



USAID
FROM THE AMERICAN PEOPLE

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



September 2019

This report was produced for review by the United States Agency for International Development under task order AID-OAA-M-12-00020. It was prepared by Sarah Liuzzi, Steven Glazerman, Nancy Murray, and Irina Cheban of Mathematica. Photo credit: Daniela Cruz, Espirállica.

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

DISCLAIMER

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

ABSTRACT

This study estimated the impacts of the EducAcción-PRI Promising Reading Intervention, a project to support the use of assessment to improve teaching and learning in primary schools in Honduras. We examined impacts on Honduras's national end-of-grade (EOG) test in reading and math for third graders. We randomly assigned 180 primary schools into three groups: a treatment group in which implementer EducAcción provided training and ongoing support to teachers and principals on using EOG assessments to develop school action plans to improve learning; a second treatment group in which EducAcción provided the EOG intervention as well as materials, training, and support to administer formative assessments (FA) and use their results to develop classroom action plans; and a control group that received the usual support from the Ministry of Education. The study was conducted in two urban areas and two predominantly rural departments. Compliance to treatment assignment was high.

The EOG component increased third grade reading scores by 0.15 standard deviations, which is equivalent to increasing the percentage of correct answers by 3 percentage points, to moving 8 percent of students to a higher one of the four performance levels reported for EOG scores, or 2 months of instruction. The FA intervention significantly improved reading scores by an additional 0.14 standard deviations. The EOG impacts were almost entirely driven by impacts in urban schools and the FA impacts were almost entirely driven by impacts in rural schools. Impacts on math were similar, but less significant.

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

CONTENTS

Glossary of terms	xi
Acronyms	xiii
Executive summary	xv
I. Introduction	1
A. The role of assessment in improving education	1
B. Literature review	2
C. The intervention and evaluation	3
1. Implementer: American Institutes for Research (AIR)	4
2. Study location and prevailing practices.....	4
3. EducAcción-PRI intervention	6
D. Evaluation questions	7
E. Summary of findings.....	8
II. Study design, data, and methods	11
A. Impact evaluation design.....	11
1. Experimental design.....	11
2. Stratified random assignment	12
3. Sample selection and cohort of interest.....	13
B. Data for impact, implementation, and cost analyses.....	14
1. Survey data	14
2. Student test data	15
3. Focus group and interview data.....	17
C. Methods for impact and implementation analysis	17
D. Secondary evaluation questions and analysis	18
III. Were the interventions implemented as intended?.....	21
A. Training and coaching	21
B. Access to EOG assessment results	24
C. Access to and use of FA materials	25
IV. Did the intervention change instructional practice?	29
A. Use of EOG test results for school management and teaching	29
B. Use of FA results to improve teaching and learning.....	30

V.	Did the intervention improve learning?	33
	A. Impacts overall.....	33
	B. Impacts by subscore.....	35
	C. Impacts by subgroup	38
VI.	Is the intervention cost effective?.....	45
	A. Methods and data for cost effectiveness analysis.....	45
	B. Cost of the interventions.....	46
	C. Cost-effectiveness and comparison to other interventions aimed at improving test scores	47
VII.	Conclusions, lessons learned, and recommendations	49
	A. Conclusions and limitations	49
	1. Conclusions on evaluation questions	49
	2. Study limitations	51
	B. Lessons learned	52
	C. Recommendations.....	53
	REFERENCES.....	55
	APPENDIX A: BASELINE EQUIVALENCE	A.1
	APPENDIX B: METHODS AND SUPPLEMENTAL INFORMATION.....	B.1
	APPENDIX C: TABLES ON IMPACTS ON INTERMEDIATE OUTCOMES, SUPPLEMENTAL TO CHAPTERS III AND IV.....	C.1
	APPENDIX D: TABLES ON IMPACTS ON TEST SCORES, SUPPLEMENTAL TO CHAPTER V	D.1
	APPENDIX E: SECONDARY QUESTION 1: HOW USEFUL IS NATIONAL CENSUS TESTING COMPARED WITH A SECURELY ADMINISTERED AND SCORED TEST IN A SAMPLE OF SCHOOLS? TEST INTEGRITY IN DIFFERENT TEST-TAKING AND SCORING CONDITIONS	E.1
	APPENDIX F: SECONDARY QUESTION 2: DOES SECURE ADMINISTRATION OF AN END-OF-GRADE TEST IMPROVE SUBSEQUENT TEST SCORES?.....	F.1
	APPENDIX G: SECONDARY QUESTION 3: WHAT ARE THE IMPACTS OF ASSESSMENT-RELATED TRAINING ON TEST SCORES?	G.1
	APPENDIX H: TEACHER AND PRINCIPAL SURVEY INSTRUMENTS: BASELINE, MIDLINE, AND ENDLINE YEARS.....	H.1
	APPENDIX I: DISCUSSION GUIDES FOR FOCUS GROUPS AND INTERVIEWS: MIDLINE AND ENDLINE YEARS.....	I.1

TABLES

I.1	Interventions by treatment group	3
II.1	Response rates for EOG testing at school and student levels	16
II.2	Secondary analysis research questions and findings.....	19
A.1	School characteristics at baseline.....	A.3
A.2	Teacher characteristics at baseline	A.4
A.3	Student characteristics at baseline	A.5
A.4	Average first-grade reading and math end-of-grade test scores, by experimental group, 2014	A.6
A.5	Teachers' self-reported use of end-of-grade and formative assessment test results at baseline	A.7
B.1	Number of focus groups and interviews, by location and treatment group.....	B.4
B.2	Characteristics of students with and without endline data (percentages except where noted)	B.7
C.1	Participation in training.....	C.3
C.2	End-of-grade training content	C.4
C.3	Formative assessment training content	C.5
C.4	Access to end-of-grade results	C.6
C.5	Formative assessment materials and application.....	C.7
C.6	Use of end-of-grade results for school action plans	C.8
C.7	Teachers' use of end-of-grade results for lesson planning.....	C.9
C.8	Use of formative assessment results to improve teaching and learning	C.10
D.1	Impacts on reading and math test scores (effects sizes): Overall and subscores.....	D.3
D.2	Impacts on reading and math test scores, by urban and rural area (effect sizes).....	D.4
D.3	Impacts on reading and math test scores, by gender (effect sizes)	D.5
D.4	Impacts on reading and math test scores (effect sizes): Overall and subscores with nonresponse weights	D.6
E.1	Characteristics of census testing and secure test administration	E.4
G.1	First-stage regressions: Relationship between MIDEH index and selection as a MIDEH target municipality	G.5
G.2	Second-stage regressions: Estimated impact of selection as a MIDEH target municipality on 2013, 2014, and 2015 end-of-grade test scores for reading and math	G.5

FIGURES

I.1.	Evaluation departments and urban areas	5
I.2.	Theory of change for EducAcción-PRI.....	7
II.1.	Evaluation design: Randomized controlled trial with three treatment arms.....	12
III.1.	Percentages of teachers and principals who received training in EOG or FA.....	21
III.2.	EOG training content.....	23
III.3.	FA training content	24
III.4.	Access to EOG results	25
III.5.	FA materials and application.....	27
IV.1.	Use of EOG results for School Action Plans (SAPs)	29
IV.2.	Teachers' use of EOG results for lesson planning	30
IV.3.	Use of FA results to improve teaching and learning	31
V.1.	Impacts on reading by intervention and overall	34
V.2.	Impacts on reading performance levels	35
V.3.	Impacts on reading—subscores	36
V.4.	Impacts on math by intervention and overall	37
V.5.	Impacts on math performance levels	38
V.6.	Impacts on reading by urban or rural area.....	39
V.7.	Impacts on math by urban and rural area	41
V.8.	Impacts on reading by gender	42
V.9.	Impacts on math by gender	43
E.1.	STA and census score distribution: 2nd grade reading, 2013.....	E.5
E.2.	STA and census score distribution: 2nd grade math, 2013.....	E.6
F.1.	Impact of secure test administration on future test scores, by grade, subject, and year.....	E.6

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 Claesy Banegas; Emiliana Tejeda, department director for La Paz; Javier Luque from the Inter-American Development Bank; and independent consultants Marco Tulio Mejía and José Salgado all shared invaluable insights into the evaluation in Honduras. The State Departments of Education in Lempira, Santa Barbara, Francisco Morazán, and Atlántida all provided essential support for the evaluation. The Ministry of Education provided essential information on schools and student enrollment.

American Institutes of Research (AIR) home office project manager Bridget Drury, research associate Mariela Goett, and research assistant Gustavo Castillo have provided important information on evaluation activities in Honduras as well as on the activities carried out by EducAcción. At EducAcción, project director Edwin Moya, pedagogical expert Ligia Águilar, and monitoring and evaluation specialist Delfina Flores have also provided valuable information and assistance throughout the process. Russbel Hernández of Mejorando el Impacto al Desempeño Estudiantil de Honduras (MIDEH) has also generously provided important information on assessment activities in Honduras and MIDEH's role in that work. Barbara Knox-Seith, senior education advisor, and Michael Lisman, education advisor, Bureau for Latin America and the Caribbean (USAID/Washington), and Kay Leherr, Mariella Ruiz, and Renan Rápalo (USAID/Honduras) have provided guidance and support to the evaluation design and its implementation.

We thank our local data collection partner, Espirállica, and its staff for high quality data collection in the study schools. César Valenzuela provided overall project management support in Honduras, oversaw data entry, led focus group discussions and interviews, and worked with the Mathematica team to spot potential risks to the evaluation's experimental design and to 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, Chantal Toledo and Kate Place provided essential support in facilitating the success of the survey and qualitative data collection and conducted skilled analysis of focus groups and interviews. Edith Felix, Alena Davidoff-Gore, Linda Molinari, Jeremy Page, Virginia Poggio, Javier Rodriguez, Sarah Chan, Patricia Costa, Jonathan McCay, and Colleen Fitts provided critical support in data cleaning, programming, analysis and presentation, and project management. Ivonne Padilla provided excellent translation and project support.

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 their students. We also are grateful to the students whose participation in end-of-grade tests made this research possible.

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

GLOSSARY OF TERMS

Administrative data: Data collected primarily for program implementation or monitoring purposes. This is distinct from data collected explicitly for research purposes.

EducAcción-PRI: Program implemented by EducAcción with funding from USAID/Honduras, which includes materials and training to support the use of end-of-grade and formative assessments in primary schools.

End-of-grade assessment: Assessment intended to evaluate student learning at the end of the year.

Formative assessment: Assessment intended to evaluate student learning on an ongoing basis to inform teachers throughout the academic year.

Instrumental variable analysis: An evaluation methodology that uses a characteristic (such as a test score or index value), known as an instrumental variable, which is correlated with receipt of a treatment, but not correlated with outcome values beyond its correlation with treatment.

Intent to treat analysis: An analysis approach in which the researcher estimates the average impact of the offer of treatment without adjusting for whether research participants received the intended treatment.

Local average treatment effect analysis: An analysis approach in which the researcher adjusts the impact estimates for the probability of receiving treatment in each treatment group.

Randomized controlled trial: Evaluation method in which units of analysis, such as schools or students, are randomly assigned to two or more groups, which receive different treatments. Random assignment to groups generates groups that are expected to be similar in the absence of treatment. The formation of groups that are similar in the absence of treatment allows for the estimation of the causal effects of differences in treatments offered to the groups.

Secure test administration: Method of test administration that involves measures to minimize the risk of indirect or direct manipulation of test scores.

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

ACRONYMS

AIR	American Institutes of Research
EducAcción-PRI	EducAcción Promising Reading Intervention
EOG	End-of-grade
FA	Formative assessment
LAC	Latin America and the Caribbean
LAC Reads	Latin America and the Caribbean Reads
MIDEH	Mejorando el Impacto al Desempeño Estudiantil de Honduras
MOE	Ministry of Education
PISA	Programme for International Student Assessment
SAP	School action plan
STA	Secure test administration
TERCE	Third Regional Comparative and Explanatory Study
USAID	United States Agency for International Development

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

EXECUTIVE SUMMARY

A. Introduction

This report presents the findings of a rigorous impact evaluation of the EducAcción Promising Reading Intervention (EducAcción-PRI), which includes a randomized control trial impact evaluation, an implementation study, cost analysis, and secondary data analysis. In 2014, the U.S. Agency for International Development (USAID) funded the development of EducAcción-PRI to improve Honduran students' early grade reading skills through the use of two types of assessment: summative end-of-grade (EOG) assessments and monthly formative assessments (FA). The American Institutes for Research (AIR) and its local partners in Honduras implemented EducAcción-PRI in four locations: two predominantly rural departments (Lempira and Santa Barbara) and two urban areas (Tegucigalpa and La Ceiba).

EducAcción-PRI involved two components: the EOG component and the FA component. Both components focused on improving early grade reading but also provided support for math.

1. **EOG component: supporting the use of end-of-grade assessment results.** The first component provided training and ongoing support to teachers and principals in the use of EOG assessments to develop school action plans to improve student learning.
2. **FA component: supporting the use of formative assessment results.** The second component provided materials, teacher training, and ongoing support to teachers to administer FAs and integrate the results into their classroom practice.

Before EducAcción-PRI, schools had limited support for EOG and FA. The Ministry of Education (MOE) first provided FA training and materials in 2009. The MOE and the *Mejorando el Impacto al Desempeno Estudiantil de Honduras* (MIDEH) project implemented the first census EOG test in 2012. However, because of budget limitations, schools had limited access to materials and support for assessment in subsequent years.

B. Evaluation questions and design

1. Were the interventions implemented as intended?
2. Did the interventions result in changes in instructional practice?
3. Did the EOG intervention improve learning?
4. Did the FA intervention improve learning?
5. How much does it cost to provide the materials and support necessary for FA and EOG assessments? Are effects large enough to justify these costs?

Impact evaluation design. We randomly assigned 180 schools to one of three groups: Group A (the EOG and FA group) received both the EOG and FA components of the intervention, Group B (the EOG-only group) received the EOG intervention only, and AIR (the implementer of EducAcción-PRI) did not intervene in Group C (the prevailing practice group). We conducted random assignment within the four locations and stratified schools based on previous use of assessments and average test score.

Impact analysis approach. We estimated impacts as the difference between groups in average outcomes at endline. To improve precision, we used a regression model that adjusted for chance differences among study schools observed at baseline as well as for the design. We estimated the impact of the EOG component by comparing outcomes from Groups B (EOG-only) and C (prevailing practice). We estimated the impact of the FA component by comparing outcomes from Groups A (EOG and FA) and B (EOG-only). This assumes that the impacts of formative assessments are the same whether or not there is also EOG support.

Cohort of interest. The evaluation followed the cohort of students enrolled in grade 2 as of the last day before the intervention began—May 31, 2015. They would receive the intervention for the second half of their grade 2 year and all of grade 3.

Data for impact, implementation, cost, and secondary data analyses. At baseline and endline, we used students' EOG test scores and the results from teacher and principal survey data that Espirállica, a Honduras-based research firm, collected for the study.

With one exception (a school closed prior to endline), all study schools participated in all three rounds of data collection. We followed students for the two school years of the evaluation and have endline test score data for 80 percent of students in the evaluation cohort.

The study team also gathered qualitative data through focus groups and interviews with principals, teachers, coaches, and EducAcción-PRI management. Cost data came from expenditure records and interviews with AIR, the prime contractor for the implementation team. The research team paired national EOG test score data with information from AIR on implementation of education programs during that period to conduct secondary data analyses that are included in appendices to this report. Chapter II of the report describes the data.

C. Summary of findings on EducAcción-PRI

Survey and qualitative data suggest that the EducAcción-PRI intervention was implemented as planned. Principals and teachers received the intended materials, training, and support according to their treatment group. Both teachers and principals reported in focus groups that they appreciated receiving EOG results at the beginning of the academic year.

Both the EOG and FA components increased teachers' likelihood of using assessment results to adjust lesson plans. Teachers in the EOG-only group were significantly more likely than teachers in the prevailing practice group to use EOG results to adjust lesson plans (99 versus 25 percent). Teachers in the FA and EOG group were significantly more likely to use FA results than teachers in the EOG-only group (90 versus 25 percent).

The EOG component of the intervention improved reading test scores. The EOG component increased reading scores by 0.15 standard deviations. This is equivalent to increasing the percentage correct by 3.0 percentage points, moving 8 percent of students to a higher one of the four performance levels reported for EOG test scores, or about 2.0 months of instruction. The EOG impacts came almost entirely from schools in urban areas.

The FA component of the intervention also improved reading test scores. The FA component increased reading scores by 0.14 standard deviations. This is equivalent to increasing

the percentage correct by 3.6 percentage points, moving 7 percent of students to a higher one of four performance levels reported for EOG test scores, or about 1.9 months of instruction. The FA impacts came almost entirely from rural areas. **Impacts on math were similar to impacts on reading.** Only the combined impact of the EOG and FA components (0.25 standard deviations) was significant at the 5 percent level.

Impacts varied between schools in urban and rural areas. The EOG intervention significantly increased reading scores in urban schools by 0.19 standard deviations, but it had no significant impact in rural schools. The FA component significantly increased reading scores in rural schools by 0.38 standard deviations, but it had no impact in urban schools. Subgroup impacts on math scores were similar to impacts on reading.

Impacts for boys and girls followed a similar pattern but were only statistically significant for boys in reading. The differences in impacts for boys and girls were not significant. The results are no longer statistically significant for girls in part because of the smaller sample size for subgroup analysis.

Cost analysis suggests that these intervention components require an investment of up to \$60 per student to generate an extra month of learning. EducAcción-PRI increased student test scores. The cost to raise reading test scores by 0.1 standard deviations, a common measure of cost-effectiveness in education programs, is \$52 per student for the EOG intervention and \$57 per student for the FA component. Including the cost of preparation before implementation began increases the cost by \$1 to \$2 for each of the EOG and FA components.

D. Overall conclusions, limitations, lessons learned, and recommendations

1. Conclusions and limitations

Our teacher and principal survey data suggest that EducAcción-PRI was implemented as planned. Principals in all schools assigned to receive the EOG component reported that they developed and implemented school action plans based on EOG results. Teachers in EOG and FA schools reported that they received the FA materials needed to administer monthly FAs, along with regular coaching on using FAs to improve teaching.

Teachers in nearly all EducAcción-PRI schools reported that they modified their teaching based on EOG and FA results. In survey data, teachers reported that they dedicated more time to subjects that challenged students on the EOG and FA tests, providing extra support to students with low scores, and identifying new teaching methods. Teachers in the EOG and FA group were significantly more likely than teachers in the EOG-only group to make such changes.

The EOG intervention significantly increased reading test scores by 0.15 standard deviations; impacts for math were similar in magnitude, but not statistically significant. The impact on reading scores is equivalent to moving a student from the 50th to the 56th percentile, to increasing by 3 points the percentage of reading questions that students answered correctly, and to moving 8 percent of students into a higher performance level. This is also equivalent to about two months of instruction.

EOG impacts varied across urban and rural schools. The impact of the EOG intervention on reading test scores was larger and significant only in urban schools. With more teachers and students to manage, principals in urban schools may have benefited more from test score data. Furthermore, given that urban principals are less likely than rural principals also to serve as classroom teachers, they may have had more time to implement a school action plan than rural principals. EOG impacts on math scores were also smaller in rural schools than in urban schools.

The intervention was not gender-biased. Differences in impacts between boys and girls were not statistically significant.

The provision of FA materials and support significantly increased reading test scores, by 0.14 standard deviations; this impact was similar in magnitude to that of the EOG intervention. Impacts on math were similar in magnitude and significance. The FA component impact is also equivalent to moving a student from the 50th to the 56th percentile. The FA component increased by 4 percentage points the percentage of reading questions students answered correctly and moved 7 percent of students to a higher performance level.

FA impacts were larger in rural schools than in urban schools. The FA component significantly raised reading test scores in rural schools by 0.38 standard deviations, but had no impact in urban schools. Rural teachers had less experience and were more likely to be first-year teachers than their urban counterparts. In addition, in urban schools, coaches had more teachers to work with, possibly limiting how much time they spent with each teacher. Impacts on math were similar to impacts on reading.

We estimated that investing in either the EOG or FA components would amount to spending \$52 or \$57 per student per 0.10 standard deviation improvement in test scores. These figures are in the middle to high range of costs per unit of improvement that have been published for other interventions. Investing in both interventions would cost about twice as much per student, but would produce about twice the impact on reading test scores. These estimates are for the full package including not only assessments, but also training and coaching.

Several study limitations should be kept in mind when interpreting findings. In particular, the FA impact estimates assume that there are no interaction effects (the impact of the FA component in a school with the EOG component is the same as the impact of FA in a school without EOG component), and that nonresponse and noncompliance, which were low and uniform across experimental groups, affected each group in the same way. Furthermore, because FA use was lower in the EOG only group than in the prevailing practice group, there is a chance the EOG estimates are underestimated and the FA estimates are overestimated. It should also be noted that the costs are incremental, meaning that we do not calculate the cost of implementing an intervention like EducAcción-PRI with no prior investments in assessments.

2. Lessons learned

Both EOG assessments and FAs can improve learning when coupled with ongoing training and support. Principals and teachers increased their capacity to analyze and respond to test results, and valued the support they received to do so. The school-level action plans and classroom-level improvement plans provided principals and teachers with a set of goals that motivated effective changes in instruction.

Training and support for EOG assessments and FAs are important if the assessments are to be effective tools to improve learning. Among schools that participated in EOG testing, we found large contrasts between the EOG-only group and the prevailing practice group in access to test score results and the likelihood of using the results.

The effectiveness of education interventions depends critically on the context. The evaluation found dramatically different effects in urban and rural schools, possibly leading to context-specific policy recommendations.

The interpretation of results is clearer when interventions under study are in their steady state. Given that the assessment instruments and the needed relationships with education authorities were already in place, EducAcción-PRI launched as soon as funds became available. These circumstances benefited the evaluation because evaluation intensity was not compromised by implementation delays or changes to the intervention.

A strong evaluation design and data tailored to the evaluation’s needs increase researchers’ ability to reach firm conclusions. The prospective experimental evaluation design used for the evaluation of EducAcción-PRI generated clean contrasts among groups and used data tailored to the evaluation.

3. Recommendations

A recommendation to policymakers in Honduras is to consider expanding and replicating EducAcción-PRI to reach more schools in the country if funding is available. To have the largest effects, the EOG component could be offered in urban schools and the FA component in rural schools.

A recommendation to policymakers throughout Latin America and the Caribbean that have already developed either EOG assessments or FAs is to consider implementing interventions similar to EducAcción-PRI. Countries considering developing an assessment system could consider the potential benefits of assessments to improve teaching and learning.

