2.1)	9.7 (2.0)
17–20	0.9 (.70)	0.7 (.28)	0.6 (.41)	1.2 (1.1)	0.5 (.32)	0.8 (.47)

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED-supported schools and comparison schools are not PAQUED-supported schools.
NOTE: Percentages are defined as number of students who correctly identified the word divided by the number of students who attempted to identify the word.
[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

Table 12. Correctly Identified Vocabulary Words Among Grade 4 Students by Group and Gender: Percentages (Standard Errors)

Number Correctly Answered	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
Across 3 Subtasks						
0	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)
1–4	5.9 (2.1)	2.4 (.76)	5.4 (2.1)	6.3 (2.9)	2.2 (.87)	2.6 (.90)
5–8	27.8 (2.6)	21.6 (2.3)	22.0[c] (3.3)	33.5[a] (3.6)	21.8 (3.0)	21.4 (2.7)
9–12	40.4 (3.2)	41.9 (2.3)	37.4 (3.6)	43.4 (4.3)	39.7 (3.1)	44.3 (3.0)

Number Correctly Answered	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
13–16	23.6 (2.5)	29.6 (2.5)	31.1[d] (3.6)	16.1[a] (2.9)	31.8 (3.6)	27.4 (3.2)
17–20	2.2 (1.0)	4.3 (1.2)	3.9[c] (1.8)	0.5[a] (.41)	4.4 (1.4)	4.3 (1.6)

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED-supported schools and comparison schools are not PAQUED-supported schools.
NOTE: Percentages are defined as number of students who correctly identified the word divided by the number of students who attempted to identify the word.
[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

As indicated in ***Table 11*** and ***Figure 13***, few Grade 2 students in both treatment and comparison schools demonstrated mastery of either 13–16 (8.9% versus 4.3%, respectively) or 17 or more (0.7% versus 0.9%, respectively) of the 20 French vocabulary words (differences between groups not statistically significant). Treatment students were more likely than comparison students to correctly identify 9–12 words (39.2% versus 23.8%, respectively, $p<0.01$), whereas comparison students were more likely to identify 5–8 words (54.4% versus 39.7%, respectively, $p<0.01$).

Similarly, as shown in ***Table 12*** and ***Figure 14***, only 4.3% of the Grade 4 students in treatment schools and 2.2% of the Grade 4 students in comparison schools had at least 17 correct answers, although somewhat more students (29.6% for treatment and 23.6% for comparison students) were able to correctly identify 13–16 words (differences between groups not statistically significant). All Grade 4 students were able to recognize at least one word.

Figure 13. Grade 2 Percentages of Total Vocabulary Scores (By Number of Words Identified Correctly)

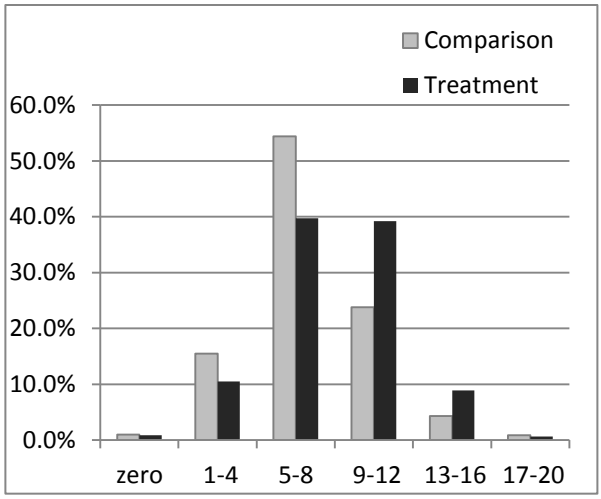
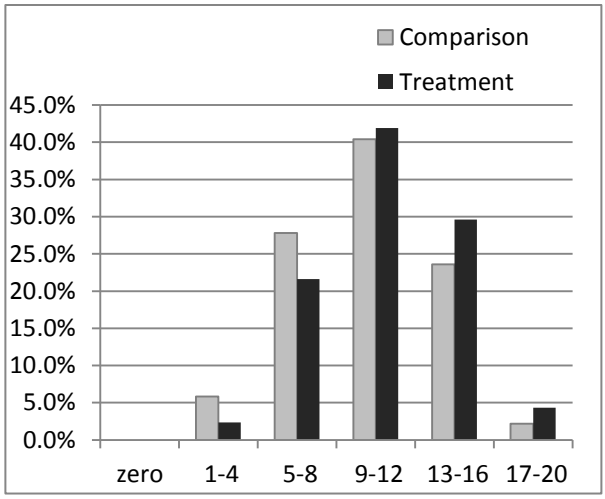


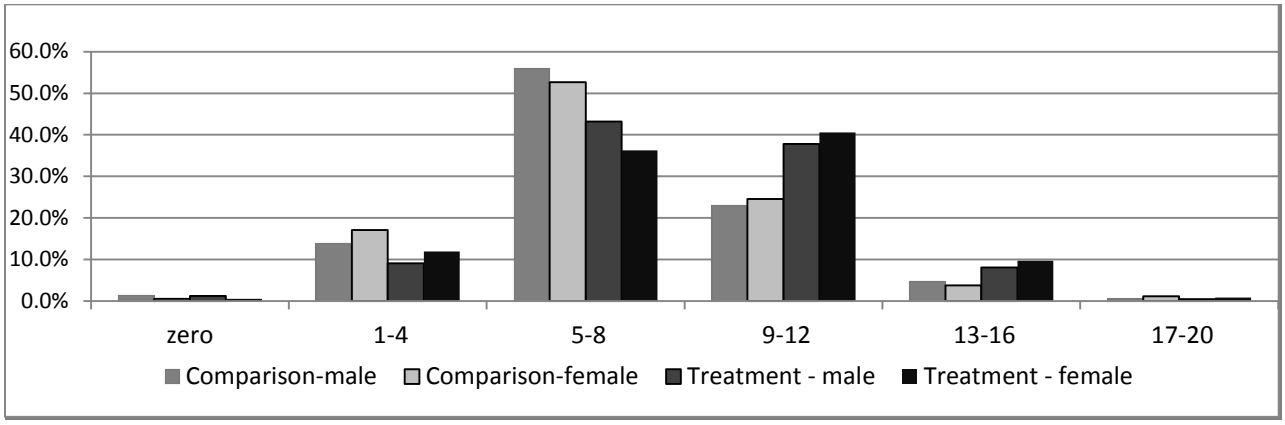
Figure 14. Grade 4 Percentages of Total Vocabulary Scores (By Number of Words Identified Correctly)



Comparisons by Grade, Treatment Condition, and Gender. Looking at Grade 2 students' total vocabulary scores by gender, ***Table 11*** and ***Figure 15*** show several areas of statistically significant differences between genders and between groups.

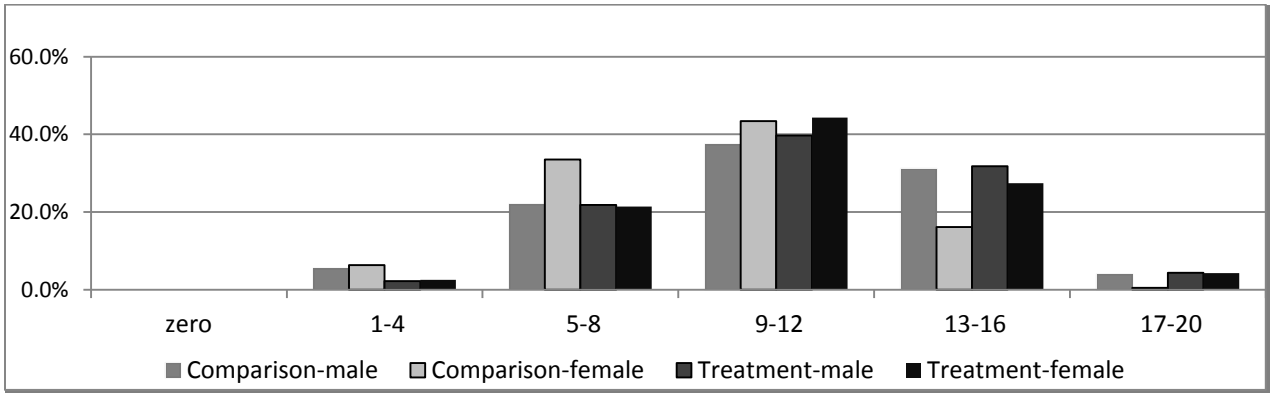
- Across treatment conditions, treatment females were more likely than comparison females to correctly identify 13–16 words (9.7% versus 3.8%, respectively, $p<0.0$) and to correctly identify 9–12 words (40.6% versus 24.6%, respectively, $p<0.01$).
- Treatment males were more likely than their comparison counterparts to identify 9–12 words correctly (37.8% versus 23.1%, respectively, $p<0.01$).
- Conversely, comparison females were more likely than treatment females to correctly identify only 5–8 words (52.7% versus 36.3%, respectively, $p<0.01$), and comparison males were more likely than treatment males to correctly identify that many words (56.0% versus 43.2%, respectively, $p<0.05$).
- Within the treatment group, males were statistically significantly more likely than females to correctly identify 5–8 words (43.2% versus 36.3%, respectively, $p<0.05$); no other within-treatment group differences were statistically significant at $p<0.05$.

Figure 15. Grade 2 Percentages of Total Vocabulary Scores by Group and Gender (By Number of Words Identified Correctly)



As displayed in ***Table 12*** and ***Figure 16***, higher percentages of students correctly identified more words than was seen among Grade 2 students. Unlike in Grade 2, however, Grade 4 males tended to outperform females. While there was no statistically significant difference between genders in the treatment group, comparison-group males outperformed females in correctly identifying 17–20 words (3.9% versus 0.5%, respectively, $p<0.05$) and 13–16 words (31.1% versus 16.1%, respectively, $p(<0.01)$. Conversely, comparison females were more likely than males to correctly identify only 5–8 words (33.5% versus 22.0%, respectively, $p<0.05$). Comparing across treatment conditions, females from treatment schools outperformed their comparison counterparts in identifying 17–20 words correctly (4.28% versus 0.5%, respectively, $p<0.05$) and in identifying 13–16 words correctly (27.4% versus 16.1%, respectively, $p<0.05$). Conversely, comparison females were more likely than treatment females to identify only 5–8 words correctly (33.5% versus 21.4%, respectively, $p<0.05$).

Figure 16. Grade 4 Percentages of Total Vocabulary Scores by Group and Gender (By Number of Words Identified Correctly)



Comparisons by Grade, Treatment Condition, and Province. ***Table 12*** and ***Table 13*** present findings of statistical analysis of Grade 2 and Grade 4 overall student vocabulary performance in treatment and comparison schools by province, with areas of statistical

difference bolded. As displayed in ***Table 13***, the following significant differences emerged.

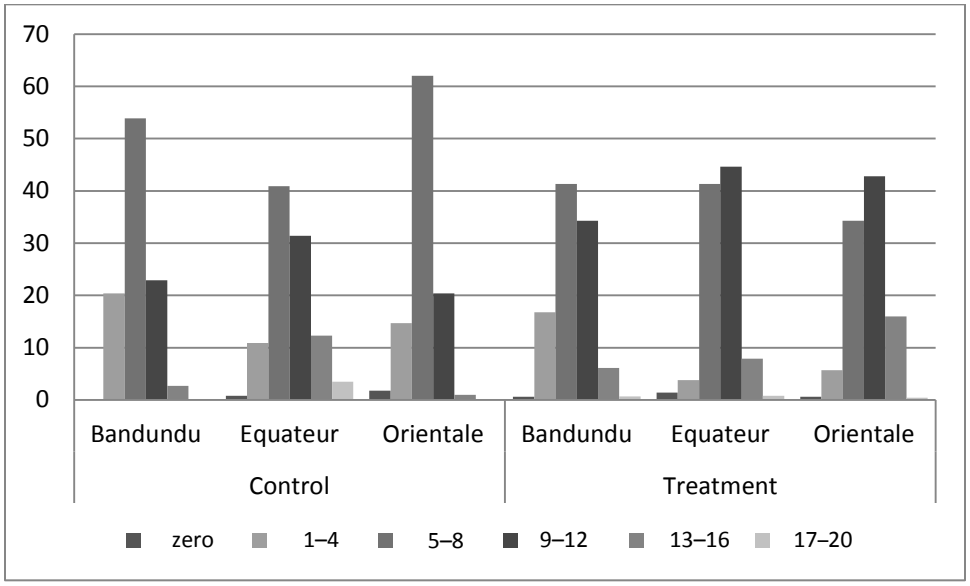
- Overall, Grade 2 students in Bandundu were more likely to identify only 1–4 words correctly than were students in either of the other two provinces (17.0% Bandundu, 6.9% Orientale, 4.2% Equateur, $p<0.05$).
- Breaking students into treatment and comparison groups shows that Bandundu comparison students were more likely than Bandundu treatment students to correctly identify 5–8 words (fewer than half of the 20 total words; 53.9% versus 41.3%, respectively, $p<0.05$).
- Similarly, Orientale comparison students were more likely than Orientale treatment students to correctly identify 5–8 words (62.0% versus 34.3%, respectively, $p<0.01$).
- Within the comparison group, Bandundu students were more likely than Equateur students to identify 5–8 words correctly (53.9% versus 40.9%, respectively, $p<0.05$), as were Orientale students (62.0% versus Equateur's 40.9%, $p<0.01$).

Table 13. Correctly Identified Vocabulary Words Among Grade 2 Students by Province and Group: Percentages (Standard Errors)

Number Correctly Answered	Province			Group by Province					
	Bandundu	Equateur	Orientale	Bandundu	Comparison Equateur	Orientale	Bandundu	Treatment Equateur	Orientale
Across 3 Subtests									
0	0.6 (.42)	1.4 (.66)	0.8 (.42)	nd (nd)	0.8 (.55)	1.8 (1.5)	0.6 (.44)	1.4 (.70)	0.6 (.42)
1–4	17.0[a,d] (4.2)	4.2 (1.9)	6.9 (1.7)	20.4 (3.6)	10.9 (3.7)	14.7 (4.5)	16.8[a,d] (4.4)	3.8 (2.0)	5.7 (1.8)
5–8	41.8 (3.9)	41.3 (4.3)	37.9 (4.6)	53.9[c,e] (3.2)	40.9[b] (5.0)	62.0[f] (4.7)	41.3 (4.1)	41.3 (4.5)	34.3 (5.0)
9–12	33.8 (5.5)	43.8 (3.0)	39.9 (4.3)	22.9 (4.0)	31.4 [e] (4.4)	20.4[f] (4.6)	34.3 (5.8)	44.6 (3.1)	42.8 (4.8)
13–16	6.0 (2.3)	8.2 (2.3)	14.0 (4.7)	2.7 (1.2)	12.3 (6.0)	1.02[f] (.71)	6.1 (2.4)	7.9 (2.5)	16.0 (5.3)
17–20	0.6 (.43)	1.0 (.54)	0.4 (.37)	nd (nd)	3.5 (2.7)	nd (nd)	0.7 (.45)	0.8 (.55)	0.4 (.42)

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED supported schools, and comparison schools are not PAQUED supported schools.
NOTE: Percentages are defined as number of students who correctly identified the word divided by the number of students who attempted to identify the word.
[a] Difference between estimate and Orientale estimate is statistically significant at the 0.05 level.
[b] Difference between estimate and Orientale estimate is statistically significant at the 0.01 level.
[c] Difference between estimate and Equateur estimate is statistically significant at the 0.05 level.
[d] Difference between estimate and Equateur estimate is statistically significant at the 0.01 level.
[e] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level
[f] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.

Figure 17. Correctly Identified Vocabulary Words Among Grade 2 Students by Province and Group: Percentages by Number of Words



As displayed in ***Table 14***, the following statistically significant differences were observed.

- Overall Grade 4 students in Equateur were more likely to correctly identify 9–12 vocabulary words than were students in either of the other two provinces (52.5% Equateur, 41.0% Bandundu, 30.5% Orientale, $p<0.05$).
- However, students in Orientale were more likely to correctly identify 13–16 words than were students in either of the other two provinces (42.2% Orientale, 28.9% Bandundu, 19.6% Equateur, $p<0.05$).
- Breaking students into treatment and comparison groups shows that Bandundu treatment students were more likely than their Bandundu comparison peers to correctly identify 13–16 vocabulary words (29.4% versus 17.5%, respectively, $p<0.05$), while Orientale treatment students were more likely than their comparison peers to correctly identify 13–16 words (43.9% versus 28.5%, respectively, $p<0.05$) and 17–20 words (10.8% versus 1.4%, respectively, $p<0.05$).
- Conversely, Orientale comparison students were more likely to correctly identify only 5–8 words than their treatment peers (28.3% versus 14.2%, respectively, $p<0.05$).
- Within the comparison group, Bandundu students were more likely to correctly identify 5–8 words than their Equateur counterparts (33.8% versus 16.5%, respectively, $p<0.01$), as were students in Orientale (28.3% versus Equateur's 16.5%, $p<0.05$).

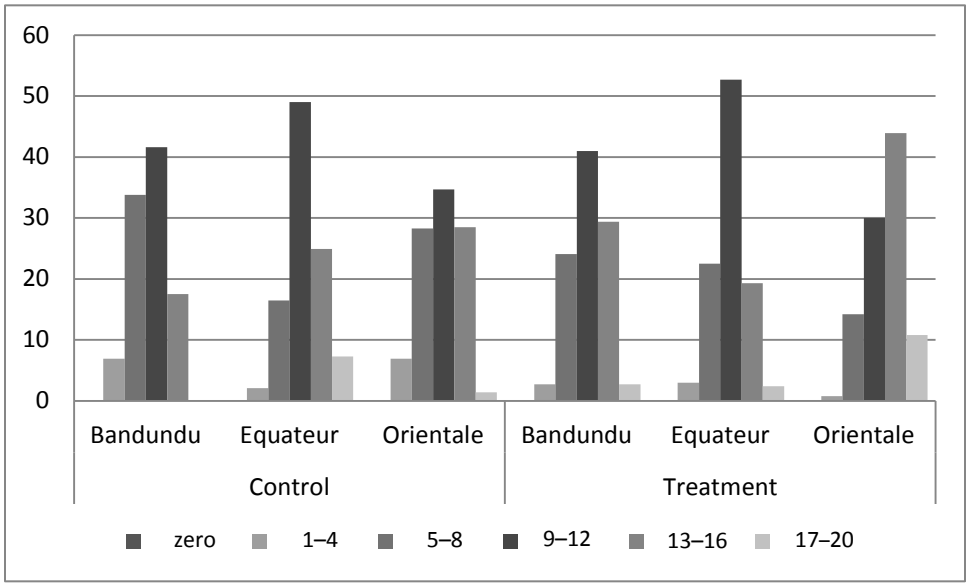
- Within the treatment group, students in Orientale were more likely to identify 13–16 words than students in either of the other two provinces (43.9% Orientale, 29.4% Bandundu, 19.3% Equateur, $p<0.05$), and Orientale students outperformed Equateur students in identifying 17–20 words correctly (10.8% versus 2.4%, respectively, $p<0.05$).

Table 14. Correctly Identified Vocabulary Words Among Grade 4 Students by Province and Group: Percentages (Standard Errors)

Number Correctly Answered	Province			Group by Province					
	Bandundu	Equateur	Orientale	Bandundu	Comparison Equateur	Orientale	Bandundu	Treatment Equateur	Orientale
Across 3 Subtests									
0	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)
1–4	2.9 (1.2)	2.9 (1.2)	1.5 (.77)	6.9 (2.0)	2.1 (1.5)	6.9 (4.5)	2.7 (1.2)	3.0 (1.2)	0.8 (.58)
5–8	24.5 (3.8)	22.2 (3.4)	15.7 (2.9)	33.8 [d](4.0)	16.5[a] (3.5)	28.3[e](4.6)	24.1 (3.9)	22.5 (3.5)	14.2 (3.1)
9–12	41.0[c] (3.3)	52.5[b] (3.2)	30.5 (4.1)	41.6 (4.0)	49.0 (7.0)	34.7 (5.7)	41.0[c] (3.5)	52.7[b] (3.3)	30.0 (4.6)
13–16	28.9[a] (3.7)	19.6[b] (4.0)	42.2 (4.0)	17.5 [e](4.1)	24.9 (4.3)	28.5[e](4.5)	29.4[a] (3.9)	19.3[b] (4.2)	43.9 (4.4)
17–20	2.6 (1.4)	2.6 (1.1)	9.9 (3.6)	nd (nd)	7.3 (4.1)	1.4[e](.86)	2.7 (1.5)	2.4[a] (1.2)	10.8 (4.0)

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED supported schools, and comparison schools are not PAQUED supported schools.
NOTE: Percentages are defined as number of students who correctly identified the word divided by the number of students who attempted to identify the word.
[a] Difference between estimate and Orientale estimate is statistically significant at the 0.05 level.
[b] Difference between estimate and Orientale estimate is statistically significant at the 0.01 level.
[c] Difference between estimate and Equateur estimate is statistically significant at the 0.05 level.
[d] Difference between estimate and Equateur estimate is statistically significant at the 0.01 level.
[e] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[f] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.

Figure 18. Correctly Identified Vocabulary Words Among Grade 4 Students by Province and Group: Percentages



Initial Sound Identification

Comparisons by Grade and Treatment Condition. In the Initial Sound Identification subtask, students listened to individual words, such as ―sak," and were asked to identify the first sound, or phoneme, of that word (in this case, /s/). This subtask comprised 10 items, for a maximum possible score of 10. The first 5 items, sac, jour, date, lac, and car were presented to all students. Students who were able to identify the first sound of at least one of these words were presented the remaining 5 items (balle, tour, par, vol, and fil).

Table 15 and ***Table 16*** present initial sound identification percentages and standard errors for Grades 2 and 4 students, respectively, by comparison and treatment group and by gender, with statistically significant differences bolded. ***Figure 13*** and ***Figure 14*** graphically present the percentage of students in Grade 2 and Grade 4, respectively, who correctly identified the first sound of the words they had attempted.

