USAID
FROM THE AMERICAN PEOPLE

Data-Driven Instruction in Honduras: An Impact Evaluation of the EducAcción Promising Reading Intervention

Baseline Report

August 2016
This publication was produced for review by the United States Agency for International Development. It was prepared by Steve Glazerman, Sarah Humpage Liuzzi, Nancy Murray, and Chantal Toledo of Mathematica Policy Research.

This page has been left blank for double-sided copying.

Data-Driven Instruction in Honduras: An Impact Evaluation of the EducAcción Promising Reading Intervention

Baseline Report

This page has been left blank for double-sided copying.

ACKNOWLEDGMENTS

The authors gratefully acknowledge numerous people and organizations that contributed to the implementation of all evaluation activities in Honduras and to the production of this report.

The Honduran Ministry of Education, Evaluation Division staff Dennis Cáceres and Claresy Banegas; and Emiliana Tejeda, department director for La Paz shared invaluable insights on evaluation in Honduras. The State Departments of Education in Lempira, Santa Barbara, Francisco Morazán and Atlántida all provided essential support for the evaluation.

Barbara Knox-Seith, senior education advisor, and Michael Lisman, education advisor, Bureau for Latin America and the Caribbean (U.S. Agency for International Development [USAID]/Washington), and Kay Leherr and Renan Rápalo (USAID/Honduras) have worked closely with us to provide important inputs into our thinking on the evaluation design and have supported its implementation.

American Institutes of Research (AIR) home office project manager Bridget Drury, research associate Mariela Goett, and research assistant Gustavo Castillo have provided key information on evaluation activities in Honduras as well as the activities carried out by EducAcción. At EducAcción, project director Edwin Moya, pedagogical expert Ligia Águilar and monitoring and evaluation specialist Delfina Flores have provided ongoing information and insights into the implementation of the EducAcción Promising Reading Intervention (EducAcción-PRI) and coordinated closely with the evaluation. Russbel Hernández of *Mejorando el Impacto al Desempeño Estudiantil de Honduras* (MIDEH) has also generously provided information on assessment activities in Honduras and MIDEH's role in that work.

We thank our local data collection partner, Espirálica, and its staff for high quality data collection in the study schools. César Valenzuela provided overall project management support in Honduras, oversaw data entry, and worked with the Mathematica team to spot potential risks to the evaluation's experimental design and identify solutions. Daniela Cruz, Leonardo Álvarez, and Lia Banegas provided excellent training, feedback on data collection instruments, and oversight of the data collection process.

At Mathematica, Alena Davidoff-Gore, Irina Cheban, Jeremy Page, Kate Place, Virginia Poggio, Javier Rodriguez, and Jonathan McCay provided critical support in data cleaning, programming, analysis, and presentation.

We are enormously appreciative of all the school principals and teachers for opening the doors of their classrooms, speaking with us, and allowing us to observe the valuable work they do with students.

This page has been left blank for double-sided copying.

CONTENTS

This page has been left blank for double-sided copying.

TABLES

FIGURES

This page has been left blank for double-sided copying.

I. INTRODUCTION

A. The Role of Assessment in Improving Education

In the last two decades, the Latin America and Caribbean (LAC) region has improved educational opportunities for many of its students. During the period from 1999 to 2012, in most LAC countries, net primary school enrollment rates increased, gender parity was achieved in many LAC countries, rates of grade repetition declined, and rates of primary school completion improved (United Nations Educational, Scientific and Cultural Organization [UNESCO] 2015). However, student achievement in the region continues to lag. In the 2012 Programme for International Student Assessment (PISA) exam, the eight participating Latin American countries scored in the bottom 30 percent of the world in reading and the bottom 25 percent in math (Bos et al. 2013). Student reading and math outcomes improved from 2006 to 2013, according to the UNESCO-administered Third Regional Comparative and Explanatory Study (TERCE), but the majority of third- and sixth-grade students across the 15 participating countries were still performing at basic levels in both reading and math (Latin American Laboratory for Assessment of the Quality of Education [LLECE] 2014; LLECE 2015).

As policymakers' focus in developing countries has shifted from *access* to education *quality*, student assessment has come to the forefront (Clarke 2012; Vegas and Petrow 2008). Some developing countries in LAC and in other regions have begun to use student assessments more systematically in an attempt to identify and address gaps in student knowledge and educational progress.

Honduras has recently been at the forefront among its peers in the use of educational assessment. The *Mejorando el Impacto al Desempeño Estudiantil* (MIDEH) project, a partnership between the U.S. Agency for International Development (USAID), the American Institutes for Research (AIR), and the Honduran Ministry of Education (MOE), is an example of the type of work begun there in recent years. MIDEH has supported the development of a new curriculum and a national testing system linked to that curriculum. This system includes monthly teacher-administered formative assessments as well as summative end-of-grade assessments, which have been applied to a national sample of schools most years since 2007 and, beginning in 2012, applied to all students in all schools in grades 1 through 11 as part of an education census. Formative assessments are designed to be used by teachers as feedback to adjust instruction. Summative assessments are designed to reach judgments about the performance of students, teachers, or schools at the end of an instructional period.

In spite of these investments and advancements in the area of assessment, hard evidence is needed to support policy. The impact of the formative or summative assessments being implemented in Honduras has never been formally evaluated, even though some evidence, reviewed below, suggests that assessment is the basis for promising educational interventions. Furthermore, the availability of testing materials such as booklets, and the number of teachers who have recently received training have both dropped off in recent years. Policymakers must decide whether to renew funding to keep these supports in place.

Despite policymakers' need for high-quality information about the use of formative and summative assessments, the evidence base for interventions that use student assessment is scant in developing countries, especially in the LAC region. However, some studies from developed and developing countries have demonstrated potential effects of formative assessments, in which teachers use frequent assessments of their students to improve their teaching, and end-of-grade summative assessments, which, like formative assessments, can be the basis for teachers and principals to adapt their teaching practices and curricula to better meet students' needs. Furthermore, end-of-grade assessments can inform schools and policymakers about schools' yearly progress and can be used for accountability or resource allocation at the school, regional, or national level.

Several systematic reviews provide evidence on the potential effectiveness of formative assessment, and the characteristics that make it most effective. Black and William (1998) reviewed more than 250 studies of interventions loosely defined as formative assessment in the primary, secondary, and college levels in the U.S. and other developed countries. They found that these activities led to student gains, on average, of 0.4 to 0.7 standard deviations, which is equivalent to moving a student from the 50th percentile to somewhere between the 66th and 76th percentiles. More recently, Kingston and Nash (2011) reviewed the literature on formative assessment in developed countries, restricting their analysis to studies that met their criteria of inclusion: studies of interventions that clearly fit the definition of formative assessment, had used a comparison group design, were from an academic K-12 setting, included statistics to calculate effect sizes, and were published in 1988 or later. The authors found a smaller average effect size of 0.2 standard deviations, or moving a student from the 50th percentile to the 58th. However, they also noted the lack of rigorous evidence, citing only 13 studies that met their criteria for inclusion. Kingston and Nash ended their paper with a call for further research to identify what characteristics make formative assessment most effective. In a recent review of the evidence on formative assessment in Africa, Perry (2013) found that the evidence is promising but quite limited. Two recent experimental evaluations in Liberia (Piper and Korda 2010) and South Africa (Piper 2009) showed that formative assessment can have large effects when paired with a detailed curriculum. In another experimental study, Muralidharan and Sundararaman (2010) found no effect on student test scores when primary school teachers in India received feedback on their students' performance and were observed periodically, but the teachers did not receive a detailed curriculum to use along with the test results. Teachers appeared to put forth more effort, but student test scores did not improve. This combination of results suggests that the activities that follow an assessment may be an important factor in the assessment's potential effect on learning.

A growing body of literature also suggests that end-of-grade assessment data could be important in efforts to improve learning outcomes, but little evidence exists on how teachers can use the results to improve their teaching. Evidence shows that summative evaluations may improve learning through an "accountability effect," even if no explicit consequences are attached to test results (low-stakes testing). Piper and Korda (2010) found that simply sharing testing data with families and schools in Liberia led to significant growth in test scores (although the effects were much smaller than when testing data were combined with giving teachers a detailed curriculum). Andrabi et al. (2015) found that disseminating test scores to families through school report cards in Pakistan led to better school quality, lower fees, and a rise in test

scores of 0.1 standard deviations, or moving a student from the 50th percentile to the 54th. In the U.S., Hanushek and Raymond (2004) found that reporting schools' test scores improved learning outcomes whether or not consequences were attached to the scores. End-of-grade assessment data could also be useful to policymakers, ministries of education, and schools in their decision making. For example, end-of-grade testing data may help schools assign students to classes at the appropriate level (Coe 1998). At a higher level, these data can also help ministries of education direct resources where they are likely to do the most good (Ravela et al. 2008; Clarke 2012).