A recommendation to policymakers in countries that provide EOG test results to principals and teachers is to provide those results early in the academic year. Principals and teachers mentioned that they valued receiving results early in the year to help them plan the year.

Another recommendation for policymakers who set national testing policies is to consider the tradeoffs between the coverage of a testing program—national census versus a sample—with the quality of the test administration. In addition to presenting impacts of EducAcción-PRI, this study included analysis of secondary data on census tests and a more secure testing administered in a random sample. As discussed in chapter 2 and detailed in Annex E, the findings suggest that secure test administration produces more credible data.

A recommendation to donors is to recognize the advantages of building in rigorous evaluations of enhancements or variations to key components of established programs. Rigorous evaluations are often particularly difficult to conduct when interventions are still under development or when education authorities are unaware of the new intervention being evaluated.

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, but student achievement has improved either slowly or not at all. During the period 1999 to 2012, in most countries in the region, net primary school enrollment rates increased, gender parity was achieved, rates of grade repetition declined, and rates of primary school completion improved (UNESCO 2015). However, in the 2012 Programme for International Student Assessment (PISA) exam, the 8 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 grade 3 and 6 students across the 15 PISA participating countries were still performing at basic levels in both reading and math (LLECE 2014; LLECE 2015).

As policymakers' focus in developing countries shifts from education *access* to education *quality*, student assessment has come to the fore (Clarke 2012; Vegas and Petrow 2008). Some 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 in the use of educational assessment. The MIDEH project (*Mejorando el Impacto al Desempeño Estudiantil de Honduras*), a partnership of 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 in Honduras in recent years. MIDEH supported the development of a new curriculum and a national testing system linked to that curriculum. The system includes two types of assessments. In the first case, formative assessments (FA) are designed for administration by teachers to gather feedback for adjusting instruction. The FAs in Honduras may be administered monthly and are aligned with the Honduras national basic curriculum. In the second case, summative assessments are designed to reach judgments about the performance of students, teachers, or schools at the end of an instructional period. In Honduras, the summative assessments are known as End-of-Grade (EOG) assessments. In nearly every year since 2007 (with the exception of 2009 and 2011), the MOE has sponsored either a national administration of EOG assessments to all students in grades 1 through 9 as part of a census or administration to a random sample of schools, depending on available funds. EOG assessments were administered to all students in 2012, 2013, and 2014 and to a random sample every year except in 2009 and 2011. MIDEH led the administration of EOG tests in the random sample of schools, with the MOE's support.

Despite these efforts to implement assessment as a general practice in basic education in Honduras, evidence is needed to support continued investment in assessment. Prior to the current study, the impact of the formative or summative assessments implemented in Honduras had not undergone formal evaluation, even though some evidence, reviewed below, suggests that assessment provides the basis for implementation of promising educational interventions. Furthermore, the availability of testing materials, such as booklets, and the number of teachers

recently trained in test administration have both declined in recent years. Education authorities and international donors must decide whether to renew funding to keep these supports in place.

Even amid policymakers' need for high quality information about the use of both formative and summative educational assessments, the evidence base for interventions that use student assessment was scant in developing countries, and in the LAC region in particular. However, some studies from both developed and developing countries have demonstrated the potential effects of formative assessments, whereby teachers use frequent assessments of their students to improve their teaching, as well as of end-of-grade summative assessments, which, like formative assessments, may provide the basis for teachers' and principals' adaptation of teaching practices and curricula to better meet the needs of students. EOG assessments focus on subject matter tests that either reveal the challenges faced by students or identify the assistance needed by students whose test scores demonstrate that students are lagging. Furthermore, end-of-grade assessments can inform schools and policymakers about schools' yearly progress and may be used for accountability or resource allocation at the school, regional, or national level.

B. Literature review

The literature to date on EOG assessments in developing countries shows its potential for improving teaching and learning, although effects documented to date have been modest. These impacts may be achieved in several ways. First, through an accountability effect, EOG assessment might motivate staff to improve teaching and learning once test results are made available, even if no explicit consequences are attached to test results. Studies from Liberia and Pakistan have found moderate effects of using assessments to improve learning; Piper and Korda (2010) found effects ranging from none to 0.21 standard deviations (SD) on grade 3 reading scores (equivalent to moving a student from the 50th to the 58th percentile) and Andrabi et al. (2015) found effects of 0.10 SD (equivalent to moving from the 50th to the 54th percentile). Second, test results provided to schools could improve learning if scores are paired with guidance on how to use the data. We found little evidence on whether support for principals and teachers on how to use test results improved teaching and learning, but in Liberia, Piper and Korda (2010) found that sharing scores had larger impacts when paired with a detailed curriculum, ranging from 0.52 to 1.23 SD (equivalent to moving a student from the 50th to the 70th percentile on the low end or the 90th on the high end). Third, EOG assessment data could also be useful to schools, policymakers, and ministries of education in their decision making. At a higher administrative level, the data can help ministries direct resources where they are likely to do the most good (Ravela et al. 2008; Clarke 2012).

FA has shown promise in systematic reviews on FA's role, and effects documented to date suggest that FA may have large impacts. However, reviews to date are based on evidence from developed countries and few of the studies included were methodologically rigorous enough to establish causation. A large-scale review of more than 250 studies of interventions loosely defined as formative assessment in the United States and other developed countries, Black and Wiliam (1998) found effects of 0.4 to 0.7 standard deviations, which is equivalent to moving a student from the 50th to somewhere between the 66th and 76th percentiles. Using a narrower definition of FA and higher standard for rigor, Kingston and Nash (2011) found a smaller average effect size of 0.2 standard deviations, or the equivalent of moving a student from the 50th to the 58th percentile.

This evaluation makes several contributions to the evidence base on the role of assessments in improving teaching and learning. First, the study adds to the small evidence base on the role of assessments in developing countries, and is the first to our knowledge to rigorously evaluate both kinds of assessments in LAC. Second, the study adds to the limited evidence on the role of support for principals and teachers in the use of assessments. Third, it provides separate estimates of the impact and cost of using EOG assessments and using FA.

C. The intervention and evaluation

To address the need for high quality evidence on interventions that improve student outcomes, particularly in reading, USAID contracted with Mathematica Policy Research as its independent evaluator to design and conduct impact evaluations and cost-effectiveness analyses, along with complementary qualitative studies of promising reading interventions and education-access interventions, in the LAC region in a project known as the Latin America and the Caribbean Reads (LAC Reads) evaluation. Through the LAC Reads evaluation project, Mathematica has conducted evaluations in El Salvador, Guatemala, Honduras, Nicaragua, and Peru. USAID/Honduras also contracted with AIR to fund a project to support formative and EOG student assessment interventions in Honduras through its existing project, EducAcción, known as the EducAcción Promising Reading Intervention (EducAcción-PRI). Under the LAC Reads evaluation contract, Mathematica worked closely with the EducAcción-PRI implementers to design a rigorous evaluation of EducAcción-PRI, its impacts on student and teacher outcomes and to analyze the implementation and the cost-effectiveness of the interventions. In this report, we present the final results of our evaluation of that program, including impact analysis, qualitative analysis, and cost-effectiveness analysis.

To estimate the impacts of EducAcción-PRI on student test scores, Mathematica implemented a randomized evaluation. We randomly assigned schools to one of three treatment groups, as described in Table I.1:

Table I.1. Interventions by treatment group

Intervention	FA and EOG (A)	EOG-only (B)	Prevailing practice (C)
Prevailing practice: EOG results available online for schools participating in testing. FAs freely available online. Training videos available online. Paper for printing FAs unavailable in most schools.	X	X	X
EOG component: Detailed reports of EOG test results and support to develop and implement institutional action plan based on EOG test results. Twice-monthly coaching visits with principals.	X	X	
FA component: FA materials and support to develop and implement classroom improvement plans based on FA results. Twice-monthly coaching visits with teachers.	X		

The design allowed us to estimate the causal impact on students' EOG test scores that may be attributed to the EOG support intervention (comparing outcomes for schools in groups B and C) and to the FA materials and support (comparing outcomes for schools in groups A and B). We had to assume that there were no interaction effects. That is, that the impacts of the FA

component are the same with or without EOG component. The design also allowed us to estimate variation in impacts for subgroups: by gender and by schools' urban or rural location.

The intervention began in June 2015 and ended at the end of the 2016 academic year (the academic year in Honduras follows the calendar year). We gathered baseline data at the end of the 2014 academic year, midline data at the end of the 2015 academic year, and the endline data on which this report is based at the end of the 2016 academic year. We describe the intervention in greater detail in the rest of this chapter and describe the evaluation design in more detail in the next chapter. For further information, see the baseline report in which results of the baseline survey are shared (Glazerman et al. 2016) as well as a slide deck in which we shared midline results with USAID/Honduras and USAID/Washington (Liuzzi et al. 2016).

1. Implementer: American Institutes for Research (AIR)

EducAcción was a USAID/Honduras-funded project carried by AIR, who worked with municipalities, districts, and schools in Honduras since 2011 to promote improved school management techniques, community involvement in schools, and teacher training on Spanish and math instruction. The EducAcción project's school supports included assistance in using FA and EOG results to improve instruction. USAID/Honduras sought to study the impacts of FAs and EOG assessments on student learning. In parallel, given the LAC Bureau's interest in using the LAC Reads evaluation contract to support an impact evaluation that could contribute to the global evidence base on what works to improve reading outcomes, the implementers of EducAcción created the Promising Reading Intervention (EducAcción-PRI), which is the focus of this evaluation.

2. Study location and prevailing practices

AIR implemented EducAcción-PRI in the two predominantly rural departments of Lempira and Santa Barbara and the two urban areas of Tegucigalpa and La Ceiba (Figure I.1). We selected these locations because they were among USAID's high-priority areas and allowed us to study a mix of rural and urban schools.

Figure I.1. Evaluation departments and urban areas



The evaluation took place in an environment in which the MOE was already promoting the use of assessment results to drive instruction. In 2003, the MOE unveiled a national curriculum (*el diseño curricular nacional básico*, DCNB). Following the DCNB's release, MIDEH—one of two projects that USAID funded AIR to implement in Honduras—developed materials to support the new curriculum, including pacing guides and FAs and EOG assessments closely aligned to the curriculum. The MOE developed and disseminated new educational materials between 2003 and 2012. The first nation-wide EOG test was administered in a sample of schools in 2007, and FA training and materials were first distributed in 2009. The MOE and MIDEH implemented the first census EOG test in 2012. Because of budget limitations, schools' access to materials and support for both types of assessment has been inconsistent.

Despite this activity, the context for the evaluation was still one of weak or limited implementation of intended reforms. Even though the MOE expected all schools to participate in census EOG testing and receive their test results in the years when census testing was conducted (2012 through 2014), responses to the baseline survey suggest that many schools had not received results, and most schools had not received training on how to interpret or use the test results from the previous year's test. Most schools did not have a sufficient number of FAs for all of their students. Yet, even though FAs were freely available on the MOE's website, most schools had not accessed or printed them for use in the classroom. Training on the use of EOGs or FAs had been sporadic. In Appendix A, we provide detailed information on schools' training and support needs at baseline.

3. EducAcción-PRI intervention

Through EducAcción-PRI, AIR provided materials and coaching¹ to primary schools for FAs and EOG assessments. The intervention included two components that support schools' use of the assessments, both of which focused on improving early grade reading but also included support for math. They are as follows:

Component 1. Supporting the use of EOG assessment results. The first component provided training and support to teachers and principals to understand and effectively use EOG assessments. It consisted of a series of capacity-building activities for schools and community members to help them understand and use end-of-grade results to improve student learning. In both intervention years (2015 and 2016), EducAcción-PRI provided a two-day training workshop in how to interpret and use the results of the end-of-grade assessments as soon as the examination reports were available. The first day of training delivered instruction to schools in how to interpret their EOG reports. During the second day, coaches worked with schools to develop school action plans (SAP) that addressed the needs identified by schools through analysis of their test score results. Schools also held a meeting with teachers and parents to share the test results with the community, including how their school compared with others, areas in which students were lagging, and how parents could help students at home. Finally, the EducAcción-PRI team assigned a coach to each school to provide ongoing technical assistance and coaching to the schools on the implementation of their action plans. Each school's coach was supposed to meet with the school's principal twice a month. The meetings were structured to address implementation of the action plan activities and decision making to improve student learning, with a particular focus on reading skills.

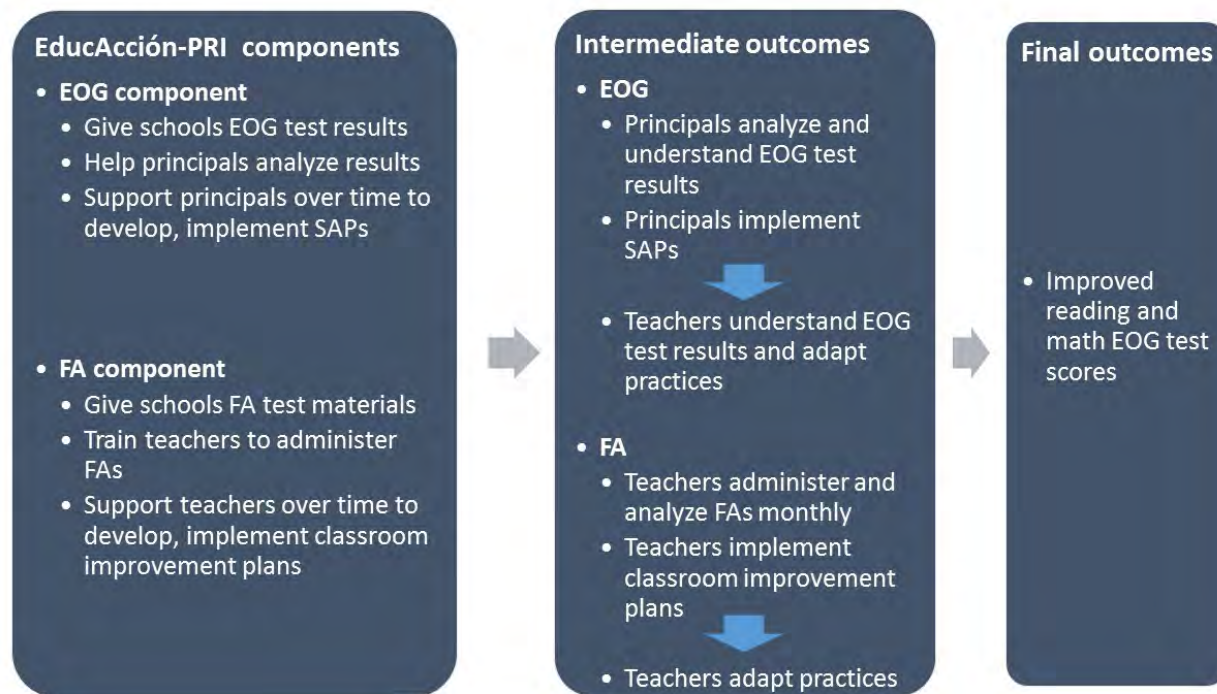
Component 2. Supporting the use of FA assessment results. The second component provided materials, training, and support to teachers to administer formative assessments and integrate the results into their classroom practice. In particular, the support helped teachers use the monthly formative assessments in reading and math to adapt their instruction according to their students' needs. The EducAcción-PRI team provided schools with reading and math formative tests for all students in grades 1 through 6. In addition, teachers received a written copy of the national curriculum standards as well as training to pace their lessons throughout the year in step with the national curriculum. The intervention included a series of training workshops (two 16-hour workshops and two 8-hour workshops) that focused on the interpretation of the formative assessment results and how to use them to guide instruction. In addition, twice a month, each school's EducAcción-PRI coach met with teachers to monitor their implementation and scoring of the formative assessments and to coach them in interpreting results and adjusting their lesson plans and educational strategies accordingly. To reduce the likelihood of contamination of treatment across groups, no coaches worked in both treatment groups.

In Figure I.2, we present the theory of change for EducAcción-PRI. It shows that the EOG component was expected to lead principals to analyze and understand EOG test results and to

¹ This was referred to as *apoyo pedagógico* during project implementation and as pedagogical support in previous project reports. Coaches were referred to as *asesores pedagógicos* during project implementation and as pedagogical advisors in previous project reports.

develop and implement SAPs, which would in turn lead teachers to adapt their teaching practices. The FA component was expected to lead teachers to administer and analyze monthly FAs and to develop and implement classroom improvement plans, which would also lead to changes in teaching practices. Both processes were expected to lead to improved EOG test scores in reading and math.

Figure 1.2. Theory of change for EducAcción-PRI



D. Evaluation questions

The purpose of this study was to estimate the impacts of providing print materials, teacher training, and ongoing coaching in the use of summative and formative assessments for teaching and learning. The study design allowed for estimation of the impacts of the EOG component on its own and the added impacts of administering FAs and providing support for them when the EOG component is already in place. Specific outcomes of interest include teachers’ use of assessment results in instruction and students’ early grade reading and math skills. We used results from the secure administration of the national EOG tests at endline as the key learning outcome. The study aimed to answer the primary evaluation questions² below through quantitative and qualitative analysis. We also used existing test score data to explore a set of secondary evaluation questions on the role of the assessments used before the start of EducAcción-PRI. These are shown in Chapter II, table II.2.

² We have rephrased the original evaluation questions to these to make them more specific, but their meaning remains the same.

Implementation

1. **Were the interventions implemented as intended?** Did the offer of focused coaching and summative assessment results increase teachers' and principals' access to these supports and materials? Did the offer of supports for administering formative assessments increase their availability and use in the target schools?
2. **Did these interventions result in changes in instructional practice?** How do support and materials influence teachers' and principals' use of EOG tests and FAs in instruction and student support?

Impact

1. **Did the EOG intervention improve learning?** What impact does providing feedback on end-of-grade test results and coaching for teachers have on student reading and math scores relative to prevailing practice?
2. **Did the FA intervention improve learning?** What added impact does providing formative assessment coaching and materials for teachers have on student reading and math scores?

Cost-effectiveness

1. **How much does it cost to provide the materials and support needed for FA and EOG assessments? Are effects large enough to justify these costs?**

We describe our methods for answering these questions in Chapter II.

E. Summary of findings

EducAcción-PRI was implemented as planned. As expected based on their treatment assignment, schools in the EOG-only or EOG and FA groups received EOG results for their schools, training on how to analyze the results, and training on how to develop and implement a school action plan based on those results. Principals reported that they implemented their school action plans, and teachers reported that they adjusted their lesson plans according to EOG results.

Teachers in EOG and FA schools received the FA materials that they needed as well as the support that they required to develop and implement classroom improvement plans based on FA results. Teachers reported that they consulted their pacing guides more frequently than before to keep up with the FAs and adjusted their teaching based on FA results.

On the endline survey, nearly all principals and teachers from schools assigned to one of the two treatment groups indicated that they had received training on EOG or FA, but almost no school staff from the prevailing practice group received training on either assessment type. Principals and teachers also indicated that they received EOG test results and FA materials as expected according to their treatment assignment.

The EducAcción-PRI intervention improved students' test scores in reading and math. The EOG intervention improved grade 3 students' scores on the EOG reading test by 0.15 standard deviations. The FA intervention improved scores on that test by 0.14 standard deviations.

Impacts on EOG math test scores were similar in magnitude but not statistically significant. However, the combined EOG and FA interventions increased reading scores by 0.25 standard deviations.

Impacts can be described in other ways. The EOG intervention increased the percentage of questions answered correctly on the reading test by 3.0 percentage points, and the FA intervention increased the percentage by 3.6 percentage points. The combined effect of the two interventions increased the percentage of math questions answered correctly by 5 percentage points. In Honduras, EOG results are typically reported by placing each student in one of four performance levels. The EOG intervention moved 8 percent of students to a higher performance level for reading and math, while the FA intervention moved 7 percent of students in both subjects. Another way to interpret impacts is to express them in terms of months of learning. The impacts of 0.15 standard deviations for EOG and 0.14 for FA correspond to 2.0 and 1.9 months of learning, respectively.³

We evaluated whether impacts varied for subgroups identified during the evaluation design phase. We found that impacts did not vary by gender but did vary by school location. The EOG intervention had significant impacts in urban schools of 0.19 standard deviations on reading and 0.18 on math, but impacts were smaller and not significant in rural schools. There are a number of explanations for this variation in impacts. First, most rural principals are also classroom teachers, perhaps limiting the time available to them to focus on implementing school action plans. Most urban principals are not classroom teachers, perhaps affording them more time to work with teachers on implementing their school's action plan. Furthermore, analyzing EOG results may be more helpful in urban areas where principals manage larger schools with more students and more teachers.

The FA intervention had large, significant impacts in rural schools of 0.38 standard deviations but no impact in urban schools. This variation in impacts could be related to differences in characteristics between urban and rural schools. On average, teachers in rural schools are less experienced than teachers in urban schools. FAs may be a more important support for the relatively inexperienced teachers in rural schools. Another possible factor is that rural teachers are less likely to work a second teaching job than urban teachers; these rural teachers might have more time to dedicate to analyzing FA results or implementing their classroom improvement plan.

³ Conversion of effect sizes to months of learning is discussed in Chapter 2.