Table 15. Initial Sound Identification Among Grade 2 Students by Group and Gender: Percentages (Standard Errors)

	Overall		By Gender			
			Comparison		Treatment	
Item	Comparison	Treatment	Male	Female	Male	Female
Sound						
Sac	23.5 (2.4)	23.8 (2.3)	19.9 (3.4)	27.2 (3.3)	23.9 (2.9)	23.7 (3.2)
Jour	13.0 (1.7)	15.8 (2.0)	11.1 (2.5)	15.0 (2.3)	15.2 (2.2)	16.4 (2.8)

Item	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
Date	9.5 (1.7)	11.2 (1.5)	8.21 (2.0)	10.7 (2.2)	12.0 (1.8)	10.3 (2.1)
Lac	8.5 (1.6)	11.6 (1.2)	8.7 (2.0)	8.2 (1.9)	12.0 (1.7)	11.3 (2.0)
Car	15.2 (3.5)	15.5 (1.8)	13.0 (4.1)	17.6 (3.6)	18.0 (2.4)	13.0 (2.3)
Balle	23.8 (5.7)	33.1 (3.5)	28.7 (7.5)	19.9 (6.1)	36.0 (4.9)	30.1 (4.7)
Tour	19.2 (5.0)	19.7 (3.1)	21.8 (6.1)	17.1 (6.1)	19.4 (3.7)	20.0 (4.6)
Par	20.9 (4.8)	29.9 (3.1)	27.0 (6.5)	16.0 (5.3)	29.8 (4.5)	30.0 (5.0)
Vol	25.3 (5.9)	36.1 (3.3)	30.4 (7.3)	21.1[a] (6.6)	34.3 (4.5)	37.9 (5.1)
Fil	22.7 (4.2)	28.6 (3.8)	28.3 (6.9)	18.2 (4.9)	25.6 (4.7)	31.7 (5.7)

nd: No students in domain
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED-supported schools and comparison schools are not PAQUED-supported schools.
NOTE: Percentages are defined as number of students who correctly identified the sound divided by the number of students who attempted to identify the sound.
[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

Table 16. Initial Sound Identification Among Grade 4 Students by Group and Gender: Percentages (Standard Errors)

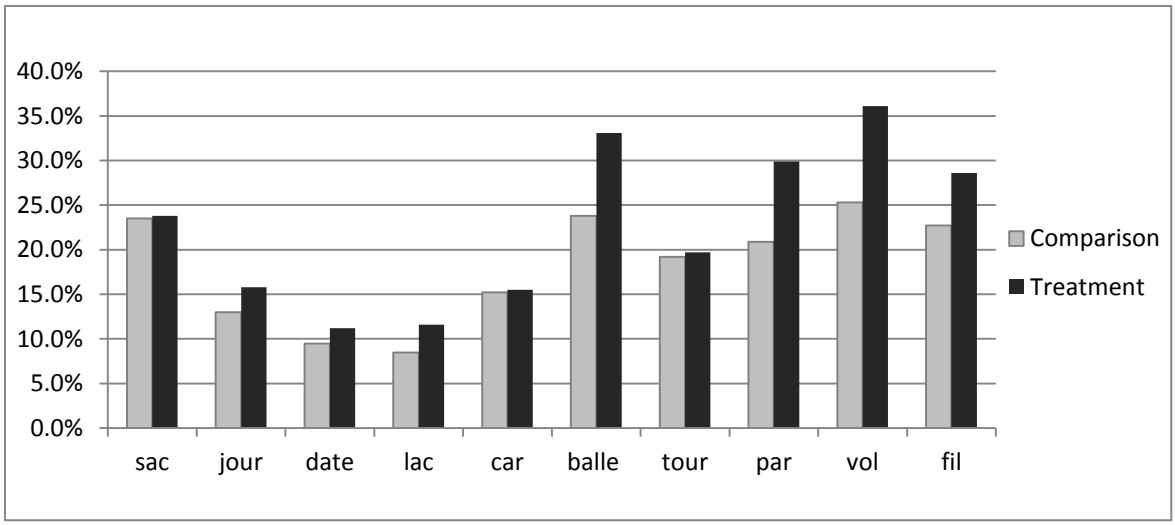
Item	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
Sound						
Sac	33.8 (2.7)	40.7 (2.5)	33.7 (3.9)	34.0 (3.0)	40.5 (3.3)	40.8 (3.0)
Jour	27.0 (2.6)	32.7 (2.6)	28.6 (3.6)	25.5 (3.1)	33.6 (3.4)	31.7 (2.9)
Date	24.0 (2.8)	30.4 (2.5)	25.8 (3.4)	22.2 (3.2)	31.3 (3.4)	29.5 (3.5)
Lac	21.1 (2.7)	27.3 (2.9)	23.4 (3.2)	18.7[a] (3.3)	24.1 (3.8)	30.7 (3.4)
Car	22.4[a] (2.8)	31.0 (2.9)	21.2[a] (3.3)	23.6 (3.5)	32.6 (3.4)	29.4 (3.3)
Balle	38.0 (3.8)	47.5 (3.6)	43.3 (4.8)	32.5[a] (5.0)	44.7 (4.4)	50.9 (5.1)
Tour	33.9 (4.8)	38.4 (3.6)	37.8 (5.3)	30.0 (5.7)	36.8 (4.5)	40.3 (4.2)
Par	37.0 (4.7)	47.4 (3.9)	42.3 (5.4)	31.6[a] (5.4)	46.8 (4.2)	48.2 (4.9)

Item	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
Vol	33.6[a] (3.5)	46.5 (4.0)	40.1[c] (5.0)	26.8[b] (4.4)	42.2 (4.8)	51.7 (4.8)
Fil	33.7 (3.4)	41.4 (4.3)	43.5 d (5.1)	23.7[b] (4.4)	37.9 (5.3)	45.6 (5.2)

nd: No students in domain
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED-supported schools and comparison schools are not PAQUED-supported schools.
NOTE: Percentages are defined as number of students who correctly identified the sound divided by the number of students who attempted to identify the sound.
[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

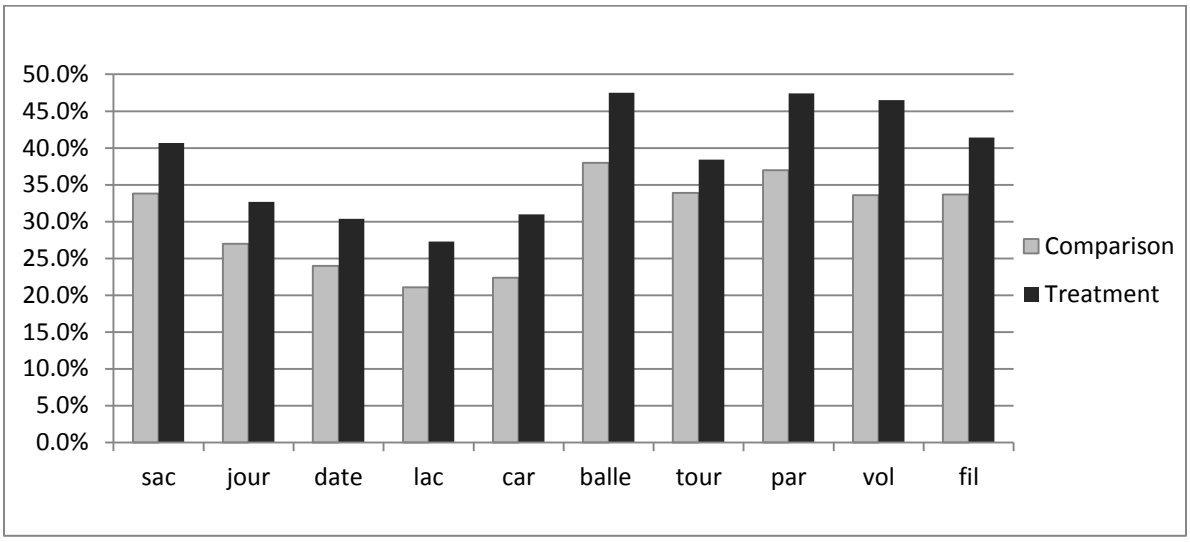
As shown in ***Table 15*** and illustrated by ***Figure 19***, overall there were no statistically significant differences between Grade 2 students from treatment schools and those from comparison schools on any of the stimuli presented as part of this subtask.

Figure 19. Grade 2 Percentages of Initial Sounds Correctly Identified, by Group



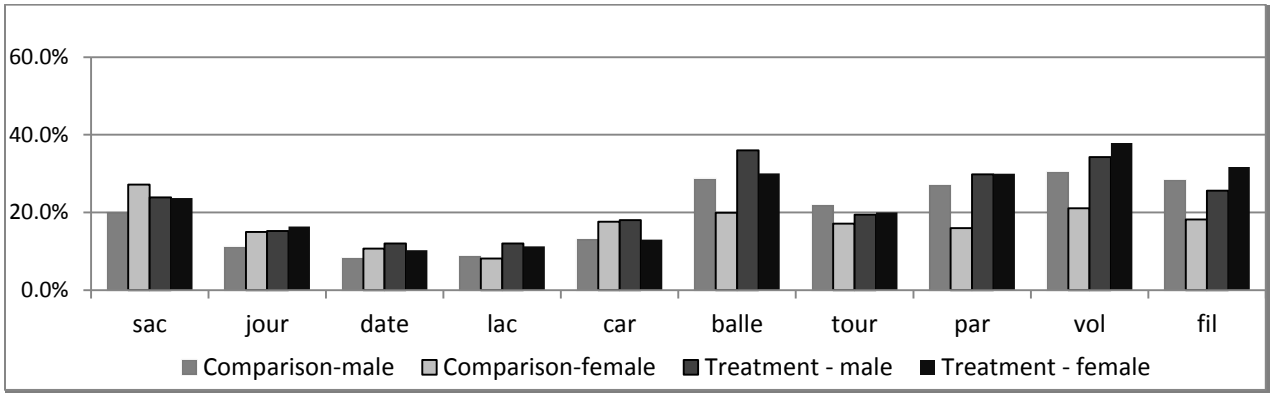
More differences reaching statistical significance at $p<0.05$ are observed among Grade 4 students, however. As shown in ***Table 16*** and illustrated by ***Figure 20***, overall, treatment students outperformed their comparison peers on two stimuli: "Car" (31.0% versus 22.4%, respectively, $p<0.05$) and "Vol" (46.5% versus 33.6%, respectively, $p<0.05$).

Figure 20. Grade 4 Percentages of Initial Sounds Correctly Identified, by Group



Comparisons by Grade, Treatment Condition, and Gender. Comparing across genders in Grade 2, as shown in ***Table 15*** and ***Figure 21***, one statistically significant difference emerged: treatment-group females outperformed their comparison counterparts on the ―Vol" stimulus (37.9% versus 21.1%, respectively, $p<0.05$).

Figure 21. Grade 2 Percentages of Initial Sounds Correctly Identified, by Group and Gender (By Stimulus)

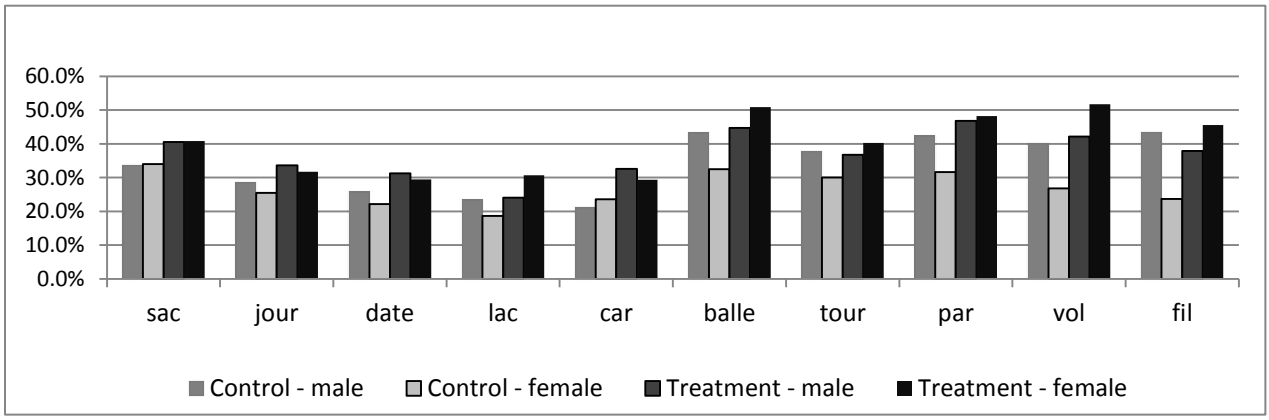


Comparing across genders in Grade 4, as shown in ***Table 16*** and ***Figure 22***, females in the treatment group outperformed females in the comparison group on several stimuli, all at $p<0.05$ or $p<0.01$ (see ***Table 16*** for individual p values).

- ―Lac" (30.7% versus 18.7%)
- ―Balle" (50.9% versus 32.5%)
- ―Par" (48.2% versus 31.6%)
- ―Vol" (51.7% versus 26.8%)
- ―Fil" (45.6% versus 23.7%)

Similarly, treatment males outperformed comparison males on the stimulus –Car" (32.6% versus 21.2%, p<0.05). Within the comparison group itself, males outperformed females on two stimuli: —Vol" (40.1% versus 26.8%, respectively, p<0.05) and –Fil" (43.5% versus 23.7%, respectively, p<0.01). No other statistically significant differences between males and females, in either group, were identified.

Figure 22. Grade 4 Distribution of Initial Sounds Correctly Identified, by Group and Gender (By Stimulus)



Comparison by Grade, Treatment Condition, and Province. ***Tables 17*** and ***18*** present findings of statistical analysis of Grade 2 and Grade 4 initial sound identification student performance in treatment and comparison schools by province, with areas of statistical difference bolded. As displayed in ***Table 17*** and illustrated in ***Figure 23***, Grade 2 students in Equateur outperformed students in the other provinces on a number of stimuli at p<0.05 or p<0.01 (see ***Table 17*** for exact p values):

- –Sac" (Equateur, 31.7%; Bandundu, 18.8%)
- –Lac" (Equateur, 16.0%; Orientale, 7.1%)
- –Balle" (Equateur, 41.0%; Orientale, 18.9%)
- –Tour" (Equateur, 27.9%; Orientale, 12.2%)
- –Par" (Equateur, 37.5%; Orientale, 14.9%)
- –Vol" (Equateur, 36.4%; Orientale, 21.0%)

Bandundu students also outperformed Orientale students on three stimuli at the same levels of statistical significance:

- –Par" (Bandundu, 31.7%; Orientale, 14.9%)
- –Vol" (Bandundu, 44.6%; Orientale, 21.0%)
- –Fil" (Bandundu, 34.4%; Orientale, 16.5%)

Comparing across treatment and comparison groups, only one statistically significant (at p<0.05) difference emerged: Orientale treatment students outperformed their comparison peers on the stimulus –Lac" (7.9% versus 2.2%, respectively).

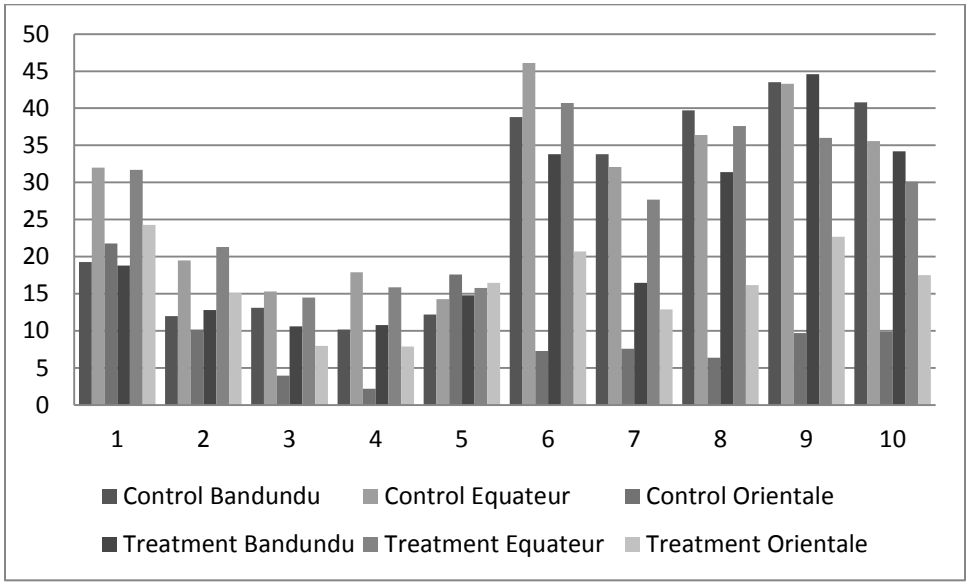
Within the comparison group, students in the Orientale province were found to significantly underperform relative to their peers in at least one of the other two provinces on eight of the ten stimuli; see ***Table 17*** for specifics of each difference. Within the treatment group, however, Orientale students underperformed their peers in other provinces on only five of the ten stimuli. Within Equateur, students in both treatment and comparison groups outperformed Bandundu students on the stimulus ―Sa ̌c ̌ (31.7% versus 18.8% for treatment; 32.0% versus 19.3% for comparison; $p<0.05$).

Table 17. Initial Sound Identification Among Grade 2 Students by Province and Group: Percentages (Standard Errors)

	Province			Group by Province					
				Comparison			Treatment		
Item	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale
Sound									
Sac	18.8[c] (3.0)	31.7 (4.6)	24.0 (3.1)	19.3[c] (4.1)	32.0 (4.7)	21.8 (3.4)	18.8[c] (3.2)	31.7 (4.9)	24.3 (3.5)
Jour	12.7 (2.7)	21.2 (3.9)	14.5 (2.1)	12.0 (3.7)	19.5[a] (3.6)	10.2 (2.1)	12.8 (2.8)	21.3 (4.1)	15.1 (2.4)
Date	10.8 (2.4)	14.6[a] (2.5)	7.50 (1.8)	13.1[a] (3.8)	15.3[b] (3.9)	4.0 (1.6)	10.6 (2.5)	14.5 (2.6)	8.03 (2.0)
Lac	10.8 (1.8)	16.0[b] (2.1)	7.1 (2.0)	10.2[a] (3.1)	17.9[b] (4.2)	2.2 [e](1.2)	10.8 (1.9)	15.9[a] (2.3)	7.9 (2.3)
Car	14.7 (2.5)	15.7 (3.5)	16.6 (2.6)	12.2 (3.4)	14.3 (3.4)	17.6 (6.8)	14.8 (2.7)	15.8 (3.7)	16.5 (2.8)
Balle	34.0 (6.3)	41.0[b] (4.5)	18.9 (5.5)	38.8[b] (10.)	46.1[b] (10.)	7.3 (4.0)	33.8 (6.5)	40.7[a] (4.8)	20.7 (6.2)
Tour	17.1 (5.8)	27.9[a] (5.1)	12.2 (4.4)	33.8[a] (10.)	32.1[a] (9.2)	7.6 (4.5)	16.5 (6.0)	27.7[a] (5.3)	12.9 (5.0)
Par	31.7[a] (5.9)	37.5[b] (4.7)	14.9 (3.3)	39.7[b] (7.3)	36.4[b] (9.6)	6.4 (3.5)	31.4[a] (6.1)	37.6[b] (4.9)	16.2 (3.7)
Vol	44.6[b] (6.0)	36.4[a] (4.8)	21.0 (4.3)	43.5[b] (8.6)	43.3[b] (9.7)	9.7 (5.6)	44.6[b] (6.2)	36.0 (5.0)	22.7 (4.7)
Fil	34.4[a] (7.1)	30.4 (6.1)	16.5 (4.1)	40.8[b] (8.9)	35.6[b] (8.8)	9.9 (3.8)	34.2 (7.3)	30.1 (6.4)	17.5 (4.7)

nd: No students in domain
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED supported schools, and comparison schools are not PAQUED supported schools.
NOTE: Percentages are defined as number of students who correctly identified the sound divided by the number of students who attempted to identify the sound.
[a] Difference between estimate and Orientale estimate is statistically significant at the 0.05 level.
[b] Difference between estimate and Orientale estimate is statistically significant at the 0.01 level.
[c] Difference between estimate and Equateur estimate is statistically significant at the 0.05 level.
[d] Difference between estimate and Equateur estimate is statistically significant at the 0.01 level.
[e] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[f] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.

Figure 23. Initial Sound Identification Among Grade 2 Students by Province and Group: Percentages



As displayed in ***Table 18*** and illustrated in ***Figure 24***, overall Grade 4 students in Orientale statistically significantly underperformed relative to their peers in the other provinces on eight of the ten stimuli (see ***Table 18*** for specifics of these differences). In addition, Bandundu students underperformed relative to Equateur students on the stimulus ―four" (37.0% versus 57.2%, respectively, $p<0.05$).

Comparing across treatment and comparison groups, only one statistically significant difference emerged: Bandundu treatment students outperformed their comparison peers on the stimulus ―four" (31.2% versus 20.5%, respectively, $p<0.05$).

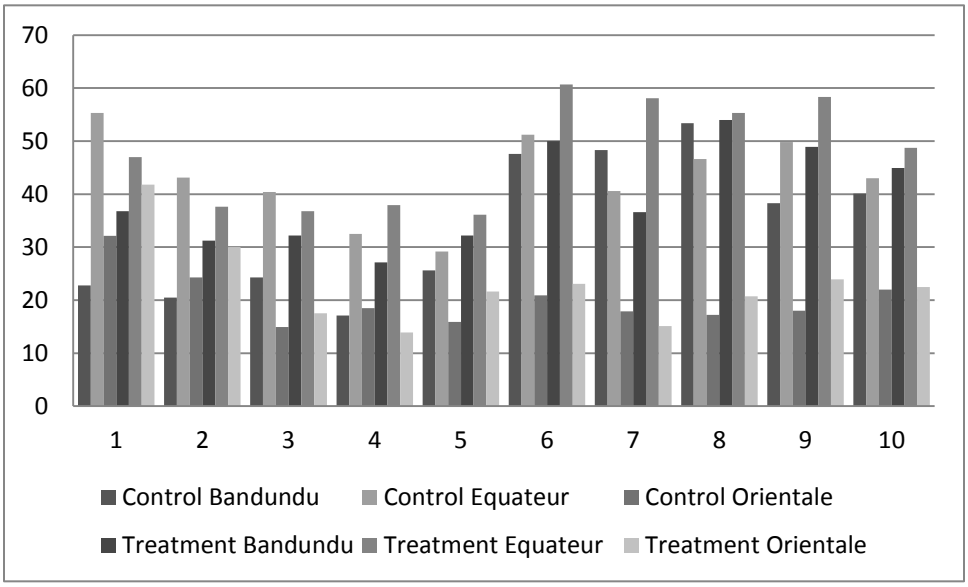
Within the comparison group, students in the Orientale province were found to significantly (at either $p<0.05$ or $p<0.01$) underperform relative to Equateur peers on all ten stimuli, and to underperform relative to Bandundu peers on six of the ten stimuli (see ***Table 18*** for specifics of these differences). Within the treatment group, Orientale students underperformed relative to peers in at least one of the other two provinces on eight of the ten stimuli (again, see ***Table 18*** for specifics of these differences). In addition, Bandundu students underperformed Equateur students on the stimulus ―four" (36.6% versus 58.1%, respectively, $p<0.01$).

Table 18. Initial Sound Identification Among Grade 4 Students by Province and Group: Percentages (Standard Errors)

	Province			Group by Province					
				Comparison			Treatment		
Item	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale
Sound									
Sac	36.2 (3.1)	47.4 (5.1)	40.7 (4.2)	22.8a df(3.1)	55.3b (7.0)	32.1 (3.1)	36.8 (3.3)	47.0 (5.3)	41.8 (4.7)
Jour	30.7 (3.7)	37.9 (5.4)	29.4 (2.9)	20.5 de(2.8)	43.1a (6.8)	24.3 (4.0)	31.2 (3.9)	37.6 (5.7)	30.0 (3.2)
Date	31.9b (3.5)	37.0b (5.2)	17.2 (2.7)	24.3c (4.8)	40.4b (6.1)	14.9 (3.1)	32.2b (3.6)	36.8b (5.5)	17.5 (3.0)
Lac	26.6a (4.5)	37.7b (4.5)	14.4 (2.4)	17.1 d(3.6)	32.5a (4.3)	18.5 (5.1)	27.1a (4.7)	37.9b (4.7)	13.9 (2.6)
Car	31.9 (4.5)	35.7a (4.9)	20.9 (3.3)	25.6 (4.9)	29.2a (3.7)	15.9 (4.3)	32.2 (4.7)	36.1a (5.1)	21.6 (3.7)
Balle	50.0b (4.3)	60.2b (7.5)	22.9 (4.9)	47.6b (6.3)	51.2b (5.3)	20.9 (6.6)	50.0b (4.4)	60.7b (7.9)	23.1 (5.4)
Tour	37.0bc(4.9)	57.2b (5.9)	15.4 (3.7)	48.3b (8.0)	40.6a (6.2)	17.9 (7.2)	36.6b d(5.1)	58.1b (6.2)	15.1 (4.0)
Par	54.0b (5.7)	54.8b (6.1)	20.4 (3.6)	53.4b (9.3)	46.6b (6.0)	17.2 (6.1)	54.0b (5.9)	55.3b (6.4)	20.7 (4.0)
Vol	48.6b (5.6)	57.9b (7.7)	23.3 (4.4)	38.3a (6.2)	49.8b (6.2)	18.0 (5.5)	48.9b (5.8)	58.3b (8.1)	23.9 (4.8)
Fil	44.8b (6.0)	48.4b (8.6)	22.4 (3.7)	40.1a (6.4)	43.0b (5.9)	22.0 (5.0)	44.9b (6.2)	48.7a (9.1)	22.5 (4.1)

nd: No students in domain
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED supported schools, and comparison schools are not PAQUED supported schools.
NOTE: Percentages are defined as number of students who correctly identified the sound divided by the number of students who attempted to identify the sound.
[a] Difference between estimate and Orientale estimate is statistically significant at the 0.05 level.
[b] Difference between estimate and Orientale estimate is statistically significant at the 0.01 level.
[c] Difference between estimate and Equateur estimate is statistically significant at the 0.05 level.
[d] Difference between estimate and Equateur estimate is statistically significant at the 0.01 level.
[e] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[f] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.