To address the need for high-quality evidence on interventions that improve student outcomes, particularly in reading, USAID contracted with MATHEMATICA to design and conduct impact evaluations, cost-effectiveness analyses, and implementation studies of promising reading interventions and education-access interventions in the LAC region. Evaluations are underway in Guatemala, Peru, Nicaragua, and Honduras. The current study in Honduras examines the impact of formative and end-of-grade student assessment interventions on student and teacher outcomes and analyzes their implementation and cost-effectiveness. This report focuses on the baseline results from the teacher and principal surveys.

B. Study Design

The purpose of this study is to estimate the impacts of providing print materials and pedagogical support for the use of assessments on teaching and learning. Specific outcomes of interest include teachers' use of assessments and or results of assessments in instruction and students' early grade reading skills. The study aims to answer two primary research questions:

1. What impact does providing feedback on end-of-grade test results and pedagogical support for teachers have on student learning, relative to usual practice?

2. What added impact does providing formative assessment materials and pedagogical support for teachers have on student learning?

The study design is a three-arm randomized controlled trial. We randomly assigned schools to one of the three approaches, described below, that vary the use of end-of-grade results and formative assessment practices. Random assignment ensures that the evaluation will provide unbiased estimates of the impact of the interventions on the outcomes of interest. When implemented correctly, random assignment ensures that there are no systematic differences between the research groups other than access to the intervention. Therefore, any differences in average outcomes between the intervention groups observed after the intervention that are too large to be the result of chance can be attributed to the differences between the interventions, not to other characteristics of students, teachers, or schools in the sample. EducAcción, a project funded by USAID/Honduras and implemented by AIR and partners, will implement the interventions we are evaluating. EducAcción has been working in Honduran schools to support the schools' use of information to improve teaching.

We collected baseline data from an initial sample of 240 potentially eligible schools selected on the basis of geography and test score data, as described below. One hundred percent of the schools selected for baseline consented to be in the study. These schools were those with the lowest average test scores in the departments of Lempira and Santa Barbara and the cities of La

Ceiba and Tegucigalpa.[1] Sixty schools were determined ineligible and eliminated from the study sample. We considered these schools ineligible for one of three reasons: having fewer than 10 students enrolled in the first grade, already working with EducAcción (the implementer), or for already using formative assessment or end-of-grade test results more intensively than the other schools in the baseline sample. The remaining 180 eligible schools were randomly assigned to one of the three groups receiving the different intervention approaches described below. Sixty schools were assigned to each of the following study groups:

- Group A: End-of-grade wrap-around support, formative assessment with support
- Group B: End-of-grade wrap-around support
- Group C: Prevailing practice

Group A: End-of-grade assessment with wrap-around support and formative assessment with materials and pedagogical support

The first approach is the full version of EducAcción's Promising Reading Intervention (EducAcción-PRI) (Group A), which provides materials and pedagogical support for end-of-grade and formative assessment. This intervention includes two separate components, both of which have a focus on improving early grade reading. They are as follows:

1. Training and support to teachers, principals, and community members to understand and effectively use end-of-grade assessments; and
2. Materials, training, and support to teachers to administer formative assessments and integrate the results into their classroom practice.

The first component consists of a series of capacity-building activities for schools and community members to understand and use end-of-grade results to improve student learning. At the end of the first year of the intervention, the EducAcción-PRI team will train principals, teachers, and community members on how to objectively administer and score the end-of-grade exams, according to the MOE protocol. In both intervention years, EducAcción PRI will provide a two-day training workshop on how to interpret and use the results of the end-of-grade exams. The second day of the workshop will include a full-day session in which schools will develop school action plans that address the needs that were identified when analyzing the test score results. Finally, the intervention team will provide technical assistance to the schools on the implementation of their action plans. Once every five weeks a pedagogical advisor will hold a meeting with a council of teachers, principals, and parents. These meetings will focus on the implementation of the action plan activities and decision making to improve student learning, with a particular focus on reading skills.

The second component consists of providing materials and training to teachers to effectively use formative assessments to adapt their instruction according to their students' needs. First, the EducAcción PRI team will provide schools with Spanish (which we will also refer to as

[1] Honduras is made up of 18 departments divided into 298 municipalities. Eligible schools were in the bottom three quintiles of performance on end-of-grade tests in 2013 and reported scores for at least 10 first-grade students that year.

"reading") and math formative tests for all students in grades 1 to 6. In addition, teachers will receive a written copy of the national curriculum standards as well as training and a written guide to pace their lessons throughout the year. Second, the intervention will include a series of training workshops (two 16-hour workshops and two 8-hour workshops) that focus on the interpretation of the formative assessment results and using them to inform instruction. Every five weeks, a member of the EducAcción PRI team will monitor teachers' implementation and scoring of the formative assessments and provide support to teachers on the scoring and interpretation of the formative assessment and the adjustment of their lesson plans accordingly.

If there are more than three participating teachers in a school, the trainings will take place in the schools. If there are fewer teachers per school, the trainings will be held for small clusters of participating schools at one of the schools in the study. Group B: EOG assessment with wrap-around support.

Group B: EOG assessment with wrap-around support

The second of the three approaches includes only the end-of-grade component of the EducAcción PRI intervention (Group B). EducAcción will offer only the training and support to teachers, principals, and community members to understand and effectively use end-of-grade assessments, without the formative assessment component. The end-of-grade component is outlined in detail in the first component described above.

EducAcción staff will not provide Group B schools with training or materials in support of the use of formative assessment. Group B teachers may choose to increase their use of formative assessment on their own as a strategy to respond to their end-of-grade test results, but they would do so without the advantage of EducAcción's training and pedagogical support. Group B teachers would also have to locate their own copies of formative assessments.

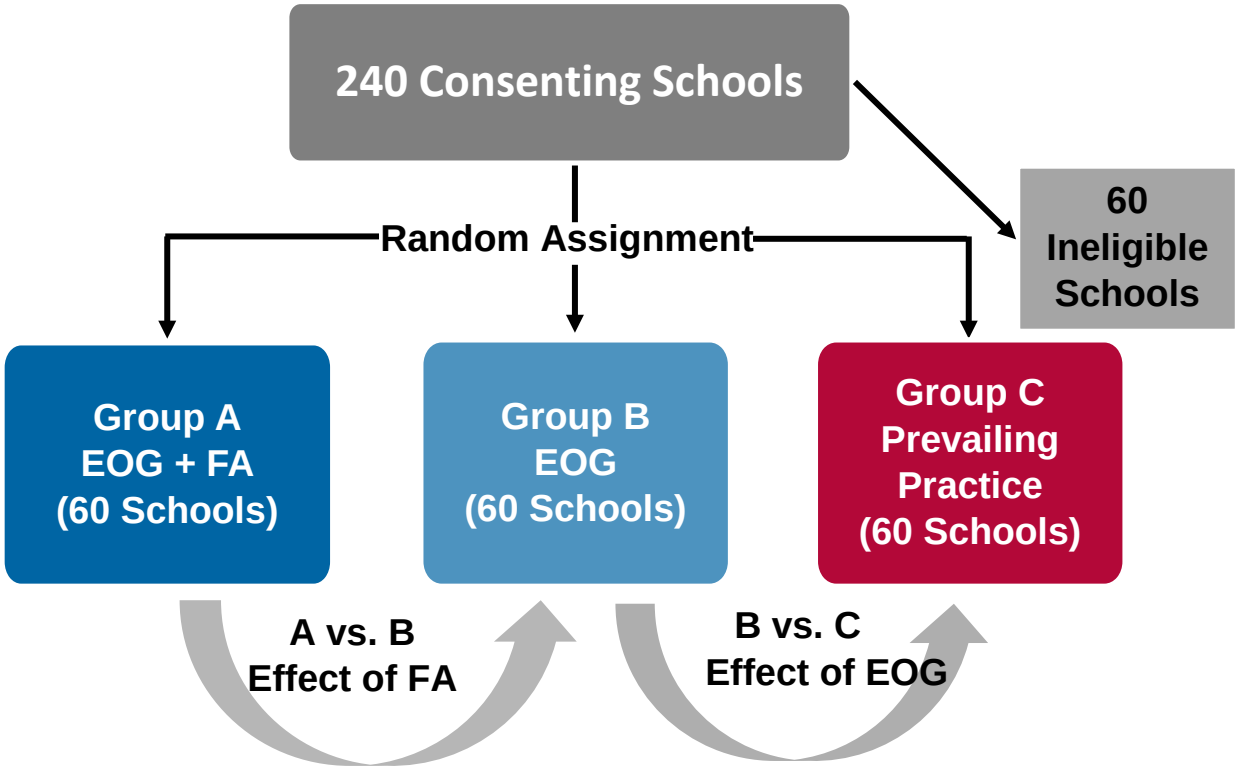
Group C: Prevailing practice

The third approach (Group C) is the continuation of the MOE's and schools' prevailing use of the formative and end-of-grade assessments. As described above, this includes monthly formative assessments that teachers may administer themselves as well as summative end-of-grade assessments, which have been applied annually to all students in grades 1 to 11 since 2012. Teachers, principals, and community volunteers have administered and scored the census end-of-grade tests. Although nearly all schools in Honduras have been exposed to both types of assessment, there is evidence that many teachers have not been able to take full advantage of either type. All schools administered end-of-grade tests at the end of the 2012, 2013, and 2014 school years, but many teachers do not have the information necessary to use test results to improve instruction. In the case of formative assessments, teachers at some schools use formative assessment as it was designed; in many other schools, teachers lack access to the formative assessment print materials or they administer the formative assessments without using their results to improve teaching.