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

II. STUDY DESIGN, DATA, AND METHODS

A. Impact evaluation design

1. Experimental design

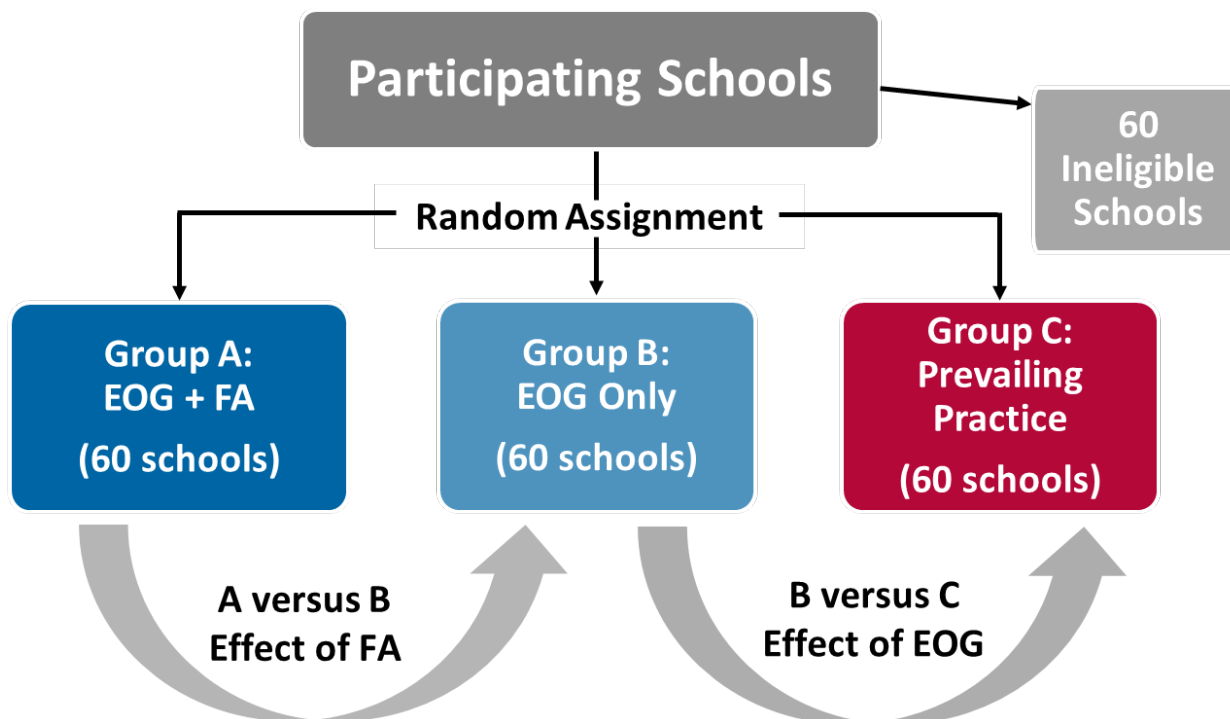
We answer the impact questions listed above by using a three-arm randomized controlled trial, complemented by qualitative analysis to understand program implementation. We randomly assigned each of the study's 180 schools to one of three groups, as described in Chapter I and shown in Figure II.1.⁴

The three-arm design makes it possible to answer the impact questions by analyzing two contrasts. First, we can compare outcomes of teachers and students in schools assigned to the EOG and FA group to those assigned to the EOG-only group (Group A versus Group B) to estimate the additional impact of the FA intervention (evaluation question 4), holding constant the impacts of the EOG intervention, which both groups received. Second, we can compare outcomes of teachers and students in schools assigned to the EOG-only group to those assigned to the prevailing practice group (Group B versus Group C) to estimate the impact of the EOG support intervention compared to the prevailing practices of the MOE and other organizations (evaluation question 3).

The MOE's ongoing support for the use of EOG and FA includes some EOG testing in a sample of schools and the distribution of school test results in print to that sample of schools and the publication of FA materials online. The MOE did not offer training on the use of assessment to schools assigned to Group C during the evaluation period.

⁴ This study is registered in the American Economic Association (AEA) RCT Registry and the unique identifying number is: AEARCTR-0000780. It was initially registered October 5, 2015.

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



Random assignment allows for the unbiased estimation of the impact of the interventions on the outcomes of interest. When implemented correctly, random assignment generates research groups with no systematic differences 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 may be attributed to the differences between the interventions, not to other characteristics of students, teachers, or schools in the sample. Even though random assignment generates groups without *systematic* differences between the groups, we took the additional step of stratification (discussed next) to minimize chance differences and therefore to increase statistical precision.

2. Stratified random assignment

To improve the expected balance of key baseline characteristics among the three treatment groups, we formed strata of schools within each of the four locations (Tegucigalpa and La Ceiba, which were mainly urban, and Santa Barbara and Lempira, which were mainly rural). We also grouped schools by the extent to which they had used formative and end-of-grade assessments at baseline and by the schools' average test scores on the 2013 EOG test, the most recent test for which results were available at the time of sample selection.⁵ We first constructed an index that characterized each school's degree of exposure to FA and EOG materials and supports as of the end of the 2014 academic year, which was eight months before random assignment. Then, we divided each evaluation area's schools into two groups of schools based on the previous exposure index and, within those groups, ranked schools according to their test score on the 2013 EOG test. We then created triplets of schools with similar test scores within the eight groups of

⁵ To measure performance, we averaged schools' reading and math test scores over grades 1, 2, and 3.

high or low use of assessments and randomized within each stratum. For example, the three schools with highest test scores among those that used assessments more intensively in Tegucigalpa would form one triplet, with each school in the triplet assigned at random to a different treatment group. We expect this approach to generate treatment groups balanced on geographic area, baseline use of assessments, and baseline school performance. We present the results of our tests of baseline equivalence in Appendix A. We found that the stratified random assignment produced three groups that were similar on school, teacher, and student characteristics at baseline.

3. Sample selection and cohort of interest

We selected an initial sample of 240 schools that were low-performing and not too small. The sample included all public schools in the four areas selected for the evaluation that, according to records from 2013, had at least 10 grade 1 students in 2013 according to records from the 2013 EOG test; and test scores in the bottom three quintiles of performance. Principals in all 240 schools consented to participate in data collection and in the experiment, permitting us to collect baseline data in all 240 schools. In addition, all the teachers invited by our data collection partner to participate in the survey consented and participated; one teacher could not be located for participation.

We then trimmed the sample to 180 schools for the study based on feasibility of implementation and previous exposure to the intervention. We first eliminated schools that were infeasible to visit because they were in areas that were highly prone to violence. Then, we eliminated schools that had worked previously with EducAcción to avoid interrupting the EducAcción project team's ongoing work in such schools, and to avoid including schools with significant previous exposure to activities similar to those in the treatment condition. For the remaining schools, we analyzed baseline data on principals' and teachers' use of formative assessments and EOG test results and selected those within the initial sample that used the formative assessment or EOG test results least intensively. The selection of schools with least intensive baseline use of assessments allowed for a cleaner contrast between the treatment groups and the control group, improving our ability to detect the interventions' impacts.

The study followed the cohort of students who were enrolled in grade 2 as of the last day before the intervention began—May 31, 2015; we chose this cohort because it was the youngest cohort for which we could collect EOG baseline data from the year before launch of the EducAcción-PRI. Any students who dropped out or transferred out before that date or enrolled after that date are not considered part of the sample. We measured learning after half a year of support in grade 2 (2015) and a full year of support in grade 3 (2016).

Treatment group crossovers among students were rare, less than 2 percent of the sample. We used students' unique ID codes when possible to find students' EOG test scores, even if students transferred to a different school. Of the 8,119 students included in the evaluation sample, 1,120 were no longer in their original schools at the end of the evaluation. Of those, 140 (1.7 percent of the full sample) were non-compliers, also referred to as crossovers, meaning that they moved to a school in a different treatment arm. The moves were symmetric, in that approximately the same number of students (8 to 13) moved between each pair of treatment arms. As noted below, we categorized all students based on the treatment status to which their original school was

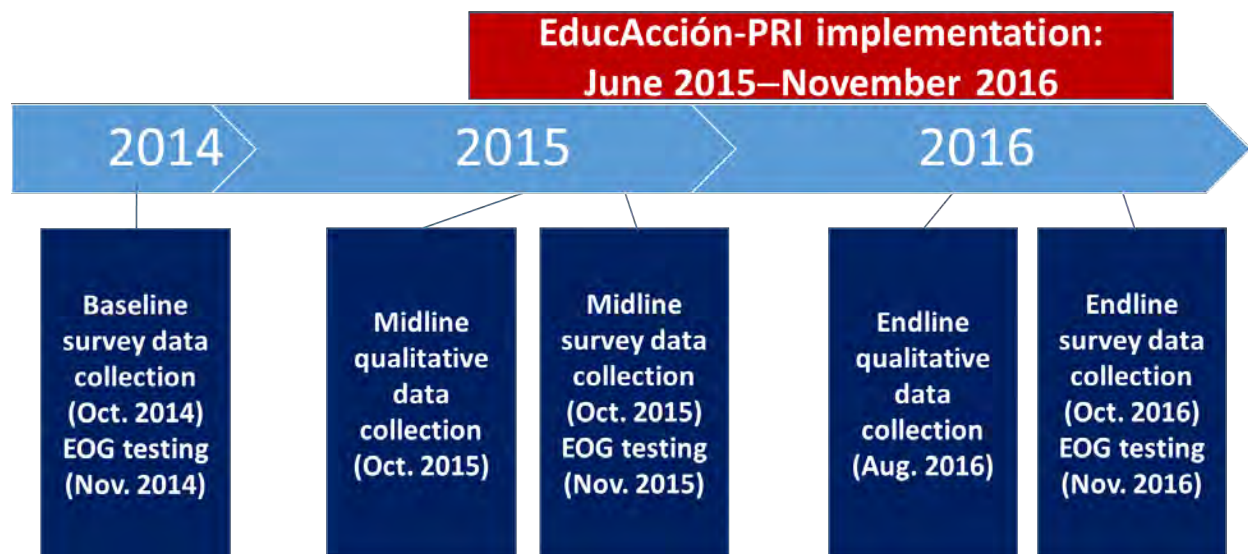
randomized. The remaining 1,011 school leavers were split between 342 dropouts who were not enrolled in any school and 669 students who transferred to a non-study school.

Our findings are generalizable to the study sample: public schools in the bottom three quintiles of performance in one of the four study areas with relatively little prior exposure to activities similar to the treatments, and that are not located in the highest risk areas. Given that the sample includes schools from diverse urban and rural areas, we believe results would be similar in other parts of Honduras and may be similar in other countries in the region.

B. Data for impact, implementation, and cost analyses

We gathered a combination of survey, student roster, and test score data to carry out the quantitative analysis. Apart from the focus groups and testing, we visited each study school toward the end of the academic year in October of 2014 (baseline), 2015 (midline), and 2016 (endline). During each visit, the data collection team surveyed the school’s principal, interviewed two teachers teaching the current grade of the cohort of interest (or one if only one teacher taught that grade), and updated student roster information. We obtained EOG test data for each year. In Figure II.2, we show the timeline of the intervention activities and data collection. EOG test administration varied from year to year, as described in more detail below. The survey instruments, data documentation, and data files used for this evaluation are available online upon request as restricted use data files at USAID’s Development Data Library website (<http://data.usaid.gov>).

Figure II.2. Timeline of intervention and data collection activities



1. Survey data

Our local data collection partner, Espirállica, contacted each school in advance to notify the principal of the visit and to find out at what time of day the principal and teachers would be available to take the surveys. During each visit, the data collection staff interviewed the principal and worked with school staff to update our list of enrolled students. The data collection staff then asked for a list of teachers in the target grade and randomly selected two of those teachers to

survey. In small schools with just one teacher in the grade, that teacher would be the only interviewee.

The principal surveys covered school characteristics (number of teachers, quantity and condition of school infrastructure), principal's and teachers' access to training on EOG and formative assessments, use of EOG and formative assessments, and principals' years of experience in education and at the school. All principals consented for their school to participate in data collection for all three rounds of data collection, and the response rates were 100 percent in each group in every round, with the exception of one Group C school that closed during the 2016 school year (the final year) and therefore did not participate in survey data collection at endline.

The teacher surveys covered teachers' access to training on EOG and formative assessments, access to printed copies of formative assessments, access to EOG assessment results, use of both types of assessments, teachers' time use, work-related expenses, and teacher characteristics (for example, demographics, education, and teaching experience). The teacher survey response rate was 99 or 100 percent in each group in every round. For every school, at least one teacher responded to the survey in every round.

2. Student test data

Test administration. In most years since 2007, the Ministry of Education, with international donor funding from several sources, has supported administration of the EOG tests. *Mejorando el Impacto al Desempeño Estudiantil de Honduras* (MIDEH) developed and has administered the tests. For this evaluation, we use EOG test data from four academic years: 2013 through 2016. The items in the tests varied from year to year; however, all items were drawn from the same item bank that MIDEH developed and considered to be equally difficult.

EOG test administration has varied from year to year. In recent years, including the entire study period, MIDEH coordinated secure test administration (STA) in a randomly selected sample of schools. In those schools, testing is made secure for the STA group by sending the exam booklets to the schools only on the day the test is administered, thereby minimizing any school staff or student opportunities to review the questions in advance. In addition, MIDEH provides a cadre of independent observers to monitor test administration and to conduct independent data entry off site. In 2013 and 2014, schools that did not participate in STA testing were expected to participate in census testing. Census testing is less secure; the test booklets arrive two to five days ahead of the test date in some cases, community volunteers observe testing, and teachers perform data entry on-site.

All test data used in the study were from secure administrations and endline testing featured extra monitoring. Study schools participated in STA and census testing in 2013 and 2014; we use those test results as baseline controls. The EducAcción and MIDEH project teams worked together to conduct testing under STA conditions in 2015 (in Groups A and B) and 2016 (in all three groups); we use results from the securely administered 2016 test as outcome measures to estimate impacts.

To verify objectivity given that the implementer organized the endline testing—which we use as our main outcome measure for the impact evaluation—representatives from Espirálca,

which is part of our research team, observed and participated at key junctures throughout the process. Espirállica staff attended randomly selected training sessions at which testing monitors received training, observed test administration at randomly selected schools, observed data entry, and replicated data entry from a randomly selected sample of hard-copy tests. Espirállica reported that everything observed by its staff was consistent with objective, unbiased administration of the tests and data entry.

Test data. The study’s main outcome measures are grade 3 EOG test scores in reading and math for the study cohort in 2016. As described in greater detail below, we controlled for baseline performance by using school-level mean test scores from the end of the study cohort’s grade 1 year, before random assignment occurred. Test score data were available for 88 percent of schools at baseline and for 80 percent of students at endline, as summarized in Table II.1 on response rates.

- **Baseline test data.** In 2014, the Ministry of Education coordinated STA testing in a sample of schools and census testing in all other schools. Nonetheless, even though all study schools’ principals reported participation in EOG testing in 2014, EOG test data were available for all but 22 of the 180 study schools that year for our cohort of interest. Of the 158 study schools for which baseline data were available, 15 had STA data and 143 had census data.
- **Endline test data.** In 2016, all but one study school, which closed during 2016, participated in EOG testing. The Ministry of Education did not offer census testing, but the EducAcción project team worked with the MIDEH project team to coordinate EOG testing in all study schools. Eighty percent of student sample participated in EOG testing. Some of the students who did not participate in testing had changed schools, though we used unique student identification codes to track students who had transferred to a different study school, thus reducing attrition in our analysis sample.

Table II.1. Response rates for EOG testing at school and student levels

Data collection round		FA and EOG (A)	EOG-only (B)	Prevailing practice (C)	All
Number in sample	Schools	60	60	60	180
	Students	2,747	2,841	2,395	7,983
Response rates					
Baseline EOG testing in 2014 (school averages used in analysis)	Percentage of schools	90	88	85	88
Endline EOG testing in 2016 (individual student scores used in analysis)	Percentage of students	79	80	80	80

Note: The table shows the number of *schools* with baseline data because baseline test score controls are at the school level. The table shows the number of *students* with endline data because the outcome test score measure is at the student level.

The sample includes students who were enrolled in a school that had been randomized to one of the three groups and who were in grade 2 as of May 31, 2015, the day before the intervention began. For any school without EOG test results from 2014, we imputed values by using average scores from the same school’s previous cohort of grade 1 students, which were available for all study schools.

Response rates listed here reflect the percentage of students included in the analysis. Students who were repeating grade 2 in 2016 were excluded from the numerator of the response rate calculation and from impact analysis because their test score results would not have been comparable to those of the cohort of interest, which was enrolled in grade 3 at endline. Of the 123 repeaters (1.5 percent of the sample), 34 percent were in the EOG and FA group, 27 percent were in the EOG-only group, and 39 percent were in the prevailing practice group. If we include these students in the calculation of the response rate, the overall response rate increases to 81 percent.

We discuss our approach to handling missing data in Appendix B (Analytic Approach).

3. Focus group and interview data

Focus groups and interviews with school and program staff shed light on how the interventions were introduced in the two treatment groups that received EducAcción-PRI support, what happened in the group of prevailing practice schools, and *why* and *how* the interventions might have influenced principal and teacher behavior and student learning.

The study team conducted focus groups and interviews with principals, teachers, coaches, coaches' supervisors, and EducAcción's senior staff at the end of the first and second academic years of the intervention. We conducted focus groups with principals and teachers separately by geographic area and treatment group. We have summarized our qualitative data collection activities in Appendix B. Appendix I includes the discussion guides used for qualitative data collection.

C. Methods for impact and implementation analysis

Because of random assignment, we were able to use simple methods, estimating impacts as the difference between groups in average outcomes at endline. To improve precision, we used a regression model that adjusted for chance differences among study schools observed at baseline and for the design. In Appendix B, we provide more detail on this approach. Regression adjustment did not meaningfully change the results.

Estimates represent the “intent to treat,” meaning that we estimated the impact of receiving the offer of intervention, not necessarily receipt of the intervention itself. However, principals and teachers in very few schools reported participating in activities that were inconsistent with their treatment assignment, and implementation of treatment occurred at high levels of fidelity in both treatment groups such that the interpretation of these effects is nearly the same as the impact of treatment on the treated.

The results presented in the body of the report are not adjusted for nonresponse because response rates were high and not correlated with treatment. However, we show in Appendix D that estimating impacts with nonresponse adjustment weights yielded results similar to the main results.

To help readers interpret findings on impacts on student test scores, we translate effect sizes, which are expressed in standard deviations, into months of growth. Such translations can be helpful to assess the size of the impacts, but require strong assumptions, so they should be interpreted with caution. They are often derived in a different context with a different population than the one intended, with estimates varying from 0.25 to 1.00 standard deviations per year (Hill et al. 2008; Baird and Pane 2018; and Araujo et al. 2018). They also assume that students learn at

a steady pace over time and have been shown to be quite sensitive to grade level (Hill et al. 2008; Baird and Pane 2018).

For the current study, we assume a conversion factor of 0.075 standard deviations per month (or 0.75 standard deviations per academic year). This is derived from a study (LoGerfo et al. 2016) using the Early Childhood Longitudinal Study, a large scale assessment that followed a nationally representative sample of U.S. children from Kindergarten through elementary school and also lies within the range of grade 2 to 3 transitions estimated by Hill et al., which varies by subject (0.60 standard deviations per year for reading and 0.89 standard deviations per year for math).

To analyze qualitative data, we organized the content of focus group and interview transcripts by specific topic and subtopic and by treatment group and geographic area. We then consolidated the findings to identify key themes and patterns in the data with quotes from participants to illustrate the themes. In this report, we present findings from the qualitative data along with the impact evaluation findings, discussing how each one informs the other. More detail on qualitative methods is provided in Appendix B.

Methods for the cost effectiveness analysis are discussed in Chapter VI.

D. Secondary evaluation questions and analysis

Related to the primary evaluation questions and as proposed in our evaluation plan (Glazerman, Liuzzi, and Murray 2016), we took advantage of the rich test score and administrative data available in Honduras to explore three secondary evaluation questions that, in our estimation, could contribute to additional learning about some aspects of the role of EOG testing and related training on the years before the EducAcción-PRI intervention began. We use EOG test data from 2013, 2014, and 2015. In 2013 and 2014, Honduras implemented EOG testing nationwide with students in grades 1 through 9 either through the census-based testing or the more rigorous STA conducted in a sample of schools.⁶ In 2015, MIDEH led STA testing in a sample of schools, but there was no census test that year.

We used the resultant data to answer—to the extent possible—a set of research questions on the importance of the method of test administration (census or STA) and on the potential impacts of MIDEH and EducAcción’s activities in schools before EducAcción-PRI began. In Table II.2, we present a summary of the methods used for the secondary research and results. Appendices E, F, and G provide detailed descriptions of background, methods, and results.

⁶ In schools selected for the STA sample, external observers are at the schools on the day the test is administered to address issues that may arise with test administration. The observers bring the tests with them on the day of the test so that teachers and students will not have the opportunity to prepare for the tests ahead of time. The observers gather the completed tests and then enter the test results off site. In contrast, in schools that participate in census testing, staff administer the tests at their own schools. They receive the tests in advance of the testing date and then enter the test results themselves in what is a less secure process than the process used in the STA schools.

Table II.2. Secondary evaluation questions and findings

Questions	Research methods	Results
1. How useful is national census testing compared to a securely administered and scored test in a sample of schools?	Natural randomized experiment. Compare test score outcomes for schools in the STA sample with similar schools that participated in census testing.	Test scores in census schools are higher than scores in STA schools on average, and the difference is large. The difference in mean scores is equal to 1.3 standard deviations (SD) for reading and 1.5 SD for math (using the SD from the STA distribution). See Appendix E for full results.
2. Does secure administration of an EOG test improve subsequent test score outcomes?	Natural randomized experiment. Use selection in 2013 for STA as the “treatment”, with participation in the 2013 Census as “control”. Compare test score outcomes on the 2014 STA test and repeat for the 2015 census test.	Impacts on the 2014 STA sample were small and not significant with the exception of impacts on grade 2 students, which were significantly negative (-0.11 SD in reading and math). Impacts on the 2015 census test were small and not significant but were positive and significant for grade 2 and 3 reading. See Appendix F for full results.
3. What are the impacts of assessment-related training on test scores?	Instrumental variables, using the index of municipality conditions as an instrument for having worked with the EducAcción or MIDEH projects.	The findings were inconclusive. The index was not a valid instrument for EducAcción participation. It was a valid instrument for MIDEH participation, but the results were unstable and highly sensitive to model specification.

We encountered several unexpected challenges as we conducted the secondary analysis. First, we learned that the STA sample was a true randomized sample only in 2013, limiting the number of years of data we could analyze as a randomized experiment. Second, we had planned to estimate the causal effects of two early assessment–related training programs on test scores by using a two-stage least squares (instrumental variables) approach. In the case of one of the programs, the variable we planned to use as an instrumental variable was not a valid instrument. In the case of the other program, the results were sensitive to model specification and were inconclusive. We present a detailed write-up of this effort in Appendix G.

For secondary question 1, we found that test scores were higher when the census method of administration and scoring was used than under the more secure administration and scoring conditions used in the STA sample. The differences were large, as shown in Table II.2, and were evident in a comparison of similar groups of primary schools. Based on this finding, we recommend that policymakers using EOG tests consider the tradeoffs between large-scale administration using lower-security administration methods and higher-security testing in a sample of schools. In light of these results, we opted to use secure testing methods for the EOG tests administered for the outcome measures for our experimental evaluation of EducAcción-PRI. We present detailed results of the analysis of census and STA scores in Appendix E.

For secondary question 2, we found that the impact of secure testing such as in the STA sample in Honduras led to mixed results on test scores in subsequent years. Impacts on census test scores one year after STA testing were negative, perhaps suggesting a reduction in the manipulation of scores. Impacts on STA scores two years after STA testing were positive for two

of three grades, suggesting a potential lagged impact on learning, which could be explained by schools making use of test score data from earlier years to improve instruction. We present detailed results in Appendix F.