Figure 24. Initial Sound Identification Among Grade 4 Students by Province and Group: Percentages



Letter Sound Knowledge

In the Letter Sound Knowledge subtask, students saw a 100-item chart containing the letters of the alphabet in random order and were required to produce the sounds for as many letters as possible within one minute. This task was discontinued prior to the end of the minute for students who were unable to produce any of the sounds for the 10 letters in the first row. Scores reported for this subtask include percentages of students able to correctly generate numbers of letters and the number of letter sounds that students could correctly generate within one minute (correct letter sounds per minute, or clspm).

Comparisons by Grade and Treatment Condition. Tables 19 and 20 present letter sound knowledge percentages and standard errors for Grades 2 and 4 students, respectively, by comparison and treatment group and by gender, with statistically significant differences bolded. Figures 25 and 26 graphically present the percentages of students in Grade 2 and Grade 4 who correctly generated letter sounds, by number of sounds generated.

Table 19. Letter Sound Knowledge Among Grade 2 Students by Group and Gender: Percentages (Standard Errors)

Number Correctly Answered	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
0	23.3 (2.6)	18.6 (2.8)	**15.5 [d](2.9)**	**31.3 (3.9)**	16.6 (2.6)	20.6 (3.8)
1–20	70.4 (2.8)	73.7 (2.8)	**76.7 [d](3.2)**	**64.0 (4.1)**	74.7 (2.6)	72.7 (3.9)
21–40	5.63 (1.4)	6.15 (1.1)	6.65 (2.3)	4.58 (1.4)	6.96 (1.6)	5.36 (1.4)
41–60	.253 (.23)	1.21 (.45)	.500 (.46)	nd (nd)	1.64 (.82)	.803 (.49)
61–80	nd (nd)	.218 (.21)	nd (nd)	nd (nd)	nd (nd)	.434 (.42)
81–100	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED-supported schools and comparison schools are not PAQUED-supported schools.
NOTE: Percentages are defined as number of students who correctly identified the letter sound divided by the number of students who attempted to identify the letter sound.
[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

Table 20. Letter Sound Knowledge Among Grade 4 Students by Group and Gender: Percentages (Standard Errors), 2012 Midterm Assessment

Number Correctly Answered	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
0	6.39 (1.1)	5.20 (1.1)	6.14 (1.8)	6.63 (1.7)	5.86 (2.0)	4.50 (1.0)
1–20	46.9 (3.0)	49.5 (2.5)	**39.8 [d](3.5)**	**54.0 (4.0)**	**42.4 [d](2.8)**	**57.1 (4.1)**
21–40	34.6 (2.7)	33.0 (2.2)	36.1 (3.3)	33.0 (4.2)	**38.2[c] (3.4)**	**27.4 (2.8)**
41–60	10.6 (1.7)	10.8 (2.0)	**15.1[c] (3.0)**	**6.21 (2.1)**	11.8 (2.3)	9.76 (2.6)
61–80	1.37 (.80)	1.23 (.50)	2.73 (1.6)	nd (nd)	1.55 (.83)	.901 (.51)
81–100	nd (nd)	.065 (.06)	nd (nd)	nd (nd)	nd (nd)	.135 (.13)

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED-supported schools and comparison schools are not PAQUED-supported schools.
NOTE: Percentages are defined as number of students who correctly identified the letter sound divided by the number of students who attempted to identify the letter sound.
[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

As displayed in ***Table 19*** and graphically illustrated in ***Figure 25***, a substantial percentage of Grade 2 students—23.3% of students in comparison schools and 18.6% of students in treatment schools—were unable to produce any of the letter sounds (difference between groups not statistically significant).

As might be anticipated, and as displayed in ***Table 20*** and ***Figure 26***, among Grade 4 students, only 6.4% students in comparison schools and 5.2% in treatment schools obtained zero scores (as with Grade 2 students, this difference between groups was not statistically significant). Overall, however, students in Grade 4 still demonstrated limited mastery of the letter sounds, as no students in comparison schools and only 0.7% of students in treatment schools were able to correctly produce between 81 and 100 letter sounds (difference not statistically significant).

Figure 25. Grade 2 Percentages of Students Correctly Identifying Letter Sounds (By Number of Letter Sounds Identified)

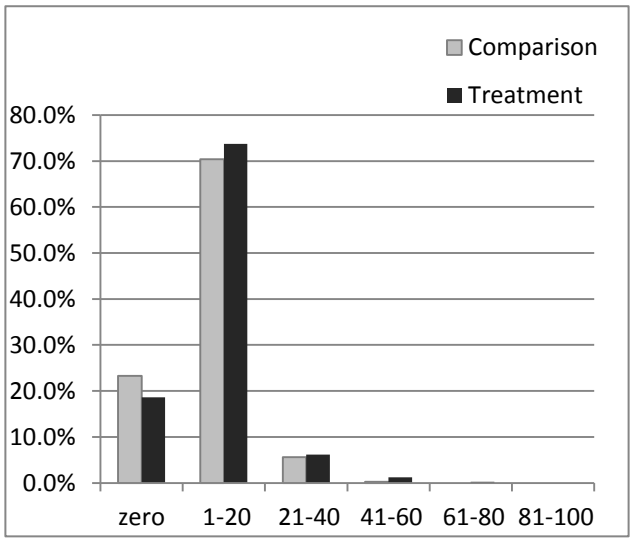
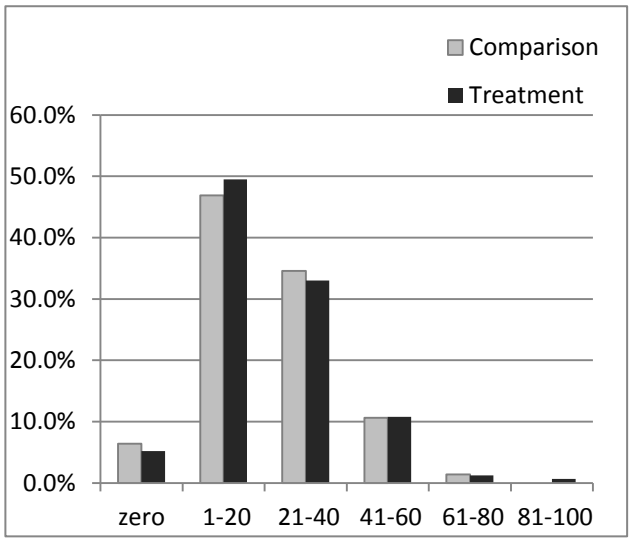
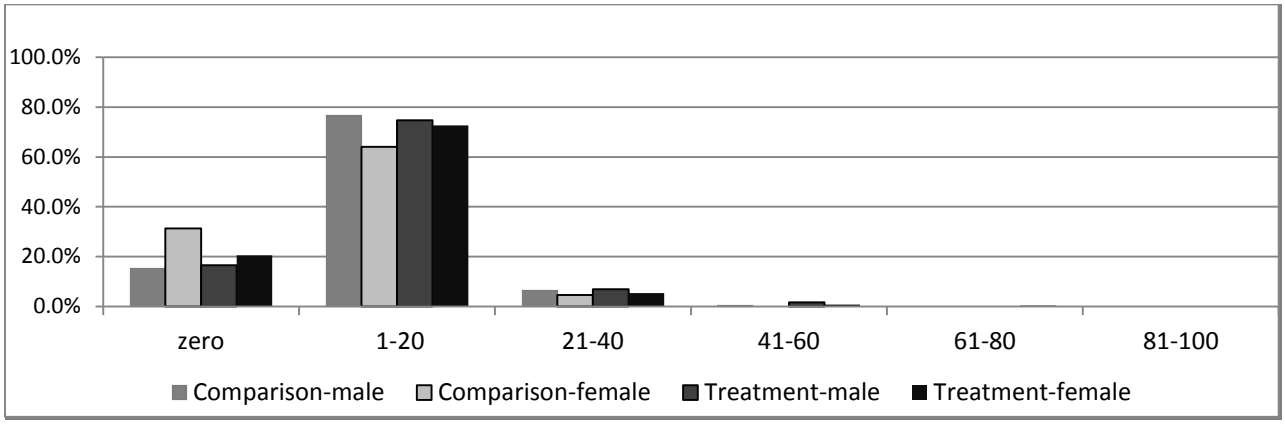


Figure 26. Grade 4 Percentages of Students Correctly Identifying Letter Sounds (By Number of Letter Sounds Identified)



Comparisons by Grade, Treatment Condition, and Gender. Figures 23 and 24 graphically display percentages of letter sounds correctly produced by group and gender for Grades 2 and 4, respectively. As indicated by ***Table 19*** and ***Figure 27***, there were no statistically significant differences across treatment and comparison groups for either Grade 2 males or females; the only statistically significant differences were observed within the comparison group, itself, in which females were more likely than males to score zero on this subtask (31.3% versus 15.5%, respectively, $p<0.01$), and males were more likely than females to correctly produce 1–20 letter sounds (76.7% versus 64.0%, respectively, $p<0.01$).

Figure 27. Grade 2 Percentages of Letter Sounds Correctly Produced by Gender (By Number of Letter Sounds Produced Correctly)



Similarly, in Grade 4—as indicated in ***Table 20*** and illustrated by ***Figure 28***—no statistically significant differences emerged across treatment and comparison groups for either males or females. Within both the treatment and comparison groups, however, males were more likely than females to correctly produce at least 21 letter sounds (treatment group: 21–40 sounds, 38.2% versus 27.4%, respectively, $p<0.05$; comparison group: 41–60 sounds, 15.5% versus 6.21%, respectively, $p<0.05$). Females, conversely, in both groups were more likely than males to correctly produce only 1–20 sounds (treatment: 57.2% versus 42.4%, respectively, $p<0.01$; comparison: 54.0% versus 39.8%, respectively, $p<0.01$).

Figure 28. Grade 4 Percentages of Letter Sounds Correctly Produced by Gender (By Number of Letter Sounds Produced Correctly)

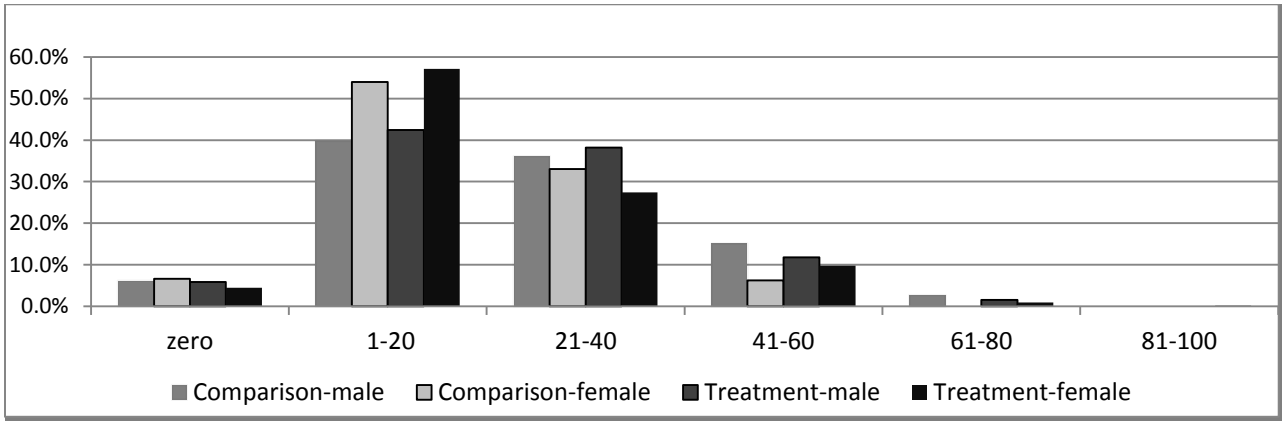


Table 21 presents—and ***Figure 29*** illustrates—the number of letter sounds correctly produced per minute (clspm) by grade, treatment group, and gender. As displayed, when comparing across gender, Grade 2 females in treatment schools outperformed their

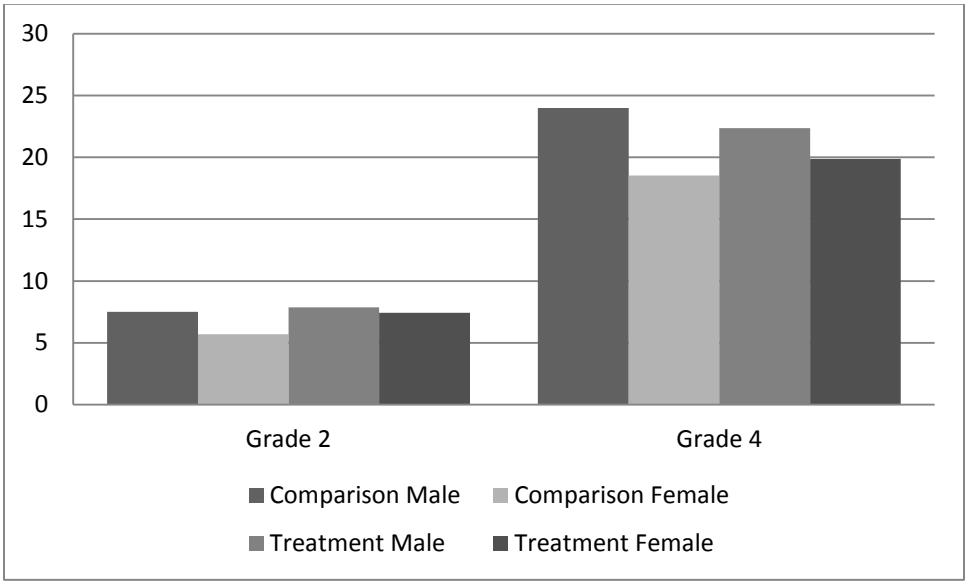
comparison counterparts (7.4 versus 5.7 clspm, respectively, p<0.05) although there was no corresponding difference in Grade 4. Within the comparison group, males outperformed females (Grade 2: 7.5 versus 5.7 clspm, respectively, p<0.05; Grade 4: 24.0 versus 18.5 clspm, respectively, p<0.01).

Table 21. CLSPM Among Grades 2 and 4 Students by Grade, Group, and Gender

	Overall		By Gender			
			Comparison		Treatment	
Grade	Comparison	Treatment	Male	Female	Male	Female
2	6.6	7.7	**7.5**[c]	**5.7**[a]	7.9	**7.4**
4	21.3	21.2	**24.0**[d]	**18.5**	22.4	19.9

[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

Figure 29. CLSPM Among Grades 2 and 4 Students by Grade, Group, and Gender



Comparisons by Grade, Treatment Condition, and Province. Tables 22 and 23 present findings of statistical analysis of Grade 2 and Grade 4 letter sound knowledge in treatment and comparison schools by province, with areas of statistical difference bolded.

As displayed in ***Table 22*** and ***Figure 30***, looking across provinces, Grade 2 Orientale students were more likely than students in either of the other provinces to correctly produce 1–20 letter sounds (Orientale, 83.7%; Bandundu, 74.1%; Equateur, 64.0%; p<0.01) while being less likely than Equateur students to score zero on this subtask (Orientale, 10.9%; Equateur, 29.5%; p<0.01). While a relatively small percentage of students overall were able to produce more than 40 sounds in the minute allotted, a

statistically significant difference did emerge between students in Bandundu (2.1%) and those in Equator (nd) at 41–60 sounds identified ($p<0.05$).

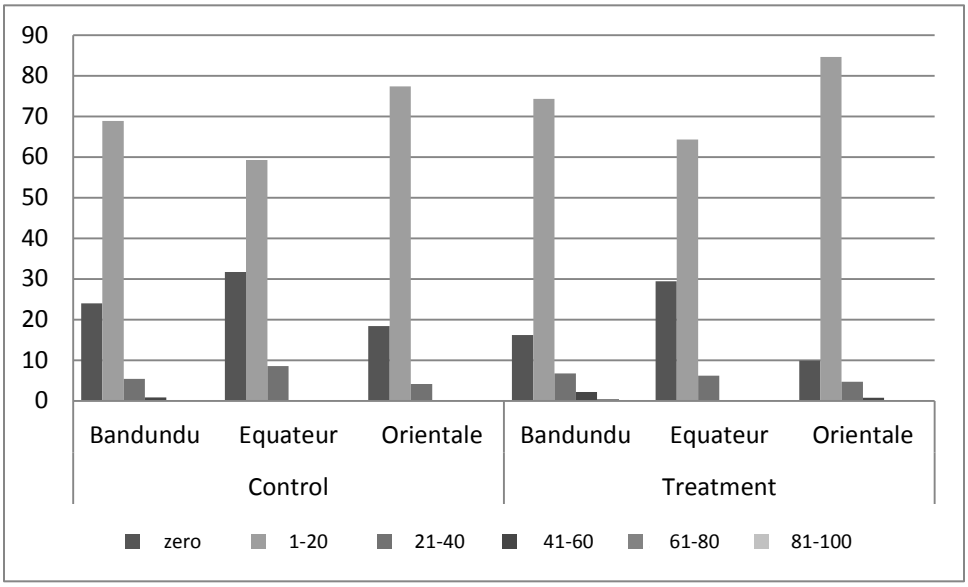
Comparing across treatment and comparison groups, no statistically significant differences emerged within any of the three provinces. Within Grade 2 treatment and comparison groups, however, students in Orientale again were more likely to outperform either one (Equateur province, within the comparison group) or both (within the treatment group) of the other provinces (see ***Table 22*** for specific differences). In both conditions, Orientale students were also less likely than Equateur students to score zero on this subtask (see ***Table 22*** for specific differences), and within the treatment group students in Bandundu outperformed students in Equateur at producing 41–60 sounds per minute (2.2% versus nd, respectively, $p<0.05$).

Table 22. Letter Sound Knowledge Among Grade 2 Students by Province and Group: Percentages (Standard Errors)

Number of Sounds Identified	Province			Group by Province					
				Comparison			Treatment		
	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale
0	16.5 (4.4)	**29.5[b] (5.6)**	**10.9 (1.9)**	24.0 (4.1)	**31.7[a] (4.9)**	**18.4 (4.2)**	16.2 (4.6)	**29.4[b] (5.9)**	**9.9 (2.1)**
1–20	**74.1[a] (4.2)**	**64.0[b] (5.7)**	**83.7 (2.1)**	68.9 (4.5)	**59.3[b] (5.1)**	**77.4 (4.4)**	**74.3[a] (4.4)**	**64.3[b] (6.1)**	**84.6 (2.3)**
21–40	6.8 (1.7)	6.4 (1.8)	4.6 (1.2)	5.4 (1.6)	8.6 (2.8)	4.2 (2.4)	6.8 (1.8)	6.2 (1.9)	4.7 (1.4)
41–60	**2.1[c] (.89)**	**nd (nd)**	0.7 (.47)	0.9 (.79)	nd (nd)	nd (nd)	**2.2[c] (.93)**	**nd (nd)**	0.8 (.54)
61–80	0.4 (.41)	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)	0.5 (.43)	nd (nd)	nd (nd)
81–100	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED supported schools, and comparison schools are not PAQUED supported schools.
NOTE: Percentages are defined as number of students who correctly identified the letter sound divided by the number of students who attempted to identify the letter sound.
[a] Difference between estimate and Orientale estimate is statistically significant at the 0.05 level.
[b] Difference between estimate and Orientale estimate is statistically significant at the 0.01 level.
[c] Difference between estimate and Equateur estimate is statistically significant at the 0.05 level.
[d] Difference between estimate and Equateur estimate is statistically significant at the 0.01 level.
[e] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
f Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.

Figure 30. Letter Sound Knowledge Among Grade 2 Students by Province and Group: Percentages



At Grade 4, overall, Equateur students were more likely than Orientale students to score zero on this subtask (8.3% versus 2.8%, $p<0.05$); no other statistically significant differences at the overall province level were observed. Comparing across treatment condition, nor did any statistically significant differences emerge within any of the three provinces when comparing across treatment and comparison groups. Within the comparison group, Orientale students were less likely than students in either other province to be able to produce 1–20 letter sounds (Bandundu, 59.1%; Equateur, 35.3%; Orientale, 42.1%; $p<0.01$). However, they were more likely than Bandundu students to produce 41–60 sounds (12.3% versus 5.5%, respectively, $p<0.05$), as were students in Equateur (16.3% versus Bandundu's 5.5%, $p<0.05$). Within the treatment group, only one statistically significant difference emerged: Equateur students were more likely than Orientale students to score zero on this subtask (8.3% versus 2.5%, respectively, $p<0.05$).

Table 23. Letter Sound Knowledge Among Grade 4 Students by Province and Group: Percentages (Standard Errors)

Number of Sounds Identified	Province			Group by Province					
				Comparison			Treatment		
	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale
0	4.7 (1.7)	**8.3[a] (2.3)**	**2.8 (.98)**	6.4 (1.7)	8.3 (2.4)	5.4 (1.8)	4.6 (1.8)	**8.3[a] (2.4)**	**2.5 (1.0)**
1–20	52.4 (3.6)	47.6 (4.1)	44.5 (4.7)	**59.1[b d] (4.1)**	**35.3 (7.2)**	**42.1 (4.7)**	52.1 (3.8)	48.2 (4.3)	44.8 (5.3)
21–40	31.8 (3.3)	33.9 (3.9)	35.0 (2.8)	28.5 (4.4)	37.6 (4.3)	38.4 (5.3)	31.9 (3.4)	33.7 (4.1)	34.6 (3.0)
41–60	9.8 (2.9)	8.5 (2.6)	16.0 (4.4)	**5.5[a c] (1.6)**	**16.3 (4.3)**	**12.3 (2.8)**	10.0 (3.0)	8.1 (2.7)	16.5 (4.9)

Number of Sounds Identified	Province			Comparison			Group by Province			Treatment
	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale	
61–80	1.3 (.72)	1.2 (.79)	1.2 (.98)	0.4 (.36)	2.4 (2.1)	1.7 (1.5)	1.3 (.75)	1.2 (.82)	1.2 (1.0)	
81–100	nd (nd)	nd (nd)	0.3 (.27)	nd (nd)	nd (nd)	nd (nd)	nd (nd)	nd (nd)	0.3 (.30)	

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED supported schools, and comparison schools are not PAQUED supported schools.
NOTE: Percentages are defined as number of students who correctly identified the letter sound divided by the number of students who attempted to identify the letter sound.
[a] Difference between estimate and Orientale estimate is statistically significant at the 0.05 level.
[b] Difference between estimate and Orientale estimate is statistically significant at the 0.01 level.
[c] Difference between estimate and Equateur estimate is statistically significant at the 0.05 level.
[d] Difference between estimate and Equateur estimate is statistically significant at the 0.01 level.
[e] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
f Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.