Figure 1 shows the three groups to which the remaining 180 schools were randomly assigned and what comparisons we will make to estimate the impacts of the two interventions. First, we will compare the outcomes of teachers and students in schools assigned to Group A to those in schools assigned to Group B. This first comparison allows us to provide unbiased

estimates of the impact of the formative assessment intervention, holding constant the end-of-grade wrap-around support intervention (research question 1). Second, we will compare the outcomes of teachers and students in schools assigned to Group B to those in schools assigned to Group C. This comparison allows us to provide unbiased estimates of the impact of the end-of-grade wrap-around support intervention compared to the ongoing programming provided by the MOE and other organizations (research question 2).

Figure 1. Evaluation design: Randomized controlled trial with three treatment arms



EOG = end-of-grade. FA = formative assessment.

A three-armed design evaluating the impacts of two interventions could have formed two treatment groups with one stand-alone intervention in each. Such an approach would have allowed for estimates of the impacts of each intervention relative to prevailing practice, as well as the estimated difference between the two. The current design allows for an estimate of the impact of the end-of-grade intervention as well as of the marginal impact of the addition of the formative assessment intervention to the end-of-grade intervention. Honduran stakeholders considered these results more relevant to their context, given an expectation that end-of-grade testing might be a permanent policy. The current design also would reveal if the formative assessment intervention has little impact in schools where the end-of-grade intervention is already in place.

After collecting data at the end of the first year of implementation in 2015, we will determine whether there is a meaningful difference in implementation between Groups A and B. EducAcción staff have been instructed not to provide group B schools training or materials in

support of the use of formative assessment, but Group B teachers may choose to increase their use of formative assessment on their own as a strategy to respond to their end-of-grade test results. Additionally, Group B facilitators may provide formative assessment copies and pedagogical support, despite being instructed against doing so. If it is common for Group B teachers to access formative assessments, either independently or through the support of their facilitators, there may not be a meaningful contrast between the interventions implemented in Group A and B schools. If we find that there is no contrast between the interventions in Groups A and B, we will redefine our treatment groups and consider Groups A and B as one treatment group, and test the difference between outcomes of schools in this combined treatment group to outcomes of schools in Group C. Similarly, if the end-of-grade intervention in Group B is similar to the prevailing practice—either because the MOE offers new support for end-of-grade testing, or because of weak implementation in intervention schools, we will combine Groups B and C into one group and compare outcomes against this large group to outcomes observed in Group A. We will determine which hypotheses to test before analyzing the final outcome data.

We will estimate each intervention's impact on process outcomes (teachers' access to assessment materials and results), intermediate outcomes (teachers' use of assessment results to modify instruction), and final outcomes (learning, as measured by end-of-grade tests and an independent reading assessment). We will estimate impacts on process and intermediate outcomes at the school level, and will estimate final outcomes on learning at the student level. We will estimate impacts on process outcomes at the end of the first and second years of the interventions, but we will only estimate final impacts on learning at the end of the interventions' second year.

C. The Baseline Report

The purpose of this report is to summarize the information obtained from baseline data collection in Honduras, which we use to assess the equivalence of the three experimental groups, and to describe the schools, teachers, and students in our sample. In this report we describe basic characteristics of the study schools, their teachers, and students. The impact analysis presented in future reports will control for any baseline differences among the groups of schools to ensure that estimates reflect the impact of the interventions being evaluated and not the effects of any chance differences between groups.

This page has been left blank for double-sided copying.

II. SAMPLE AND DATA

A. Sample Selection

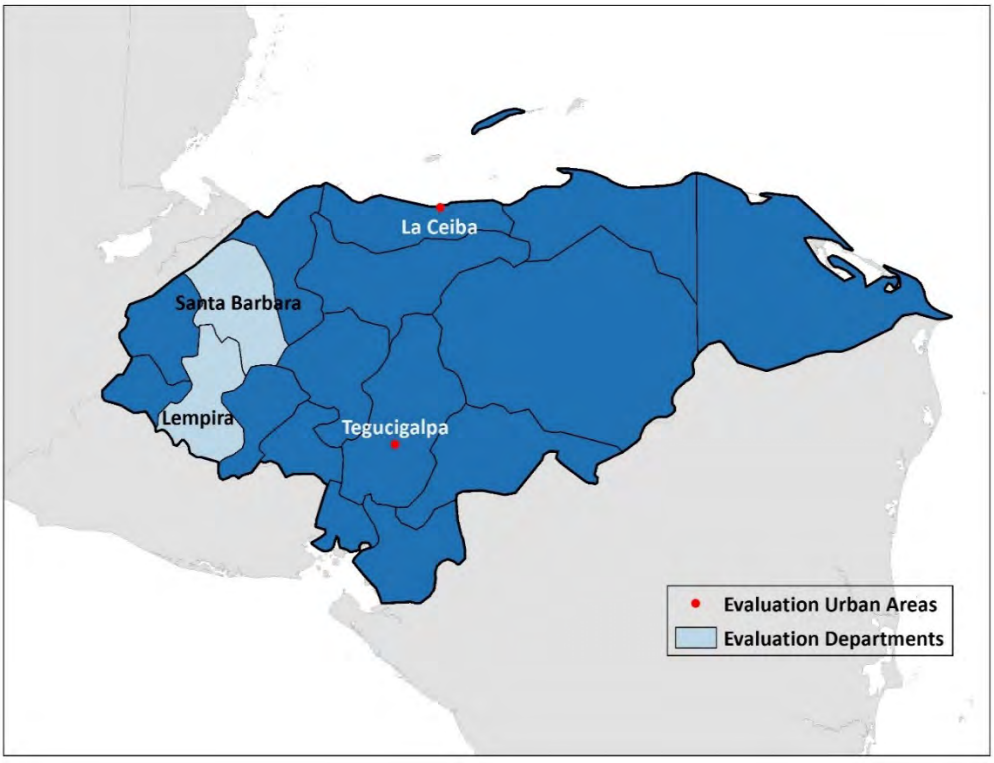
1. Selecting the initial sample

In the first step of the sample selection, we selected four geographic areas for the evaluation, based on the following criteria:

- They include a combination of rural and urban areas.
- They avoid areas that would be considered either very advantaged or disadvantaged relative to the rest of the country.
- They are concentrated geographically to make it efficient to supervise the data collection and administer the treatments.
- They are priority areas for USAID assistance.

The areas selected for the study were two mostly rural departments, Lempira and Santa Barbara, and two mostly urban municipalities, Tegucigalpa (the capital) and La Ceiba. These areas are shown in Figure 2. Because Lempira and Santa Barbara are relatively large departments, with widely dispersed schools, we worked with EducAcción to identify a subset of municipalities within each department within which to focus the evaluation. This made both implementation and supervision feasible.

Figure 2. Evaluation departments and urban areas



2. Selecting the randomized sample

To select individual schools within the evaluation areas, we used 2013 end-of-grade test score data to identify public schools that had a minimum of ten students enrolled in the first grade, and that were in the bottom three quintiles of performance on a combined measure of first-, second- and third-grade reading test scores. Our cohort of interest was the 2014 first-grade class; however, because data on this cohort were not yet available in early 2014 when we selected the sample, we used data from the 2013 first grade cohort. We worked with USAID/Honduras and EducAcción staff to identify schools that were located in areas that would be dangerous to enter. We eliminated schools in such areas to avoid putting implementer or evaluation data collection staff at risk, and to minimize the chance that such schools would be dropped from the sample later. After applying the eligibility criteria, there were more schools in Lempira and Santa Barbara than in the urban municipalities of Tegucigalpa and La Ceiba. Because we sought to have a sample that was as balanced as possible between the urban and rural areas, we included all eligible schools in Tegucigalpa and La Ceiba, and a random selection of schools in Lempira and Santa Barbara to achieve a total of 240 schools.

We collected baseline data in all 240 schools selected for the initial sample. The data was then analyzed to trim the sample to identify the final randomized sample of 180 schools. Our objective in trimming the initial sample was to identify and eliminate schools that, at baseline, had substantially more exposure to the activities that comprise the EducAcción PRI intervention, as well as any schools that did not have 10 students enrolled in the 2014 first-grade cohort. First, we eliminated 27 schools for having fewer than 10 students enrolled in first grade. After baseline data collection, EducAcción identified an additional 12 schools where they were already working. We also eliminated those schools, leaving a remaining sample of 201 eligible schools. Baseline survey data was used to create an index that measured each school's prior exposure to activities similar to EducAcción PRI to identify an additional 21 schools to trim from the sample.