Results for the quasi-experimental analysis for question 3 were ambiguous. Because EducAcción’s municipal index was not a valid instrumental variable, we were unable to estimate the impacts of being selected as one of EducAcción’s target municipalities. MIDEH’s index was a valid instrument, but we found that impact estimates were sensitive to the specification of the instrumental variable. We present detailed results for secondary question 3 in Appendix G.

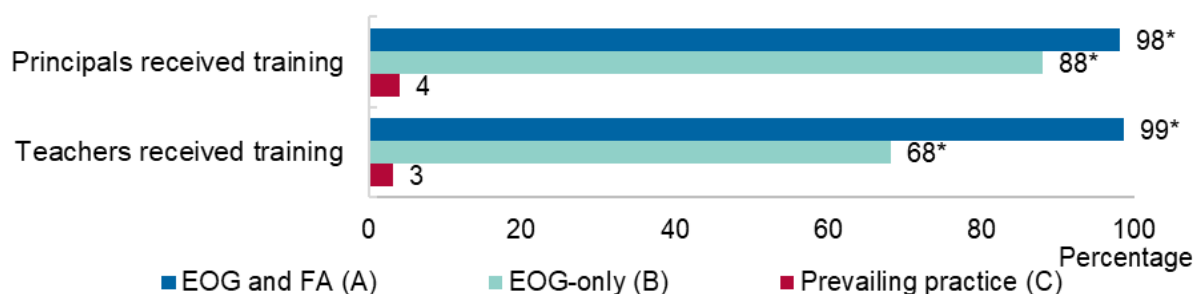
III. WERE THE INTERVENTIONS IMPLEMENTED AS INTENDED?

In this chapter, we discuss the degree to which teachers and principals in schools assigned to EducAcción-PRI received the intervention as designed, and we compare the results to prevailing practice (using the control group). The results presented in this chapter are based on data collected at endline and refer to principals’ and teachers’ experiences in 2016, the second academic year of the intervention and the only year in which the intervention took place throughout the full academic year. During 2016, treatment schools relied on the results of the end-of-grade (EOG) test administered at the end of the previous academic year (2015). We learned about the implementation of EducAcción-PRI from the endline surveys of teachers and principals, from monthly meetings with the implementer during the intervention period, and from the qualitative data collection involving the implementer, principals, and teachers discussed in Chapter II. We present group means, adjusted for stratified random assignment and baseline differences. For every figure presented here, Appendix C includes a table with complete results, including means, differences, and *p*-values associated with the differences.

A. Training and coaching

Most principals and teachers in both treatment groups reported receiving training on EOG assessments or formative assessments FA. EducAcción’s implementation plans called for offering training for principals and teachers in all schools in Groups A (EOG and FA) and B (EOG-only), collectively referred to as EducAcción schools. Most principals and teachers in EducAcción schools indicated that they had received training on EOG or FA. Almost no principals and teachers in Group C (control) schools indicated that they received any training on EOG or FA in the 2016 academic year (those that did receive such training indicated that they received training from groups other than EducAcción). Principals and teachers in Group A (EOG and FA) schools were significantly more likely to have received training than were their counterparts in Group B (EOG-only) schools. The difference was larger for teachers than for principals, probably because of the fact that the EOG intervention focused on principals. In Figure III.1, we present the results for principals and teachers.

Figure III.1. Percentages of teachers and principals who received training in EOG or FA



Source: Principal and Teacher Surveys—Endline 2016.

Note: Bars represent group means that are adjusted for the stratification design with a regression. An asterisk on bar A (B) indicates that the difference in means between groups A and B (B and C) is significant at the 5 percent level, two-tailed test.

Teacher results have been averaged to the school level so each school's teacher data carry the same weight in the analysis.

Principals and teachers report receiving training less often than expected.

EducAcción’s implementation plan called for providing intensive training for principals and teachers at the beginning of each academic year⁷ and continuous support in the form of coaches’ twice-monthly visits to all EducAcción schools. Nonetheless, principals and teachers in EducAcción schools reported that they participated in fewer than one training session on average during the last two months before administration of the endline principal and teacher surveys.

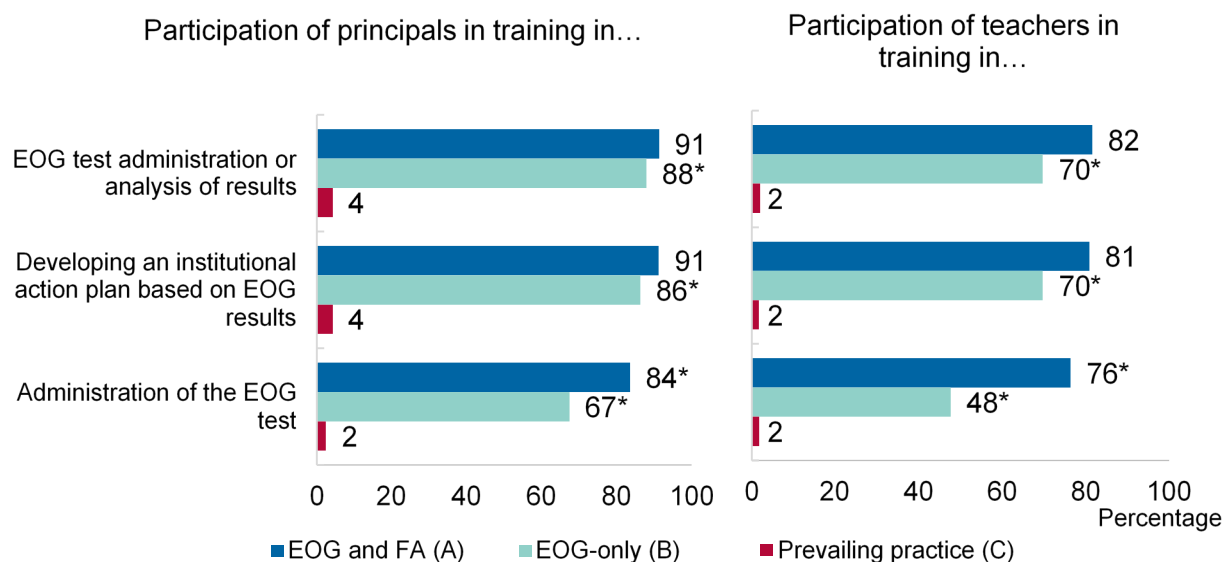
Principals’ and teachers’ comments in focus groups are consistent with visits being less frequent than twice each month, with visits becoming less frequent toward the end of the intervention. Principals and teachers referred to coaches’ visits as *monthly*, suggesting that they did not expect visits twice a month. In addition, principals and teachers noted that they had learned how to analyze and use test results on their own, which could be consistent with a reduced need for frequent visits. One principal from the EOG-only group commented on her improved ability to analyze EOG test results. “I know that initially my knowledge was just about 20 percent, but now, thanks to this whole process, it’s more like 80 percent. I think that before, we really lacked the training to be able to analyze the results.”

Qualitative data also suggest that low participation could be attributable to low levels of principal engagement. In focus groups, coaches noted that it was challenging to meet with principals at some schools. One advisor remarked, “What’s more difficult is to be able to track down and charm the principals so that they might develop their plan. . . They’re absent a lot. I call and call and they are always out. . . never at their school.” Coaches in the EOG-only group also noted that they were limited by the expectation that they work with principals, but not with teachers.

Consistent with the assigned interventions, almost all principals and most teachers in EducAcción schools reported the receipt of training on key elements related to using EOG test results. Close to 90 percent of principals in EducAcción schools reported that they received training related to EOG testing at some point in the second year of program rollout. In addition, most teachers in those schools reported that they received training on EOG testing. Given that the EOG intervention focused on principals, it was unsurprising that teachers were less likely to have received any training on EOG or any specific elements of the training (Appendix C presents full results for principals and teachers). Survey results shown in Figure III.2 suggest that, for nearly all principals and teachers who participated in any training related to EOG assessment, their training included the development of a school action plan and the adaptation of lesson plans based on EOG results. The majority also received training in administering the EOG assessment. Although principals and teachers in the EOG-only group should have received the same EOG-related training as principals and teachers in the EOG and FA group, fewer reported receiving the training; the difference is greater among teachers. Almost no principals and teachers from the prevailing practice group received training related to EOG assessments.

⁷ The academic year runs from February through November in Honduras. EducAcción offered the initial training in June 2015, the month it began EducAcción-PRI. For the 2016 academic year, the initial training sessions took place in February.

Figure III.2. EOG training content



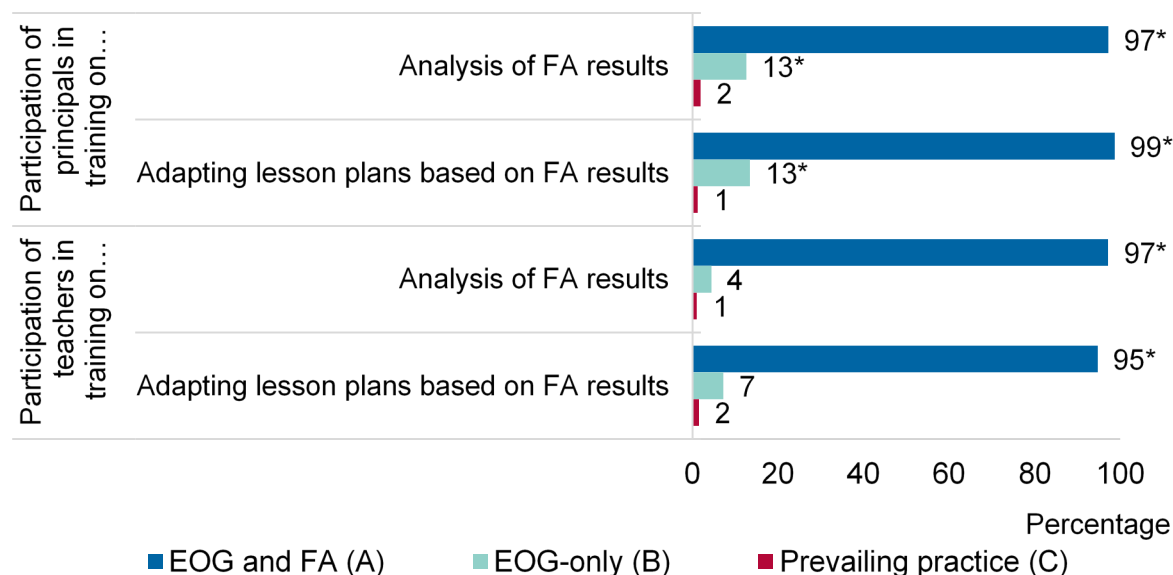
Source: Principal Survey—Endline 2016. The prevailing practice school that closed during the endline year did not participate in endline data collection. Responses refer to training during the 2016 academic year.

Note: Bars represent group means that are adjusted for the stratification design with a regression. An asterisk on bar A (B) indicates that the difference in means between groups A and B (B and C) is significant at the 5 percent level, two-tailed test.

Teacher results have been averaged to the school level so that each school's teacher data carry the same weight in the analysis.

Consistent with the assigned intervention, almost all principals and teachers in the EOG and FA group (over 94 percent) reported that they received training on key elements related to using FA in the current year. However, some principals and teachers in the EOG-only group (up to 13 percent) also reported that they received the same training. In Figure III.3, we show the results for all groups. Principals and teachers in the EOG-only group were not targeted to receive training related to FA. However, it could be that, in some cases, principals and teachers from EOG-only schools touched on themes related to formative assessments as part of their training on EOG testing. It is also possible that respondents may have indicated that they received training on FA when they were unsure of what specific topics had been covered. In focus groups, no principals or teachers from EOG-only schools indicated that they received training on or other support for formative assessment. Coaches from the EOG-only group who participated in focus groups emphasized that they were careful not to encourage or support the use of FA in their schools, in some cases going so far as to discourage interested teachers from using FA.

Figure III.3. FA training content



Source: Principal and Teacher Surveys—Endline 2016. The prevailing practice school that closed during the endline year did not participate in endline data collection.

Note: Bars represent group means that are adjusted for the stratification design with a regression. An asterisk on bar A (B) indicates that the difference in means between groups A and B (B and C) is significant at the 5 percent level, two-tailed test.

Teacher results have been averaged to the school level so each school's teacher data carry the same weight in the analysis.

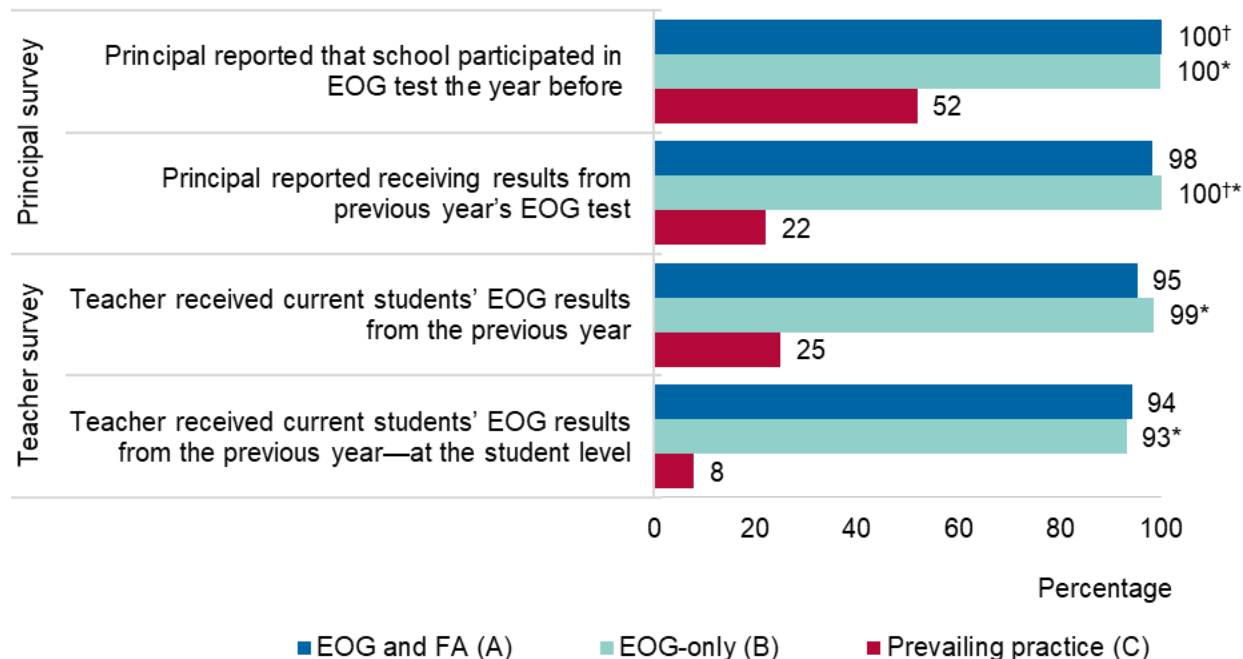
B. Access to EOG assessment results

At endline, nearly all the EducAcción schools indicated that they participated in EOG testing at the end of the previous year, which was the first year of the intervention, but only about half of the prevailing practice schools did so. The difference was consistent with the program implementation plans because EducAcción organized EOG testing in all treatment schools in 2015 to generate EOG data for EducAcción schools in 2016, the second year of EducAcción-PRI, whereas the only prevailing practice schools that participated in EOG testing in 2015 were those that were randomly selected as part of the national STA sample.

Of schools that participated in EOG testing in the previous year, principals and teachers in EducAcción schools were more likely to receive results of those tests in 2016. In 2016, nearly all principals and teachers in EducAcción schools reported that they received the results of the previous year's test, but less than half of the principals or teachers in prevailing practice schools that participated in EOG testing reported that they received the results. In focus groups, the majority of principals and teachers in prevailing practice schools stated that they had not seen their school's results and did not know how to access them, although some were aware that the results were available online. One teacher from a prevailing practice school expressed frustration at not having the results, saying, "Ever since we started with these tests with our children, we haven't seen results. They make demands of teachers and of students to perform on these tests, but then there aren't results to motivate you because with these tests." In Figure III.4,

we show access to EOG scores by study group for principals and teachers. Detailed results, including *p*-values, appear in Appendix Table C.4.

Figure III.4. Access to EOG results



Source: Principal and Teacher Surveys—Endline 2016. The prevailing practice school that closed during 2016 did not participate in endline data collection.

Note: Bars represent group means that are adjusted for the stratification design with a regression. An asterisk on bar A (B) indicates that the difference in means between groups A and B (B and C) is significant at the 5 percent level, two-tailed test.

Teacher results have been averaged to the school level so each school's teacher data carry the same weight in the analysis.

[†] Indicates that the regression-adjusted score was rounded down to 100.

Teachers and principals in EducAcción schools considered it important to receive test results by the beginning of the year. In 84 percent of EducAcción schools, principals indicated that they received EOG test results in either February or March. In teacher and principal focus groups, participants mentioned that they would find EOG results most useful if they received them before the start of the school year (for example, by January), thereby allowing them to incorporate the results into their planning before the start of the school year and develop student-specific learning plans earlier in the year. One principal remarked that she would like to be able to access results online.

C. Access to and use of FA materials

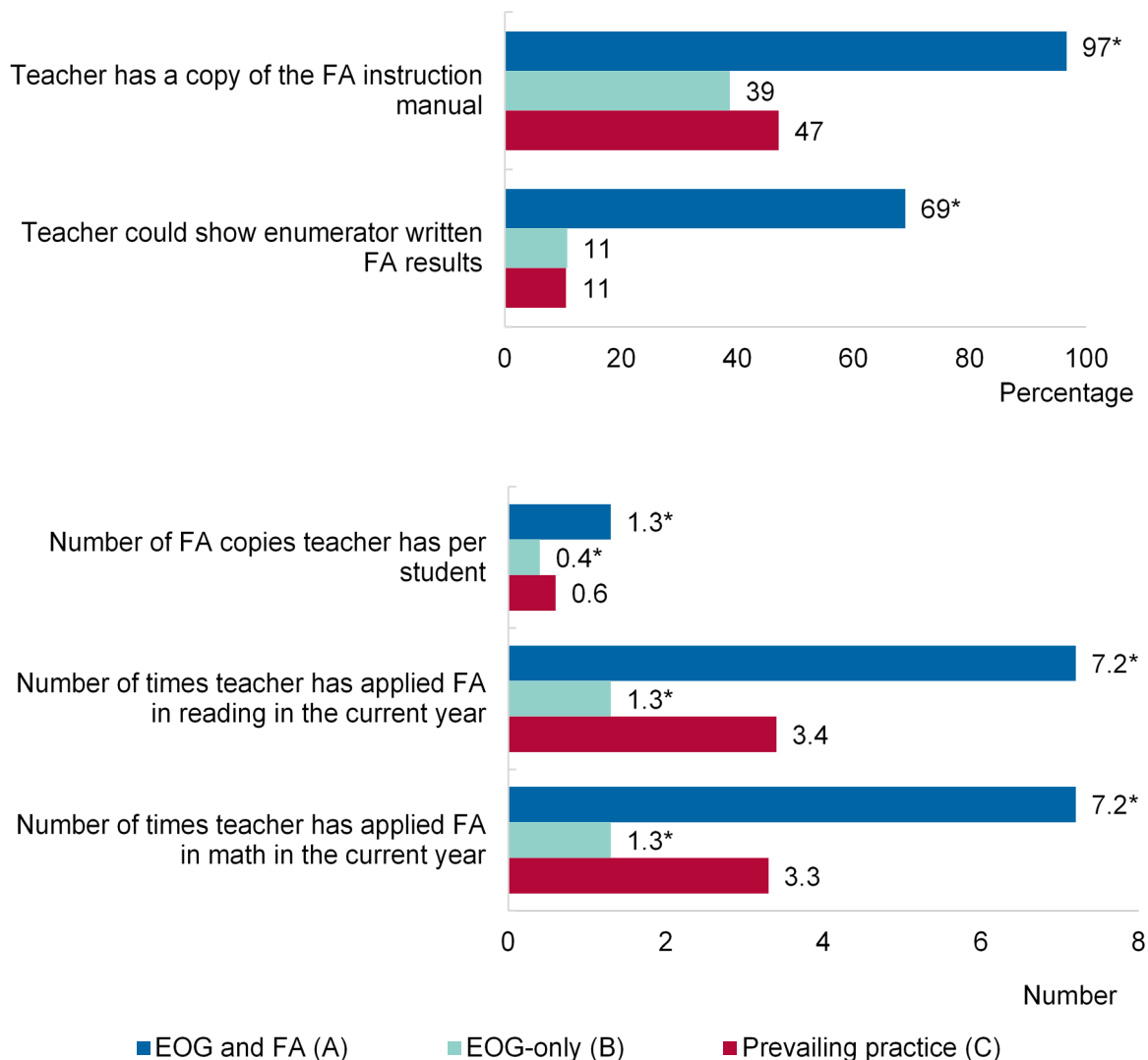
Nearly all teachers in the EOG and FA group had access to FA materials. Almost all teachers in schools receiving both the EOG and FA interventions had the instruction manual and sufficient copies of the FA tests for their students. In contrast, fewer than half of the teachers in

EOG-only and prevailing practice schools had the instruction manual, and they had roughly one copy of the set of FA tests per every two students (Figure III.5).

Teachers in EOG and FA schools who administered formative assessments were on track to administer all the FAs scheduled (or recommended) that year. On average, teachers in the EOG and FA group had administered FA 7.2 times in both reading and math—the amount expected given that there are eight FAs per school year and that data collection took place in the last weeks of the school year. Teachers in the EOG and FA schools were more likely than teachers in the EOG-only or prevailing practice groups to be able to show enumerators where they had recorded FA results, suggesting that teachers without access to training on FA were less likely to administer or record FA results. Teachers in the EOG-only and prevailing practice groups completed less than half of the FAs expected.

Teachers in the EOG-only group had less access to FAs and used FAs less frequently than teachers in the prevailing practice group. In training, neither the evaluation team nor EducAcción-PRI staff asked coaches in EOG-only schools to discourage the use of FAs but rather not to promote the use of FAs actively; nonetheless, some coaches—aware of the evaluation’s experimental design—revealed that they did in fact discourage teachers in their schools from the use of FAs. Such behavior by the coaches likely led to the significant differences reported with respect to access to and use of FA in schools in the EOG-only as compared to prevailing practice groups (Figure III.5). We return to this issue in discussing the study’s conclusions in Chapter VII.

Figure III.5. FA materials and application



Source: Teacher Survey—Endline 2016. The Group C school that closed during the endline year did not participate in endline data collection.