Figure 31. Letter Sound Knowledge Among Grade 4 Students by Province and Group: Percentages

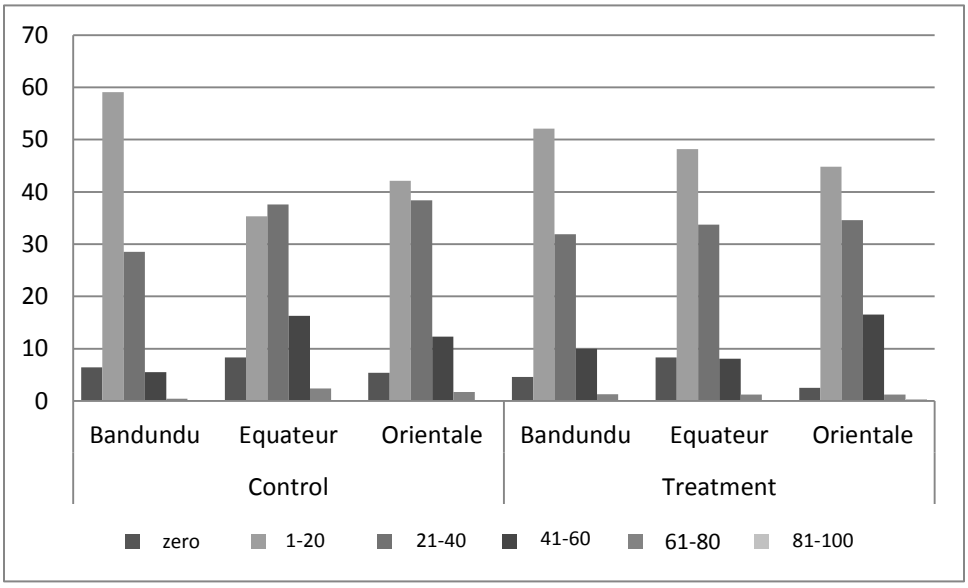


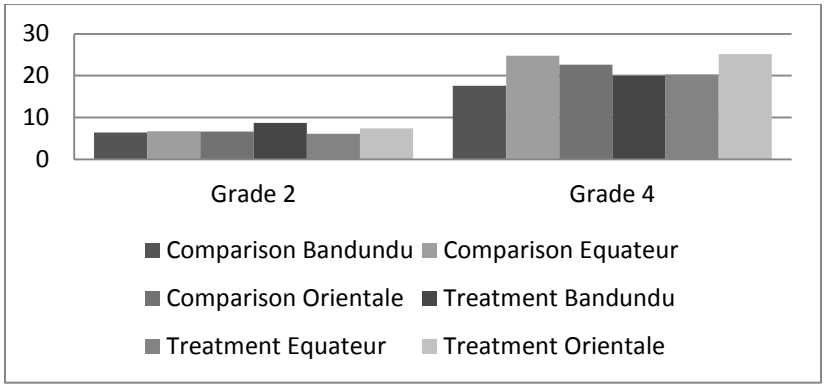
Table 24 presents—and ***Figure 32*** illustrates—the number of letter sounds correctly produced per minute (clspm) by grade, province, and treatment group. For both Grades 2 and 4, there is no statistically significant difference across treatment or comparison groups for any of the three provinces.

Table 24. CLSPM Among Grades 2 and 4 Students by Province and Group

	Group by Province					
	Comparison			Treatment		
Grade	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale
2	6.5	6.7	6.6	8.7	6.1	7.4
4	17.6	24.8	22.7	20.0	20.3	25.2

[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

Figure 32. CLSPM Among Grades 2 and 4 Students by Grade, Group, and Province



Familiar Word Reading

The Familiar Word Reading subtask was administered to only students in Grade 4. In this task, students were shown a chart of 50 familiar words (e.g., tu, ami) and were required to read as many words as they could within 1 minute. This subtask was discontinued before the end of 1 minute for students who were unable to read any of the first five words. Scores reported for this subtask include percentages of students able to correctly identify words and the number of words that students could correctly identify within one minute (correct words per minute, or cwpm).

Comparisons by Treatment Condition. ***Table 25*** presents familiar word reading percentages and standard errors for Grade 4 students by comparison and treatment group and by gender, with statistically significant differences bolded. ***Figure 33*** graphically presents the percentages of students who correctly read words, by number of words read.

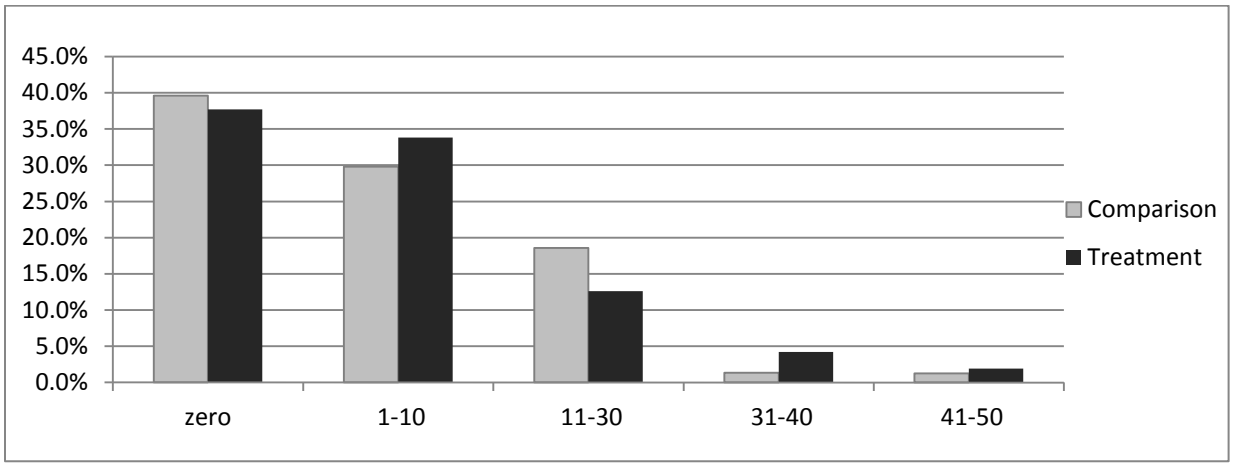
Table 25. Reading Familiar Words Among Grade 4 Students by Group and Gender: Percentages (Standard Errors)

Number of Words Read	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
0	39.6 (2.9)	37.8 (2.8)	**34.2 [d] (3.5)**	**45.1 (3.4)**	34.5 (3.9)	41.3 (3.8)
1–10	29.8 (2.4)	33.8 (1.9)	26.4 (3.3)	33.1 (3.0)	34.2 (3.2)	33.4 (2.8)
11–30	**18.6 [a] (2.1)**	**12.6 (1.6)**	**22.8 [ac] (2.6)**	**14.4 (3.1)**	**13.8 (2.3)**	11.3 (2.1)
31–40	**1.4 [a] (.63)**	**4.2 (1.0)**	1.7 (1.1)	**1.0 [a] (.62)**	3.3 (.96)	**5.3 (1.7)**
41–50	1.3 (.68)	1.9 (.60)	2.2 (1.3)	0.3 (.28)	2.2 (.91)	1.6 (.65)

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED-supported schools and comparison schools are not PAQUED-supported schools.
NOTE: Percentages are defined as number of students who correctly identified the familiar word divided by the number of students who attempted to identify the familiar word.
[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

As displayed in ***Table 25*** and illustrated in ***Figure 33***, Grade 4 comparison students were more likely than treatment students to accurately read 11–30 words (18.6% versus 12.6%, respectively, $p<0.05$), while treatment students were more likely to read 31–40 words (4.2% versus 1.4%, respectively, $p<0.05$). Comparing across treatment conditions by gender, comparison males were more likely than treatment males to correctly read 11–30 words (22.8% versus 13.8%, respectively, $p<0.05$), while within the comparison group males were also more likely to correctly read 11–30 words than their female comparison peers (22.8% versus 14.4%, respectively, $p<0.05$). Comparison females were less likely than treatment females to correctly read 31–40 words (1.0% versus 5.25%, respectively, $p<0.05$) and were more likely than their comparison male peers to score zero on this subtask (45.1% versus 34.2%, respectively, $p<0.01$).

Figure 33. Grade 4 Percentages of Familiar Words Read Correctly by Group



Comparison by Treatment Condition and Gender. As shown in ***Table 25*** and illustrated in ***Figure 34***, when comparing across gender, comparison males were more likely than treatment males to correctly read 11–30 words (22.8% versus 14.9%, respectively, $p<0.05$); no other statistically significant differences across groups when breaking results out by gender emerged. Within the comparison group, however, males were more likely to read 11–30 words than were females (22.8% versus 14.9%, respectively, $p<0.05$), while females were more likely than males to score zero on this subtask (45.1% versus 34.2%, respectively, $p<0.01$).

Figure 34. Grade 4 Percentages of Familiar Words Read Correctly, by Group and Gender

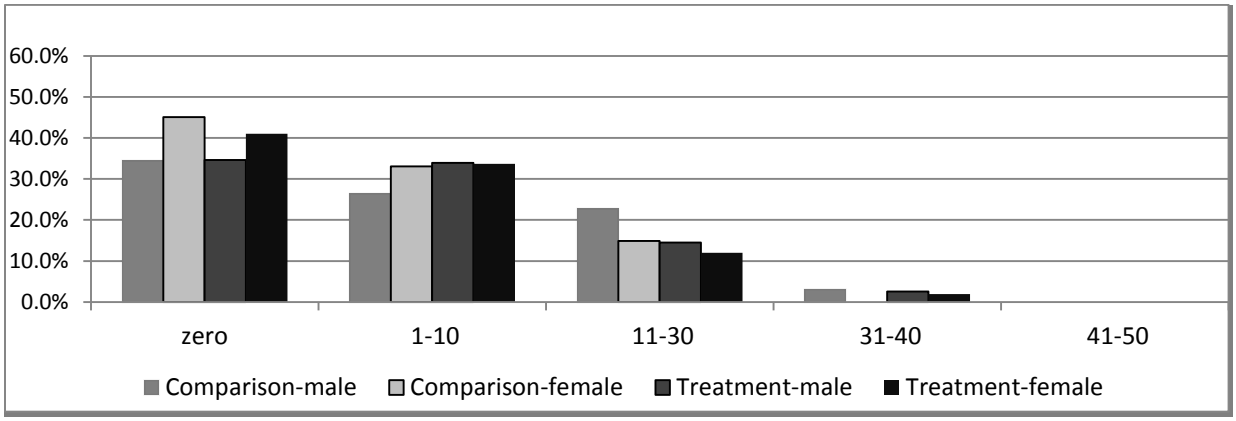


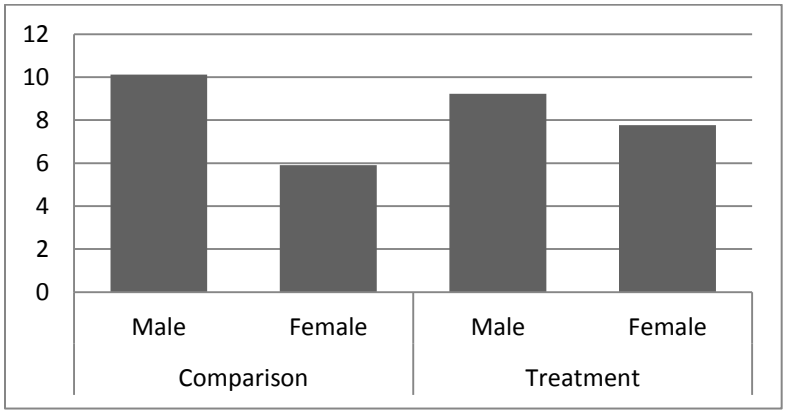
Table 26 presents—and ***Figure 35*** illustrates—the number of familiar words correctly read per minute (cwpm) by grade, treatment group, and gender. No statistically significant difference emerged between Grade 4 treatment and comparison groups, overall, or between males and females within the treatment group. However, males from comparison schools did read more cwpm than did their female peers (10.1 versus 5.9, $p<0.01$).

Table 26. CWPM Among Grade 4 Students by Group and Gender

	Overall		By Gender			
			Comparison		Treatment	
Grade	Comparison	Treatment	Male	Female	Male	Female
4	8.0	8.5	10.1[d]	5.9	9.2	7.8

[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

Figure 35. CWPM Among Grade 4 Students by Group and Gender



Comparisons by Treatment Condition and Province. ***Table 27*** and ***Figure 36*** present findings of statistical analysis of Grade 4 familiar word reading in treatment and comparison schools by province, with areas of statistical difference bolded. Overall, students in Orientale were significantly more likely than those in Bandundu to read 41–50 words (3.5% versus 0.7%, respectively, $p<0.05$) and less likely to score zero on this subtask (28.3% versus 42.3%, respectively, $p<0.05$).

Comparing across treatment conditions, Orientale treatment students were more likely than their comparison peers to accurately read 31–40 words (9.2% versus 1.0%, respectively, $p<0.05$), while Bandundu comparison students were more likely than treatment students to score zero on this subtask (55.8% versus 41.7%, respectively, $p<0.05$). Within the comparison group, Bandundu students were more likely than students in the other provinces to score zero on this subtask (Bandundu, 55.7%; Equateur, 33.6%; Orientale, 28.3%; $p<0.05$), while students in Orientale were more likely than those in Bandundu to read 11–30 words (23.3% versus 11.9%, respectively, $p<0.05$). No significant differences were observed within the treatment group.

Table 27. Reading Familiar Words Among Grade 4 Students by Province and Group: Percentages (Standard Errors)

Number of Words Read	Province			Group by Province					
				Comparison			Treatment		
	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale
0	**42.3[a] (4.0)**	37.5 (5.0)	**28.3 (4.8)**	**55.8[bde](4.1)**	**33.6 (5.9)**	**28.3 (5.0)**	**41.7 (4.1)**	37.7 (5.3)	28.3 (5.4)
1–10	36.1 (2.9)	33.8 (2.9)	27.6 (3.0)	29.5 (4.0)	30.7 (4.6)	29.5 (4.0)	36.4 (3.0)	33.9 (3.1)	27.4 (3.3)
11–30	11.0 (2.3)	14.1 (2.8)	16.1 (2.6)	**11.9[a] (2.2)**	21.0 (4.7)	**23.3 (3.7)**	10.9 (2.4)	13.7 (2.9)	15.2 (2.9)
31–40	2.9 (1.0)	2.8 (1.4)	8.3 (3.3)	1.2 (1.1)	2.2 (1.2)	**1.0[e] (.88)**	3.0 (1.1)	2.8 (1.5)	**9.2 (3.7)**
41–50	**0.7[a] (.62)**	2.9 (1.3)	**3.5 (1.2)**	0.4 (.38)	nd (nd)	2.7 (1.5)	0.7 (.64)	3.1 (1.4)	3.6 (1.4)

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED supported schools, and comparison schools are not PAQUED supported schools.
NOTE: Percentages are defined as number of students who correctly identified the familiar word divided by the number of students who attempted to identify the familiar word.
[a] Difference between estimate and Orientale estimate is statistically significant at the 0.05 level.
[b] Difference between estimate and Orientale estimate is statistically significant at the 0.01 level.
[c] Difference between estimate and Equateur estimate is statistically significant at the 0.05 level.
[d] Difference between estimate and Equateur estimate is statistically significant at the 0.01 level.
[e] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[f] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.

Figure 36. Reading Familiar Words Among Grade 4 Students by Province and Group: Percentages

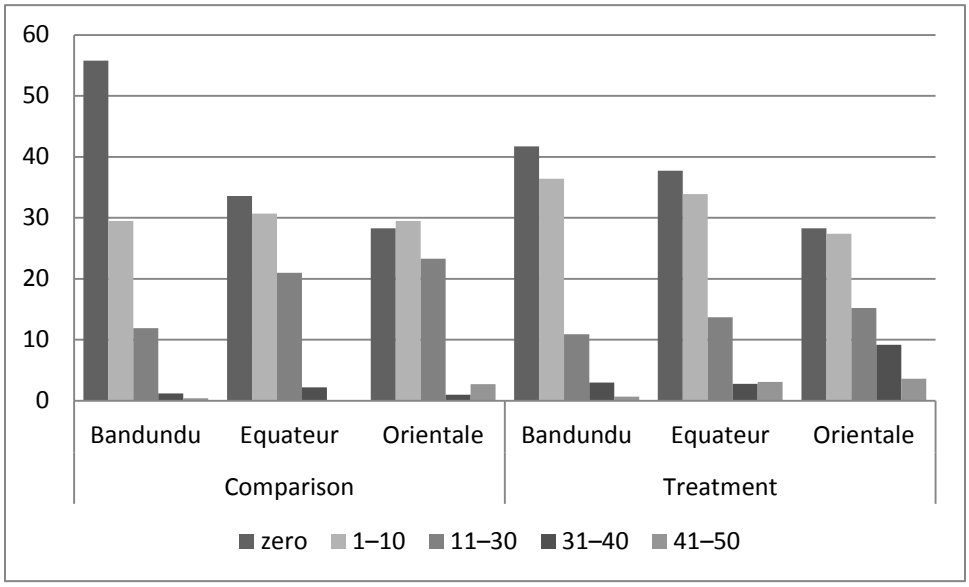


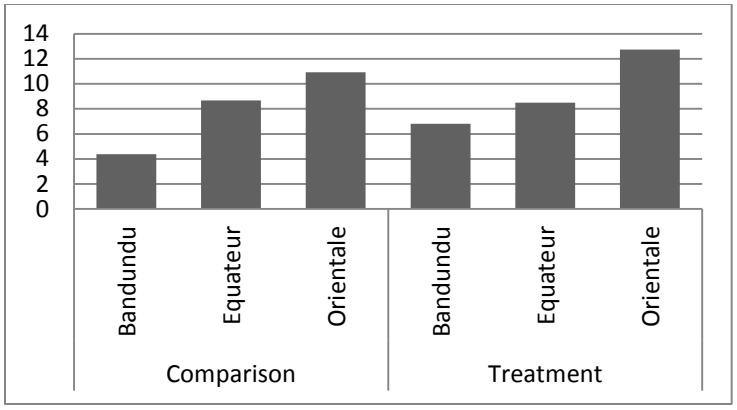
Table 28 presents—and ***Figure 37*** illustrates—the number of familiar words read correctly per minute (cwpm) by province and treatment group. No significant differences across groups were observed in any of the provinces.

Table 28. CWPM Among Grade 4 Students by Province and Group

	Group by Province					
	Comparison			Treatment		
Grade	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale
4	4.4	8.7	10.9	6.8	8.5	12.7

[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

Figure 37. CWPM Among Grade 4 Students by Province and Group



Invented Word Reading

The Invented Word Reading subtask was also administered to only students in Grade 4. In this task, students were given a chart of 50 invented words (e.g., tal, vor) and were required to read as many words as they could within 1 minute. This subtask was discontinued before the end of the minute for students who were unable to read any of the first five words. Scores reported for this subtask include percentages of students able to correctly identify words and the number of words that students could correctly identify within one minute (correct words per minute, or cnonwpm).

Comparisons by Treatment Condition. ***Table 29*** presents invented word reading percentages and standard errors for Grade 4 students by comparison and treatment group and by gender, with statistically significant differences bolded. ***Figure 38*** graphically presents the percentages of students who correctly read invented words, by number of words read.

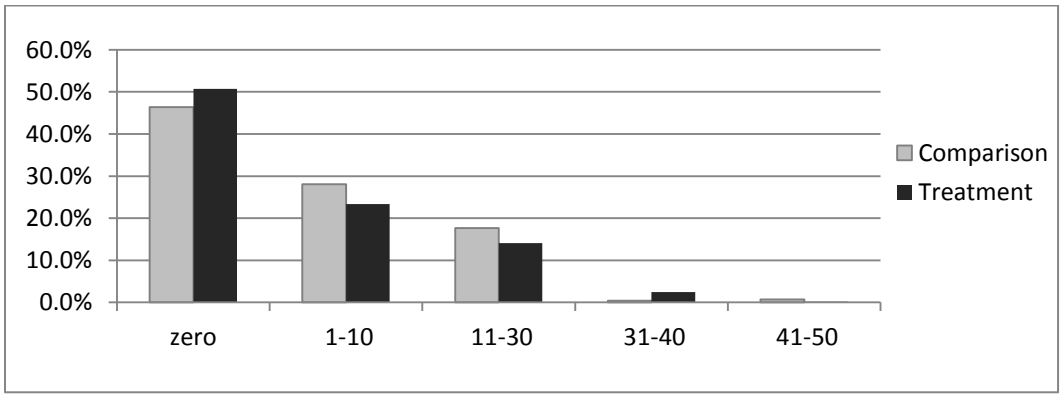
Table 29. Reading Invented Words Among Grade 4 Students by Group and Gender: Percentages (Standard Errors)

Number Correctly Answered	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
0	46.4 (2.8)	50.7 (3.2)	**39.2 [d] (3.5)**	**53.6 (3.5)**	45.1 (4.1)	56.6 (4.4)
1–10	28.1 (2.7)	23.4 (2.0)	27.1 (3.4)	29.1 (3.3)	25.0 (2.9)	21.7 (2.7)
11–30	17.7 (2.3)	14.1 (1.6)	**21.8 [c] (2.9)**	**13.5 (3.1)**	**17.0 [c] (2.2)**	**11.0 (1.9)**
31–40	**0.4 [b] (.28)**	**2.5 (.71)**	0.8 (.56)	nd (nd)	3.1 (1.1)	1.8 (.84)
41–50	0.7 (.64)	0.2 (.12)	1.4 (1.3)	nd (nd)	0.3 (.24)	nd (nd)

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED-supported schools and comparison schools are not PAQUED-supported schools.
NOTE: Percentages are defined as number of students who correctly answered the question divided by the number of students who attempted to answer the question.
[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

As indicated in ***Table 29*** and illustrated in ***Figure 38***, very few Grade 4 students were able to read more than 40 invented words: 0.2% for treatment students and 0.7% for comparison students (difference between groups not statistically significant). Fewer than 3% of students in each group were able to read even 31–40 words (2.5% for treatment students and 0.4% for comparison students, difference not statistically significant). Treatment students did outperform their comparison counterparts in reading 31–40 words (2.47% versus 0.4%, respectively, $p<0.01$), although no other statistically significant differences were observed between treatment conditions overall.

Figure 38. Grade 4 Percentages of Invented Words Read Correctly



Comparisons by Treatment Condition and Gender. As indicated in ***Table 29*** and illustrated in ***Figure 39,*** no significant differences were observed across treatment conditions when considering gender. Within the comparison group, however, females were more likely than males to score a zero on this subtask (53.6% versus 39.2%, respectively, $p<0.01$), whereas across both conditions males were more likely than females to read 11–30 words (treatment: 17.0% versus 11.0%, respectively, $p<0.05$; comparison: 21.8% versus 13.5%, respectively, $p<0.05$).