The index took on values from zero to seven, with higher values indicating higher exposure to activities similar to EducAcción PRI. Schools were assigned one point each for the following characteristics:

- On average, surveyed teachers had participated in more than eight hours of training on formative assessment or end-of-grade assessment in the 2013 school year.
- At least one teacher had received end-of-grade assessment results from the previous year.
- At least one teacher surveyed had applied a formative assessment in reading at least once in 2013.
- On average, surveyed teachers had applied formative assessment more than four times that year.
- At least one surveyed teacher could show where they kept written records of administering the formative assessments.
- At least one surveyed teacher had a copy of the formative assessment instruction book.
- At least one teacher had more than one copy of the formative assessments.

Table 1 provides a summary of the characteristics describing the use of formative assessment and end-of-grade assessments at the 240 schools that were and were not selected for randomization. The differences are not as great as one might anticipate since 39 of the 60 schools that were dropped were dropped for either having too few first-grade students or for already working with EducAcción.

Table 1. Sample school characteristics at baseline

School characteristic	Schools in randomized sample	Schools initially sampled but dropped before randomization
Number of schools	180	60
Average number of hours of training first grade teachers had received that year on FA or EOG was greater than 8 hours	5.0	11.7
Percent of schools with at least one surveyed teacher who had received EOG assessment results from the previous year	58.3	70.0
First grade teachers have applied FA in reading at least once in current year	60.0	71.7
Average number of times teacher has applied FA in reading in the current year	2.1	3.7
Percent of schools with at least one surveyed teacher who could show where they kept written records of FA results	22.8	40.0
Percent of schools with at least one surveyed teacher who had a copy of the FA instruction book	75.0	75.0
Percent of schools with at least one surveyed teacher who had more than one copy of the FAs	29.4	56.7

Source: EducAcción PRI Impact Evaluation Teacher Survey—Baseline 2014, initial sample.

Note: We surveyed up to two first-grade teachers in each of the 240 schools in the initial sample. Later, we selected 180 of these schools to include in the randomized sample. This table presents the characteristics we used to construct an index, which was the basis for our selection of the final sample to randomize. Of the 60 schools that were eliminated, 27 were eliminated for having fewer than 10 students in the first grade, 12 were eliminated because EducAcción was already working with them, and 21 were eliminated for their index scores.

EOG = end-of-grade. FA = formative assessment.

As shown in Table 1, schools selected to be randomized for the evaluation had less exposure to training, were less likely to have received end-of-grade results, and had administered formative assessments fewer times than the schools that were trimmed from the initial sample. As we discuss further in Chapter III, schools selected to be randomized did have some prior exposure, as reported by teachers at baseline, to activities that are part of EducAcción PRI.

3. Random assignment

Once we identified eligible schools, we conducted stratified random assignment at the school level. We stratified schools on geographic location (location in one of the four study areas: Lempira, Santa Barbara, La Ceiba, and Tegucigalpa) and baseline use of activities that are part of EducAcción PRI. We divided schools into “high” and “low” usage based on the same index that was used for sample selection. This resulted in eight strata. Within each stratum, we ordered schools by 2013 end-of-grade test scores, and placed each into a triplet. We randomly

assigned schools within each triplet to each of the three experimental groups. Table 2 presents average 2013 test scores at each stage of sample selection. Test scores were reported as the percent of questions answered correctly. For ease of interpretation, we have converted the scores for each grade and subject combination such that a score of 50 represents the national average score for every subject and grade combination, and an additional 20 points represents a difference of one standard deviation. For example, a score of 50 would be an average score, a score of 70 would be one standard deviation above average, and a score of 40 would be half of one standard deviation below average, relative to the national average. Schools ultimately included in the study had, on average, lower reading and math end-of-grade test scores than the national sample[2]. We expected this result since we drew our sample of schools from among schools with test scores in the bottom three quintiles.

Table 2. Average reading and math end-of-grade test scores, first through third grades, 2013

Subject	Grade	Honduras	Evaluation Areas	Initial Sample	Randomized Sample
Reading	1	50.0	48.8	44.9	45.5
	2	50.0	48.6	43.6	44.0
	3	50.0	49.1	44.6	45.4
Sample size (students with reading score in any grade)		406,549	89,758	24,747	19,351
Math	1	50.0	48.8	46.7	47.0
	2	50.0	48.7	43.0	43.9
	3	50.0	49.0	43.0	43.8
Sample size (students with math score in any grade)		406,892	89,923	24,788	19,389

Source: 2013 end-of-grade test scores.

Note: Test score data come from sample-based and census-based test administration. For each source (sample-based and census-based), we standardized test scores to have a mean of fifty and a standard deviation of twenty in the national population for each subject-grade combination.

We repeated the analysis shown in Table 2 using 2014 data and found similar results. Both the 2013 and 2014 end-of-grade tests were administered before we conducted random assignment (February 2015). Because we drew our initial sample from 2013 end-of-grade test score records, we have no missing data in 2013. We have not been able to locate 2014 data for approximately 18 percent of the schools in the study because these schools either did not participate in end-of-grade testing in 2014, or did not upload their results to the online system. Results using 2014 data are shown in Appendix A. The results based on 2014 test scores demonstrate the same patterns as those based on 2013 test score data, confirming our expectations for how these samples relate to one another: evaluation scores are lower than

[2] For each subject-grade combination of first-, second-, and third-graders in math and reading, the evaluation sample had significantly lower test scores than the national sample. This difference is significant at the 1 percent level for all combinations other than first-grade math, which is significant at the 5 percent level.

national averages, and scores from the initial and randomized samples are lower than those in the evaluation areas.

Table 3 presents the distribution of schools by geographic area and experimental group in the randomized sample. Because random assignment was stratified on geographic area, the three experimental groups are balanced on geographic distribution. The allocation of sample between the two urban municipalities is unbalanced, reflecting the size difference between the capital, Tegucigalpa, and La Ceiba, which is a smaller city.

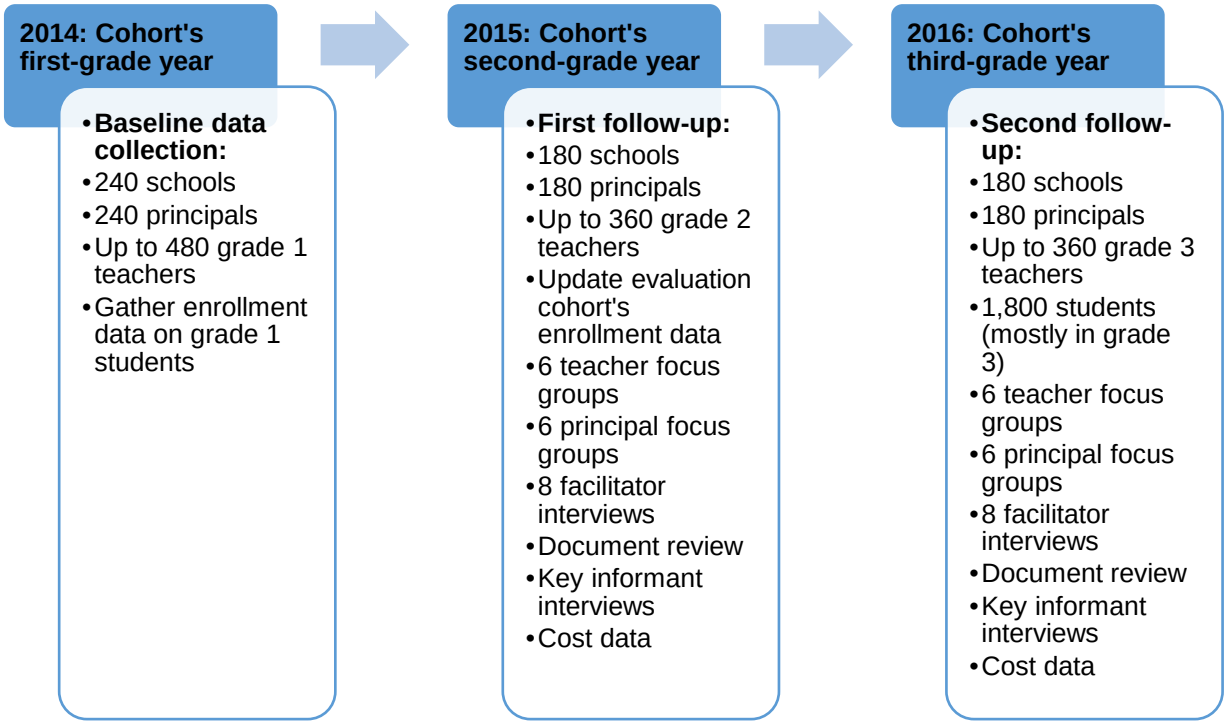
Table 3. Geographic distribution of schools in the study, by treatment group

Location	End of Grade and Formative Assessment (A)	End of Grade Only (B)	Prevailing Practice (C)	Total Randomized Sample
Urban municipalities	**28**	**28**	**28**	**84**
La Ceiba	3	3	3	9
Tegucigalpa	25	25	25	75
Departments	**32**	**32**	**32**	**96**
Lempira	16	16	16	48
Santa Barbara	16	16	16	48
Total randomized sample	**60**	**60**	**60**	**180**

B. Data Collection Instruments and Survey Response Rates

We will collect data at the end of each academic year—which corresponds roughly to the calendar year—for three years: at baseline (October 2014, before randomization, which took place in February 2015), after the first year of the intervention (October 2015), and at the end of the second year of the intervention (October 2016). Figure 3 shows the timeline for data collection.