Note: Bars represent group means that are adjusted for the stratification design with a regression. An asterisk on bar A (B) indicates that the difference in means between groups A and B (B and C) is significant at the 5 percent level, two-tailed test.

Teacher results have been averaged to the school level so each school's teacher data carry the same weight in the analysis.

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

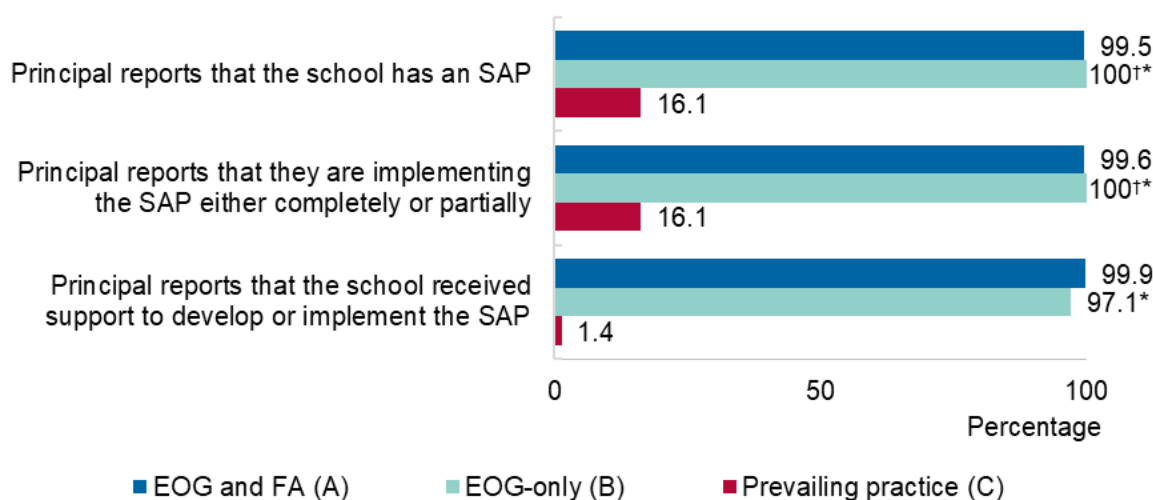
IV. DID THE INTERVENTION CHANGE INSTRUCTIONAL PRACTICE?

In Chapter III, we saw that principals and teachers in nearly all the EducAcción-PRI schools reported that they received end-of-grade (EOG) test results and administered formative assessments (FA) in accordance with the intervention plans. In this chapter, we present results on whether principals and teachers changed their instructional practices based on their analysis of EOG and FA results and how they did so. More details on primary findings related to changes in instructional practices appear in Appendix Tables C.6, C.7, and C.8.

A. Use of EOG test results for school management and teaching

Principals in nearly all EducAcción-PRI schools reported that their school had a school action plan (SAP) and that they were implementing at least some part of the plan. Principals in nearly all EducAcción-PRI schools indicated that they had received support to develop or implement the plan. Few prevailing practice schools indicated that they had implemented or were implementing an SAP (Figure IV.1). Almost no principals in prevailing practice schools indicated that they received support to develop or use SAPs.

Figure IV.1. Use of EOG results for School Action Plans (SAPs)



Source: Principal Survey—Endline 2016. The prevailing practice school that closed during the endline year did not participate in endline data collection.

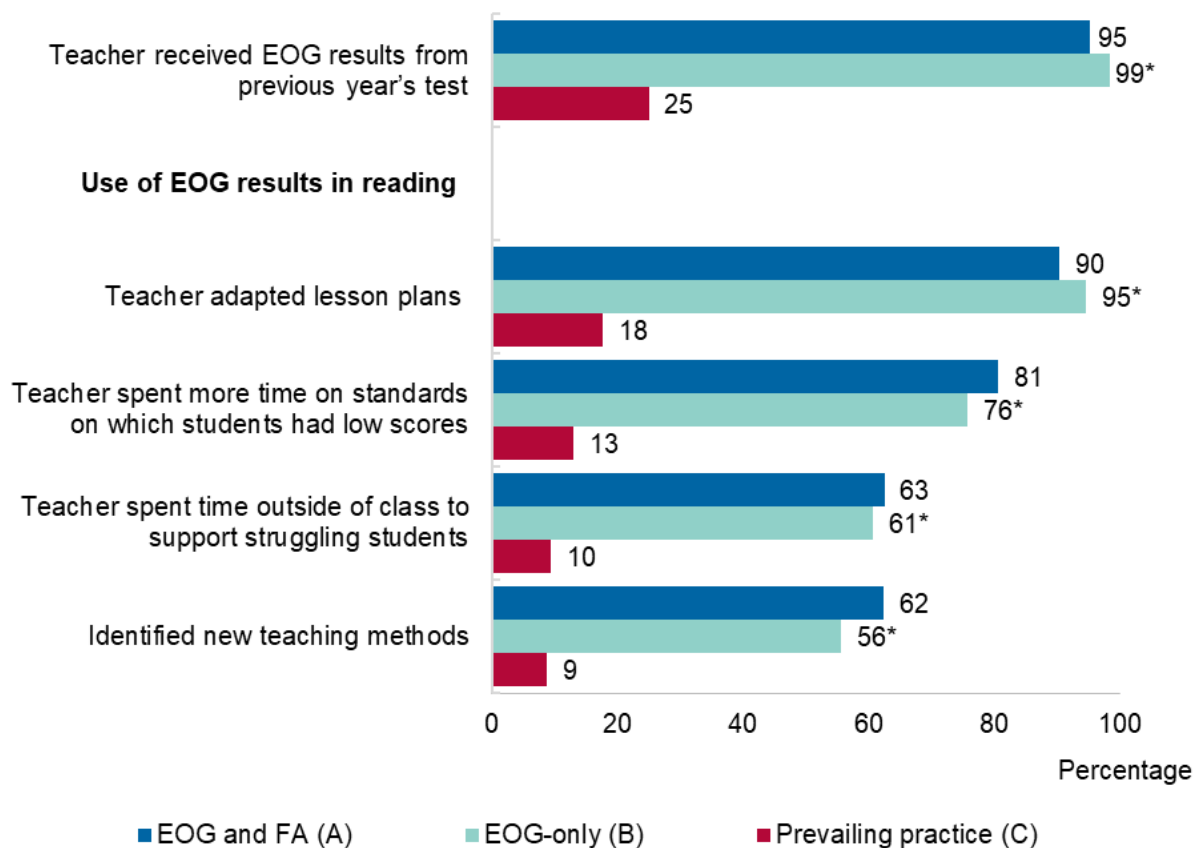
Note: Bars represent group means that are adjusted for the stratification design with a regression. An asterisk on bar A (B) indicates that the difference in means between groups A and B (B and C) is significant at the 5 percent level, two-tailed test.

[†] Indicates that the regression-adjusted score was rounded down to 100.

Nearly all teachers in EducAcción-PRI schools also indicated that they received EOG results and have used those results. In Figure IV.2, we show that more than 90 percent of schools adapted lesson plans to reflect the test results. Teachers were most likely to indicate that they focused their teaching time on the standards on which their students performed poorly—roughly four in five teachers in EducAcción-PRI schools did so. A smaller majority of teachers

spent time outside of class with struggling students or identified new teaching methods or materials based on the EOG results. In Figure IV.2, we show how teachers used the EOG reading results. Teachers used EOG math results in similar ways (Appendix C). Coaches noted in focus groups that principals and teachers valued the EOG test results. They also commented that some school staff indicated that they had manipulated student test scores in the past but would no longer do so now that they understood the purpose of the test.

Figure IV.2. Teachers’ use of EOG results for lesson planning



Source: Teacher Surveys—Endline 2016. The prevailing practice school that closed during the endline year did not participate in endline data collection.

Note: Bars represent group means that are adjusted for the stratification design with a regression. An asterisk on bar A (B) indicates that the difference in means between groups A and B (B and C) is significant at the 5 percent level, two-tailed test.

Teacher results have been averaged to the school level so each school's teacher data carry the same weight in the analysis.

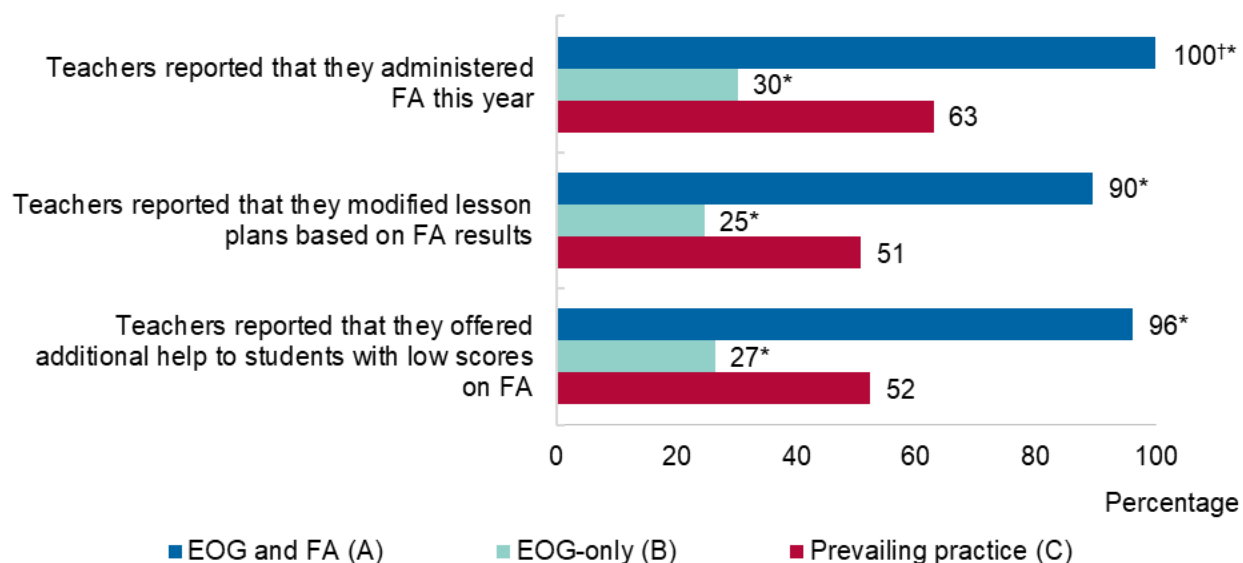
B. Use of FA results to improve teaching and learning

The group that was targeted for EOG and FA used the assessments. Teachers in nearly all schools in the EOG and FA group reported that they administered the FA during the current school year (Figure IV.3). Of those, nearly all reported that they used the test results either to modify lesson plans or provide targeted assistance to students who performed poorly on the FAs.

Also consistent with the implementation plan, teachers in EOG-only schools were much less likely than teachers in EOG and FA schools to use FAs.

Teachers in prevailing practice schools were significantly *more* likely than teachers in EOG-only schools to administer FAs and use the results. The intervention targeted neither of these groups to receive FA support. As noted in the previous chapter, the EOG-only group’s reduced access to and use of FAs appears to reflect the efforts of the coaches assigned to the EOG-only schools to discourage the use of FAs in their schools in order to comply with their understanding of the study protocol, which actually called for advisors to “refrain from encouraging” use of FAs; it did not explicitly call for discouraging their use.

Figure IV.3. Use of FA results to improve teaching and learning



Source: Teacher Survey—Endline 2016. The prevailing practice school that closed during the endline year did not participate in endline data collection.

Note: Bars represent group means that are adjusted for the stratification design with a regression. An asterisk on bar A (B) indicates that the difference in means between groups A and B (B and C) is significant at the 5 percent level, two-tailed test.

Teacher results have been averaged to the school level so each school’s teacher data carry the same weight in the analysis.

[†] Indicates that the regression-adjusted score was rounded down to 100.

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

V. DID THE INTERVENTION IMPROVE LEARNING?

In this chapter, we use vertical bar charts to represent the impacts of EducAcción-PRI on EOG test scores for each of the three experimental groups. The impact on the prevailing practice group (Group C) is defined as zero because this is the reference group for the evaluation and the EducAcción project team did not intervene in those schools. Green bar segments represent the impact of the EOG intervention in the EOG-only (Group B) and EOG and FA (Group A) schools relative to outcomes in the prevailing practice schools. A blue bar segment on top of the green bar segment represents the added impact of the FA intervention in the EOG and FA schools relative to the EOG-only schools. In two cases, we found small negative impacts, which we represent with white shading within bar segments and explain in notes below the figures. Other figures in this chapter represent impacts on performance levels, as measured by EOG test scores (Figures V.2 and V.5).

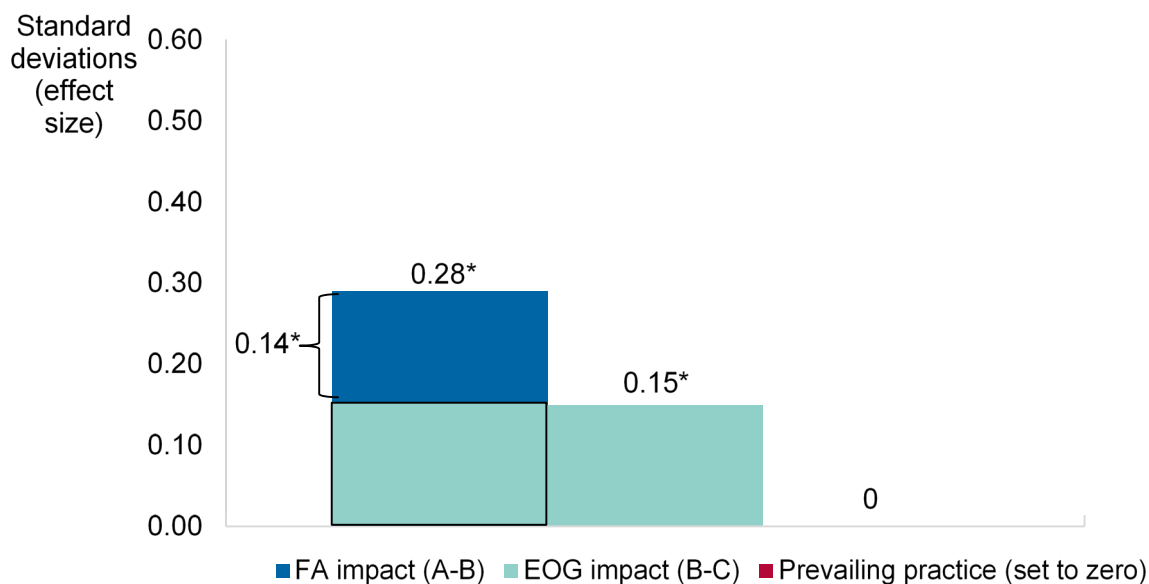
A. Impacts overall

The end-of-grade (EOG) intervention improved students' test scores in reading. The formative assessment (FA) intervention further improved students' reading scores. The EOG-only intervention improved students' reading scores by 0.15 standard deviations (Figure V.1; Appendix D presents more detailed results, including impacts on subscores). Students in schools that received the FA intervention in addition to the EOG intervention improved by an additional 0.14 standard deviations compared to students who received only the EOG intervention. These differences are equivalent to roughly 2.0 and 1.9 months of learning, respectively, if we use an estimate of 0.075 standard deviations per month.⁸ In terms of percentage of items correct, the impact estimates were equivalent to moving students in the prevailing practice group from 58 percent correct on the reading test to 61 percent correct with EOG component and 65 percent correct.⁹

⁸ This estimate of 0.075 standard deviations per month comes from large scale standardized testing in the United States on vertically equated scale, as discussed in Chapter II.

⁹ All impact estimates are regression-adjusted, so the regression-adjusted means for each group are not the same as the raw means. Also, impacts on percentage correct may not align with effect size estimates, which were calculated using scale scores. Scale scores take into account the item response patterns.

Figure V.1. Impacts on reading by intervention and overall



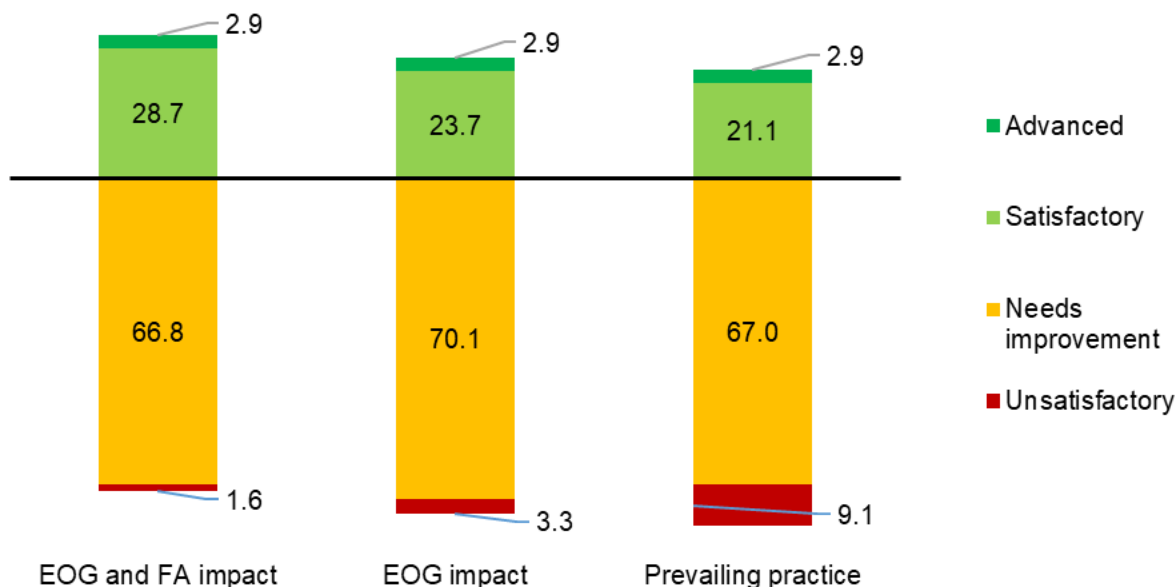
Source: EOG reading assessment 2016.

Note: The mean for the prevailing practice group has been normalized to 0. Impact estimates from a regression that includes baseline controls and adjusts for stratified randomization are shown as effect sizes (standard deviations). * $p < 0.05$, two-tailed test. The separate impact estimates for the EOG and FA interventions do not sum to the impact estimate for the combined impact of both interventions because of rounding.

The impact of the EOG-only intervention alone was large enough to move students to higher reading performance levels. Reporting the impact results as standard deviations is useful for comparisons across studies, but in Honduras, student success is typically described in terms of the percentage of questions answered correctly or a performance level derived from the EOG test. The four performance levels are unsatisfactory, needs improvement, satisfactory and advanced. We translated the impacts estimated above into impacts on changes in performance levels, quantified as the proportion of students moving from one performance level to a higher one. For example, we estimate that the EOG intervention moved 8 percent of students to a higher performance level. The FA intervention moved another 7 percent of students to a higher performance level. In Figure V.2, we show the EOG and FA interventions' impacts on students' performance levels.¹⁰ The EOG intervention moves students out of the unsatisfactory performance level to a higher one but does not affect the percentage of students in the advanced category.

¹⁰ Under different initial conditions, such as having a student population with more or fewer students near the thresholds for passing into a higher level, these estimates would be different, so the calculation should be considered illustrative of this case.

Figure V.2. Impacts on reading performance levels



Source: EOG reading assessment 2016.

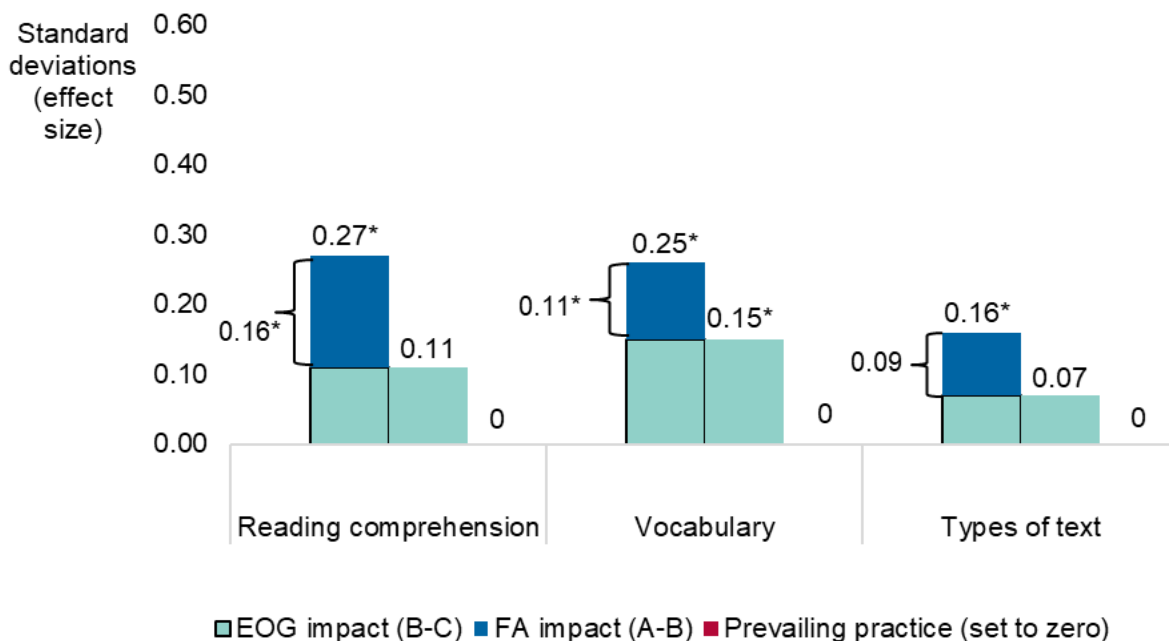
Note: Performance levels for the prevailing practice group are the actual distribution of results for that group. Performance levels for the EOG-only and EOG and FA groups are estimated by adding the estimated impact to scores observed in the prevailing practice group.

B. Impacts by subscore

The EOG intervention had a significant positive impact on students’ vocabulary; impacts on reading comprehension and types of text were positive but not significant. The reading subscores are based on the percentage of questions on vocabulary, comprehension, and types of text that students answered correctly. The measure of types of text refers to students’ ability to read and use diverse types of text, including narrative, descriptive, expository, and persuasive, among others. The EOG intervention increased the percentage of vocabulary questions students answered correctly by 3.0 percentage points, which is equivalent to increasing the subscore by 0.15 standard deviations (see Figure V.3 for subscore results). The EOG intervention did not have a significant impact on other subscores.