Figure 39. Grade 4 Percentages of Invented Words Read Correctly, by Gender

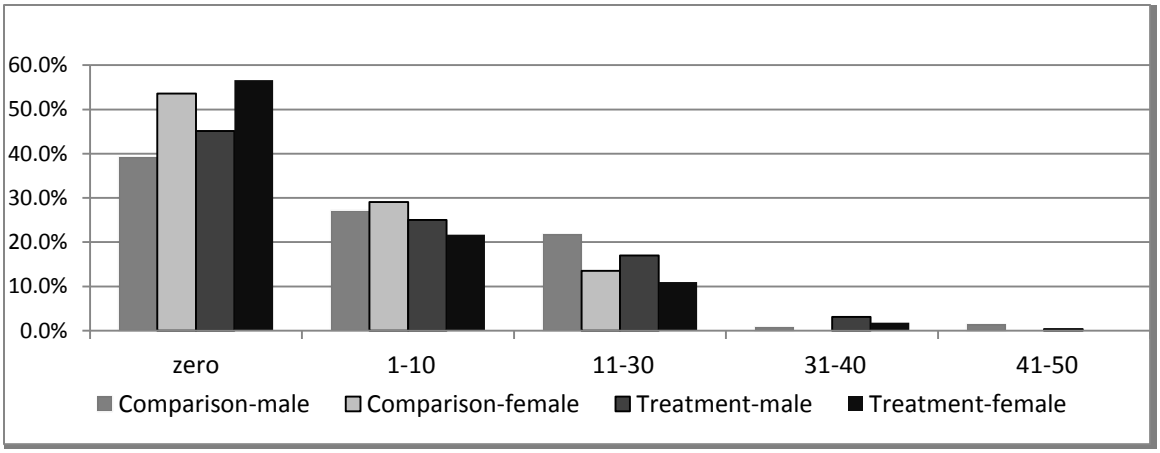


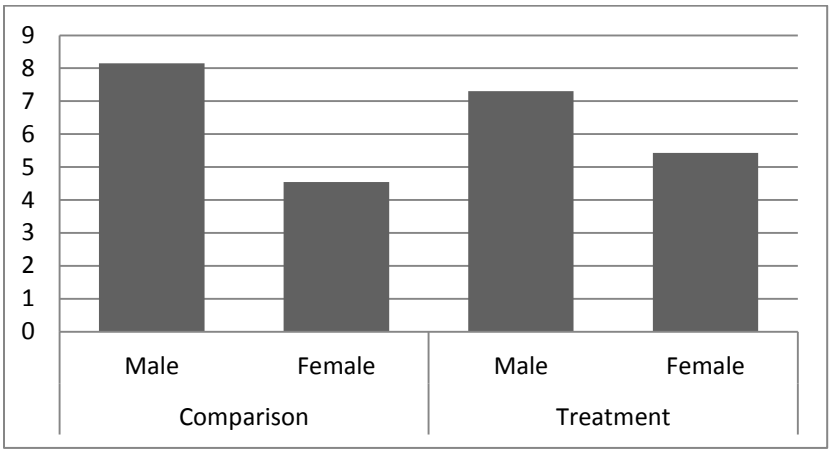
Table 30 presents—and ***Figure 40*** illustrates—the number of invented words correctly read per minute (cnonwpm) by treatment group and gender. No statistically significant difference emerged across Grade 4 treatment and comparison groups, although in both treatment and comparison groups males outperformed their female peers (treatment: 7.3% versus 5.4%, respectively, $p<0.05$; comparison: 8.2% versus 4.5%, respectively, $p<0.01$).

Table 30. CNONWPM Among Grade 4 Students by Group and Gender

	Overall		By Gender			
			Comparison		Treatment	
Grade	Comparison	Treatment	Male	Female	Male	Female
4	6.4	6.4	**8.2**[d]	**4.5**	**7.3**[c]	**5.4**

[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

Figure 40. CNONWPM Among Grade 4 Students by Group and Gender



Comparisons by Treatment Condition and Province. ***Table 31*** presents findings of statistical analysis of Grade 4 invented word reading in treatment and comparison schools by province, with areas of statistical difference bolded. Overall across the provinces, few students were able to read more than 40 invented words: 0.6% in Equator and 0.2% in Orientale (difference between groups not statistically significant). That said, students in Orientale outperformed their peers in Bandundu and Equateur in reading 11–30 invented words (Orientale, 23.2%; Bandundu, 12.0%; Equateur, 11.2%; $p<0.01$). Conversely, Orientale students were statistically significantly less likely than Bandundu students to score zero on this subtask (37.9% versus 56.2%, respectively, $p<0.01$).

Comparing provinces across treatment conditions again shows no statistically significant differences across treatment groups. Within the comparison group, however, students in Bandundu were more likely than students in either other province to get a score of zero on this subtask (Bandundu, 63.7%; Equateur, 39.4%; Orientale, 34.5%; $p<0.01$). Conversely, students in Orientale were more likely than those in Bandundu to read 11–30 invented words (24.8% versus 8.2%, respectively, $p<0.01$). A similar trend was seen within the treatment group, with Bandundu students more likely than Orientale students to score zero on this subtask (55.9% versus 38.3%, respectively, $p<0.05$), and students in Orientale more likely than those in either of the other two provinces to read 11–30 words correctly (Orientale, 23.0%; Bandundu, 12.2%; Equateur, 10.8%; $p<0.05$).

Table 31. Reading Invented Words Among Grade 4 Students by Province and Group: Percentages (Standard Errors)

Number of Words Read	Province			Group by Province					
				Comparison			Treatment		
	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale
0	**56.2[b] (4.4)**	49.8 (5.9)	**37.9 (5.0)**	**63.7[bd](4.7)**	**39.4 (5.6)**	**34.5 (4.9)**	**55.9[a] (4.6)**	50.4 (6.2)	**38.3 (5.6)**
1–10	23.3 (2.5)	26.3 (4.8)	21.3 (2.2)	25.3 (4.6)	30.5 (6.7)	29.5 (4.0)	23.3 (2.6)	26.0 (5.0)	20.3 (2.5)
11–30	**12.0[b] (2.3)**	**11.2[b] (2.6)**	**23.2 (3.4)**	**8.2[b] (2.1)**	20.2 (5.8)	**24.8 (4.3)**	**12.2[a] (2.4)**	**10.8[a] (2.7)**	**23.0 (3.7)**

Number of Words Read	Province			Group by Province					
				Comparison			Treatment		
	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale
31–40	1.8 (1.0)	2.0 (1.0)	4.1 (1.5)	nd (nd)	nd (nd)	1.0 (.69)	1.9 (1.0)	2.1 (1.0)	4.5 (1.6)
41–50	nd (nd)	0.6 (.43)	0.2 (.17)	nd (nd)	nd (nd)	1.7 (1.5)	nd (nd)	0.6 (.46)	nd (nd)

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED supported schools, and comparison schools are not PAQUED supported schools.
NOTE: Percentages are defined as number of students who correctly answered the question divided by the number of students who attempted to answer the question.
[a] Difference between estimate and Orientale estimate is statistically significant at the 0.05 level.
[b] Difference between estimate and Orientale estimate is statistically significant at the 0.01 level.
[c] Difference between estimate and Equateur estimate is statistically significant at the 0.05 level.
[d] Difference between estimate and Equateur estimate is statistically significant at the 0.01 level.
[e] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[f] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.

Figure 41. Reading Invented Words Among Grade 4 Students by Province and Group: Percentages

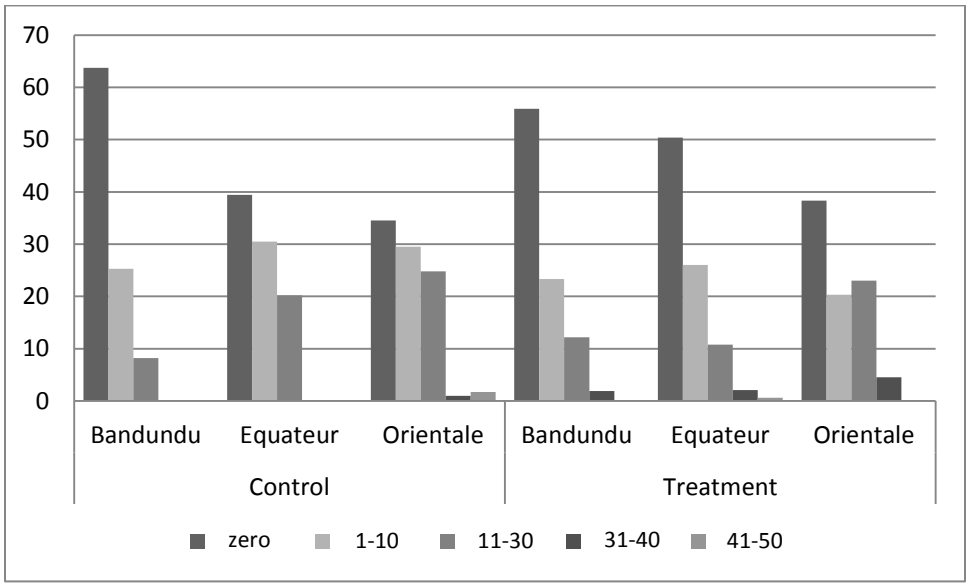


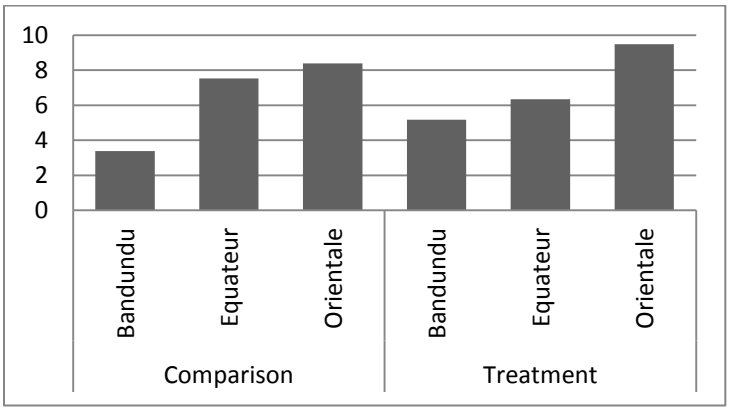
Table 32 presents—and ***Figure 42*** illustrates—the number of invented words read correctly per minute (cnonwpm) by province and treatment group. No significant differences across treatment and comparison groups were observed.

Table 32. CNONWPM Among Grade 4 Students by Province and Group

	Group by Province					
	Comparison			Treatment		
Grade	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale
4	3.4	7.5	8.4	5.2	6.3	9.5

[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

Figure 42. CNONWPM Among Grade 4 Students by Province and Group



Oral Reading Fluency

The Oral Reading Fluency subtask was administered to only students in Grade 4. In this task, students were given a passage containing 50 words and were required to read as much of the passage as they could within 1 minute. Scores reported for this subtask include percentages of students able to orally read words and the number of words that students could correctly read within one minute (oral reading fluency, or orf).

Comparisons by Treatment Condition. ***Table 33*** presents oral reading fluency percentages and standard errors for Grade 4 students by comparison and treatment group and by gender, with statistically significant differences bolded. ***Figure 43*** graphically presents the percentages of students who correctly read words orally, by number of words read.

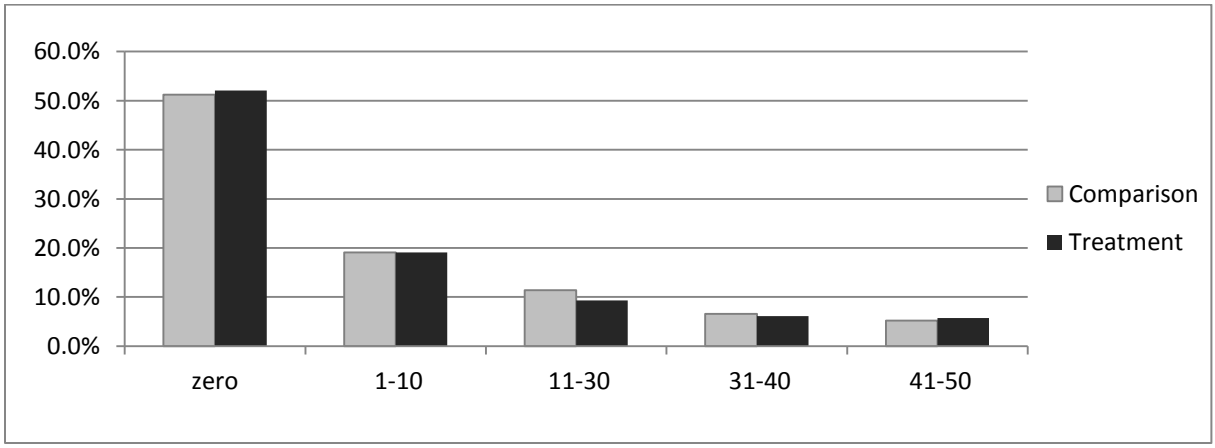
Table 33. Oral Reading Test Among Grade 4 Students by Group and Gender: Percentages (Standard Errors)

Number Correctly Answered	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
0	51.2 (3.6)	52.1 (3.1)	**43.7 [d](4.0)**	**58.7 (4.4)**	48.5 (3.7)	55.9 (4.2)
1–10	19.1 (2.2)	19.1 (1.8)	20.8 (2.8)	17.3 (2.7)	19.4 (2.1)	18.8 (2.5)
11–30	11.4 (1.5)	9.33 (1.2)	12.2 (2.3)	10.6 (2.4)	10.7 (2.0)	7.87 (1.4)
31–40	6.60 (1.5)	6.09 (1.1)	**10.4 [d](2.6)**	**2.76 (1.0)**	6.03 (1.3)	6.15 (1.8)
41–50	5.17 (1.5)	5.72 (1.2)	6.53 (2.1)	3.81 (1.3)	6.22 (1.7)	5.17 (1.3)

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED-supported schools and comparison schools are not PAQUED-supported schools.
NOTE: Percentages are defined as number of students who correctly answered the question divided by the number of students who attempted to answer the question.
[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

As shown in ***Table 33*** and ***Figure 43***, just over half of the students (51.2% in comparison schools and 52.1% in treatment schools) were unable to read a single word (difference between groups not statistically significant). Overall, no statistically significant differences emerged between treatment and comparison groups.

Figure 43. Grade 4 Percentages of Words Read Correctly in the Oral Reading Test



Comparisons by Treatment Condition and Gender. ***Table 33*** and ***Figure 44*** display results by gender. Again, no statistically significant differences were observed between groups. Within the comparison group, females were more likely than males to score a zero on this subtask (58.7% versus 43.7%, respectively, $p<0.01$), while males were more

likely than females to read 31–40 words (10.4% versus 2.8%, respectively, p<0.01). No significant differences were observed between males and females in the treatment group.

Figure 44. Grade 4 Percentages of Words Read Correctly in the Oral Reading Test, by Gender

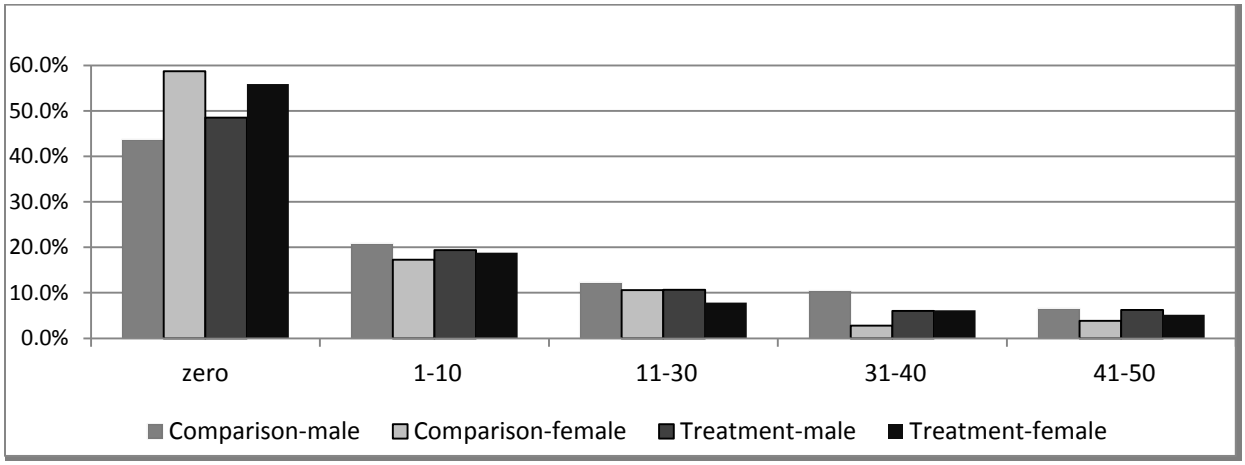


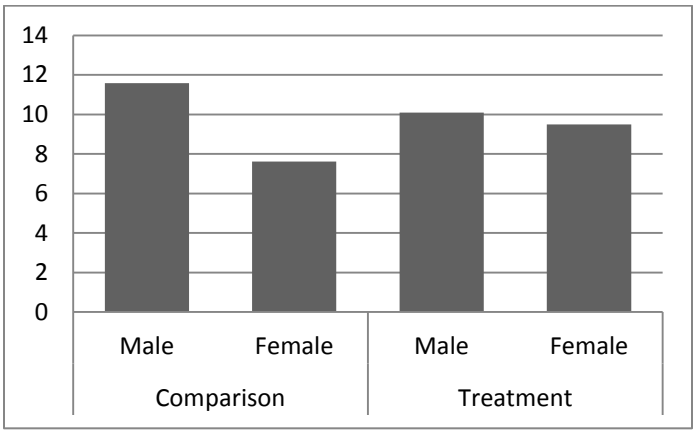
Table 34 presents—and ***Figure 45*** illustrates—the oral reading fluency (orf) achieved by treatment group and gender. No statistically significant differences were observed between treatment groups for either males or females. Comparison-group males did outperform their female counterparts (11.6% versus 7.6%, p<0.01), although there was no statistically significant difference between treatment males and females.

Table 34. ORF Among Grade 4 Students by Group and Gender

	Overall		By Gender			
			Comparison		Treatment	
Grade	Comparison	Treatment	Male	Female	Male	Female
4	9.6	9.8	**11.6**[d]	**7.6**	10.1	9.5

[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

Figure 45. ORF Among Grade 4 Students by Group and Gender



Comparisons by Treatment Condition and Province. ***Table 35*** presents findings of statistical analysis of Grade 4 oral reading fluency in treatment and comparison schools by province, with areas of statistical difference bolded. Overall, students in Orientale were less likely than their peers in either of the other two provinces to score zero on this subtask (Equateur, 55.9%; Bandundu, 55.8%; Orientale, 38.6%; $p<0.05$), while they were more likely than students in Bandundu to read at least 41 words (11.8% versus Bandundu's 2.7%, $p<0.01$).

As indicated in ***Table 35*** and illustrated in ***Figure 46***, no statistically significant differences emerged when comparing across treatment conditions. Within the comparison group, however, Bandundu students were significantly more likely than students in the other two provinces to score zero on this subtask (Bandundu, 68.1%; Equateur, 43.0%; Orientale, 40.3%; $p<0.01$), while Orientale students were significantly more likely than students in the other two provinces to read 31–40 words (Orientale, 12.3%; Bandundu, 3.0%; Equateur, 2.1%; $p<0.05$). Students in Orientale were also more likely than those in Bandundu to read 11–30 words (15.2% versus Bandundu's 6.3%, $p<0.01$). Similarly, within the treatment group, Orientale students were less likely than those in the other two provinces to score zero on this subtask (Equateur, 56.5%; Bandundu, 55.3%; Orientale, 38.4%; $p<0.05$), while they were more likely than students in Bandundu to read at least 41 of the 50 total words (12.6% versus Bandundu's 2.8%, $p<0.01$).

Table 35. Oral Reading Comprehension Among Grade 4 Students by Province and Group: Percentages (Standard Errors)

Number Correctly Answered	Province			Group by Province					
				Comparison			Treatment		
	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale
0	**55.8[a] (4.3)**	**55.9[a] (5.3)**	**38.6 (5.7)**	**68.1[bd] (5.6)**	**43.0 (5.1)**	**40.3 (5.7)**	**55.3[a] (4.5)**	**56.5[a] (5.6)**	**38.4 (6.3)**
1-10	21.6 (3.2)	15.9 (2.4)	17.5 (2.0)	20.9 (4.4)	18.8 (4.6)	17.6 (2.9)	21.6 (3.3)	15.8 (2.5)	17.5 (2.3)
11-30	9.2 (1.7)	9.9 (2.4)	9.5 (2.0)	**6.3[b] (1.9)**	12.9 (3.4)	**15.2 (2.5)**	9.4 (1.8)	9.7 (2.5)	8.8 (2.2)

Number Correctly Answered	Province			Group by Province					
	Bandundu	Equateur	Orientale	Bandundu	Comparison Equateur	Orientale	Bandundu	Treatment Equateur	Orientale
31-40	5.2 (1.4)	4.2 (2.0)	10.6 (2.6)	**3.0[a] (1.8)**	**2.1[b] (1.3)**	**12.3 (3.2)**	5.3 (1.4)	4.3 (2.0)	10.3 (2.9)
41-50	**2.7[b] (1.5)**	6.3 (2.1)	**11.8 (2.6)**	nd (nd)	13.7 (4.4)	5.2 (2.4)	**2.8[b] (1.6)**	5.9 (2.2)	**12.6 (3.0)**

nd: No students in domain.
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED supported schools, and comparison schools are not PAQUED supported schools.
NOTE: Percentages are defined as number of students who correctly answered the question divided by the number of students who attempted to answer the question.
[a] Difference between estimate and Orientale estimate is statistically significant at the 0.05 level.
[b] Difference between estimate and Orientale estimate is statistically significant at the 0.01 level.
[c] Difference between estimate and Equateur estimate is statistically significant at the 0.05 level.
[d] Difference between estimate and Equateur estimate is statistically significant at the 0.01 level.
[e] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
f Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.