Figure 3. Sample sizes for data collection, by year



The data for the evaluation includes surveys, focus groups, interviews, and administrative data. Baseline data consists of just surveys and administrative data. [3] We used the end-of-grade test score data provided by the MOE to select the initial sample of potentially eligible study schools. These data, as well as survey data collected before random assignment and the beginning of the intervention, are the basis of the analysis in this report, in which we describe the sample and assess baseline equivalence. The baseline data will also be used as control variables in the impact analysis to be presented in a future report. Each of the data sources are summarized below in Table 4. Focus groups and interviews with program beneficiaries, program implementers and other stakeholders will take place during both midline and endline data collections and will be discussed in future reports.

[3] Survey instruments for all rounds of data collection will be made available along with restricted use data files once the final report is finalized and published on the USAID Development Experience Clearinghouse.

Table 4. Data sources and sampling for baseline and follow-up data collection

Data Source	Contents	Sample
MOE end-of-grade test score data	Student test scores in math and reading	Baseline: All students enrolled in first through third grade in initial sample First and second follow-up: All students enrolled in first through third grade in randomized schools
Teacher survey	Teacher characteristics, teacher participation in FA, EOG training, teacher use of FA and EOG assessments and results	Baseline: Up to two first-grade teachers in each of the initial sample schools First follow-up: Up to two second-grade teachers in each of the randomized schools Second follow-up: Up to two third-grade teachers in each of the randomized schools
Principal survey	Principal characteristics, school characteristics, school participation in FA, EOG training and assessment	Baseline: Principals in initial sample schools First and second follow-up: Principals in randomized schools
Student roster	Names, dates of birth, enrollment status, and parent names	Baseline: All students enrolled in the first grade in initial sample schools First follow-up: All students enrolled in second grade in randomized schools

EOG = end-of-grade. FA = formative assessment.

1. End-of-grade test score data

The MOE has conducted annual summative end-of-grade census testing in primary schools nationwide since 2012. Tests are offered in Spanish (reading) and math, and are designed to measure concepts covered in the national curriculum. The monthly formative assessments, which are part of this evaluation, are aligned with the same national curriculum.

Each year, in a randomly selected sample of schools, external test administrators are responsible for securing the testing environment. In these schools, referred to as "secure testing site schools," test administrators bring printed copies of the tests to the schools on the day the tests are administered, observe test administration, and take the completed tests with them to conduct scoring and data entry off site. In all other schools, referred to as "census schools," test administration is less secure. Schools receive the tests days or weeks before the testing date; teachers may administer the tests to their own students with minimal supervision; and teachers do their own scoring and data entry of their students' tests. Although these are not high stakes tests—teachers' employment and salary do not depend on their students' performance—the testing environment in census schools is not secure and teachers in this group have the opportunity to influence their students' scores. Census schools and secure testing site schools are given different versions of the same test, but scores in the census are higher than scores in the sample on average. When using these data for analysis, we standardize sample and census scores according to the mean and distribution of each test's scores to improve the comparability of the results.

EducAcción PRI facilitators use end-of-grade test score data to work with schools to analyze student performance, and to develop their Institutional Action Plans—plans for school-wide improvement based on end-of-grade test results.

We used 2013 end-of-grade reading test scores as part of our sample selection and stratification of schools for random assignment. The test scores are also used to assess baseline equivalence in this report. We plan to use 2014 end-of-grade test data as a baseline measure in our impact analysis, and plan to use 2016 end-of-grade testing results as our outcome data on student achievement.

2. Teacher survey

A teacher survey will be administered in all three rounds of survey data collection—baseline, midline, and endline. Because the evaluation focuses on the cohort of students who are in second grade during the first year of the intervention, the teacher survey data collection follows the study cohort's teachers. At baseline, while most students from the study's cohort of interest were in first grade, we surveyed up to two first-grade teachers per school. For the first and second rounds of follow-up data collection, we will survey up to two second- and third-grade teachers, respectively. In schools with only one teacher working with the target grade, we will only survey that teacher; in schools with two or more teachers in the target grade, we will survey two randomly selected teachers.

The baseline teacher survey measures teachers' exposure to training on the use of formative assessments and end-of-grade assessments, their access to formative assessment materials and end-of-grade assessment results, and their use of the test results. The survey also measures teachers' educational and professional background and basic demographic characteristics. We were able to survey teachers in all 240 schools in the initial sample, achieving a response rate of 100 percent.

3. Principal survey

The baseline principal survey measures principals' exposure to training on the use of formative assessments and end-of-grade assessments, and their perception of their teachers' use of formative assessments and the results of end-of-grade assessments. The survey also captures basic school characteristics, such as school size, structure, and infrastructure. We were able to survey principals in all 240 schools in the initial sample, for a response rate of 100 percent.

4. Student roster information

Enumerators gathered student roster information for all students in the cohort of interest—the majority of students in this cohort were first-graders in 2014. This roster information included student names, gender, age, enrollment status, and parent names. This information will be updated at midline data collection. We were able to collect student roster information in all 240 schools in the initial sample.

C. Data Collector Training and Quality Assurance

The data collection team piloted the instruments in several primary schools to identify any survey questions that were not well understood by respondents. The team then consulted with

Honduran data collection experts at Espirálica and with implementers of the program within EducAcción. The researchers incorporated data collectors' and implementers' comments, and the findings from the pilot before enumerator training began.

Espirálica recruited a pool of enumerators with whom they had worked in the past and qualified new recruits who responded to ads posted in newspapers. Enumerators participated in five full days of training at a training center and in schools. The training covered the following material:

- **Formative assessment and end-of-grade assessments:** The training covered formative assessment and end-of-grade assessment to ensure that the enumerators understood the intent of the questions they would be asking principals and teachers.
- **Randomized evaluations and study design:** The training included several sessions on randomized evaluations and the evaluation design. The sessions were designed to prepare enumerators to clearly explain the evaluation design to school principals and teachers and answer questions about the schools' role in the study.
- **Courteous and ethical behavior with respondents:** Enumerators learned about the importance of being respectful and courteous with all survey respondents, as well as the importance of being transparent about the study and schools' potential role. This involved clarifying that schools participating in baseline data collection would be randomly selected to participate in one of two interventions, or in none. The training also emphasized the importance of seeking informed consent from all respondents.
- **Survey content:** Espirálica staff reviewed each survey question with the enumerators, describing the intent and any associated skip patterns of each question, and responding to enumerators' questions.

Following the training sessions, enumerators spent two mornings visiting nearby schools to practice administering the survey with principals and teachers. After the school visits, enumerators met to discuss any questions that arose, challenges they encountered, and strategies to address them.

Espirálica used several strategies to ensure the quality of data collection in the field. Field supervisors, who participated in and led portions of the training, supervised enumerators throughout the entire field period. Supervisors reviewed completed surveys every night, reviewing any incorrect or incomplete surveys with enumerators immediately. Supervisors also called a randomly selected sample of schools to verify that the enumerators had visited the schools, to confirm answers to a set of key questions, and to ask about enumerators' behavior at the school.

Espirálica used a data entry program that allowed for a series of data checks that provided an alert when out of range values were entered. Data entry began while data collection continued in the field. This early data entry provided another opportunity to identify data collection problems early on. Finally, Espirálica conducted double data entry to identify and correct data entry errors.

This page has been left blank for double-sided copying.

III. BASELINE CHARACTERISTICS OF SCHOOLS, TEACHERS, AND STUDENTS

Background characteristics of participating schools, teachers, and students, measured at baseline before random assignment, provide useful context for the study. Here we compare these characteristics for each of the three groups of schools formed by random assignment.

Throughout this chapter we report regression-adjusted means for each of the three treatment groups. These are averages that adjust for the stratified random assignment process described in the previous section. We calculated most group means by averaging student and teacher outcomes to the school level first.