The FA intervention had significant positive impacts on students’ reading comprehension and vocabulary. The percentage of reading comprehension questions that students in EOG and FA schools answered correctly was, on average, four points higher than that of students in EOG-only schools, which is equivalent to 0.16 standard deviations. Students in EOG-only schools also scored higher than students in prevailing practice schools, but the difference was not significant at the 5 percent level ($p = 0.064$). The FA intervention’s impact on types of text was positive but not significant.

Figure V.3. Impacts on reading—subscores

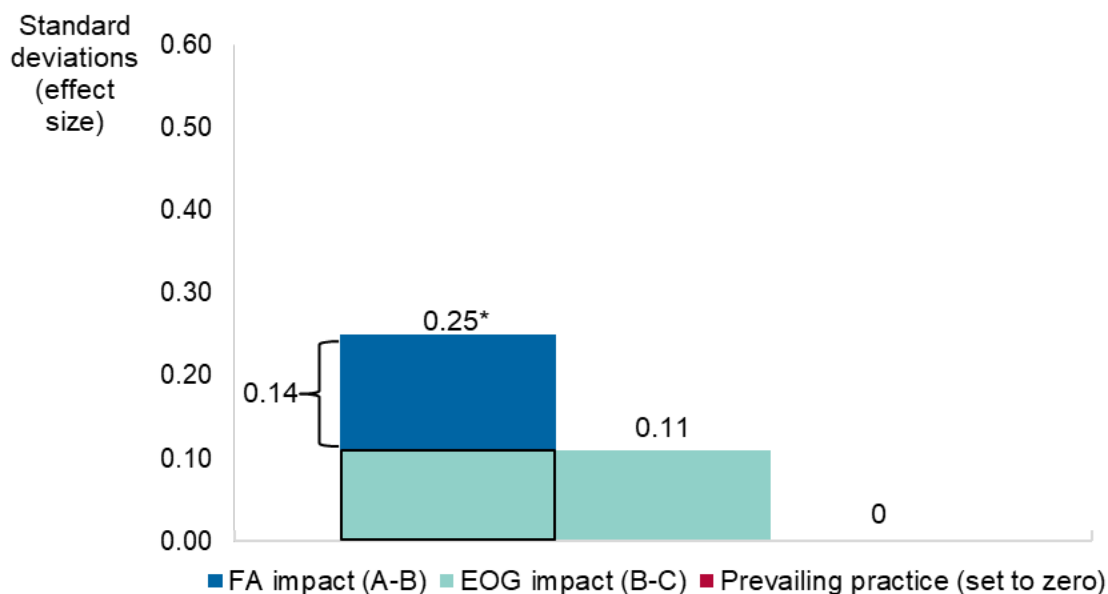


Source: EOG reading assessment 2016.

Note: The mean for the prevailing practice group has been normalized to 0. Impact estimates from a regression that includes baseline controls and adjusts for stratified randomization are shown as effect sizes (standard deviations). * $p < 0.05$, two-tailed test. For the vocabulary subscore, the separate impact estimates for the EOG and FA interventions do not sum to the impact estimate for the combined impact of both interventions because of rounding.

Impacts on math were similar to impacts on reading. Impact estimates for math were of similar magnitude as those for the reading estimates. However, given that the estimated impacts were slightly smaller and less precisely estimated (standard errors were slightly larger), the impacts of Group A relative to B and Group B relative to C were not significant at the 5 percent level. Nonetheless, we believe that the interventions’ impacts on math were substantively similar to their impacts on reading. Recognizing that coaches were aware that the intervention focused on reading and therefore directed their efforts to reading, we are not surprised that impacts on math were smaller. However, the combined impacts of the FA and EOG interventions (Group A compared to C) on math performance were significant. In Figure V.4, we show impacts by interventions and on math overall.

Figure V.4. Impacts on math by intervention and overall

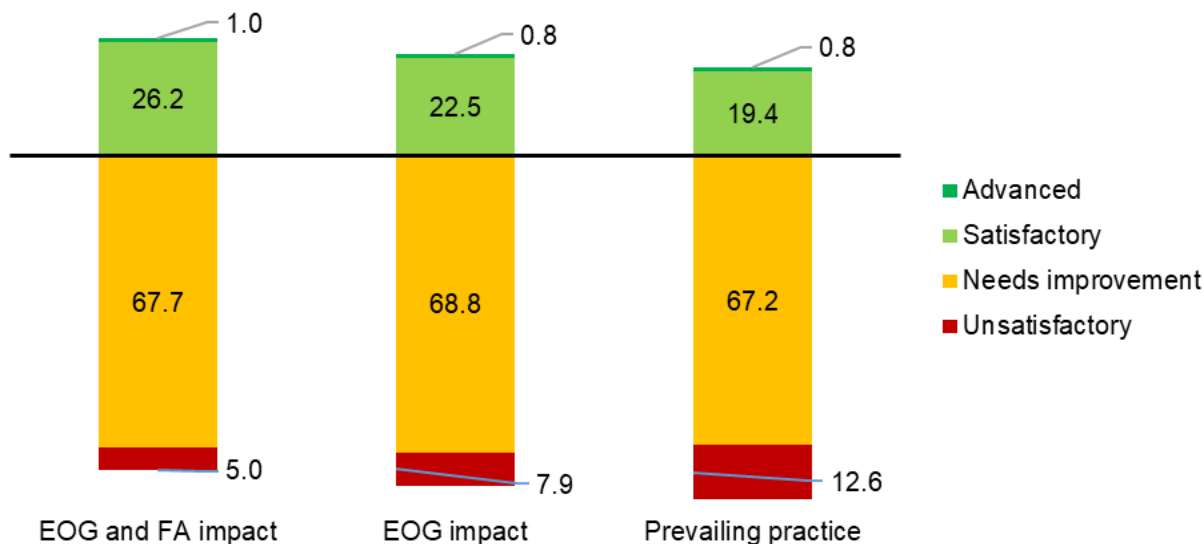


Source: EOG math assessment 2016.

Note: The mean for the prevailing practice group has been normalized to 0. Impact estimates from a regression that includes baseline controls and adjusts for stratified randomization are shown as effect sizes (standard deviations). * $p < 0.05$, two-tailed test.

EducAcción-PRI moved some students to higher math performance levels. Similar to what we found in the case of reading performance, the EOG intervention moved 8 percent of students into a higher performance level in math. The FA intervention moved another 7 percent of students into a higher performance level in math. In Figure V.5, we show the EOG and FA interventions’ impacts on students’ performance levels. We present full results for impacts on test scores, including subscores for math and reading, in Appendix Table D.4.

Figure V.5. Impacts on math performance levels



Source: EOG math assessment 2016.

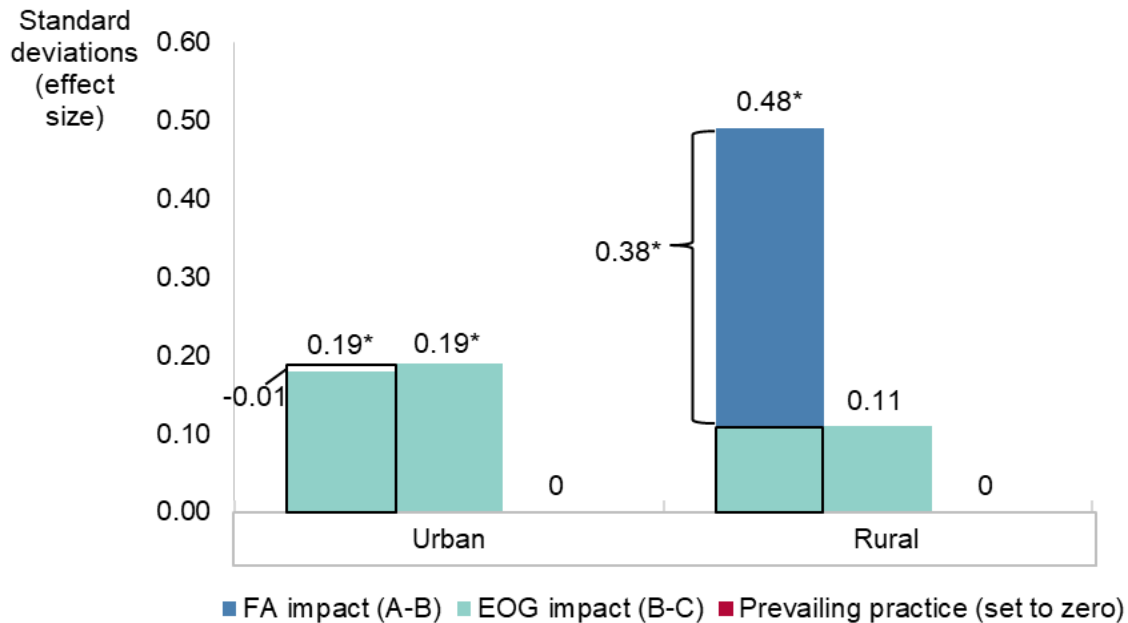
Note: Performance levels for the prevailing practice group are the actual distribution of results for that group. Performance levels for the EOG-only and EOG and FA groups are estimated by adding the estimated impact to scores observed in the prevailing practice group.

C. Impacts by subgroup

At the outset of the study, USAID identified two policy-relevant subgroups for which we were to examine impacts. One subgroup is defined by its urban or rural status. The urban schools include all schools in the urban areas of Tegucigalpa and La Ceiba, and the rural schools include all schools in the predominantly rural departments of Lempira and Santa Barbara. The other subgroup of interest is gender. We found that the impacts of the EducAcción-PRI interventions varied significantly by setting (urban or rural) but that the variation in impacts by gender was not significant.

The EOG intervention had a significant impact on reading in urban areas, but not in rural areas. In urban areas, the EOG intervention increased reading scores by 0.19 standard deviations; the impact was significant. In rural areas, however, the impact was smaller—0.11 standard deviations—and not significant. The difference between the EOG intervention’s impacts in urban and rural schools was not significant (Figure V.6; full results in Appendix D).

Figure V.6. Impacts on reading by urban or rural area



Source: EOG reading assessment 2016.

Note: The mean for the prevailing practice group has been normalized to 0. Impact estimates from a regression that includes baseline controls and adjusts for stratified randomization are shown as effect sizes (standard deviations). * $p < 0.05$, two-tailed test.

The impact estimate of the FA component was negative for the EOG and FA group in urban schools. The white shading in the group A column for urban schools represents the negative impact estimate.

The p -value on the test for the significance of difference in A-B impacts between urban and rural students was 0.002 for reading.

The p -value on the test for the significance of difference in B-C impacts between urban and rural students was 0.448 for reading.

For both the urban and rural subgroups, the separate impact estimates for the EOG and FA interventions do not sum to the impact estimate for the combined impact of both interventions because of rounding.

The EOG intervention focused on principals, whose responsibilities differ between urban and rural schools; the differences may explain why the EOG intervention had a smaller impact on learning in rural schools. Among principals in the study sample, 90 percent of rural principals also worked as classroom teachers in their schools at baseline; in urban schools, only 17 percent worked as classroom teachers.¹¹ In view of their teaching obligations, rural principals may have less time to dedicate to the development and implementation of a school action plan. Furthermore, principals who are involved in teaching their own classes during the school day have less time to supervise their teachers' implementation of the school action plan.

School size is another difference between urban and rural schools that may influence the EOG intervention's impacts on learning. Analysis of EOG results may be less important in small schools, which are more common in rural areas. At baseline, the average rural school had enrolled 145 students, whereas the average urban school had enrolled 374 students. Understanding and acting on EOG test score results may be more effective in large schools than in small schools where principals and teachers may be able to track individual students' progress more easily in the absence of EOG test data.

The FA intervention had a significant impact on reading test scores in rural areas, but not in urban areas. In rural areas, the FA intervention had a significant impact on reading test scores of 0.38 standard deviations; the FA intervention had no impact on reading scores in urban schools. The difference between the impacts in urban and rural schools was significant ($p = 0.002$) (Figure V.6; full results in Appendix D). We offer possible explanations for the difference in impacts between urban and rural schools.

First, coaches are expected to work individually with grade 1 through 3 teachers, which means that they had to work with more than three times as many teachers in urban schools, on average, compared to coaches in rural schools. At baseline, urban study schools employed an average of 18 teachers each, whereas rural study schools employed only 5 teachers. Accordingly, urban coaches in the FA and EOG group were assigned to fewer schools; nonetheless, the number of teachers may have made it more difficult to spend as much time per teacher as might have been needed or desired. Second, urban teachers were more likely than rural teachers to have to leave immediately at the end of the school day to work at a second teaching job, limiting the time available to some urban teachers to work with for coaches.¹² Finally, in crowded urban schools, coaches might have faced difficulty in finding a quiet place to work effectively with teachers.

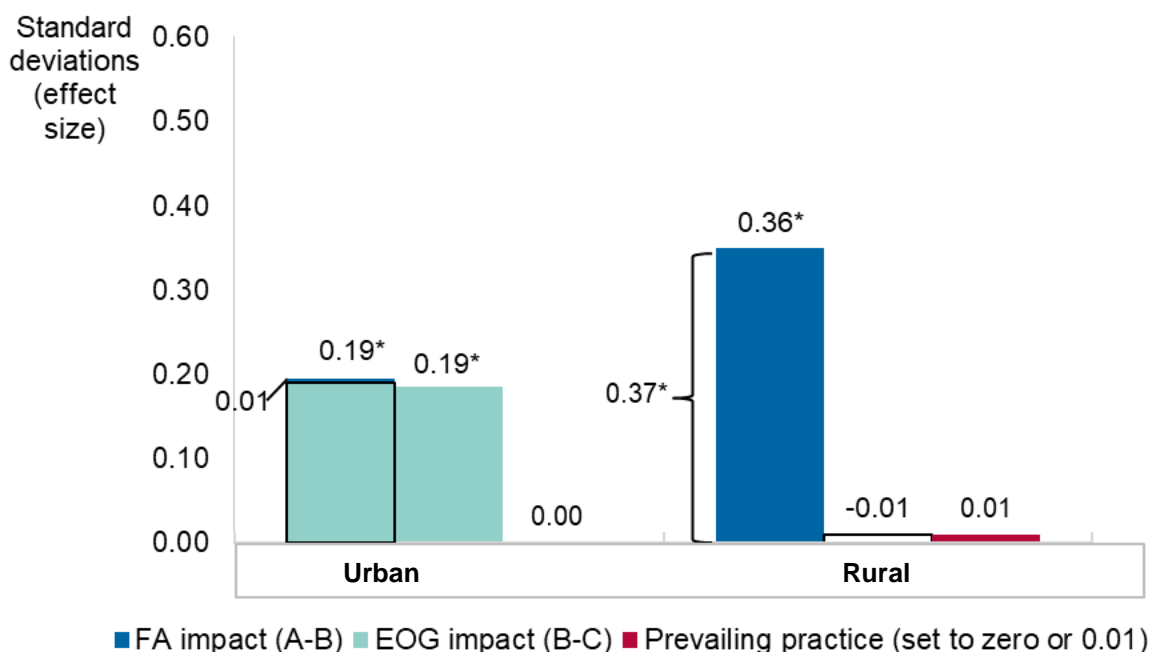
Given that they were less likely than urban teachers to hold a second teaching job, rural teachers, by contrast, may have had more time to analyze FA results and modify their teaching based on those results as noted above. In focus groups, teachers reported spending between 10 and 15 hours per month on activities related to FAs (including urban and rural teachers).

¹¹ At endline, 91 percent of rural principals also worked as teachers, and 12 percent of urban principals did so.

¹² Of urban teachers, 19 percent indicated at baseline that they taught at another school versus 9 percent of rural teachers. Similar percentages of teachers taught at a second school at endline.

Both interventions’ impacts on math were similar to impacts on reading in urban and rural areas. The EOG intervention had a significant impact of 0.19 standard deviations on math test scores in urban areas, and the FA intervention had a large, significant impact of 0.37 standard deviations on math scores in rural areas. The difference between EOG impacts in urban and rural areas was not significant, but the difference between FA impacts was significant. In Figure V.7, we show impacts on math by urban and rural area. The potential explanations for the differences in reading, as described above, apply similarly to the variation in impacts on math scores.

Figure V.7. Impacts on math by urban and rural area



Source: EOG math assessment 2016.

Note: The mean for the prevailing practice group has been normalized to 0 or 0.01. Impact estimates from a regression that includes baseline controls and adjusts for stratified randomization are shown as effect sizes (standard deviations). * $p < 0.05$, two-tailed test.

The prevailing practice group’s impacts are set to 0.01 instead of 0 in the rural group to show the EOG intervention’s impact of -0.01 without going below the horizontal axis. The estimated impact of the EOG intervention in rural schools was negative. The white shading in the EOG-only column for rural schools represents this negative impact estimate.

The p -value on the test for the significance of difference in A-B impacts between urban and rural students was 0.033 for math.

The p -value on the test for the significance of difference in B-C impacts between urban and rural students was 0.264 for math.

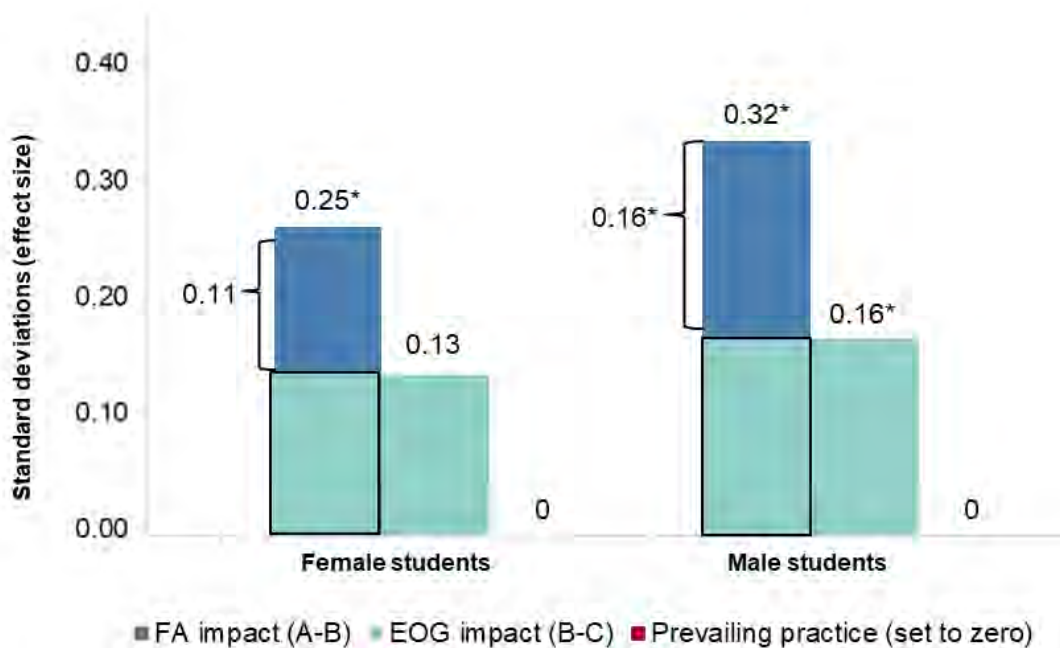
For the urban subgroup, the separate impact estimates for the EOG and FA interventions do not sum to the impact estimate for the combined impact of both interventions because of rounding.

Both the EOG and FA interventions had significant impacts on reading test scores for boys, but not for girls. Each intervention improved boys’ test scores by 0.16 standard deviations. The EOG intervention improved girls’ scores by 0.13 standard deviations, and the FA intervention improved their scores by 0.11 standard deviations. The difference between impacts on reading for boys and girls was not significant.

The variation in impacts appears to be driven by a catch-up effect for boys. Boys’ reading test scores were lower than girls’ scores in all three treatment groups, but the gap was smaller in the EOG-only group than in the prevailing practice group and smaller still in the EOG and FA groups (Figure V.8; Appendix D for full results).

Coaches trained teachers to use EOG and FA test results to identify students who were struggling, and teachers reported that they used the results of both tests to identify and support struggling students. In addition, coaches received training on gender equity. They learned how to analyze test score results by gender and how to work with teachers and principals to address any disparities in learning by gender. The focus on struggling students in general and on the gender issue in particular may explain the smaller gender gap in reading scores in EducAcción schools.

Figure V.8. Impacts on reading by gender



Source: EOG reading assessment 2016.

Note: The mean for the prevailing practice group has been normalized to 0. Impact estimates from a regression that includes baseline controls and adjusts for stratified randomization are shown as effect sizes (standard deviations). * $p < 0.05$, two-tailed test.

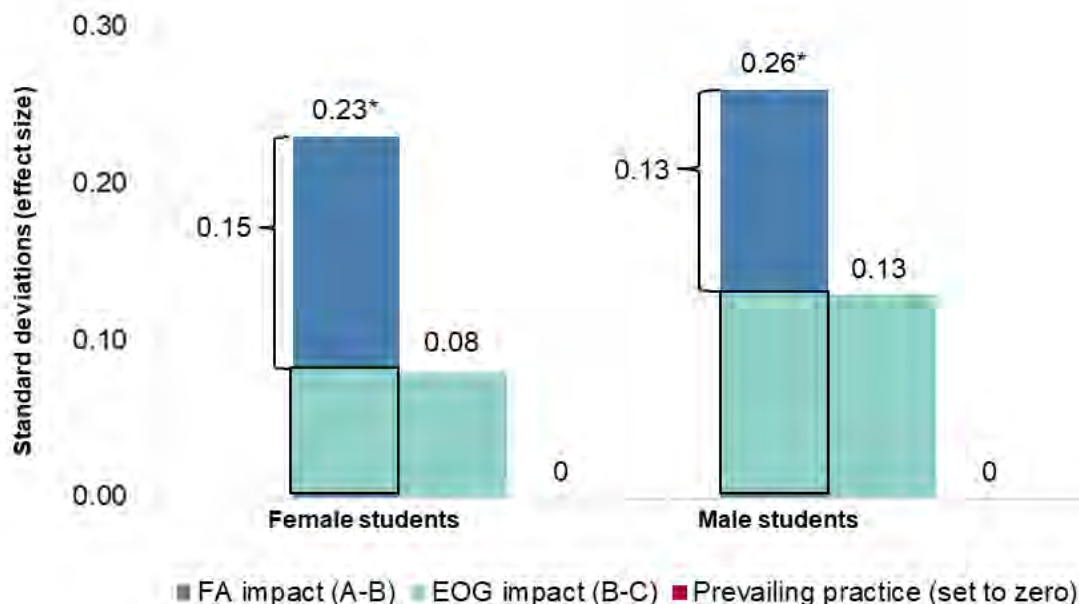
The p -value on the test for the significance of difference in A-B impacts between female and male students was 0.426 for reading.