Figure 46. Oral Reading Comprehension Among Grade 4 Students by Province and Group: Percentages

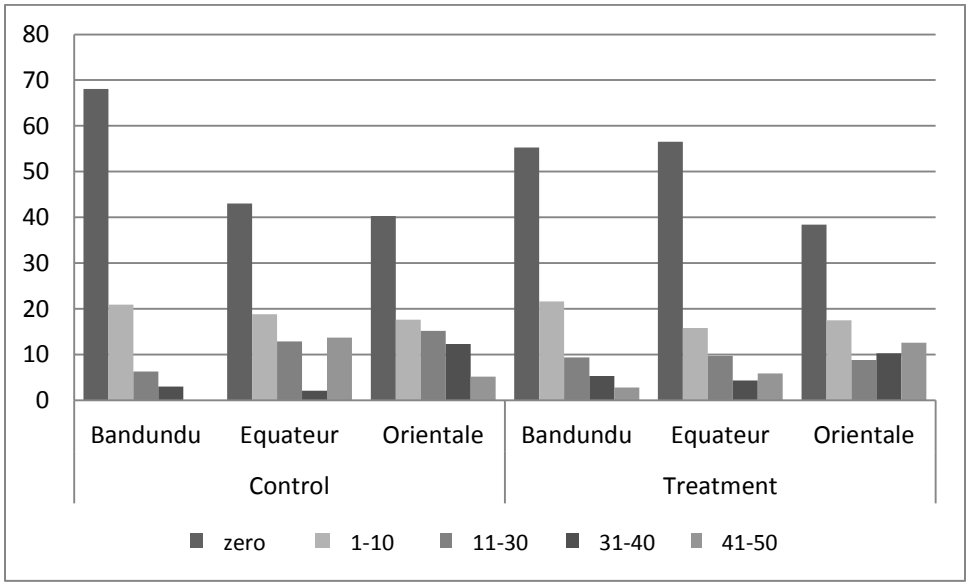


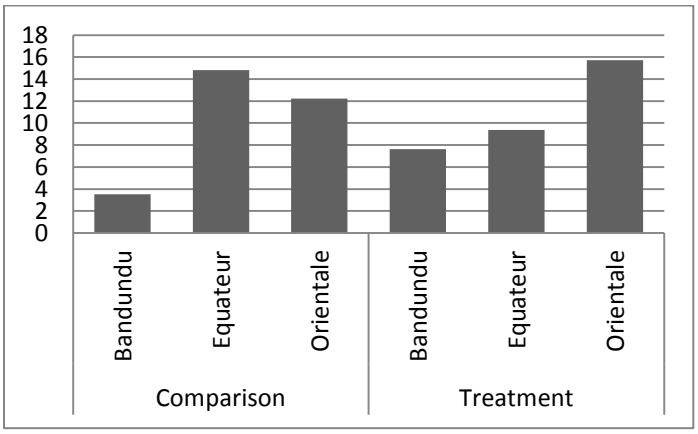
Table 36 presents—and ***Figure 47*** illustrates—student oral reading fluency (orf) by province and treatment group. Within Bandundu, treatment students outperformed their comparison peers (7.6 versus 3.5 orf, respectively, $p<0.05$); no significant differences were observed in the other two provinces.

Table 36. ORF Among Grade 4 Students by Province and Group

Grade	Group by Province					
	Comparison			Treatment		
	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale
4	**3.5**[a]	14.8	12.2	**7.6**	9.4	15.7

[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

Figure 47. ORF Among Grade 4 Students by Province and Group



Reading Comprehension

After reading the passage in the Oral Reading Fluency subtask, Grade 4 students were asked five questions that assessed their basic comprehension of that passage. Scores reported for this subtask include percentages of students able to answer comprehension questions, based upon the number of students who attempted to answer each question.[14]

Comparisons by Treatment Condition. ***Table 37*** presents reading comprehension percentages and standard errors for Grade 4 students by comparison and treatment group and by gender, with statistically significant differences bolded. ***Figure 48*** presents this information graphically.

14 Students were only asked questions that corresponded with the lines of the passage that they were able to read within one minute.

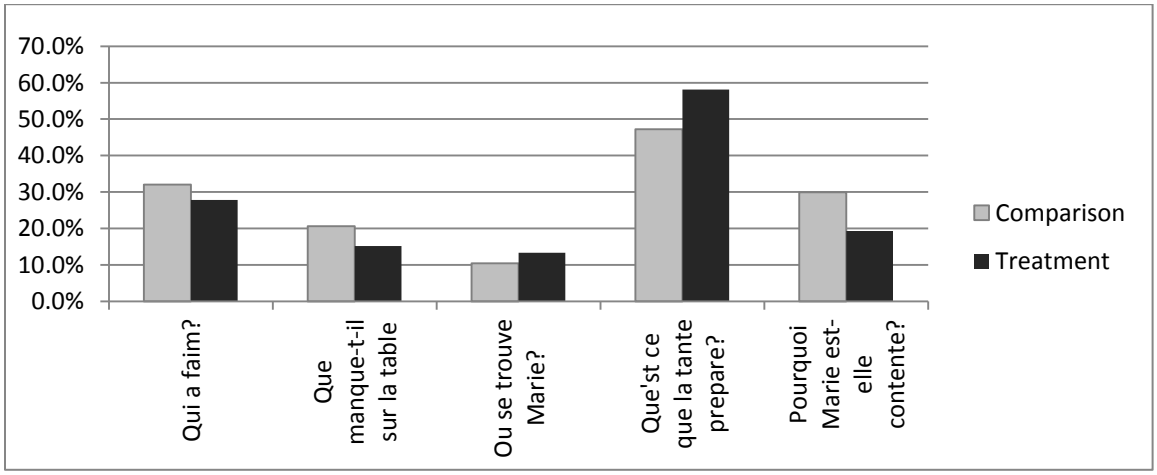
Table 37. Percentages of Students Responding Correctly to Each Reading Comprehension Question Among Grade 4 Students by Group and Gender: Percentages (Standard Errors)

By Question Number	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
1	32.0 (4.2)	27.8 (3.4)	38.3 (5.9)	24.4 (5.8)	24.6 (3.5)	32.1 (5.8)
2	20.6 (4.1)	15.2 (2.7)	22.2 (5.6)	18.1 (6.1)	16.4 (3.6)	13.4 (3.8)
3	10.4 (4.9)	13.3 (3.2)	10.3 (5.4)	10.7 (6.6)	14.7 (4.2)	11.4 (4.2)
4	47.2 (8.2)	58.1 (4.9)	49.2 (10.)	43.3 (11.)	56.8 (7.5)	59.4 (7.0)
5	29.9 (14.)	19.3 (7.4)	30.2 (16.)	29.4 (14.)	13.1 (6.5)	26.4 (12.)

nd: No students in domain
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED-supported schools and comparison schools are not PAQUED-supported schools.
NOTE: Percentages are defined as number of students who correctly answered the question divided by the number of students who attempted to answer the question.
[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

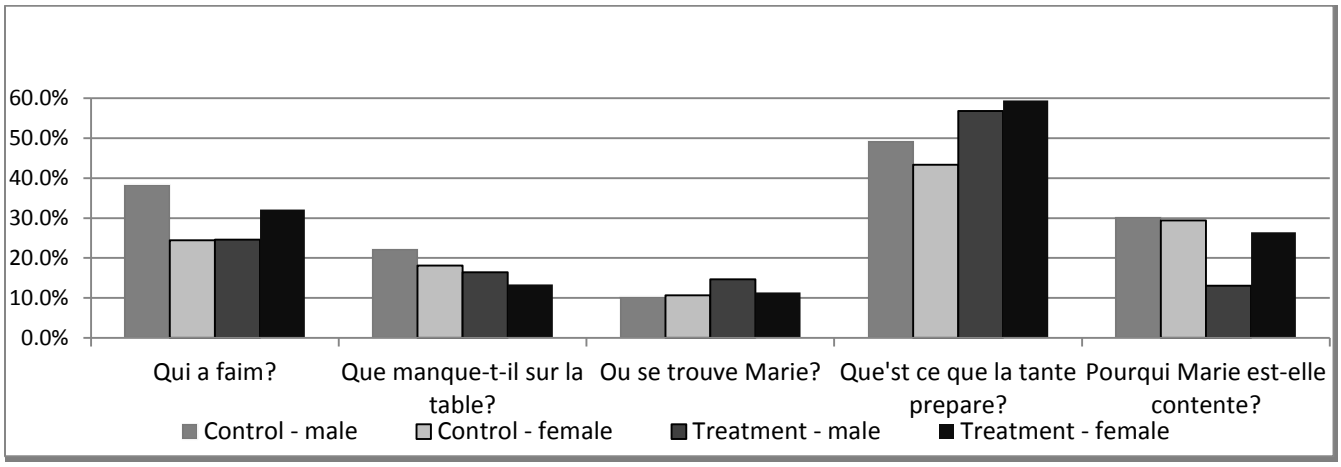
As shown in ***Table 37***, no statistically significant differences between treatment and comparison groups emerged. While differences between groups were not statistically significant, however, Grade 4 students in both treatment and comparison groups tended to perform better on one question in particular, "Que'est ce que la tante prepare?," with nearly half (47.2%) of comparison students and over half (58.1%) of treatment students answering it correctly.

Figure 48. Grade 4 Percentages of Students Responding Correctly to Each Reading Comprehension Question



Comparisons by Treatment Condition and Gender. As shown in ***Table 37*** and ***Figure 49***, no statistically significant differences emerged when comparing genders across treatment conditions.

Figure 49. Grade 4 Percentages of Reading Comprehension Questions Answered Correctly, by Gender



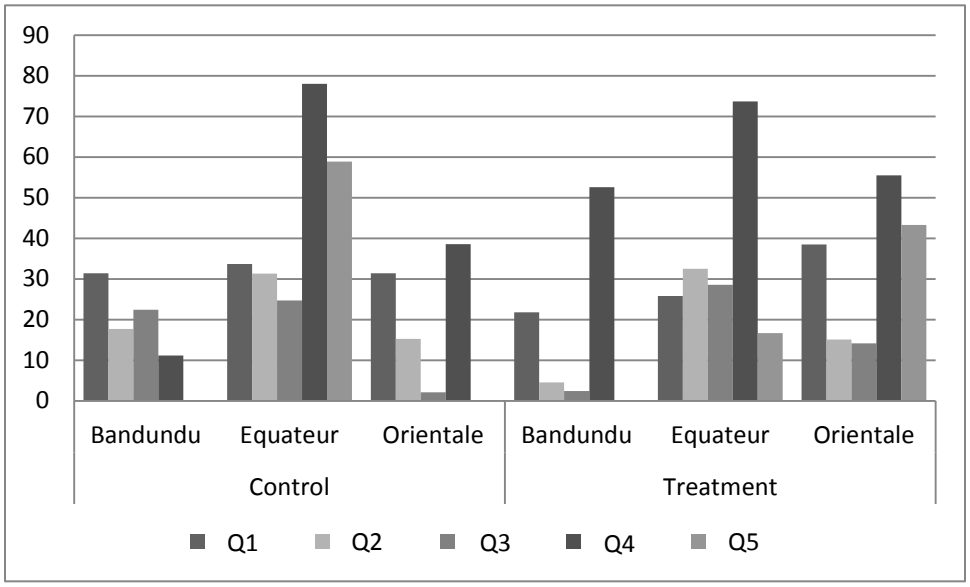
Comparisons by Treatment Condition and Province. ***Table 38*** presents findings of statistical analysis of Grade 4 reading comprehension in treatment and comparison schools by province, with areas of statistical difference bolded.

Table 38. Reading Comprehension among Grade 4 Students by Province and Group: Percentages (Standard Errors)

By Question Number	Province			Group by Province					
	Bandundu	Equateur	Orientale	Bandundu	Comparison Equateur	Orientale	Bandundu	Treatment Equateur	Orientale
1	**22.1[a] (4.5)**	26.3 (5.9)	**37.8 (6.0)**	31.4 (7.1)	33.7 (10.)	31.4 (5.1)	**21.8[a] (4.6)**	25.8 (6.2)	**38.5 (6.6)**
2	**4.9[ad](3.0)**	**32.4[a] (7.0)**	**15.1 (2.9)**	17.7 (6.3)	31.3 (9.2)	15.3 (4.9)	**4.6[ad](3.1)**	**32.5[a] (7.5)**	**15.1 (3.2)**
3	2.7 [d](2.3)	28.4 (7.7)	13.1 (5.3)	22.4 (15.)	24.7 (14.)	**2.1[e] (1.7)**	**2.4[d] (2.3)**	**28.6 (8.2)**	**14.2 (5.8)**
4	52.0 (9.0)	**74.1[a] (7.1)**	**54.2 (6.2)**	**11.2[adf](9.3)**	**78.0[a] (11.)**	**38.6 (9.8)**	**52.6 (9.1)**	73.7 (7.7)	55.5 (6.7)
5	nd (nd)	21.1 (11.)	40.7 (13.)	nd (nd)	**58.9[b] (20.)**	**nd (nd)**	nd (nd)	16.7 (11.)	43.3 (14.)

nd: No students in domain
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED supported schools, and comparison schools are not PAQUED supported schools.
NOTE: Percentages are defined as number of students who correctly answered the question divided by the number of students who attempted to answer the question.
[a] Difference between estimate and Orientale estimate is statistically significant at the 0.05 level.
[b] Difference between estimate and Orientale estimate is statistically significant at the 0.01 level.
[c] Difference between estimate and Equateur estimate is statistically significant at the 0.05 level.
[d] Difference between estimate and Equateur estimate is statistically significant at the 0.01 level.
[e] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
f Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.

Figure 50. Reading Comprehension Among Grade 4 Students by Province and Group: Percentages



As shown in ***Table 38***, while no Grade 4 Bandundu students responded correctly to Question 5, 21.1% of Equateur and 40.7% of Orientale students did (differences between groups not statistically significant). Students in Equateur were statistically significantly more likely than students in Orientale to respond correctly to Question 4 (74.1% versus 54.2%, respectively, $p<0.05$) and more likely than students in both other provinces to respond correctly to Question 2 (Equateur, 32.4%; Orientale, 15.1%; Bandundu, 4.9%; $p<0.05$). Students in Orientale were also significantly more likely than students in Bandundu to respond correctly to Question 1 (37.8% versus 22.1%, respectively, $p<0.05$). Equateur students were more likely than Bandundu student to respond correctly to Question 2 (32.4% versus 4.9%, $p<0.01$).

Comparing across treatment conditions, Bandundu treatment students were more likely than comparison students to respond correctly to Question 4 (52.6% versus 11.2%, respectively, $p<0.01$); similarly, Orientale treatment students were more likely than comparison students to respond correctly to Question 3 (14.2% versus 2.1%, respectively, $p<0.05$). Within the comparison group, itself, Equateur students were more likely than students in the other two provinces to correctly answer Question 4 (Equateur, 78.0%; Orientale, 38.6%; Bandundu, 11.2%; $p<0.05$); they were also more likely than students in Orientale to respond correctly to Question 5 (58.9%, while no Orientale students fell into this category, $p<0.01$). Within the treatment group, Equateur students were more likely than students in the other two provinces to respond correctly to Question 2 (Equateur, 32.5%; Orientals, 15.1%; Bandundu, 4.6%, $p<0.05$), and they were more likely than treatment students in Bandundu to respond correctly to Question 3 (28.6% versus Bandundu's 2.4%, $p<0.01$). Treatment students in Orientale were more likely than their peers in Bandundu to respond correctly to Question 1 (38.5% versus 21.8%, $p<0.05$).

Listening Comprehension

For the Listening Comprehension subtask, Grade 2 and Grade 4 students listened to a short passage and were asked five questions that assessed their basic comprehension of that passage. Scores reported for this subtask include percentages of students able to answer the comprehension questions, based upon the number of students who attempted to answer each question.[15]

Comparisons by Grade and Treatment Condition. ***Tables 39*** and ***40*** present listening comprehension percentages and standard errors for Grades 2 and 4 students by comparison and treatment group and by gender, with statistically significant differences bolded. ***Figures 51*** and ***52*** present information graphically by grade and group.

Table 39. Percentages of Students Responding Correctly to Each Listening Comprehension Question Among Grade 2 Students by Group and Gender: Percentages (Standard Errors)

By Question Number	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
1	2.9 (.92)	5.7 (1.2)	**2.4[a] (1.0)**	3.3 (1.3)	**6.3 (1.6)**	5.2 (1.5)
2	7.76 (1.5)	12.3 (1.9)	9.6 (2.0)	5.9 (1.7)	13.5 (2.4)	11.1 (2.0)
3	**4.6[a] (1.0)**	**9.0 (1.6)**	5.5 (1.3)	**3.6[a] (1.3)**	8.8 (1.8)	**9.1 (1.8)**
4	**13.3[a] (2.3)**	**21.8 (2.5)**	15.3 (3.1)	**11.3[b] (2.4)**	20.0 (2.9)	**23.6 (3.1)**
5	**4.3[a] (1.1)**	**7.9 (1.3)**	5.4 (1.4)	**3.2[a] (1.0)**	8.9 (1.6)	**7.0 (1.4)**

nd: No students in domain
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED-supported schools and comparison schools are not PAQUED-supported schools.
NOTE: Percentages are defined as number of students who correctly answered the question divided by the number of students who attempted to answer the question.
[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

As shown in ***Table 39*** and illustrated in ***Figure 51***, overall the Grade 2 treatment-group students were more likely than comparison students— with all differences statistically significant at $p<0.05$—to respond correctly to Question 3 (9.0% versus 4.6%, respectively), 4 (21.8% versus 13.3%, respectively), and 5 (7.9% versus 4.3%, respectively).

15 Students were only asked questions that corresponded with the lines of the passage that they were able to read within one minute.

Figure 51. Grade 2 Percentages of Students Responding Correctly to Each Listening Comprehension Question

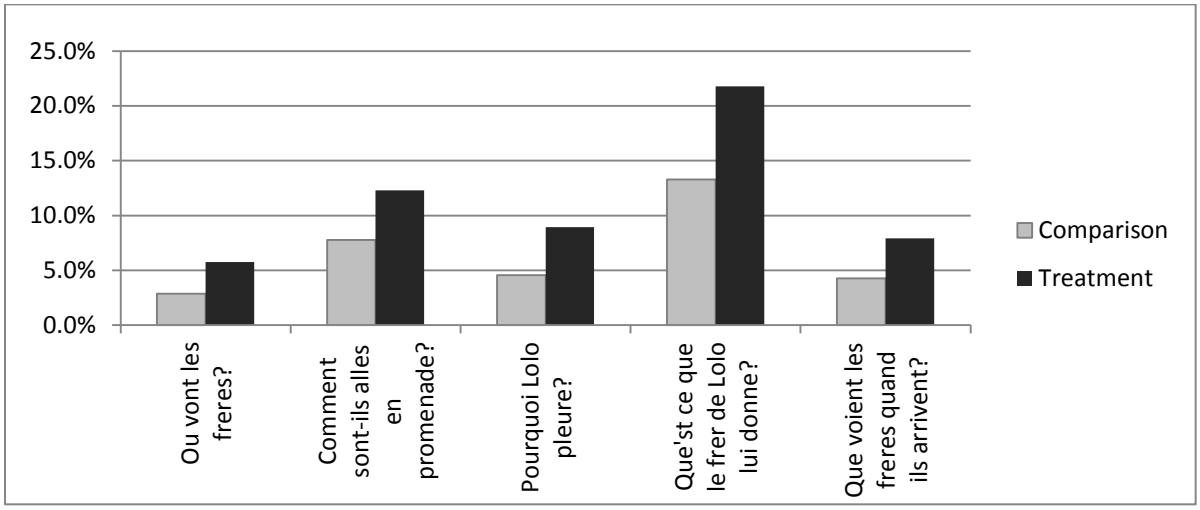


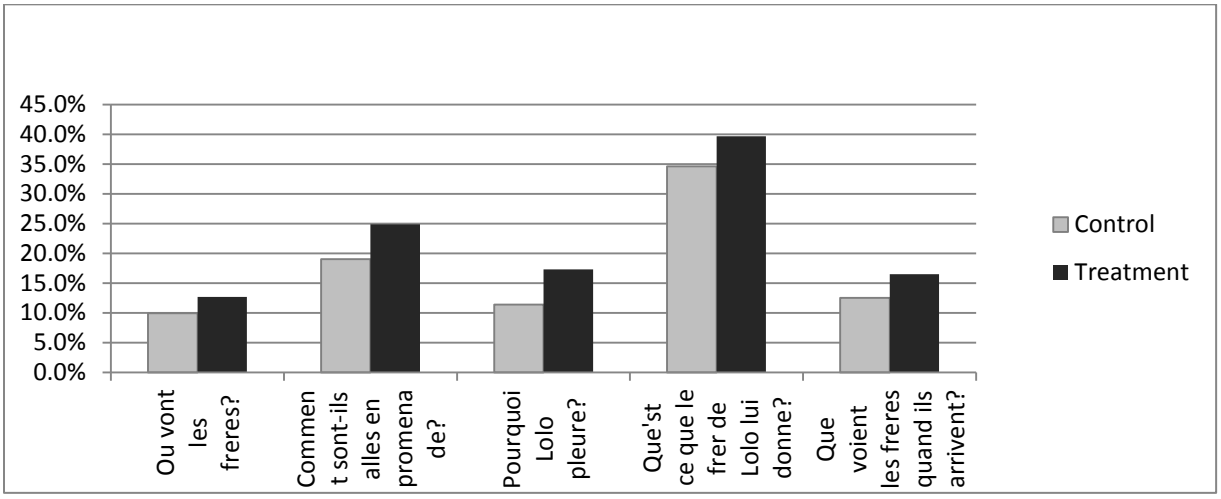
Table 40. Percentages of Students Responding Correctly to Each Listening Comprehension Question Among Grade 4 Students by Group and Gender: Percentages (Standard Errors)

By Question Number	Overall		By Gender			
			Comparison		Treatment	
	Comparison	Treatment	Male	Female	Male	Female
1	9.91 (2.1)	12.7 (2.1)	12.7 (3.0)	7.00 (2.1)	14.6 (2.5)	10.6 (2.5)
2	19.0 (3.4)	24.9 (2.2)	**23.0[c] (4.3)**	**14.9 (3.1)**	**30.2[c] (2.9)**	**19.2 (2.5)**
3	11.4 (2.3)	17.3 (2.5)	12.6 (2.9)	10.1 (2.4)	17.3 (2.7)	17.4 (2.9)
4	34.6 (3.7)	39.7 (2.8)	**39.6[c] (4.8)**	**29.6 (3.8)**	41.2 (3.3)	38.0 (3.4)
5	12.5 (2.3)	16.5 (2.1)	14.6 (3.1)	**10.4[a] (2.3)**	14.6 (2.2)	**18.5 (2.7)**

nd: No students in domain
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED-supported schools and comparison schools are not PAQUED-supported schools.
NOTE: Percentages are defined as number of students who correctly answered the question divided by the number of students who attempted to answer the question.
[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.
[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.
[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

As indicated in ***Table 40*** and illustrated in ***Figure 52***, no statistically significant differences emerged between Grade 4 treatment and comparison students overall.

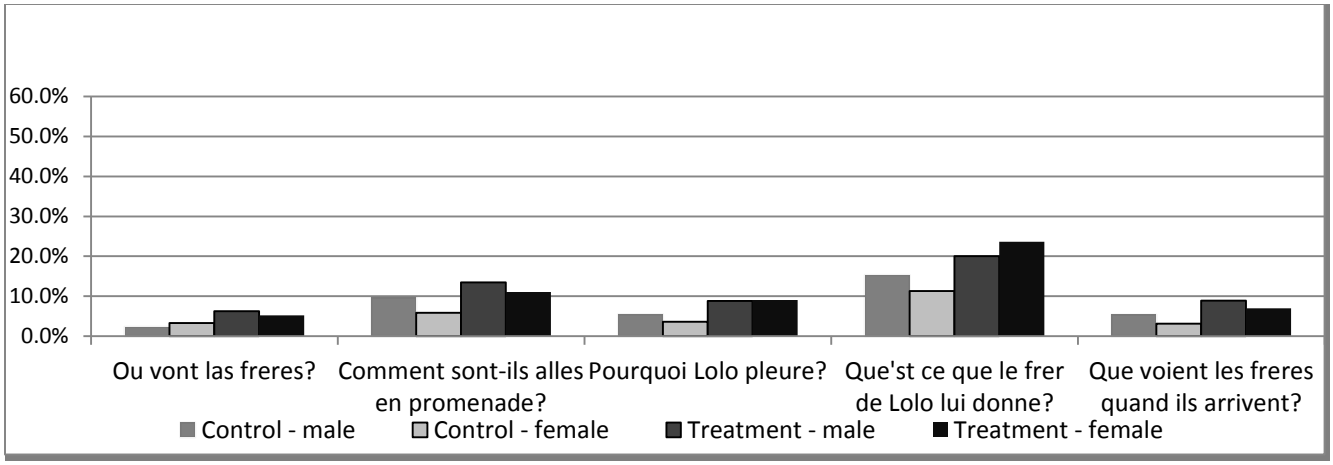
Figure 52. Grade 4 Percentages of Students Responding Correctly to Each Listening Comprehension Question



Comparisons by Treatment Condition and Gender. ***Figures 53*** and ***54*** graphically display results by treatment group and gender for Grades 2 and 4, respectively.