The regression can be expressed as follows:

(1) $y_s = a + \lambda_A \mathrm{T}_{As} + \lambda_C \mathrm{T}_{Cs} + \alpha_1 ... \alpha_{r-1} + \varepsilon_s$

where y_s is the outcome of interest (such as teachers' use of end-of-grade results, averaged at the school level) for school *s* at baseline, and we include fixed effects for the *r* strata used in random assignment, as represented by α_1 to α_r. The variables T_{As} and T_{Cs} are indicators equal to one for schools assigned to Groups A or C, respectively, and zero otherwise. Group B is the omitted group. The term ε_s is a random error term for school *s* observed in the base year. The parameters λ_A and λ_C represent differences between schools in Groups A and B, and differences between schools in Groups B and C, respectively. In this report, we present the results of two tests: whether $\lambda_A = 0$ and whether $\lambda_C = 0$.

By including the stratum fixed effects, we capture the design effect generated by the stratified randomization. In other words, stratification (conducting random assignment within groups of similar schools) improves efficiency by reducing the amount of variation between schools of different treatment status, but at the same time, it imposes a small penalty associated with the need to implicitly estimate the block effect associated with each group (stratum) of similar schools.

A. Schools

Random assignment produced three groups of schools that are balanced on enrollment, class size, gender, school structure, infrastructure, and resources. Across the sample, nearly half of schools are multigrade, meaning that in at least one classroom, one teacher is teaching more than one grade level simultaneously. On average, and across all three groups, schools and class sizes are moderate in size. Across all three groups, schools have approximately 252 students, and first-grade classes have 26 students on average.

Schools in the sample are lacking essential infrastructure overall and across treatment groups. On average, schools are lacking roughly half of five essential classroom infrastructure components, such as indoor classrooms or a chalkboard; three out of five school infrastructure components, such as potable water or electricity; and two out of five components of learning

infrastructure, such as a library or playground (see notes to Table 5 for complete lists of index components).

The test score data included below in Table 5 are from 2013 first graders, the cohort that completed first grade before the 2014 first grade study cohort began school. There is no significant difference in test scores between schools in the end-of-grade + formative assessment group and the end-of-grade only group, or between schools in the end-of-grade only group and the prevailing practice schools in 2013 or 2014. [4] The school characteristics presented here suggest that teachers and their students face numerous challenges. More than 40 percent of schools are multigrade, meaning teachers not only must teach to students of varying levels of preparation, but must also cover the curricula of multiple grades. Low scores on the school infrastructure and learning resources indices suggest that schools lack basic resources, such as classrooms, chalkboards, or desks.

Table 5. School characteristics at baseline

Characteristic	EOG and FA (A)	EOG Only (B)	Prevailing Practice (C)	A – B	B – C
Multigrade school (percent)[a]	46.7	41.7	48.3	5.0 (0.474)	-6.7 (0.341)
Number of students enrolled, grades 1-6	260.1	270.1	225.1	-10.0 (0.767)	45.0 (0.183)
Class size, grade 1	26.1	26.0	25.0	0.1 (0.938)	0.9 (0.577)
Number of primary grade sections in the school	8.5	8.9	7.6	-0.4 (0.729)	1.2 (0.246)
Classroom infrastructure index (average; 0-100)[b]	49.0	47.0	46.0	2.0 (0.478)	1.0 (0.723)
School infrastructure index (average; 0-100)[c]	62.0	60.0	56.7	2.0 (0.471)	3.3 (0.231)
Learning resources index (average; 0-100)[d]	39.0	38.0	36.7	1.0 (0.765)	1.3 (0.690)
EOG reading test score, grade 1, 2013[e]	45.6	46.2	44.8	-0.6 (0.485)	1.4 (0.123)
EOG math test score, grade 1, 2013[e]	47.0	46.8	47.2	0.2 (0.927)	-0.4 (0.835)
Total number of schools	**60**	**60**	**60**		

Sources: EducAcción PRI Impact Evaluation Principal survey—Baseline 2014; EducAcción/PRI Impact Evaluation Teacher Survey—Baseline 2014; 2013 EOG test scores. Results are based on the 180 schools selected for the randomized sample.

Note: Columns A, B, and C present group means that are adjusted for the stratification design with a regression. The *p*-values from tests of differences between group means are presented in parentheses.

[a]In schools with multigrade classrooms, several grades are taught in the same classroom.

[4] Test score results for 2014, the first year that the study cohort was enrolled, are shown in table A2 in the Appendix. As noted after Table 2, we selected schools that reported 2013 test scores; for this reason, we have no missing data in 2013. At the time of the baseline analysis, we were missing baseline (2014) EOG data for 33 of the 180 schools in the randomized sample of schools. Through continuous efforts to locate the missing data, we located baseline EOG data for an additional 11 schools by the time we conducted endline analysis.

[b]The classroom infrastructure index is the average percent of five school-level classroom characteristics found at each school; this ranges from zero to 100. A school that has two of the five would have a score of 40, while a school with all five would have a score of 100. The five school-level classroom characteristics are: no class sections are held in an informal area, such as outdoors; no sections are exposed to weather extremes, such as heat or rain; all sections have a blackboard or other writing surface; all sections have enough desks and chairs; and no sections share a classroom with another section. The standard deviation is 14.5.

[c]The school infrastructure index is constructed in the same way as the classroom infrastructure index, and also ranges from zero to 100 (see table note b). The five characteristics included in this index are: school has piped plumbing, a source of potable water, electricity, functioning bathrooms (excluding latrines), and a health center. The standard deviation is 16.9.

[d]The school learning resource index is constructed in the same way as the classroom infrastructure index, and also ranges from zero to 100 (see table note b). The five characteristics included in this index are: school has a library resource room, computers for teachers, Internet access, a music or art room, and a playground. Standard deviation is 20.6.

[e]EOG test scores are standardized to have a mean of 50 and a standard deviation of 20 relative to the entire nation of test-takers.

EOG = end-of-grade. FA = formative assessment.

The schools in the randomized sample are balanced in their participation in education and social programs at baseline in five out of six programs we tested for baseline equivalence (see Table 6). Schools in Group A are significantly more likely to participate in the Honduran Community-based Education Program (PROHECO) than schools in Group B. PROHECO is a program through which schools practice community-based management. For most programs, few schools participate, and we find no significant differences in participation in the other programs: Educatrachos (a program that promotes the use of the One Laptop Per Child computers in schools), Intercultural Bilingual Education (an MOE program to promote bilingual, intercultural education in Honduras), Healthy Schools (a program of the Office of the First Lady in cooperation with several other government agencies, which promotes healthy lifestyles by identifying and addressing environmental, health, and nutrition risks), EducAcción (the implementing agency, which promotes educational quality through community and school-based initiatives), or the Honduran Radio Education Institute (IHER, a radio education program). EducAcción programs in Group C schools included limited training on formative assessments prior to PRI, and have since been discontinued.

Table 6. School participation in education and social programs at baseline

	EOG and FA (A)	EOG Only (B)	Prevailing Practice (C)	A – B	B – C
PROHECO	8.3	0.0	3.3	8.3* (0.020)	-3.3 (0.349)
Educatrachos	16.7	13.3	11.7	3.3 (0.596)	1.7 (0.791)
Intercultural Bilingual Education	0.0	1.7	0.0	-1.7 (0.223)	1.7 (0.223)
Healthy Schools	40.0	43.3	35.0	-3.3 (0.651)	8.3 (0.259)
EducAcción (services provided prior to PRI)	8.3	8.3	5.0	0.0 (1.000)	3.3 (0.321)
IHER	5.0	1.7	6.7	3.3 (0.356)	-5.0 (0.168)
Total number of schools	**60**	**60**	**60**	**120**	**120**

Source: EducAcción PRI Impact Evaluation Principal survey—Baseline 2014.

Note: Group means have been adjusted to reflect the stratified randomization strategy. Means may differ slightly from unadjusted means.

[a] The EducAcción programs in Group C (prevailing practice) schools are programs that included limited training of FAs prior to PRI, and have since been discontinued.

EOG = end-of-grade. FA = formative assessment.

B. Teachers

As shown in Table 7, teachers in sample schools are balanced across treatment groups on gender, age, education, and experience, but not ethnicity. A strong majority of teachers are women, and the average teacher is in his or her late thirties. Teacher self-identified ethnicity varies significantly between groups. Teachers in Group A are more likely to identify as Mestizo or Ladino, whereas teachers in the Groups B and C are more likely to identify with Lencas or another ethnic group. The Lenca ethnic group, Honduras's largest indigenous group, is found in southwestern Honduras and El Salvador.

Teachers' most common education level is an associate or technical degree in elementary education. Teachers' educational attainment is balanced across treatment groups with the exception of the "other" group, which includes college degrees in areas other than education, and advanced degrees. Teachers in the prevailing practice group are most likely to fall into this category, while teachers in Group B are least likely. This difference is significant at the 5 percent level when examined one category at a time. A chi-squared test for education levels between Groups A and B and between Groups B and C revealed that neither was significant ($p \geq .186$). The chi-square test is more appropriate for a categorical variable like this, but does not account for the study design.