The p -value on the test for the significance of difference in B-C impacts between female and male students was 0.645 for reading.

For both the female subgroup, the separate impact estimates for the EOG and FA interventions do not sum to the impact estimate for the combined impact of both interventions because of rounding.

Impacts on math were not significant for boys or girls. The estimated impact for the EOG intervention was greater for boys than for girls, though the difference in impacts was not significant. As was the case with reading, boys’ math test scores were lower than girls’ scores in the prevailing practice group, and the gender gap in scores was smaller in EducAcción schools. However, the magnitude of the gender gap in prevailing practice schools was smaller for math than for reading (0.04 standard deviations for math compared to 0.17 for reading), leaving little room for the interventions to address the gap. The smaller impacts on math when split by gender are consistent with the smaller impacts on math scores overall (Figure V.9; Appendix D for full results).

Figure V.9. Impacts on math by gender



Source: EOG math assessment 2016.

Note: The mean for the prevailing practice group has been normalized to 0. Impact estimates from a regression that includes baseline controls and adjusts for stratified randomization are shown as effect sizes (standard deviations). * $p < 0.05$, two-tailed test.

The p -value on the test for the significance of difference in A-B impacts between female and male students was 0.819 for math.

The p -value on the test for the significance of difference in B-C impacts between female and male students was 0.426 for math.

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

VI. IS THE INTERVENTION COST EFFECTIVE?

Both the end-of-grade and formative assessment (EOG and FA) supports and materials had positive impacts on student achievement, but policymakers need to know if the impacts are sufficiently large to justify the costs. Thus, we estimated the average cost per student in order to compare that to the average impact per student presented in Chapter V. We then express cost-effectiveness as the cost in current U.S. dollars per unit of impact per student.

We estimated that the cost-effectiveness of the EOG intervention was \$52 per 0.10 standard deviation increase in reading scores and the cost-effectiveness of the FA intervention was \$57 per 0.10 standard deviation increase in reading scores. In this chapter, we explain how we obtained these estimates as well as the assumptions that underlie them, and then we compare the estimates to the cost-effectiveness of alternative interventions.

A. Methods and data for cost effectiveness analysis

To measure average costs per student, we obtained expenditure data from AIR, the implementer of EducAcción, and interviewed key program staff. The interviews allowed us to categorize costs correctly and account for any use of resources that were not explicitly included in program documentation (such as the value of time contributed by participating teachers or principals). We also gathered detailed data on program participation and student enrollment in order to include the correct number of students, which is essential for calculating the cost per student.

The general principle guiding the cost analysis was to estimate the incremental costs between study treatment arms in a parallel manner to the way we estimated impacts between those same treatment arms. We estimated impacts in units of reading test score performance per student as well as costs on a per student basis. EducAcción incurred costs for a much larger group than the study sample, which encompassed just one cohort among four that were exposed to the interventions; the intervention targeted grades 1, 2, and 3 in each of the two years (2015 and 2016) it was implemented. The cohort in our evaluation was enrolled in grade 2 in 2015 and, for those not repeating grade 2, enrolled in grade 3 in 2016. We assumed that the average cost per student was the same for all cohorts served. We believe that our assumption is reasonable because the intervention was applied at the school and teacher levels, not at the student level, and there was no evidence that it would vary by grade level. Thus, we calculated an average cost per student over all students exposed to the intervention, and compared that cost to the average impact for students in the study sample.

In an approach consistent with the way we estimated impacts, we did not need to estimate the total cost of providing all services in all three study groups (prevailing practice, EOG-only, and EOG and FA). Instead, we estimated the incremental costs, that is, the *differences* in the incurred costs for Group A (EOG and FA) relative to Group B (EOG-only), and Group B (EOG-only) relative to Group C (prevailing practice). Thus, it was important to allocate program expenditures to the two types of interventions, EOG and FA support. At the beginning of the intervention, we worked with AIR and EducAcción to develop a template that allowed us to record costs for each treatment group separately and thereby estimate incremental costs.

B. Cost of the interventions

Variations in cost by context. Cost for future implementers may vary, and estimates of the cost to implement a similar intervention in a different context will depend on location-specific costs and available resources. Here, we highlight costs that we did not include in our cost analysis but that may accrue to others to implement a similar intervention.

- *Development of EOG assessments, FAs, and methods for analyzing and reporting results.* The assessments and related analysis and reporting methods were already in place in Honduras, but, in locations where assessments and reporting methods would need to be developed, the cost estimates would need to reflect the cost of these activities.
- *Administration and scoring of EOG tests.* Others considering implementation of a similar intervention would also need to account for the administration and scoring costs of EOG tests if such assessments are not already offered.
- *Coordination with national and regional education authorities.* When planning for EducAcción-PRI began, the intervention also benefited from EducAcción’s long-term involvement with national and regional authorities in Honduras. Our cost estimates account for time spent to coordinate with national and local education authorities, but organizations without existing relationships with education authorities would likely need to plan to spend more time to coordinate the intervention.

Measured costs. Of the measured costs for providing the program to Groups A and B, we estimate that nearly two-thirds of expenditures went toward hiring and paying local staff, such as the coaches. About 5 percent went to the “home office” and to travel for management support. Printing and materials represented about 5 percent of total expenditures, with the rest, about 26 percent, allocated to equipment, vehicles, office supplies, and other expenses.

When we divide by the numbers of students served, we arrive at estimates of \$57 per student per 0.10 standard deviations for the FA intervention and \$52 for the EOG intervention.

Unmeasured costs. It is possible that unmeasured costs could raise the price of implementing a similar intervention in the future. Such costs include teachers’ unmeasured labor or the unmeasured labor if substitute teachers or volunteers covered teachers’ classes so that the regularly assigned teachers could participate in EducAcción-PRI activities, such as training or working with small groups. We do not believe that these factors add up to a large total unmeasured cost. We did not measure the cost of teachers’ overtime hours because we did not find a significant difference by group in teachers’ reported overtime hours. We also did not measure the cost of substitute teachers because only a small number of teachers (4 of 232) indicated that another teacher covered their class while they participated in training. We did not account for any donated materials; however, neither we nor the implementer have found evidence of donated materials.

Beginning in the month before implementation, AIR staff kept a detailed accounting of costs associated with EducAcción-PRI. However, AIR did not include in its cost estimates any preparation work completed before that period. According to AIR’s estimate of the level of effort expended for preparation work, we estimate that accounting for the cost of preparation time

would increase the cost per student by 2 to 3 percent, or roughly \$1 to \$2 per child per 0.1 standard deviations for the FA or EOG component.

C. Cost-effectiveness and comparison to other interventions aimed at improving test scores

The cost-effectiveness of approximately \$57 (for the FA intervention) and \$52 (for the EOG intervention) per student per 0.10 standard deviations of improvement places the EducAcción-PRI interventions in the middle to high end of the range of costs per unit of improvement that have been published for other interventions. We compare our estimates to those found in a systematic review of the impacts and cost-effectiveness of educational investments in developing countries by Evans and Ghosh (2008).¹³ The estimates in that review that were based on randomized controlled trials range from \$1 to \$296 per student per 0.1 standard deviations. Among those randomized controlled trials are two interventions that, like EducAcción-PRI, adjust instruction to students' abilities. First, in the least expensive intervention in the review, low-cost tutors worked with lagging primary school students in India for two hours per school day for a cost of \$1.11 per 0.1 standard deviations in test score improvement. In another intervention from the same study in India, students spent two hours per week using educational computer programs that adjusted to students' abilities for \$8.25 per 0.1 standard deviations. The randomized evaluation with costs closest to those of EducAcción-PRI was an educational voucher program in Colombia, which cost \$47 per 0.1 standard deviations. The most expensive intervention was an early childhood program in Jamaica, which cost \$296 per 0.1 standard deviations. As is always the case when reviewing cost estimates from programs implemented in sometimes very different country contexts and perhaps target populations, policymakers must consider how costs would vary based on differences in labor, transportation, and other expenses between the country under study and their own country.

The Evans and Ghosh review also included several estimates based on nonexperimental evaluations, which revealed an even wider range of values for cost effectiveness, including two estimates from a study conducted in Honduras and involved teacher training. Bedi and Marshall (1999) used observational data from 33 schools in rural Honduras to estimate education production functions. They estimated that improving learning by reducing class size cost \$23 dollars per child per 0.1 standard deviation increase in test scores. Using the same data, Bedi and Marshall also estimated the cost of increasing test scores with teacher training, finding that providing three teacher training seminars could increase test scores with a cost of about \$0.75 per 0.1 standard deviation increase in test scores. These estimates are especially relevant to policymakers from Honduras since they are based on evidence from the same country. However, the impact estimates are based on observational data and a small sample size of only rural schools that should be interpreted with caution. Because the findings are based on only rural schools, they do not reflect the impacts such interventions would have in urban settings. Furthermore, the

¹³ For ease of comparison between our estimates and those presented in Evans and Ghosh, all estimates use constant 2015 US dollars. We used the Consumer Price Index – Urban (CPI-U) to adjust estimates presented in Evans and Ghosh. Evans and Ghosh present estimates with and without adjustments for the deadweight loss associated with the cost of raising money through taxation; we refer to the estimates from their review that do not include the deadweight loss adjustment to make those results comparable to ours, which are also not adjusted for deadweight loss.

study is observational and its claims on the potential causal relationship between the teacher training and class size and student achievement are based on strong assumptions that cannot be tested.

VII. CONCLUSIONS, LESSONS LEARNED, AND RECOMMENDATIONS

A. Conclusions and limitations

In addition to documenting the implementation of EducAcción-PRI and contrasting it with prevailing practices, the study posed five questions about impacts and cost-effectiveness. We consider each conclusion in response to the questions.

1. Conclusions on evaluation questions

1. **Were the interventions implemented as intended? Specifically, did the offer of focused coaching and summative end-of-grade (EOG) assessment results increase teachers' and principals' access to these supports and materials? Did the offer of supports for administering formative assessments (FA) increase the assessments' availability and use in the targeted schools?**

Our teacher and principal survey data suggest that a large majority of the targeted teachers and principals received the intended services. Principals in all EducAcción schools (EOG-only as well as EOG and FA) reported that they developed and implemented school action plans based on EOG results. Teachers and principals also reported that they received training and materials in accordance with their treatment assignment, although, on average, teachers and principals reported working with their coaches less frequently than expected.

2. **Did the interventions result in changes in instructional practice? Specifically, how do support and materials influence teachers' and principals' use of EOG tests and FAs in instruction and student support?**

Teachers in nearly all EducAcción schools reported that they modified their teaching based on EOG results. In survey data, teachers reported that they dedicated more time to subjects that challenged students on the EOG tests, providing extra support to students with low scores, and identifying new teaching methods. In focus groups, teachers mentioned that their principals supervised them more closely once school action plans were in place, suggesting a potential mechanism for how the principal-focused intervention led to changes in teacher behavior. Both teachers and principals reported in focus groups that the EOG results would be most useful if received by the beginning of the academic year.

After receiving FA materials and coaching in their use, teachers in the EOG and FA group reported that they modified their lesson plans based on FA results and provided extra support to students whose FA scores identified them as lagging. Teachers in the EOG and FA group were significantly more likely than teachers in the EOG-only group to make such changes.

In focus groups, teachers described how FAs helped them recognize the knowledge their students had acquired in the previous month and the areas in which they needed to provide more support. Teachers also reported that they acted on FA results as soon as they had tabulated them because they had only four weeks before administering the next scheduled FA. Coaches noted that teachers spent more time on lesson planning and deliberately structuring their lessons around the needs of their students once they began analyzing FA results.

3. **Relative to prevailing practice, what is the impact of providing feedback on EOG test results and teacher coaching on student learning?**

EOG test results and feedback significantly increased reading test scores by 0.15 standard deviations; impacts for math were similar in magnitude, but they were not statistically significant. The impact on reading scores is equivalent to moving a student from the 50th to the 56th percentile. This impact is also equivalent to increasing by 3 percentage points the percentage of reading questions that students answered correctly and moving 8 percent of students into a higher performance level. Finally, it is equivalent to 2.0 months of instruction, if we use an estimate of 0.075 standard deviations per month.

EOG impacts varied across urban and rural schools. The impact of the EOG intervention was larger and significant only in urban schools. Many explanations for this subgroup difference are possible. With more teachers and students to manage, principals in urban schools may have benefited more from test score data. Furthermore, given that urban principals are less likely than their rural counterparts also to serve as classroom teachers, they may have had more time than rural principals to dedicate to developing and implementing a school action plan. Impacts on math test scores were also lower in rural schools than in urban schools; the EOG intervention had no impact in rural schools. When we presented these findings in preliminary form to stakeholders—educators and administrators at all levels in the four study areas—the meeting participants’ written and oral responses corroborated these hypotheses. For example, one participant, whose comments were echoed by others, noted that “[EOG] had more impact in the urban area because of the number of students who were served in single-grade classrooms, where it’s more feasible to apply the school improvement plan.”

4. **What added impact does providing teachers with FA materials and coaching have on student learning?**

Providing FA materials and support significantly increased reading test scores, by 0.14 standard deviations; the impact was similar in magnitude to that of the EOG intervention. Impacts on math were similar in magnitude and significance. The impact of the FA intervention is also equivalent to moving a student from the 50th to the 56th percentile. The FA intervention increased by 4 percentage points the percentage of reading questions that students answered correctly and moved 7 percent of students to a higher performance level. The equivalent impact in terms of months of instruction would be roughly 1.9 months. It should be noted that this impact estimate may overstate the net effect of the FA component relative to prevailing levels of formative assessment usage and training experience in Honduras, which were higher in the prevailing practice experimental group than they were in the EOG-only group (as shown in Chapter III).

FA impacts were larger in rural schools than urban schools. The FA intervention significantly raised test scores in rural schools by 0.38 standard deviations, but it had no impact in urban schools. Survey data reveal differences between urban and rural teachers’ characteristics and responsibilities that may explain the difference in impacts. First, unlike their rural counterparts, rural teachers had less experience and were more likely to be first-year teachers. Second, coaches in urban versus rural schools had to work with many more teachers, perhaps limiting how much time they spent working with each teacher. During our discussion of the

findings in country, several stakeholders in both urban and rural areas noted that the rural areas had fewer material resources like booklets and paper, which were provided as part of EducAcción-PRI, whereas these inputs in urban areas made less of a difference. Other factors noted by stakeholders included the much closer ties between parents and teachers in rural areas than urban areas, suggesting that the FA component was more effective in that setting in creating feedback and followup loops.

5. **How much does it cost to provide the materials and support needed for EOG assessments and FAs? Are the effects sufficiently large to justify the costs?**

We estimated that investing in either the EOG assessments or FAs would amount to spending \$52 and \$57 per student per 0.10 standard deviation improvement in test scores (1.3 months of instruction). These figures fall in the middle to high end of the range of costs per unit of improvement that have been published for other interventions. Investing in both interventions would have the same cost-effectiveness, costing about twice as much per student but producing about twice the impact on reading test scores.

2. Study limitations

There are several caveats to consider when interpreting the findings from this study. These have been noted throughout the text, but we summarize them here as well. First, conclusions about the impact of formative assessment rest on an assumption of “no interaction effects.” That is, we must assume that the effect of adding the FA component to a school that already receives the EducAcción-PRI’s EOG component is the same as the impact of adding the FA component to a school that does not have the EOG component.

Second, the impacts of the FA component estimated in this study may be larger than the impact that we would have observed had we compared the component to prevailing practice in Honduras. We noted that program implementers’ awareness of the study protocol may have limited the degree of usage of FA materials in Group B schools compared to Group C schools. This makes Group B more like a pure control (no-treatment control group) than was intended by the study design.

Third, as with any longitudinal field experiment, nonresponse, attrition, and non-compliance could have introduced bias. In the current study, these issues were minimized and unlikely to influence the results. The response rates were 99 percent or higher for all teacher and principal surveys in each experimental group. For student endline test scores, they were 80 percent, as anticipated in the study design, in each experimental group. Students missing from endline test score data were mostly students who left school or were held back, as described in Chapter II. The rates at which students moved between schools, either within or between experimental groups, were low and symmetric with respect to experimental group. We reported impacts based on “intent to treat”, which means that students were classified according to the school they were attending at the point of random assignment even if they moved. Compliance rates, defined as the percentage who remained with their original treatment assignment, were about 98 percent.

One final caveat, which is also true of most evaluations, is that there are costs and benefits that could not be measured or could not be measured precisely. These include the potential long term benefits, impacts felt beyond the end of the study’s observation period, as well as the

division of costs between the two program components, which we had to estimate based on implementing program staff's estimates of allocation of time between various activities. In particular, the cost effectiveness analysis focused on incremental costs of each intervention relative to its comparison, not relative to no program at all. Thus, we did not estimate or include the costs of developing or producing assessments, because this work had been done previously in Honduras and tests were available for download from the Ministry's website for schools in all three groups.

Another quantity that cannot be measured directly is the conversion rate between effect sizes and months of learning. This metric is merely provided to aid in interpretation, but the conversion requires several strong assumptions. As discussed in Chapter II, many reasonable researchers disagree on the most appropriate rate to use.

B. Lessons learned

Both EOG assessments and FAs can improve learning when coupled with ongoing training and support. Principals and teachers valued the support they received for use of the assessments, recognizing that it increased their capacity to analyze and respond to test results. With the school-level action plans and classroom-level improvement plans, principals and teachers could work toward goals based on student achievement data, which led to effective changes in instruction.

Training on and support for EOG assessments and FAs are important if these tools are to improve learning. Among schools that participated in EOG testing, we found large contrasts between the EOG-only group and the prevailing practice group in terms of access to test score results, and principals' and teachers' likelihood of using those results to improve teaching and learning. Interventions such as EducAcción-PRI could help schools take advantage of education ministries' existing investments in EOG assessments and FAs.

The effectiveness of education interventions depends critically on context. The evaluation found dramatically different effects in urban and rural schools, potentially leading to context-specific policy recommendations.

The interpretation of results is clearer when interventions under study are in their steady state. Given that the key components of assessment instruments and the relationships with national and local education authorities were already in place, EducAcción was able to launch EducAcción-PRI as soon as funds became available. The rapid start-up benefited the evaluation because the evaluation intensity was not compromised by implementation delays or changes to the intervention.

A strong evaluation design and data tailored to the evaluation's needs increase researchers' ability to reach firm conclusions. The prospective experimental evaluation design used for the evaluation of EducAcción-PRI generated clean contrasts between groups of schools with high-intensity treatments (the two intervention groups) and the control group of schools that received no intervention. The data for the evaluation were collected at the school and student levels, further maximizing our ability to detect impacts. In contrast, in our secondary analysis of the impacts of MIDEH and EducAcción's work in targeted municipalities (see Appendix G), we were limited by aspects of the retrospective quasi-experimental design we used and by the

administrative data available to us. Given that contrasts between municipalities in the target group and other municipalities were less stark and that implementation data were not available at the school level, our findings were less conclusive than those of the experimental evaluation. Had we conducted only a quasi-experimental evaluation, we might have failed to identify the assessments' potential impacts on learning that we identified in our evaluation of EducAcción-PRI.

C. Recommendations

A recommendation to policymakers in Honduras is to expand and replicate EducAcción-PRI to reach more schools in the country if funding is available. Our findings show that, to have the largest effects, the EOG intervention could be offered in urban areas and the FA intervention in rural schools. One caveat for rolling out the FA component on its own is the fact that it was tested under conditions where EOG component was already in place. This recommendation assumes that there are no interaction effects, meaning that the FA component impacts are unaffected by the presence or absence of the EOG component.

A recommendation to policymakers in other countries in the Latin America and Caribbean region that have already developed either EOG assessments or FAs is to consider implementing interventions similar to EducAcción-PRI. Countries considering whether to develop an assessment system could review the potential benefits of using assessments to improve teaching and learning, as EducAcción-PRI was able to do in Honduras.

A recommendation to policymakers in countries that provide EOG test results to principals and teachers is to provide those results early in the academic year. In focus groups, principals and teachers mentioned that they valued receiving results early in the year to help them plan the year.

Another recommendation for policymakers who set national testing policies is to consider the tradeoffs between the coverage of a testing program—national census versus a sample—with the quality of the test administration. In addition to presenting impacts of EducAcción-PRI, this study included analysis of secondary data on census tests and a more securely administered random sample. The findings suggest that secure test administration produces more credible data.

A recommendation to donors and policymakers is to recognize the advantages (or value added) to building in rigorous evaluations of enhancements or variations to key components of established programs. The success of the EducAcción-PRI evaluation may have been in part due to the fact that implementers took advantage of well-tested and accepted national assessment tools produced by the project (EOG and formative assessments) and existing relationships with national and local education authorities. Rigorous evaluations are often more difficult when interventions are still in development or when education authorities are unaware of the new intervention being evaluated.

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

REFERENCES

- Abeberese, Ama Baafr, Todd J. Kumler, and Leigh L. Linden. “Improving Reading Skills by Encouraging Children to Read in School: A Randomized Evaluation of the Sa Aklat Sisikat Reading Program in the Philippines.” *Journal of Human Resources*, vol. 49, no. 3, 2014, pp. 611–633.
- 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.
- Araujo, M. Caridad, Pedro Carneiro, Yyannú Cruz-Aguayo, and Norbert Schady. “Teacher Quality and Learning Outcomes in Kindergarten.” *The Quarterly Journal of Economics*, vol. 131, no. 3, pp. 1415–1453.
- Baird, Matthew D. and John F. Pane. “Translating Standardized Effects of Education Programs into More Interpretable Metrics.” Santa Monica, CA: Rand Corporation, March 2018.
- Banerjee, Abijit V., Shawn Cole, Esther Duflo, and Leigh Linden. “Remedying Education: Evidence from Two Randomized Experiments in India.” Research Working Paper 11904. Cambridge, MA: National Bureau of Economic Research, December 2005.
- Bedi, A.S., and J.H. Marshall. “School Attendance and Student Achievement: Evidence from Rural Honduras.” *Economic Development and Cultural Change*, vol. 47, no. 3, 1999, pp. 657–682.
- 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?” Washington, DC: 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.