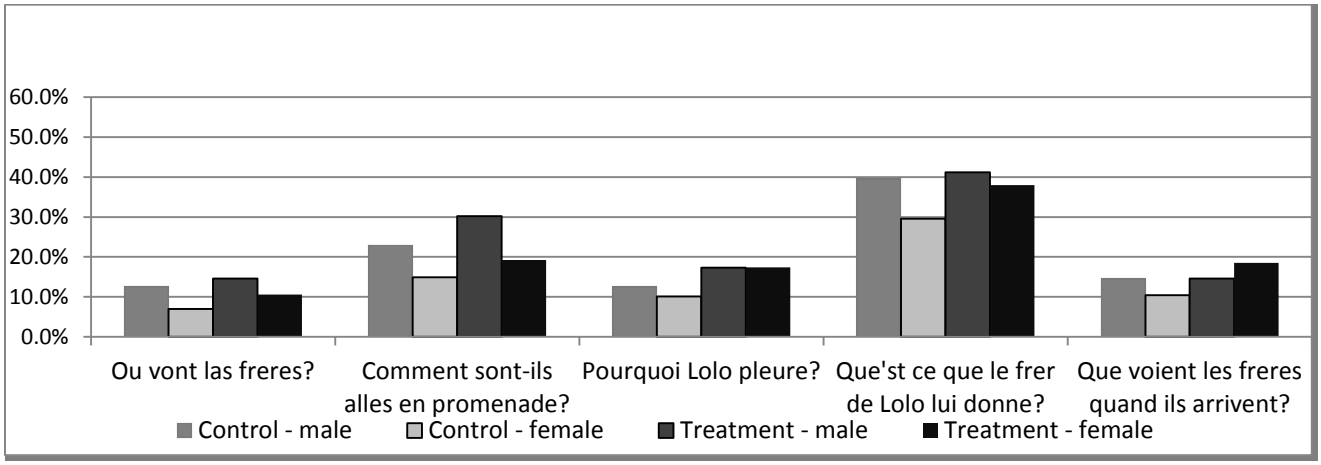
As shown in ***Table 39*** and ***Figure 53***, females in the treatment group were more likely than those in the comparison group to respond correctly to Questions 3 (9.1% versus 3.6%, respectively, $p<0.05$), 4 (23.6% versus 11.3%, respectively, $p<0.01$), and 5 (7.0% versus 3.2%, respectively, $p<0.05$). This pattern did not emerge between male treatment and comparison students. However, male treatment students were more likely than their comparison peers to respond correctly to Question 1 (6.3% versus 2.4%, respectively, $p<0.05$).

Figure 53. Grade 2 Percentages of Students Responding Correctly to Each Listening Comprehension Question, by Gender



In Grade 4, comparing across genders, female treatment students were more likely than their comparison counterparts to correctly respond to Question 5 (18.5% versus 10.4%, respectively, $p<0.05$), as shown in ***Table 40*** and ***Figure 54***. Within the comparison group, males were more likely than females to respond to Questions 4 (39.6% versus 29.6%, respectively, $p<0.05$) and 2 (23.0% versus 14.9%, respectively, $p<0.05$); within the treatment group, males were also more likely than females to respond correctly to Question 2 (30.2% versus 19.2%, respectively, $p<0.05$).

Figure 54. Grade 4 Percentages of Students Responding Correctly to Each Listening Comprehension Question, by Gender



Comparisons by Grade, Treatment Condition, and Province. ***Tables 41*** and ***42*** present findings of statistical analysis of Grades 2 and 4 listening comprehension in treatment and comparison schools by province, with areas of statistical difference bolded. ***Figures 55*** and ***56*** display this information graphically.

Table 41. Listening Comprehension Among Grade 2 Students by Province and Group: Percentages (Standard Errors)

By Question Number	Province			Group by Province					
				Comparison			Treatment		
	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale
1	3.6[c] (1.1)	**10.6[a] (3.1)**	**3.3 (1.9)**	**3.4[a] (1.6)**	**7.5[a] (3.0)**	**nd (nd)**	**3.6[c] (1.2)**	**10.8 (3.2)**	3.7 (2.1)
2	9.7 (2.1)	15.6 (4.0)	12.2 (3.1)	7.9 (2.3)	11.0 (3.4)	5.9(2.1)	9.8 (2.2)	15.9 (4.2)	13.1 (3.5)
3	6.1 (1.9)	12.3 (3.1)	9.3 (3.0)	**6.1[a] (1.9)**	**8.8[a] (3.0)**	**1.4[e](.98)**	6.1 (2.0)	12.5 (3.2)	**10.4 (3.5)**
4	**15.6[a] (3.4)**	22.9 (4.1)	**29.9 (4.2)**	11.5 (3.0)	17.1 (6.3)	**12.4[f](3.3)**	**15.8[b] (3.6)**	23.3 (4.4)	**32.6 (4.8)**
5	5.5 (1.7)	9.5 (2.2)	9.7 (2.4)	**7.7[a] (2.4)**	4.9 (1.9)	**1.7[f](1.4)**	5.4 (1.8)	9.7 (2.4)	**10.9 (2.8)**

nd: No students in domain
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED supported schools, and comparison schools are not PAQUED supported schools.
NOTE: Percentages are defined as number of students who correctly answered the question divided by the number of students who attempted to answer the question.
[a] Difference between estimate and Orientale estimate is statistically significant at the 0.05 level.
[b] Difference between estimate and Orientale estimate is statistically significant at the 0.01 level.
[c] Difference between estimate and Equateur estimate is statistically significant at the 0.05 level.
[d] Difference between estimate and Equateur estimate is statistically significant at the 0.01 level.
[e] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
f Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.

Grade 2 students in Orientale were significantly more likely that those in Bandundu to correctly answer Question 4 (29.9% versus 15.6%, respectively, $p<0.05$), whereas Equateur students were more likely to answer Question 1 than were Orientale students (10.6% versus 3.3%, respectively, $p<0.05$). Comparing across treatment conditions, Orientale treatment students were more likely than comparison students to correctly answer Questions 3 (10.4% versus 1.4%, respectively), 4 (32.6% versus 12.4%, respectively), and 5 (10.9% versus 1.7%, respectively) ($p<0.05$).

Within the comparison group, both Bandundu and Equateur students were more likely than Orientale students to answer Questions 1 and 3, while Bandundu students were also more likely than Orientale students to answer Question 5 (see Table 41 for specific differences). Within the treatment group, Equateur students outperformed Bandundu students on Question 1 (10.8% versus 3.6%, respectively, $p<0.05$), while Orientale students outperformed Bandundu students on Question 4 (32.6% versus 15.8%, respectively, $p<0.01$).

Figure 55. Listening Comprehension Among Grade 2 Students by Province and Group: Percentages

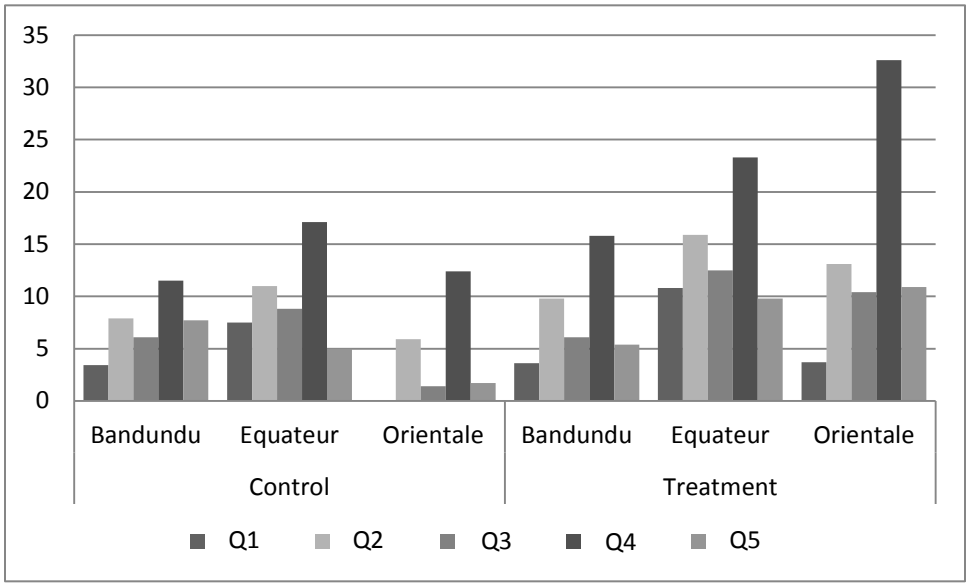


Table 42. Listening Comprehension Among Grade 4 Students by Province and Group: Percentages (Standard Errors)

By Question Number	Province			Group by Province					
				Comparison			Treatment		
	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale
1	7.59 (2.0)	15.6 (4.0)	20.0 (6.2)	7.7 (2.0)	18.3 (8.0)	7.4(1.9)	7.6(2.0)	15.5 (4.2)	21.6 (6.9)
2	21.6 (3.0)	24.8 (3.9)	31.0 (4.0)	**14.4[c] (3.4)**	**36.9[a] (10.)**	**13.5[f](3.5)**	**22.0[a](3.1)**	24.2 (4.0)	**33.1 (4.4)**
3	**12.6[a] (2.9)**	16.2 (4.0)	**28.1 (6.3)**	**8.2[c] (2.2)**	**25.8[a] (7.2)**	**6.5[f](2.1)**	**12.8[a](3.0)**	15.7 (4.2)	**30.7 (7.0)**
4	**32.5[b] (3.2)**	**36.6[b] (6.3)**	**58.4 (4.6)**	**23.2[a] (5.1)**	42.2 (10.)	**40.8[f](5.3)**	**32.9[b](3.3)**	**36.3[b](6.6)**	**60.6 (5.1)**
5	**10.8[b] (1.9)**	17.8 (4.3)	**26.7 (5.3)**	**8.0[c] (2.6)**	**26.0[a] (6.9)**	**9.4[f](2.7)**	**10.9[b](2.0)**	17.3 (4.5)	**28.9 (6.0)**

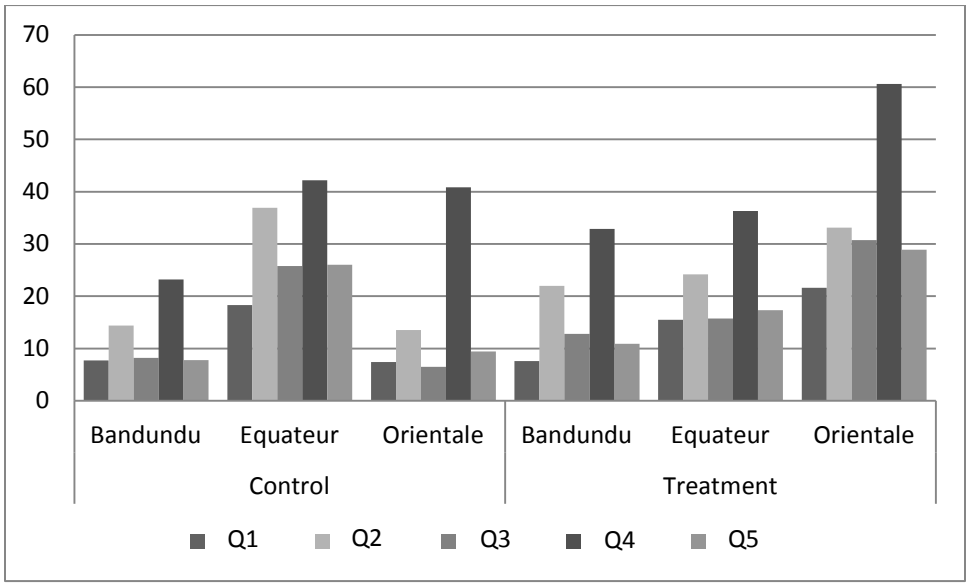
nd: No students in domain
NOTE: Group is defined as comparison schools and treatment schools, where treatment schools are PAQUED supported schools, and comparison schools are not PAQUED supported schools.
NOTE: Percentages are defined as number of students who correctly answered the question divided by the number of students who attempted to answer the question.
[a] Difference between estimate and Orientale estimate is statistically significant at the 0.05 level.
[b] Difference between estimate and Orientale estimate is statistically significant at the 0.01 level.
[c] Difference between estimate and Equateur estimate is statistically significant at the 0.05 level.
[d] Difference between estimate and Equateur estimate is statistically significant at the 0.01 level.
[e] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
f Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.

Among Grade 4 students, as displayed in ***Table 42*** and illustrated in ***Figure 56***, overall, Grade 4 students in Orientale outperformed students in Bandundu on Questions 3 and 5 and students in both Bandundu and Equateur on Question 4 (see ***Table 42*** for specifics on differences). Comparing across treatment groups, Orientale treatment students

outperformed their comparison counterparts on Questions 2 (33.1% versus 13.5%, respectively), 3 (30.7% versus 6.5%, respectively), 4 (60.6% versus 40.8%, respectively), and 5 (28.9% versus 9.4%, respectively) ($p<0.05$). No differences between treatment and comparison groups were observed for either of the other two provinces.

Within the comparison group, Equateur students outperformed students in both other provinces on questions 2, 3, and 5; Orientale students outperformed Bandundu students on Question 4 (see ***Table 42*** for specifics on differences). Within the treatment group, Orientale students outperformed Bandundu students on all questions except Question 1, while they also outperformed Equateur students on Question 4 (see ***Table 42*** for specifics on differences).

Figure 56. Listening Comprehension Among Grade 4 Students by Province and Group: Percentages



Dictation

For the Dictation subtask, examiners read a short sentence to Grade 4 students (“Mon ami s’appelle Bola.”), and students attempted to write the sentence. The three target words that were scored were mon, ami, and appelle. Scores reported for this subtask include the number of words spelled correctly.

Comparisons by Grade and Treatment Condition. ***Table 43*** presents numbers of words spelled correctly and standard errors for Grades 2 and 4 students by comparison and treatment group and by gender, with statistically significant differences bolded. ***Figures 32*** and ***33*** present information graphically by grade and group.

Table 43. Numbers of Words Spelled Correctly Among Grade 4 Students by Group and Gender: Means (Standard Errors)

	Overall		By Gender			
			Comparison		Treatment	
Grade	Comparison	Treatment	Male	Female	Male	Female
4	1.0 (0.1)	1.0 (0.1)	1.1 (0.1)	0.9 (0.1)	1.1 (0.1)	0.9 (0.1)

[a] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.file://rtifile02/cidprojectshares/0211537_(El_Salvador_CVPP)[b] Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.file://rtifile02/cidprojectshares/0211537_(El_Salvador_CVPP)[c] Difference between female estimate and male estimate is statistically significant at the 0.05 level.file://rtifile02/cidprojectshares/0211537_(El_Salvador_CVPP)[d] Difference between female estimate and male estimate is statistically significant at the 0.01 level.

As displayed in ***Table 36***, treatment and comparison students performed the same on the dictation subtask, with students in each group correctly spelling, on average, one word.

Comparisons by Grade, Treatment Condition, and Gender. As indicated in ***Table 43*** and displayed in ***Figure 57***, no statistically significant difference between genders emerged within either group.

Figure 57. Grade 4 Number of Words Spelled Correctly, by Gender

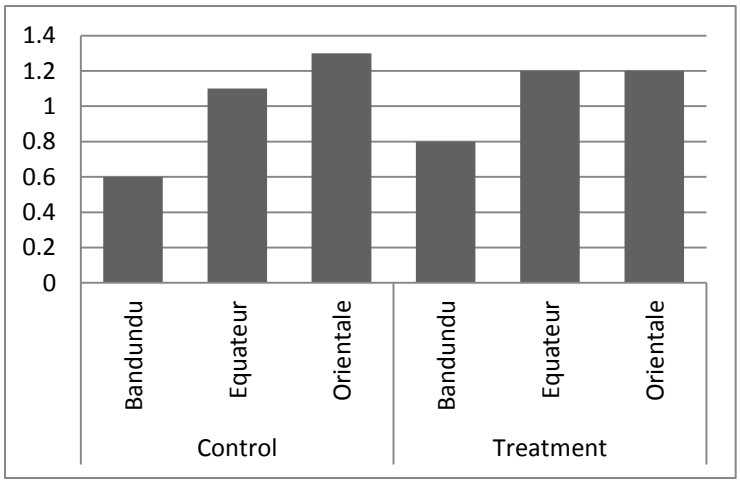
Comparisons by Grade, Treatment Condition, and Province. As indicated in ***Table 44*** and shown in ***Figure 58***, Bandundu students performed statistically significantly worse than their counterparts in the other two provinces (Equateur and Orientale, 1.2; Bandundu, 0.8; $p<0.05$). Comparing across treatment groups, however, this statistical difference disappears.

Table 44. Numbers of Words Spelled Correctly Among Grade 4 Students by Province and Group: Means (Standard Errors)

	Province			Group by Province					
				Comparison			Treatment		
Grade	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale	Bandundu	Equateur	Orientale
4	0.8 (0.1)[da]	1.2 (0.1)	1.2 (0.1)	0.6 (0.1)	1.1 (0.1)	1.3 (0.1)	0.8 (0.1)	1.2 (0.1)	1.2 (0.1)

[a] Difference between estimate and Orientale estimate is statistically significant at the 0.05 level.
[b] Difference between estimate and Orientale estimate is statistically significant at the 0.01 level.
[c] Difference between estimate and Equateur estimate is statistically significant at the 0.05 level.
[d] Difference between estimate and Equateur estimate is statistically significant at the 0.01 level.
[e] Difference between comparison estimate and treatment estimate is statistically significant at the 0.05 level.
f Difference between comparison estimate and treatment estimate is statistically significant at the 0.01 level.

Figure 58. Numbers of Words Spelled Correctly Among Grade 4 Students by Group and Province: Means



Summary of Findings

To be interpreted within the constraints presented in the Introduction of this report, analyses suggest the following.

- As may be expected, overall Grade 4 students tended to produce higher scores on EGRA subtasks than did Grade 2 students, although these differences were not tested for statistical significance.
- Results in either Grade 2 or Grade 4 did not consistently favor treatment-group students, although where statistically significant differences did emerge the general trend was for students in treatment schools to outperform students in comparison schools at the higher levels of performance within a subtask (e.g., treatment students more likely than comparison students to score 5 out of 5, with comparison students more likely to produce lower scores).
- While not a consistent trend, participation in the treatment group at times seems to have eliminated differences between males and females.

- No consistent patterns emerged between or within provinces.

A summary of each subtask's results follows.

Vocabulary

- Grade 2: Treatment-group students appear to have outperformed their comparison peers, for both males and females.
- Grade 4: Treatment-group students appear to have outperformed their comparison peers, for both males and females. Furthermore, while comparison males tended to outperform comparison females, no gender differences were observed within the treatment group.
- No meaningful differences were observed between or within provinces in Grade 2. In Grade 4, Orientale students seem to have outperformed those in other provinces, with Orientale treatment students outperforming their comparison peers.

Initial Sound Identification

- Grade 2: No meaningful trends were observed among males or females.
- Grade 4: Overall, treatment students outperformed comparison students on two of the ten stimuli, although when comparing across genders female treatment students outperformed female comparison students on five stimuli, with male treatment students outperforming male comparison students on one.
- Across provinces, in both Grade 2 and Grade 4, students in Equateur tended to generate higher scores where significant differences were observed. Within each grade, only one significant difference between treatment and comparison groups emerged, in each instance favoring the treatment group.

Letter Sound Knowledge

- Grade 2: No significant percentage differences emerged across treatment and comparison groups. Significant percentage differences favoring males appeared within the comparison group, while no gender differences were observed within the treatment group. Sounds identified per minute, however, showed a treatment effect for females and a leveling of gender effect within the treatment group.
- Grade 4: No significant percentage differences emerged across treatment and comparison groups. Within each treatment group, two significant percentage differences were observed, with the trend to favor males. Looking at sounds identified per minute, while there was no gender effect within the treatment group, comparison males outperformed females.
- Looking at either percentages or letter sounds identified per minute, no significant treatment effects emerged within any of the three provinces for Grade 2 or Grade 4.

Familiar Word Reading

- Grade 4: Considering both percentage and fluency scores, significant differences favoring the comparison group emerged on this subtask; in addition, within the comparison group, a trend emerged favoring males over females, while no such trend was observed within the treatment group.
- Within each of two provinces a significant difference emerged that favored the treatment group; no other significant difference across treatment conditions surfaced.

Invented Word Reading

- Grade 4: Looking at both percentage and fluency scores, there were no significant differences across treatment conditions although, within both groups, gender differences favoring males emerged.
- No significant differences across treatment conditions were observed for any of the three provinces.

Oral Reading Fluency

- Grade 4: Looking at both percentage and fluency scores, there were no significant differences across treatment conditions although, within the comparison group, gender differences favoring males emerged.
- No significant differences across treatment conditions were observed for any of the three provinces.

Reading Comprehension

- Grade 4: No significant differences were observed across or within treatment conditions.
- Looking at results by province, on two questions treatment students outperformed comparison peers: one for Bandundu students and one for Orientale students.

Listening Comprehension

- Grade 2: Treatment students outperformed their comparison peers on three questions. No gender differences emerged within either group.
- Grade 4: Treatment females outperformed comparison females on one question; while within the comparison group males outperformed females on two of the five questions, this occurred on only one question within the treatment group.
- Within Orientale, treatment students outperformed comparison students on three questions. No differences between treatment and comparison groups emerged in the other two provinces.

5. Comparison of Baseline and Midterm Assessments

While the primary intent of this report is to explore differences between treatment and comparison groups at the 2012 midterm assessment, this subsection summarizes comparisons between 2012 midterm student performance and student performance within the reduced 2010 sample (reduced sample baseline) in order to identify possible student gains over time. It is important to note the following constraints when interpreting these results.