On average, teachers have over eight years of experience at their current school, and relatively few teachers are first-year teachers. The low percentage of first-year teachers and relatively high average years of experience at their current school suggest that teacher turnover will not be a major limitation to the intervention. In most schools, EducAcción's facilitators should be able to build on the teachers' training from the first year during the second year of the intervention. There are no significant differences in teacher experience or probability of being a first-year teacher across the three groups.

Table 7. Teacher characteristics at baseline

	EOG and FA (A)	EOG Only (B)	Prevailing Practice (C)	A – B	B – C
Female (percentage)	74.2	80.8	84.2	-6.7 (0.345)	-3.3 (0.637)
Age in years	38.1	36.6	37.2	1.4 (0.326)	-0.6 (0.693)
Years of teaching experience	9.8	8.5	8.5	1.3 (0.214)	0.0 (0.970)
First-year teachers (percentage)	2.5	8.3	7.5	-5.8 (0.166)	0.8 (0.843)

	EOG and FA (A)	EOG Only (B)	Prevailing Practice (C)	A – B	B – C
Highest level of education (percentage)[a]					
Associate or technical degree in elementary education	44.2	44.2	40.0	0.0 (1.000)	4.2 (0.587)
College degree in elementary education	32.5	34.2	27.5	-1.7 (0.823)	6.7
College degree in secondary education	10.8	14.2	10.8	-3.3 (0.464)	3.3 (0.464)
Other[b]	12.5	7.5	21.7	5.0 (0.418)	-14.2* (0.023)
Ethnic group (percentage)					
Mestizo or Ladino (mixed race)	70.8	50.8	50.8	20.0* (0.007)	0.0 (1.000)
Lenca	21.7	32.5	35.8	-10.8* (0.049)	-3.3 (0.542)
Other	7.5	18.3	13.3	-10.8 (0.054)	5.0 (0.371)
Total number of teachers	**60**	**60**	**60**	**120**	**120**

Source: EducAcción PRI Impact Evaluation Teacher Survey—Baseline 2014.

Note: Columns A, B, and C present group means that are adjusted for the stratification design with a regression. The *p*-values from tests of differences between group means are presented in parentheses.

Regression-adjusted values have been rounded up from -0.0 to 0.0 where applicable.

[a] A chi squared test for equal distribution among these four education levels between Groups A and B had a *p*-value of 0.970, and of 0.186 between Groups B and C.

[b] This category includes college degrees in other areas, as well as advanced degrees (masters and doctorates).

EOG = end-of-grade. FA = formative assessment.

C. Students

Using enrollment records, we collected a limited amount of data on student characteristics. We present this information in Table 8. Just under half of students are female; students are balanced across treatment groups on gender. Students were 6.5 years of age at the start of the school year in all three treatment groups—this is approximately what we would expect given that students are expected to be six years old when they begin first grade. Schools are also balanced in the percentage of overage students, or the percent of students who either began primary school late, or have repeated a grade.

Table 8. Student characteristics at baseline

	EOG and FA (A)	EOG Only (B)	Prevailing Practice (C)	A – B	B – C
Female (percentage)	49.7	46.5	46.5	3.2 (0.589)	0.0 (0.589)
Age[a] (years)	6.5	6.5	6.5	0.0 (0.589)	0.0 (0.589)
Overage for grade[b] (percentage)	13.4	13.8	14.8	-0.4 (0.589)	-1.0 (0.589)
Total number of schools[c]	**57-60**	**55-60**	**58-60**	**112-120**	**113-120**

Source: EducAcción PRI Impact Evaluation Student Roster—Baseline 2014.

Note: Columns A, B, and C present group means that are adjusted for the stratification design with a regression. The p-values from tests of differences between group means are presented in parentheses.

Regression-adjusted values have been rounded up from -0.0 to 0.0 where applicable.

[a] Age is as of February 1, 2014, the beginning of the 2014 school year.

[b] We considered a first-grade student to be overage for grade if he or she was seven or older as of the official start date of the school year when students in the study attended first grade. The official age to enter first grade in Honduras is age six.

[c] Ten schools did not report date of birth for students so we were unable to calculate average age and overage for those schools. We were able to obtain student gender information for all schools.

EOG = end-of-grade. FA = formative assessment.

D. Prior Exposure to Treatment

The teacher surveys included questions about teachers' access to and use of end-of-grade and formative assessment results at baseline. These results capture the baseline level of utilization of assessments through support from the MOE or other organizations prior to the start of EducAcción PRI's activities (see Table 9).

Survey results show that relatively few teachers participated in training on end-of-grade or formative assessment during the base year. Furthermore, the average number of hours of training is low—teachers received fewer than two hours of training on average in each group. There are no significant differences across groups on prior exposure to training on end-of-grade or formative assessment.

Teachers' survey responses indicate that it is relatively common for teachers to have received end-of-grade results from the previous year. At baseline, teachers in schools assigned to Group C were less likely to report having received the end-of-grade test scores from the previous year than Group B teachers; however, this difference is not statistically significant at the 5 percent level. Teachers in Group C were also less likely to report having used end-of-grade results for planning, but this difference is not significant.

While Group C teachers are less likely than those in the end-of-grade only group to receive and use end-of-grade results at baseline, they are more likely to use formative assessment results; teachers at nearly half of prevailing practice schools indicated that they modified their teaching based on formative assessment results. Based on an index of formative assessment use to adapt

teaching, teachers in the prevailing practice group were more likely to use formative assessments to adapt teaching than teachers in the end-of-grade only group.

Table 9. Teachers' self-reported use of EOG and FA test results at baseline

	EOG and FA (A)	EOG Only (B)	Prevailing Practice (C)	A – B	B – C
Training on EOG and FA					
Participated in EOG training last year (percent)	10.0	5.0	12.5	5.0 (0.328)	-7.5 (0.143)
Participated in FA training last year (percent)	8.3	5.0	12.5	3.3 (0.498)	-7.5 (0.128)
Duration of teachers' most recent training on EOG or FA last year (hours)	1.5	0.5	1.9	0.9 (0.278)	-1.4 (0.106)
Use of EOG results					
Received results of EOG test last year (percent)	53.3	54.2	40.0	-0.8 (0.922)	14.2 (0.097)
Used results of EOG test for planning (percent)	34.2	35.8	25.8	-1.7 (0.835)	10.0 (0.212)
Use of FA results					
Has at least one copy of FA per student (percent)	22.5	17.5	19.2	5.0 (0.440)	-1.7 (0.797)
Has modified lessons in practice based on FA results (percent)	40.8	36.7	49.2	4.2 (0.524)	-12.5 (0.057)
Index of FA use to adapt teaching (from 0-100)[a]	38.6	35.7	50.1	2.9 (0.592)	-14.4* (0.009)
Total number of schools	**60**	**60**	**60**	**120**	**120**

Sources: EducAcción PRI Impact Evaluation Teacher Survey—Baseline 2014. Results are based on the 180 schools selected for the randomized sample.

Note: Columns A, B, and C present group means that are adjusted for the stratification design with a regression. The p-values from tests of differences between group means are presented in parentheses.

[a] The 6 components of the index of FA use to adapt teaching (from 0-1) are: whether the teacher graded or reviewed the tests; whether the teacher could show records of test scores; whether based on test scores, the teacher could identify standards for which students needed additional support; whether the teacher said that she or he modified teaching based on test scores; whether the teacher could identify which students needed additional help based on test results; and whether the teacher said that she or he was able to provide additional support to those students. Because these components are highly correlated with one another, we do not present results for each component separately.

* Difference in group means is statistically significant at the .05 level.

EOG =end-of-grade. FA = formative assessment.

Teacher survey results reveal that in fewer than half of schools in the randomized sample, teachers report grading and using formative assessments, while only 18 percent of teachers were able to show documentation of students' formative assessment results. This suggests that teachers may grade formative assessments without tracking students' results, limiting teachers' ability to track students' progress. Another possibility is that some teachers gave answers about formative assessment use that they thought were "socially desirable," or that would portray them in a positive light, when possible—this type of exaggeration would not have been possible when asked to show written formative assessment results.

Based on conversations with education authorities and teachers at schools we visited, we expected that some teachers used formative assessment and some used end-of-grade results. These findings confirm this expectation. Our research questions reflect the prevalence of the use of formative assessments and end-of-grade test results, by focusing on the impact of the full package of support provided by EducAcción, including complete sets of assessment materials and ongoing pedagogical support to maximize the impact of formative and EOG assessment.

IV. CONCLUSION

Schools in the study face a number of challenges. Many lack essential infrastructure, such as plumbing and classrooms protected from the elements; classroom resources such as sufficient desks and chairs for students; and learning resources such as libraries. In roughly one third of schools, teachers indicate that they have used the end-of-grade results for planning purposes, while in nearly half of schools, teachers report that they use formative assessment results for planning. Relatively few teachers, however, had access to training on end-of-grade or formative assessment during the base year. It is unclear whether the resource-scarce environment in the study schools will lead to smaller or greater impacts than we would expect in better-resourced schools. We confirmed our expectation that some schools already used formative assessments and end of grate test results; the contrast provided by EducAcción will result from the complete materials and pedagogical support the intervention will provide.

