Differences in Reading and Mathematics Comprehension Between Students in Paraguay Educa’s “Phase I” OLPC Deployment and Students in Nearby Schools

Morgan G. Ames
Stanford University
October 3, 2012

Summary

This document presents the results of a between-subjects comparison of the exam scores of third- and sixth-grade students in the nine public schools part of Paraguay Educa’s “Phase I” laptop program, which provided OLPC-designed “XO-1” laptops to all students and teachers in grades one through six in May 2009, and third- and sixth-grade students in public schools not in “Phase I.” This research also provides a benchmark measurement for students in “Phase II” of the program who received laptops in May 2011 for a follow-up test in November 2013.

Exam Development and Administration

The education division of Paraguay Educa and the author collaborated in developing two multiple-choice tests of 30 questions each, one for third-grade students and one for sixth-grade students. We selected third and sixth grades as suitable benchmarks in the Paraguayan schooling system: by the end of third grade, all students are expected to read Spanish, and at the end of sixth they are promoted to junior high (and out of the official laptop program, though there have been efforts to train seventh-grade teachers). Roughly one third of the questions related to reading comprehension –

---

1 Citation: Ames, Morgan G. (2012). Differences in Reading and Mathematics Comprehension Between Students in Paraguay Educa’s “Phase I” OLPC Deployment and Students in Nearby Schools. Working paper v.2012.10.3, Department of Communication, Stanford University.
which consisted of one fiction passage and one nonfiction passage – and two thirds related to mathematical reasoning (equal parts numeracy, patterns/algorithms, and spatial reasoning). These tests were modeled after cognitive reasoning tests for students two grade-levels lower in the United States\(^2\) and after the 2006 SNEPE, a standardized test previously used in Paraguay and elsewhere in the world. Questions were developed to specifically not depend on any specific curriculum (e.g. words or concepts taught in a particular grade), but on more general cognitive reasoning.

We validated the third-grade test once and the sixth-grade test twice, with the help of a Paraguayan education evaluation expert, Dr. Oscar Serafin. We administered both tests to pilot groups of third-grade and sixth-grade students, and Dr. Serafin conducted a validation of the responses in the following areas:

1) Content validity
2) Reliability
3) Analysis questions:
   a) Skill Level
   b) Presentation
   c) Plausibility of distractors
   d) Discriminatory power
   e) Appropriateness to the target group

At the end of the validation, each test had a Chronbach’s Alpha score of over .7, which is the minimum threshold for reliability of the instrument.

Paraguay Educa’s teacher trainers and other employees administered the test to 2085 students in 52 schools between November 22 and 29, 2010, at the end of the 2010 school year in Paraguay. There were three groups of students who took the test: 576 students in Phase I schools who have had laptops since 2009 (treatment), 787 students in

\(^2\) We chose to model on tests two grade-levels lower because that is the level at which Paraguayan students have historically performed on standardized tests such as the SNEPE compared to students in the United States.
Phase II schools who did not have laptops at the time of the test but received them in May 2011 (for this analysis, part of the control), and 722 students in the nearby towns of Atyra and San Bernardino which have no plans for joining the laptop program (control). In addition to providing a comparison of results to date, this test provides a solid baseline measure for a follow-up test with third- and sixth-grade students at the same schools at the end of 2013, five years into laptop use for Phase I schools and three years in for Phase II schools, at which point this latter group will be “treatment 2.” Paraguay Educa’s employees entered and cross-validated the 2085 tests.

Exam Results

The author conducted a Welch two-sample T-test\(^3\) to evaluate whether the mean exam scores between the treatment group and the control group have a statistically significant difference. These results are broken down by subject (language and mathematics) and by grade (third, sixth, and combined). The author used the statistical analysis package R to conduct the analysis and to generate density plot graphs to visually observe the difference in means and variances (see Appendix 1). Table 1 summarizes the results for each subject and grade.

The author notes that these are preliminary results and acknowledges the many issues of between-group experimental design. However, at this point only between-subjects analysis is possible; after the follow-up test in 2013 the author will conduct a within-groups comparison. To reduce the differences between the groups, the author

---

\(^3\)The author chose this method of analysis because it provides a correction for when the samples in question are not guaranteed to have equal variances (when, for example, the analysis is done between-groups).
restricted the analysis to public schools, as public schools are funded and provisioned similarly across the sample.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Treatment mean</th>
<th>Control mean</th>
<th>Difference in means</th>
<th>p-value (significant at p&lt;0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd grade reading</td>
<td>55.128%</td>
<td>49.386%</td>
<td>5.743%</td>
<td>0.000601</td>
</tr>
<tr>
<td>3rd grade math</td>
<td>50.154%</td>
<td>42.921%</td>
<td>7.233%</td>
<td>0.0000002214</td>
</tr>
<tr>
<td>3rd grade overall</td>
<td>52.431%</td>
<td>45.648%</td>
<td>6.783%</td>
<td>0.0000002404</td>
</tr>
<tr>
<td>6th grade reading</td>
<td>55.533%</td>
<td>51.230%</td>
<td>4.302%</td>
<td>0.003086</td>
</tr>
<tr>
<td>6th grade math</td>
<td>49.101%</td>
<td>48.891%</td>
<td>0.210%</td>
<td>0.8954</td>
</tr>
<tr>
<td>6th grade overall</td>
<td>51.674%</td>
<td>49.827%</td>
<td>1.847%</td>
<td>0.1741</td>
</tr>
</tbody>
</table>

Table 1. Summary of results and significance for all experimental conditions. Statistically significant differences are shown in bold.

The results of the third-grade test are most promising in showing a payoff of the sustained efforts of Paraguay Educa. The 288 third-grade students in the treatment group had scores that were on average higher than the 507 students in the control group in both reading (5.743% higher than the control, p<0.05) and mathematics (7.233% higher than the control, p<0.05). Sixth-grade test results are more mixed. The 241 sixth-grade students in the treatment group (Phase I of the laptop program) had scores that were higher than the 481 students in the control group in reading (4.302% higher than the control, p<0.05), but not mathematics. Because of the lack of difference in mathematics, the difference in the mean scores of reading and mathematics combined are also not statistically significant for the sixth grade.

Overall, these findings, while preliminary, suggest that there are incremental but statistically significant benefits to being part of Paraguay Educa’s OLPC laptop program, as measured by a standardized cognitive exam. Benefits are more marked for the third grade and for reading.
Appendix 1: Exam Result Density Plots

Figures 1 and 2: density plots of reading scores for third- and sixth-grade test takers. Scores for the treatment group (green dotted line) clearly trend higher than the control group (red solid line) in both graphs.

Figures 3 and 4: density plots of mathematics scores for third- and sixth-grade test takers. Scores for the treatment group (green dotted line) clearly trend higher than the control group (red solid line) in the third-grade graph, but not sixth-grade.

Figures 5 and 6: density plots of combined reading and math scores for third- and sixth-grade test takers. The treatment group is a green dotted line, and the control group a red solid line.