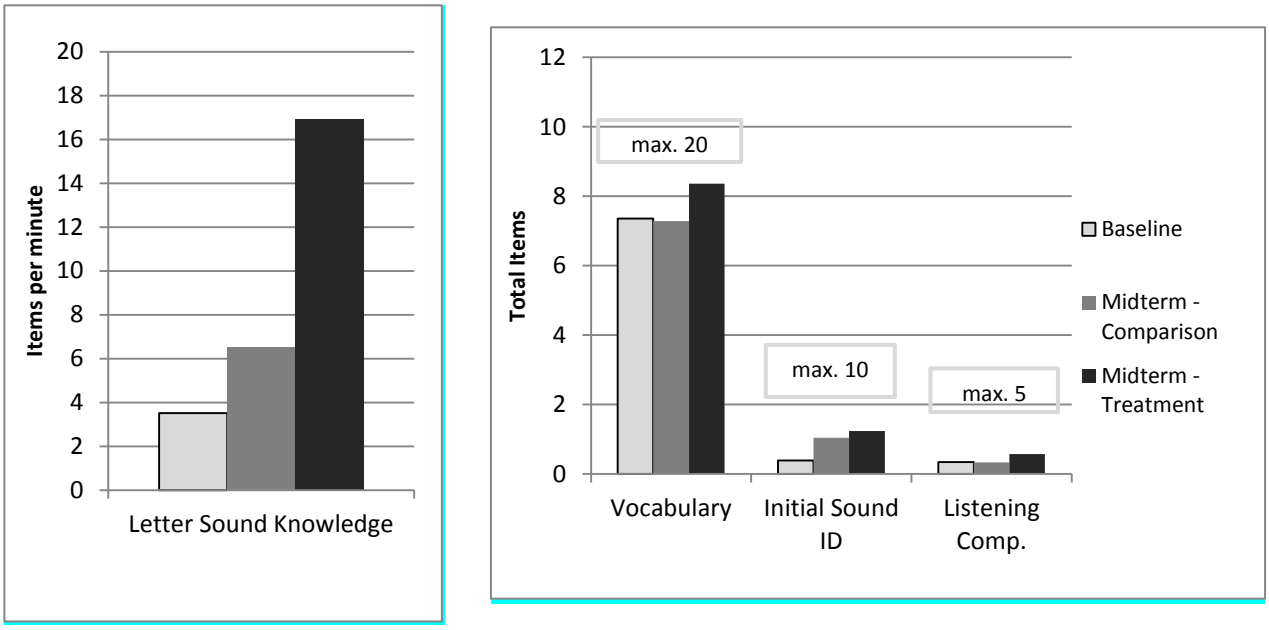
- Because the schools represented in the reduced 2010 sample were selected post hoc based on 2012 sampling criteria, these analyses make no claims regarding the generalizability of results beyond this immediate sample.
- Because individual schools were not tracked from 2010 to 2012, results are not to be interpreted as a cohort study.
- Because schools in neither treatment nor comparison groups had yet received the intervention in 2010, all 16 schools in the reduced 2010 sample have been collapsed into one baseline group.

Students in Grade 2 were administered only four of the nine subtasks whereas students in Grade 4 were administered all nine. Therefore, grade-level results are reported separately.

Program Impact Comparing Baseline Grade 2 and Midterm Grade 2

Student performance on the Grade 2 EGRA subtasks at midterm evaluation, as a function of whether they attended treatment or comparison schools, is compared to reduced sample baseline performance. ***Figure 59*** and ***Table 45*** highlight key trends. Students in treatment schools had significantly higher scores on each of the Grade 2 EGRA measures at midterm than at baseline. Thus, treatment schools yielded higher student performance at midterm in vocabulary knowledge, phonemic awareness, letter sound knowledge, and listening comprehension than did comparable schools at baseline. Comparison schools yielded midterm scores that exceeded those of baseline schools in the two tasks more closely related to beginning reading achievement, initial sound identification, and letter sound knowledge; however, the gains shown by comparison schools were not as large as those shown by treatment schools.

Figure 59. Comparing Grade 2 Baseline and Midterm Assessment Scores for Treatment and Comparison Schools



Note: Separate scales were used for the two parts of ***Figure 59***.

The graph on the left side of ***Figure 59*** shows student performance on the timed task, Letter Sound Identification, and uses clpm as the unit of measurement. The graph on the right shows student performance on the tasks that were untimed, and had a restricted range for possible scores (vocabulary had 20 items, initial sound identification had 10 items, and the listening comprehension subtask had 5 questions).

Table 45. Program Impact, Baseline and Midterm Assessments, for Grade 2

Item	School Type	Baseline, Grade 2	Midterm Assessment, Grade 2	Program Impact		
		Mean	Mean	Gains Over Baseline	Increase Over Comparison	Percent Increase Over Baseline
Vocabulary	Comparison	7.36	7.28	-0.08		-1.08%
	Treatment		8.36	1.00*	1.08	13.59%
Initial Sound Identification	Comparison	0.39	1.04	0.65*		166.67%
	Treatment		1.24	0.85*	0.20	217.95%
Letter Sound Knowledge	Comparison	3.51	6.53	3.02*		86.03%
	Treatment		7.65	4.14*	1.12	117.95%
Listening Comprehension	Comparison	0.34	0.33	-0.01		2.94%
	Treatment		0.57	0.23*	0.24	70.59%

"Gains over baseline" indicates the difference in scores between baseline and midterm assessment as an absolute difference
"Increase over comparison," shows the difference in the gains between baseline and midterm assessment for treatment schools, less the gains between baseline and midterm assessment for the comparison schools
"Percent increase over baseline," converts the "gains over baseline" column to a percentage increase against the baseline scores
Note: Only those items marked with an asterisk (*) are statistically significant. SD = standard deviation

Table 48 suggests that, among the schools represented by these samples, the PAQUED program led to gains in Grade 2 students' performance in terms of their oral language proficiency in French and their knowledge of the letter sounds. Examination of the ~~"p~~ercent increase over baseline" column reveals gains over baseline for treatment schools in phonemic awareness (218%) and in letter sound knowledge (118%), moderate gains in listening comprehension (71%), and small gains in vocabulary knowledge (14%). Gains over baseline were smaller for comparison schools, and limited to initial sound identification (167%) and letter sound knowledge (86%).

Program Impact Comparing Baseline Grade 4 and Midterm Grade 4

Student performance on the Grade 4 EGRA subtasks at baseline and midterm evaluation as a function of whether they attended treatment or comparison schools is presented in ***Table 46*** and graphically in ***Figure 60***. Compared to baseline, students in treatment schools had higher scores on the Grade 4 EGRA measures that assessed oral language skills in French (vocabulary knowledge, initial sound identification, and listening comprehension) and in letter sound identification. In contrast, comparison schools did not show gains in midterm scores on any of the EGRA tasks, with the exception of letter

sound identification. Students in both treatment and comparison schools had lower dictation scores than at baseline.

Table 46. Program Impact, Baseline and Midterm Assessments, for Grade 4

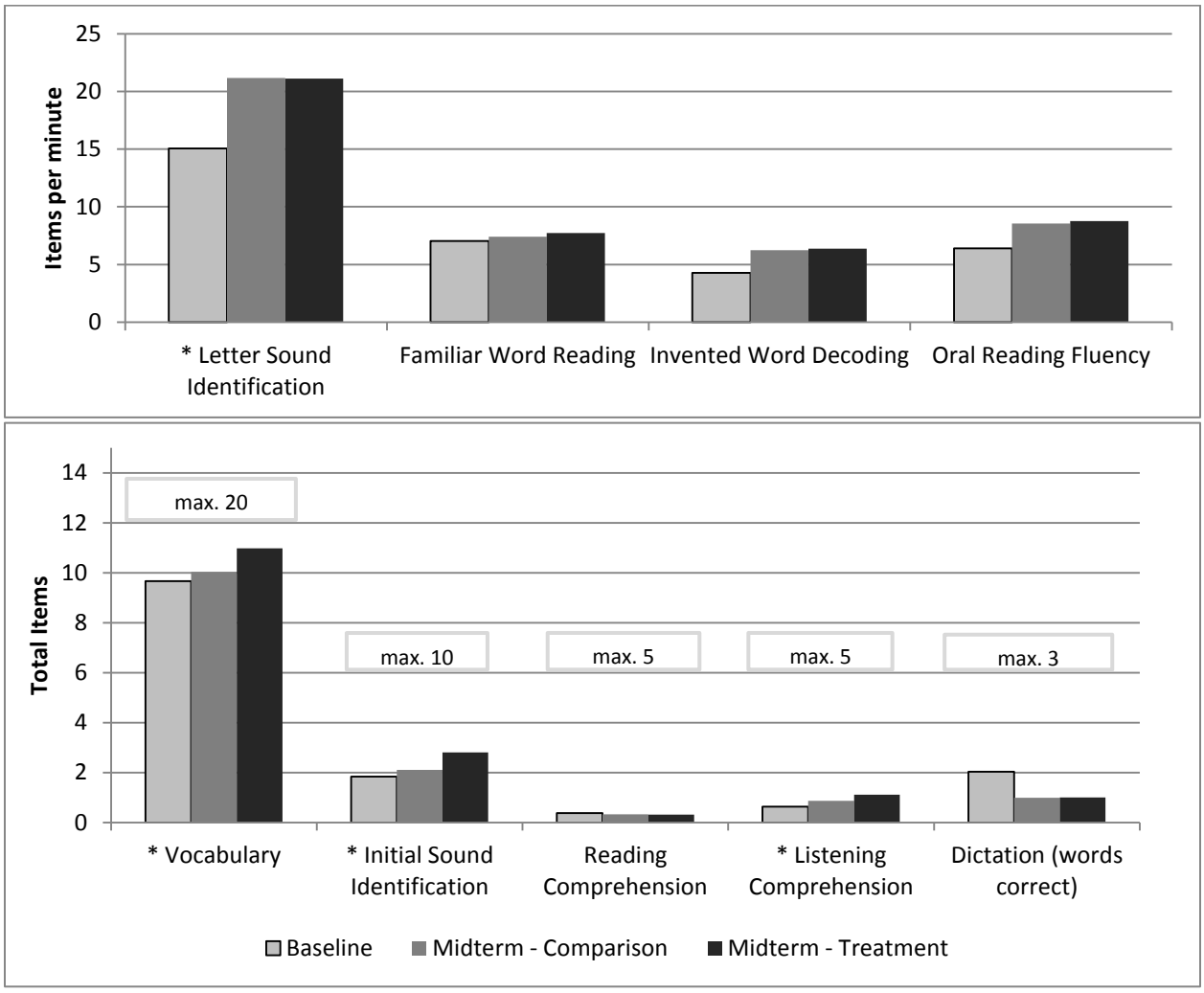
Item	School Type	Baseline, Grade 4	Midterm Assessment, Grade 4	Program Impact		
		Mean	Mean	Gains Over Baseline	Increase Over Comparison	Percent Increase Over Baseline
Vocabulary	Comparison	9.67	10.03	0.36		3.72%
	Treatment		10.98	1.31*	0.95	13.55%
Initial Sound Identification	Comparison	1.84	2.11	0.27		14.67%
	Treatment		2.81	0.97*	0.70	52.72%
Letter Sound Knowledge	Comparison	15.07	21.16	6.09*		40.41%
	Treatment		21.12	6.05*	-0.04	40.14%
Familiar Word Reading	Comparison	7.03	7.42	0.39		5.54%
	Treatment		7.73	0.70	0.31	9.96%
Invented Word Decoding	Comparison	4.27	6.23	1.96		45.90%
	Treatment		6.37	2.10	0.14	49.18%
Oral Reading Fluency	Comparison	6.41	8.56	2.31		36.04%
	Treatment		8.76	2.35	0.04	36.66%
Reading Comprehension	Comparison	0.37	0.33	-0.04		-10.81%
	Treatment		0.31	-0.06	-0.02	-16.22%
Listening Comprehension	Comparison	0.64	0.87	0.23		35.94%
	Treatment		1.11	0.47*	0.24	73.44%
Dictation	Comparison	2.04	0.99	-1.05*		-51.47%
	Treatment		1.00	-1.04*	0.01	-50.98%

"Gains over baseline" indicates the difference in scores between baseline and midterm assessment as an absolute difference
"Increase over comparison," shows the difference in the gains between baseline and midterm assessment for treatment schools, less the gains between baseline and midterm assessment for the comparison schools
"Percent increase over baseline," converts the "gains over baseline" column to a percentage increase against the baseline scores
Note: Only those items marked with an asterisk (*) are statistically significant. SD = standard deviation

Table 46 shows that the PAQUED program led to gains in Grade 4 students' performance in terms of their oral language proficiency in French and their knowledge of the letter

sounds, but not in their reading skills. Examination of the "percent increase over baseline" column revealed that students in both treatment and comparison schools showed moderate gains over baseline in letter sound knowledge (40%). However, only students in treatment schools showed significant gains over baseline in phonemic awareness (53%), listening comprehension (73%), and vocabulary knowledge (14%). However, neither treatment nor comparison schools showed significant gains in students' decoding of invented words, reading of real words in lists and passages, and reading comprehension. Further, both treatment and comparison schools yielded dictation scores that had declined by 51% from baseline.

Figure 60. Comparing Grade 4 Baseline and Midterm Assessment Scores for Treatment and Comparison Schools



Note: Separate scales were used for the two parts of ***Figure 60***. Only those items marked with an asterisk (*) are statistically significant.

The graph on the top portion of ***Figure 60*** shows student performance on the timed tasks, and uses items/minute as the unit of measurement. The graph at the bottom shows student

performance on the tasks that were untimed, and had a restricted range for possible scores (vocabulary had 20 items, initial sound identification had 10 items, the reading comprehension and listening comprehension subtasks had 5 questions, and 3 words were given in dictation).

6. Conclusions and Recommendations

Interpretation of these midterm results must take into consideration key logistic and sampling constraints. Numerous technological and logistical problems delayed and impeded implementation of planned interventions, with the result that very few schools received the planned complement of services and support. In addition, the assessment's sampling frame was modified to focus on those schools that, by virtue of being ―accessible" (defined as being within approximately 20 km of an urban center), might be more likely to have received the intervention as planned and might, therefore, present a more accurate portrait of the intervention's potential impact. As a result, schools could no longer be tracked longitudinally from 2010 through 2012, and the sampling methodology transitioned from a longitudinal comparison-led study to a cross-section sampling design. While the findings of the baseline report were initially intended to serve as a point of comparison against which midterm and final results could be compared, because it was determined that it would not be possible to maintain the same set of schools, the midterm sample was drawn from a revised list of participating schools and potential comparison schools. This, in turn, reduced the ability to make statistically valid comparisons between midterm and baseline, reducing the confidence with which we can isolate and draw conclusions about program impacts on student performance. (See Sampling Methodology for a fuller discussion of the approach and its implications for analysis.) Finally, as reliable information regarding fidelity of implementation is unavailable, the extent to which the assessment accurately reflects the impact of the intervention, itself, cannot be quantified.

Reporting Constraints

As indicated earlier in this report, the intent of the midterm assessment was to identify changes in indicators of reading performance between the 2010 and 2012 assessment periods and, if changes were identified, to determine whether they were attributable to PAQUED program intervention. However, when the revised accessibility criterion for inclusion for the 2012 sample was applied to the schools assessed in 2010 (in order to find schools suitable for a 2010–2012 longitudinal comparison), only 16 schools (13 treatment and 3 comparison) and 417 students (339 treatment and 78 comparison) from the 2010 sample (―reduced 2010 sample") qualified for inclusion. Given this dramatically reduced 2010 sample size, EDC and RTI opted to constrain the scope of this report to comparisons between treatment and comparison schools within the 2012 sample, which are the most meaningful analyses available and the ones with the greatest statistical

power. Any comparisons between the 2010 schools (16 in total) that meet the 2012 sampling criteria and the 2012 sample that are deemed by EDC and RTI to be appropriate will be addressed in a separate report.

Overall Conclusions

The results of the midterm assessment suggest that the challenges involved in the implementation of the PAQUED program may have limited effectiveness in increasing students' early reading achievement.

Overall, student performance on many of the EGRA subtasks was higher at midterm than it had been at baseline and stronger among treatment than comparison groups. Among Grade 2 students, the midterm scores were up to 167% higher than at baseline in comparison schools and were between 14% and 218% higher at midterm in treatment schools. Similarly, Grade 4 students' scores at midterm were higher than baseline in measures of oral language (vocabulary, initial sound identification), and in letter sound knowledge. However, there were no overall gains in Grade 4 students' scores in reading measures (such as oral reading fluency or comprehension). It is also important to note that despite these gains, students at midterm continued to have poor reading skills, regardless of group (treatment or comparison) or gender. For example, at midterm, Grade 4 students in both treatment and comparison schools read passages with very limited fluency (on average, fewer than nine correct words per minute) and with very limited comprehension on average.

However, results suggest that students in schools receiving the PAQUED program may have improved their proficiency in French. More specifically, students in treatment schools tended to perform better than their comparison school peers in French vocabulary, initial sound identification in Grade 4, and the Grade 2 listening comprehension scores. Further, students at both grade levels in treatment schools showed stronger performance in vocabulary, phonemic awareness, and listening comprehension when compared to their peers in comparison schools. These findings are encouraging, as they suggest the potential of the PAQUED program to promote students' proficiency in French, a critical foundation for skilled reading.

Recommendations

Recommendations, by subtask, are presented below, with the goal of suggesting potential areas of improvement for PAQUED interventions targeting reading.

1. Whereas the midterm assessments suggest that the PAQUED intervention improved students' vocabulary knowledge, overall students' understanding of French vocabulary is weak. Additional interventions may be implemented to support vocabulary acquisition in French. These should occur early in the primary school cycle to ensure that students understand French and can transition to learning to read in French.

2. Although the PAQUED program led to moderate gains in initial sound identification, students in Grades 2 and 4 showed very limited phonological awareness. Phonological 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. Teachers should explicitly teach phonological awareness by drawing students' attention first to more salient sounds, such as individual words within sentences, progressing to rhymes or syllables, then moving to more challenging sounds, such as the initial and final letter sounds in words.
3. Scores for invented words were also low. These results suggest that many students struggle to decode words, a key reading skill. Higher scores on familiar word reading in comparison with invented word reading suggest that some students may be relying on memorization of sight words, rather than decoding words to read. This is a characteristic of struggling readers and indicates possible deficiencies in their mastery of the alphabetic principle (that speech sounds map onto print) as well as a lack of word-attack skills. Explicit instruction in how students can and should begin to identify words that they do not know is clearly necessary.
4. Students in Grade 4 were largely unable to read text accurately, quickly, and with the correct expression, or with automaticity. The combination of these factors is called fluency, a skill that is essential to support reading comprehension. It may be difficult to ensure reading fluency if students have poor mastery of the letter sounds, weak word attack skills, and limited knowledge of French vocabulary. At the same time, teachers may begin promoting oral reading fluency from the first day of school by modeling fluent reading through oral read-alouds. As students begin to read, a key teaching strategy that promotes oral reading fluency is frequent guided oral reading performed by the students themselves.
5. Students exhibited very little reading comprehension. Clearly, a major barrier to reading comprehension is the low level of reading skills among the students sampled. Focusing on developing more basic early reading skills such as letter sound knowledge, phonological awareness, and fluency will promote reading comprehension. These results suggest that teachers could also teach complementary strategies to support overall comprehension, such as the characteristics of narratives, common vocabulary in stories, and ways to make a prediction. An important method for teaching comprehension is for teachers to build student knowledge before reading a story and consolidate that knowledge after reading a story (regardless of whether it is the student or teacher reading the story). At the same time, student listening comprehension—which does not require reading—was low. The low listening comprehension scores, coupled with the low vocabulary scores, indicate that students lack the proficiency in French required to read grade-level texts.
6. Students' writing skills were poor. When asked to write a dictation in the language of instruction, most students were unable to write a single word. If students are unable to identify letters, writing words will be challenging. Engaging students in letter-writing activities can foster letter identification and phonological awareness skills as well as provide the opportunity to practice

writing. Building phonological awareness will give students the ability to attempt to write words based on their understanding of letter-sound relationships, rather than relying on memorization of specific words.

Annex 1: Percent Attempted and Mean Scores by Group and Gender and by Group and Province

Table 1-1. EGRA Subtasks: Percent Attempted and Mean Scores by Group and Gender

Grade	Group	Male		Female	
		Percent Attempted	Mean	Percent Attempted	Mean
Total Vocabulary					
2	Comparison	36.4%	7.28	36.4%	7.28
	Treatment	41.6%	8.32	42.0%	8.40
4	Comparison	53.2%	10.64	47.2%	9.43
	Treatment	55.5%	11.10	54.2%	10.85
Initial Sound Identification					
2	Comparison	10.1%	0.97	11.8%	1.10
	Treatment	13.1%	1.28	12.3%	1.21
4	Comparison	23.1%	2.31	19.4%	1.91
	Treatment	28.3%	2.879	28.4%	2.82
Letter Sound Knowledge					
2	Comparison	31.7%	7.48	25.2%	5.61
	Treatment	33.4%	7.87	31.3%	7.43
4	Comparison	64.1%	23.92	55.5%	18.39
	Treatment	61.6%	22.34	58.5%	19.77
Familiar Word Reading					
4	Comparison	37.7%	9.71	25.5%	5.83
	Treatment	34.7%	9.05	29.6%	7.63
Invented Word Reading					
4	Comparison	32.3%	7.98	21.5%	4.48

Grade	Group	Male Percent Attempted	Male Mean	Female Percent Attempted	Female Mean
	Treatment	28.9%	7.29	21.2%	5.36
Oral Word Reading					
4	Comparison	33.2%	10.5	23.5%	6.6
	Treatment	30.6%	11.47	25.3%	7.95
Reading Comprehension					
4	Comparison	12.5%	0.44	7.7%	0.22
	Treatment	8.5%	0.32	8.2%	0.31
Listening Comprehension					
2	Comparison	7.7%	0.38	5.5%	0.27
	Treatment	11.5%	0.58	11.2%	0.56
4	Comparison	20.5%	1.03	14.4%	0.72
	Treatment	23.5%	1.17	20.6%	1.03

Table 1-2. EGRA Subtasks: Percent Attempted and Mean Scores by Group and Province

Grade	Group	Bandundu Percent Attempted	Bandundu Mean	Equateur Percent Attempted	Equateur Mean	Orientale Percent Attempted	Orientale Mean
Total Vocabulary							
2	Comparison	34.5%	6.90	43.8%	8.76	33.7%	6.73
	Treatment	38.7%	7.74	43.6%	8.72	46.0%	9.21
4	Comparison	46.8%	9.36	56.2%	11.24	50.0%	10.00
	Treatment	53.6%	10.71	51.7%	10.35	62.2%	12.44
Initial Sound Identification							
2	Comparison	11.9%	1.03	17.1%	1.66	7.0%	0.70
	Treatment	11.0%	1.09	17.1%	1.66	10.6%	1.03
4	Comparison	20.0%	2.00	35.0%	3.43	15.0%	1.50
	Treatment	28.3%	2.81	35.7%	3.54	18.9%	1.81
Letter Sound Knowledge							
2	Comparison	27.2%	6.43	27.2%	6.59	29.9%	6.61
	Treatment	33.6%	8.72	28.2%	6.09	35.1%	7.39
4	Comparison	53.5%	17.42	64.9%	24.71	62.8%	22.60
	Treatment	57.3%	19.93	60.0%	20.25	67.0%	25.06
Familiar Word Reading							
4	Comparison	19.0%	4.19	34.9%	8.67	41.2%	10.52
	Treatment	27.6%	6.70	33.5%	8.40	42.8%	12.41
Invented Word Reading							
4	Comparison	15.5%	3.33	32.2%	7.48	34.5%	8.17
	Treatment	21.1%	5.17	25.2%	6.32	35.1%	9.31
Oral Word Reading							
4	Comparison	13.7%	3.2	35.5%	11.9	37.6%	11.6
	Treatment	22.5%	6.6	27.7%	8.3	42.1%	14.7

Grade	Group	Bandundu Percent Attempted	Bandundu Mean	Equateur Percent Attempted	Equateur Mean	Orientale Percent Attempted	Orientale Mean
Reading Comprehension							
4	Comparison	7.4%	0.13	15.4%	0.63	9.4%	0.34
	Treatment	4.6%	0.17	10.0%	0.35	14.7%	0.59
Listening Comprehension							
2	Comparison	7.3%	0.37	10.1%	0.49	4.3%	0.21
	Treatment	8.1%	0.41	14.5%	0.72	14.2%	0.71
4	Comparison	12.3%	0.61	29.8%	1.49	15.5%	0.77
	Treatment	17.0%	0.85	21.8%	1.09	34.9%	1.75