Overall, the schools in the study's three experimental groups are similar on baseline characteristics. Given the number of comparisons we examined, the number of statistically significant differences was roughly consistent with random chance (5 out of 78, or 6 percent, of differences we tested were significant at the 5 percent level). For example, more Group A teachers than Group B teachers identify as mestizo or mixed race, and fewer as Lenca.

The advantage of having collected baseline data is that in future analysis, we will be able to control for observed baseline differences between groups, improving the precision of our estimates of the impact of EducAcción PRI activities.

This page has been left blank for double-sided copying.

REFERENCES

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets." Policy research working paper, No. WPS 7226. Impact Evaluation Series. March 2015. Available at http://documents.worldbank.org/curated/en/2015/03/24220579/report-cards-impact-providing-school-child-test-scores-educational-markets. Accessed March 31, 2015.

Bennett, R.E. "Formative Assessment: A Critical Review." *Assessment in Education: Principles, Policy, and Practice,* vol. 18, no. 1, 2011, pp. 5–25.

Black, Paul, and Dylan Wiliam. "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy, and Practice*, vol. 5, no. 1, 1998, pp. 7–73.

Bos, María Soledad, Alejandro J. Ganimian, and Emiliana Vegas. "América Latina en PISA 2012: ¿Cómo le fue a la región?" Inter-American Development Bank, 2013. Available at https://publications.iadb.org/handle/11319/698. Accessed September 15, 2015.

Clarke, Marguerite. "What Matters Most for Student Assessment Systems: A Framework Paper." Washington, DC: The World Bank, 2012.

Coe, Robert. "Feedback, Value Added, and Teachers' Attitudes: Models, Theories, and Experiments." 1998. Available at http://etheses.dur.ac.uk/1059/. Accessed December 2, 2013.

Hanushek, Eric A., and Margaret E. Raymond. "The Effect of School Accountability Systems on the Level and Distribution of Student Achievement." *Journal of the European Economic Association*, vol. 2, no. 2–3, April–May 2004, pp. 406–415.

Kingston, Neal, and Brooke Nash. "Formative Assessment and a Call for Research." Educational Measurement: Issues and Practice, vol. 30, no. 4, winter 2011, pp. 28–37.

Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación. "Comparison of Results between the Second and the Third Regional Comparative and Explanatory Studies: SERCE and TERCE 2006-2013." Santiago, Chile: United Nations Educational, Scientific and Cultural Organization/Regional Bureau of Education for Latin America and the Caribbean, 2014. Available at http://www.unesco.org/new/fileadmin/MULTIMEDIA/FIELD/Santiago/pdf/Firts-Release-results-TERCE.pdf. Accessed September 5, 2015.

Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación. "Informe de resultados Tercer Estudio Regional Comparativo y Explicativo: Cuadernillo N°2, Logros De Aprendizaje." Santiago, Chile: United Nations Educational, Scientific and Cultural Organization/Regional Bureau of Education for Latin America and the Caribbean, July 2015. Available at http://unesdoc.unesco.org/images/0024/002435/243532S.pdf. Accessed September 8, 2015.

Muralidharan, Karthik, and Venkatesh Sundararaman. "The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India." *The Economic Journal,* vol. 120, August 2010, pp. F187–F203.

Perry, Lindsey. "Review of Formative Assessment Use and Training in Africa." *International Journal of School and Educational Psychology*, vol. 1, 2013, pp. 94–101.

Piper, Benjamin. "Integrated Education Program: Impact Study of SMRS Using Early Grade Reading Assessment in Three Provinces in South Africa." Research Triangle Park, NC: RTI, July 2009.

Piper, Benjamin, and Medina Korda. "EGRA Plus: Liberia Program Evaluation Report." Research Triangle Park, NC: RTI International, October 31, 2010.

Ravela, Pedro, Patricia Arregui, Gilbert Valverde, Richard Wolfe, Guillermo Ferrer, Felipe Martínez Rizo, Mariana Aylwin, and Laurence Wolff. "The Educational Assessments That Latin America Needs." PREAL working paper no. 40. Washington, DC: Programa de Promoción de la Reforma Educativa en América Latina y el Caribe, March 2008.

United Nations Educational, Scientific and Cultural Organization. "Education for All 2000-2015: Achievements and Challenges." Education for All Global Monitoring Report. Paris: United Nations Educational, Scientific and Cultural Organization, 2015. Available at http://unesdoc.unesco.org/images/0023/002322/232205e.pdf. Accessed September 7, 2015.

Vegas, Emiliana, and Jenny Petrow. "Raising Student Learning in Latin America: The Challenge for the 21st Century." Washington, DC: The World Bank, 2008.

APPENDIX A:

BASELINE TEST DATA FROM 2014 CENSUS

This page has been left blank for double-sided copying.

In the report, we presented test score data from 2013, the year before the study cohort began school. In this appendix, we present test score results for 2014, when the study cohort was in the first grade. Because we selected schools from among those that reported 2013 test scores, we have no missing data in 2013. In 2014, we are missing data for 33 of 180 schools in the randomized sample (we are continuing to attempt to locate results for these schools).

Table A1 shows test score results for all primary schools in Honduras, for schools in the evaluation areas, for the 240 schools in the initial sample, and for the 180 schools in the randomized sample. The 2014 results are similar to the 2013 results shown in the report. Table A2 shows that there is no significant difference in test scores between schools in the end-of-grade and formative assessment group and the end-of-grade only group, or between schools in the end-of-grade only group and the prevailing practice schools in 2014; this was also the case with the 2013 data.

Table A3 shows the percentage of schools that have missing end-of-grade 2014 data by treatment status. Overall, 18 percent of schools in the randomized sample are missing 2014 end-of-grade test scores for first-graders. Twelve percent of schools are missing 2014 end-of-grade test scores for all grades. The percent of schools missing data in the prevailing practice group is higher than in the other groups. This is likely due to the fact that EducAcción helped schools in Groups A and B upload their test scores if they had not already done so. We are working with our data collectors in Honduras to try to locate scores for schools with missing data. When we conduct impact analysis at follow-up, we will control for baseline performance using the 2014 scores rather than the 2013 scores presented here. For schools missing 2014 scores, we will impute their scores using the 2013 results.

Table A.1. Average reading and math EOG test scores, first through third grades, 2014

Subject	Grade	Honduras	Evaluation Areas	Initial Sample	Randomized Sample
Reading	1	50.0	47.3	45.8	47.1
	2	50.0	47.7	46.6	48.6
	3	50.0	48.2	49.4	50.8
Sample size (students with reading score in any grade)		375,594	87,871	20,086	15,514
Math	1	50.0	47.3	46.5	48.0
	2	50.0	46.8	46.4	47.8
	3	50.0	47.4	46.2	47.8
Sample size (students with math score in any grade)		374,732	87,566	20,065	15,502

Source: 2014 EOG test scores. Initial sample results are based on the 240 schools identified as potentially eligible (44 schools did not provide first grade test data and were not include in these analyses). Randomized sample results are based on the 180 schools selected for the evaluation (33 schools did not provide first grade test data and were not included in these analyses).

Note: Test score data come from census-based test administration. We standardized test scores to have a mean of fifty and a standard deviation of twenty in the national population for each subject-grade combination.

EOG = end-of-grade. FA = formative assessment.

Table A.2. Average first-grade reading and math EOG test scores, by experimental group, 2014

Characteristic	EOG and FA (A)	EOG Only (B)	Prevailing Practice (C)	A - B	B - C
Reading[a]	47.1	46.1	48.6	1.0 (0.789)	-2.5 (0.526)
Math[a]	47.9	47.2	48.1	0.7 (0.836)	-1.0 (0.788)
Total number of schools	**51**	**50**	**46**		

Source: 2014 EOG test scores. Results are based on the 180 schools selected for the randomized sample (33 schools did not provide first-grade test data and were not included in these analyses).

Note: Columns A, B, and C present group means that are adjusted for the stratification design with a regression. The *p*-values from tests of differences between group means are presented in parentheses.

[a] EOG test scores are standardized to have a mean of 50 and a standard deviation of 20 relative to the entire nation of test-takers.

EOG = end-of-grade. FA = formative assessment.

Table A.3. Schools with missing end-of-grade 2014 data, by treatment status

Type of missing data	EOG and FA (A)	EOG Only (B)	Prevailing Practice (C)	Total
Schools missing first-grade EOG 2014 data (percentage)	15	17	23	**18**
Schools missing EOG 2014 data for all grades (percentage)	7	10	18	**12**

Source: 2014 end-of-grade test scores.

EOG = end-of-grade. FA = formative assessment.

This page has been left blank for double-sided copying.

U.S. Agency for International Development
1300 Pennsylvania Avenue, NW
Washington, DC 20523
Tel: (202) 712-0000
Fax: (202) 216-3524
www.usaid.gov