- Evans, David K., and Arkadipta Ghosh. "Prioritizing Educational Investments in the Developing World." 2008. RAND Working Paper 587. Available at https://www.rand.org/pubs/working_papers/WR587.html. Accessed March 1, 2018.
- Glazerman, Steve, Sarah Liuzzi, and Nancy Murray. "Data-Driven Instruction in Honduras: An Impact Evaluation of the EducAcción Promising Reading Intervention. Evaluation Plan." Washington, DC: Mathematica Policy Research, June 2016.
- Glazerman, Steve, Sarah Liuzzi, Nancy Murray, and Chantal Toledo. "Data-Driven Instruction in Honduras: An Impact Evaluation of the EducAcción Promising Reading Intervention. Baseline Report." Washington, DC: Mathematica Policy Research, August 2016.
- 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.
- Hill, Carolyn J., Howard S. Bloom, Alison Rebecka Black, and Mark W. Lipsey. "Empirical Benchmarks for Interpreting Effect Sizes in Research. MDRC Working Paper on Research Methodology." July 2007. Available at https://www.mdrc.org/sites/default/files/full_84.pdf. Accessed May 6, 2019.
- 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.
- Kremer, Michael, Conner Brannen, and Rachel Glennerster. "The Challenge of Education and Learning in the Developing World." *Science*, vol. 340, no. 6130, April 19, 2013, pp. 297–300.
- 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.
- Liuzzi, Sarah, Steven Glazerman, Nancy Murray, and Irina Cheban. "Data-Driven Instruction: Research on the Role of Evaluation in Education in Honduras. Highlights of the midline results of the randomized controlled trial." Presented to USAID, August 10, 2017.

- LoGerfo, Laura, Austin Nichols, and Sean Reardon. “Achievement Gains in Elementary and High School.” Washington, DC: Urban Institute, March 2006.
- McEwan, Patrick. “Improving Learning in Primary Schools of Developing Countries. A Meta-Analysis of Randomized Experiments.” *Review of Educational Research*, vol. 85, no. 3, September 2015, pp. 353–394.
- Muralidharan, Karthik, and Venkatesh Sundararaman. “The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India.” *The Economic Journal*, vol. 120, no. 546, 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, no. 2, 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, France: 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.

This page has been left blank for double-sided copying

APPENDIX A:
BASELINE EQUIVALENCE

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

We reported on characteristics of schools, principals, teachers, and students at baseline in our Baseline Report (Glazerman et al. 2016). As we reported in detail in the baseline report, we found that our randomization created three groups of similar schools with similar teachers and students.

Schools had similar characteristics at baseline. The principal surveys conducted at baseline reveal that the randomization successfully formed three similar groups of schools, and that the schools in the sample faced numerous challenges. Table A.1 shows that more than 40 percent of schools are multigrade, meaning that 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 A.1. School characteristics at baseline

Characteristic	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Difference between FA + EOG and EOG only groups		Difference between EOG only and prevailing practice groups	
				(A) – (B)	p-value	(B) – (C)	p-value
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)
Multigrade school (percent) ^a	46.7	41.7	48.3	5.0	(0.474)	–6.7	(0.341)
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)
Total number of schools	60	60	60	120		120	

Source: 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.

^b The classroom infrastructure index is the percentage of five school-level classroom characteristics found at each school; this ranges from 0 to 100. A school that has two of the five would have a score of 40, whereas a school with all five would have a score of 100. The five school-level classroom characteristics are as follows: (1) no class sections are held in an informal area, such as outdoors; (2) no sections are exposed to weather extremes, such as heat or rain; (3) all sections have a blackboard or other writing surface; (4) all sections have enough desks and chairs; and (5) 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 0 to 100 (see table note b). The five characteristics included in this index are (1) school has piped plumbing, (2) a source of potable water, (3) electricity, (4) functioning bathrooms (excluding latrines), and (5) a health center. The standard deviation is 16.9.

^dThe school learning resource index is constructed in the same way as the classroom infrastructure index, and also ranges from 0 to 100 (see table note b). The five characteristics included in this index are (1) school has a library resource room, (2) computers for teachers, (3) Internet access, (4) a music or art room, and (5) a playground. Standard deviation is 20.6.

EOG = end-of-grade.

FA = formative assessment.

Teachers were similar at baseline. As shown in Table A.2, teachers are balanced across treatment groups on gender, age, and teaching experience, but we found some differences in education and ethnicity. The majority of teachers are women, and the average age of teachers is in the upper thirties. On average, teachers have nearly 10 years of experience; few teachers are in their first year. Group C teachers were more likely than Group B teachers to have an education level other than an associate's or college degree. There were some differences in ethnic group between teachers in Groups A and B.

Table A.2. Teacher characteristics at baseline

	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Difference between FA + EOG and EOG only groups		Difference between EOG only and prevailing practice groups	
				(A) – (B)	p-value	(B) – (C)	p-value
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)
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	(0.371)
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.

Students were similar at baseline. Using enrollment records gathered at baseline and updated at midline and endline for students in the study sample, we collected a limited amount of data on student characteristics. We present this information in Table A.3. Students are balanced across treatment groups on gender, age, and overage (which reflects students who began primary school later than expected or who were repeating first grade). Just under half of students are female. 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. Just under 15 percent of students were older than we would expect for the grade.

Table A.3. Student characteristics at baseline

	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Difference between FA + EOG and EOG only groups		Difference between EOG only and prevailing practice groups	
				(A) – (B)	p -value	(B) – (C)	p -value
Female (percentage)	49.7	46.5	46.5	3.2	(0.130)	0.0	(0.989)
Age ^a (years)	6.5	6.5	6.5	0.0	(0.690)	0.0	(0.365)
Overage for grade ^b (percentage)	13.4	13.8	14.8	-0.4	(0.832)	-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.

Student test scores were similar at baseline. Table B.4 shows baseline test scores for the evaluation cohort of interest in their first grade year (2014). Scores were standardized to have a mean of 50 and a standard deviation of 20 relative to the entire nation of test-takers in that year. Students in study schools had average test scores below the national average. Differences among the treatment groups were not statistically significant.

Table A.4. Average first-grade reading and math end-of-grade test scores, by experimental group, 2014

Characteristic	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Difference between FA + EOG and EOG only groups		Difference between EOG only and prevailing practice groups	
				(A) – (B)	p-value	(B) – (C)	p-value
Reading test score ^a	46.8	45.6	46.8	1.2	(0.712)	-1.2	(0.705)
Math test score ^a	48.9	46.0	48.8	2.8	(0.363)	-2.7	(0.379)
Total number of schools	60	60	60	120		120	

Source: 2014 EOG test scores. Results are based on the schools selected for the randomized sample that had baseline data. For 22 schools, baseline data were imputed using 2013 EOG test score data for first graders.

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.

Teachers’ prior exposure to treatment was similar at baseline. The teacher surveys included questions about teachers’ prior training on, access to, and use of end-of-grade (EOG) and formative assessment results at baseline. These results capture the baseline level of utilization of assessments through support from the Ministry of Education or other organizations prior to the start of EducAcción-PRI’s activities (see Table A.5).

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

Teachers’ survey responses indicated that roughly half of teachers had received EOG test results from the previous year; fewer used them. Nearly half of teachers had modified lesson plans based on formative assessment results, but only about one in five teachers had enough copies of the assessments for all their students. Differences among treatment groups suggest that the prevailing practice group teachers were less likely to report receiving EOG results in the previous year, but difference was only statistically significant at the 10 percent level (difference = 14 percentage points, p-value = 0.097). Differences in reported prior use of formative assessments to modify instruction or adapt teaching went in the other direction, with higher rates

in the prevailing practice group (difference of 14 percentage points in the index of FA use, p -value = 0.009). However, the two experimental groups used to test the impacts of the FA component, groups A and B, had similar rates of FA use.

Table A.5. Teachers' self-reported use of end-of-grade and formative assessment test results at baseline

	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Difference between FA + EOG and EOG only groups		Difference between EOG only and prevailing practice groups		
				(A) – (B)	p -value	(B) – (C)	p -value	
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.

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

APPENDIX B:

METHODS AND SUPPLEMENTAL INFORMATION

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

1. Qualitative data analysis

In the first and second years of the intervention, Espirállica conducted focus groups with principals, teachers, and coaches, and Mathematica staff conducted interviews with coaches and EducAcción management (staff responsible for overseeing implementation of EducAcción-PRI). The Year 1 midline qualitative data were collected in October 2015, and the Year 2 endline qualitative data were collected in August 2016.

Sample selection. In each year, Espirállica conducted six focus groups with teachers. On average, there were 10 teacher participants in the midline focus groups and 7 teachers in the endline focus groups. In the first year, the teachers were from the Tegucigalpa and Santa Barbara study regions, and in the second year, the teachers were from the Tegucigalpa and Lempira study regions. Within each region, Espirállica held three focus groups, each with teachers from one of the three study groups (Group A, Group B, and Group C).

Schools in Tegucigalpa were selected through a combination of random selection and convenience sampling. Fourteen schools that were too far from the location of the focus group were excluded. In the first year, schools were randomly selected after excluding the schools that were too far geographically. In the second year, all schools that were randomly selected to participate in the midline qualitative data collection were excluded. One teacher from each of the remaining Tegucigalpa schools (after the exclusions) was invited to participate. In Santa Barbara and Lempira, one teacher from each Group A, Group B, and Group C study school was also invited to participate. The project staff notified the selected the schools, and each school chose one teacher to take part in the focus group.

Espirállica also conducted six focus groups with principals each year. On average, there were 8 principal participants in the midline focus groups and 7 principals in the endline focus groups. Like the teacher focus groups, three focus groups were held within each region, and each focus group had principals (or representatives sent by principals) from one of the study groups. The principals were invited from the same schools as the teachers, and the school selection process in Tegucigalpa, Santa Barbara, and Lempira is described above.

Finally, Espirállica conducted interviews or focus groups with the coaches, known in Spanish as *asesores*. Each year a total of eight coaches participated in the focus groups or interviews. The coaches represented all four study regions: Tegucigalpa (two coaches from Group A and one from Group B), Santa Barbara (one coach from Group A and one from Group B), Lempira (one coach from Group A and one from Group B), and La Ceiba (one coach from Group B). In Year 1, all coaches were interviewed. In year 2, the seven coaches from Tegucigalpa, Santa Barbara, and Lempira participated in the two focus groups, while the one coach from La Ceiba took part in the in-depth interview.

EducAcción staff selected the coaches, based on a set of criteria provided by the evaluation team. The selection guidelines included:

- Coaches who had not participated in the midline qualitative study;
- Coaches who were from each of the four study regions;
- An equivalent number of males and females;
- An equivalent number from Groups A and B.

Because of the limited number of coaches in each region, the final sample of focus group and interview participants for the endline met only three of the four selection criteria. Three of the eight coaches had also participated in the midline qualitative study.

Protocols. The qualitative focus group and interview protocols were developed by Mathematica in English. The protocols were then translated into Spanish. Espirática staff reviewed the protocols for language use and content and provided feedback. The revised versions of these protocols were used for the midline data collection. For the endline data collection, a few questions were added to address issues that had emerged in the project during the period between the midline and the endline qualitative data collection. For example, the endline protocols included questions on the distribution of formative assessments to the study schools, the revisions that were made to the formative assessments, and the annual participatory evaluation. Protocols and discussion guides are presented in Appendix I.

Data analysis. Espirática prepared literal transcriptions of the focus groups and interviews to facilitate analysis. To ensure that the transcripts were accurate, Mathematica randomly selected one or two transcripts from each type of focus group or interview and checked the transcripts against the audio recording. Mathematica developed an initial thematic coding scheme based on our qualitative data collection protocols to track common themes we anticipated, and we added codes for unexpected themes. We summarized responses in Excel to analyze their content, organized by our anticipated and additional frequently mentioned themes. We organized the analysis and reporting around these themes; we also used relevant information gleaned from these conversations to inform our interpretations of primary program impacts, as well as differential impacts by subgroups. Each transcript was coded by only one research analyst or researcher, but Mathematica staff who attended the focus groups and interviews in person checked the coding for accuracy. Table B.1 summarizes the qualitative data collection activities by round, and Table B.2 summarizes the sample by gender and round.

Table B.1. Number of focus groups and interviews, by location and treatment group

Participants	Data collection mode	Location	FA and EOG (A)	EOG only (B)	Prevailing practice (C)	All
Year one of intervention						
Principals	Focus groups	Santa	2	2	2	6
Teachers	Focus groups	Barbara and Tegucigalpa	2	2	2	6
Coaches	Interviews	All	4	4	—	8
EducAcción management	Interviews	N/A	—	—	—	3
Year two of intervention						
Principals	Focus groups	Lempira and	2	2	2	6
Teachers	Focus groups	Tegucigalpa	2	2	2	6
Coaches	Focus groups and interview	All	4	4	N/A	8
EducAcción management	Interviews	N/A	—	—	—	3

Note: In year two, we interviewed the Group B coach from La Ceiba because it was not feasible for the coach from La Ceiba to travel to participate in a focus group with other coaches. The other seven coaches participated in focus groups.

Table B.2. Number of focus group participants, by year and gender

Participants	Females	Males	Total
Year 1 of intervention			
Teachers	49	8	57
Principals	25	21	46
Coaches ¹	4	4	8
Year 1 Total	78	33	111
Year 2 of intervention			
Teachers	28	14	42
Principals	24	18	42
Coaches	4	4	8
Year 2 Total	56	36	92
Total	134	69	203

¹In Year 1, the coaches participated in individual interviews, rather than focus groups.

2. Impact analysis: regression model

Throughout this report, we reported regression-adjusted means for each of the three treatment groups. These are averages that adjust for the stratified random assignment process and for any baseline characteristics that may differ between treatment groups by chance. The school was the unit of analysis for principal and teacher outcomes; for teacher outcomes, we averaged individual teachers’ responses to the school level. The unit of analysis for impacts on learning was the student. Standard errors were clustered at the school level for all regressions.

For analysis of principal and teacher outcomes, the regression can be expressed as follows:¹⁴

$$(1) y_s = a + \lambda_A T_{As} + \lambda_C T_{Cs} + \sum_{j=1}^p \beta_j x_{sj} + \alpha_1 \dots \alpha_{r-1} + \varepsilon_s$$

where y_s is the outcome of interest (such as teachers’ use of EOG results, averaged at the school level) for school s , and the model includes 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 vector X represents a set of p baseline controls and β is a vector of p coefficients on each of those controls.¹⁵ The term ε_s is a random error term for school s . The parameters λ_A and λ_C represent differences between schools in Groups A and B, and differences between schools in

¹⁴ The notation for this equation is different from the notation used in the baseline report. The equation itself is unchanged.

¹⁵ Baseline controls, all defined at the school level, include whether the school has multigrade classrooms; first grade teachers’ average years of experience at the study school, age, and education; binary variables indicating whether teachers or principals had received training on formative or end-of-grade assessment; duration in hours of most recent training session; whether first grade teachers had at least one copy of the formative assessments per student; the number of times teachers had applied formative assessment during the baseline year; whether teachers had used formative assessment results to adapt lesson plans; whether teachers had used formative assessment results to identify and support lagging students; whether the teacher adapted lesson plans based on EOG results; the average class size in grades 1 through 3; and average first grade reading and math scores.

Groups B and C, respectively. In this report, we present the results of two tests: whether $\lambda_A = 0$ and whether $\lambda_C = 0$.

We conduct our analysis of student learning at the student level. We use the baseline controls used in the teacher and principal analysis, in addition to student age in months and a gender dummy. Again, we cluster standard errors at the school level.

In both regression types, we account for the stratified randomization by including the stratum fixed effects. 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 estimate the block effect associated with each group (stratum) of similar schools.

Our impact estimates represent intent to treat (ITT) analysis. ITT estimates reflect the impact of offering a treatment in the schools assigned to the treatment group, without adjusting for whether some schools in the treatment group chose not to participate, or for any schools in the control group that gained access to the treatment. To the extent that schools do not participate, ITT estimates would be lower than local average treatment effect (LATE) estimates because ITT estimates are diluted by the presence of those who choose not to participate and for whom we would assume there is no treatment effect. Policymakers are likely to be interested in the average impact for those offered the treatment, including the null effect for those who decline to participate. In the case of EducAcción-PRI, participation rates were high and the LATE estimates would be similar to ITT estimates. In Group A and B schools, 97 and 86 percent of principals, respectively, indicated having received training from EducAcción, whereas no principals in Group C schools indicated they received training from EducAcción.

3. Missing data in test score analysis

We had high response rates on survey data and did not make adjustments for survey nonresponse (shown in Table II.1).

For the students at the 22 schools that had no baseline EOG test scores, we imputed school mean reading and math test scores with first graders' EOG reading and math test scores from the previous year, averaged at the school level.

We exclude the students with no endline test data from our main analysis. Given that attrition does not vary significantly by treatment group, we did not adjust for nonresponse in our main results. Nonetheless, as a robustness check, we estimated impacts on endline test scores with nonresponse weights. Nonresponse adjustment weights account for students' predicted likelihood of having endline test score data. Students who are estimated to be unlikely to have endline test score data are given a larger weight to compensate for the fact that the sample used for impact estimation excludes students with missing test score data.

We estimated students' probabilities of having endline test score data with the following variables: Student age in months, number of students in the sample from the student's school (a proxy for school size), and binary indicator variables for whether the student was over age for his or her grade at baseline, gender, study location (Tegucigalpa, Lempira, or Santa Barbara; La Ceiba was the excluded group), whether student had baseline test score data, whether principal considered EOG test important at baseline, and whether teachers used EOG test data at baseline.

The results from the analysis using nonresponse adjustment weights, shown in Appendix D, are similar to the unweighted main results.

We also analyzed how student characteristics varied between students with endline data (the analytic sample) and those without endline data (the attrition group). Table B.3 presents these two groups' treatment status and characteristics on which we conducted subgroup analysis: rural or urban location, gender, and age. The two samples differ on location, age, and gender. The students missing endline EOG data are more likely to be from La Ceiba, less likely to be from Tegucigalpa, more likely to be male, and are older on average.

Table B.3. Characteristics of students with and without endline data (percentages except where noted)

Characteristic	Analytic sample <i>n</i> = 6,375	Attrition group <i>n</i> = 1,608	Difference
Treatment groups			
Group A	34.2	35.3	-1.3
Group B	35.7	35.3	0.4
Group C	30.1	29.4	0.7
Geographic area			
La Ceiba	7.4	10.3	-2.9*
Tegucigalpa	61.2	57.0	4.2*
Lempira	13.5	14.6	-1.1
Santa Barbara	18.0	18.2	-0.2
Age and gender			
Female (percentage)	49.2	44.4	5.2*
Average age (years)	6.4	6.7	-0.3*

Source: Reporting on the percentage of study schools in the randomized sample, and the number of students in the sample. The sample includes students who were enrolled in a school that had been randomized to one of the three groups, who was in the second grade as of May 31, 2015, the day before the intervention began.

**p* < 0.05.

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

APPENDIX C:

TABLES ON IMPACTS ON INTERMEDIATE OUTCOMES,
SUPPLEMENTAL TO CHAPTERS III AND IV

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

Table C.1. Participation in training

Characteristic	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Difference between FA + EOG and EOG only groups		Difference between EOG only and prevailing practice groups	
				(A) – (B)	p-value	(B) – (C)	p-value
Principals' participation in training							
Received training on EOG or FA during the current school year (percentage)	97.4	85.5	0.5	11.9*	(0.007)	85.0*	(0.000)
Number of FA or EOG training sessions in which principal participated over last two months	0.5	0.4	0.0 [†]	0.0	(0.730)	0.4*	(0.003)
Hours of training on FA or EOG in which principal participated over last two months	2.8	2.7	0.0 [†]	0.1	(0.919)	2.8*	(0.007)
Teachers' participation in training							
Received training on EOG or FA during the current school year (percentage)	98.4	65.3	2.1	33.1*	(0.000)	63.2*	(0.000)
Number of FA or EOG training sessions in which teacher participated over last two months	0.6	0.1	0.0 [†]	0.5*	(0.000)	0.1	(0.433)
Total hours of training on FA or EOG in which teacher participated over last two months	4.2	0.6	0.0 [†]	3.5*	(0.000)	0.7	(0.444)
Sample size (schools)	60	60	59	120		119	

Source: Principal and Teacher Surveys—Endline 2016. The Group C school that closed during the endline year did not participate in endline data collection.

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.

* $p < 0.05$.

[†] Indicates that the regression-adjusted score was rounded from –0.0 to 0.0.

EOG = end-of-grade.

FA = formative assessment.

Table C.2. End-of-grade training content

Characteristic	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Difference between FA + EOG and EOG only groups		Difference between EOG only and prevailing practice groups	
				(A) – (B)	p-value	(B) – (C)	p-value
Principals' training on key components (percentage)							
Participation in any training related to EOG test administration or analysis of results	91.3	87.9	4.2	3.4	(0.513)	83.8*	(0.000)
Participation in training on developing a school action plan based on EOG results	91.2	86.3	4.2	4.9	(0.360)	82.2*	(0.000)
Participation in training on adapting lesson plans based on EOG results	89.5	85.1	3.7	4.4	(0.441)	81.4*	(0.000)
Participation in training on administration of the EOG test	83.6	67.4	2.4	16.2*	(0.014)	65.0*	(0.000)
Teachers' training on key components (percentage)							
Participation in any training related to EOG test administration or analysis of results	81.6	69.7	2.0	12.0	(0.081)	67.6*	(0.000)
Participation in training on developing a school action plan based on EOG results	80.9	69.7	1.9	11.2	(0.107)	67.8*	(0.000)
Participation in training on adapting lesson plans based on EOG results	80.9	69.7	1.7	11.9	(0.088)	67.3*	(0.000)
Participation in training on administration of the EOG test	76.4	47.7	1.8	28.7*	(0.000)	45.8*	(0.000)
Sample size (schools)	60	60	59	120		119	

Source: Principal and Teacher Surveys—Endline 2016. The Group C school that closed during the endline year did not participate in endline data collection.

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.

* $p < 0.05$.

EOG = end-of-grade.

FA = formative assessment.

Table C.3. Formative assessment training content

Characteristic	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Difference between FA + EOG and EOG only groups		Difference between EOG only and prevailing practice groups	
				(A) – (B)	p-value	(B) – (C)	p-value
Principals' training on key components in the current year							
Participation in any training related to FA administration or analysis	98.7	13.4	1.2	85.3*	(0.000)	12.2*	(0.011)
Participation in training on FA administration	96.9	12.8	0.3	84.0*	(0.000)	12.5*	(0.008)
Participation in training on analysis of FA results	97.2	12.6	1.9	84.6*	(0.000)	10.7*	(0.030)
Participation in training on adapting lesson plans based on FA results	98.7	13.4	1.2	85.3*	(0.000)	12.2*	(0.011)
Teachers' training on key components in the current year							
Participation in any training related to FA administration or analysis	96.7	7.0	1.3	89.7*	(0.000)	5.7	(0.107)
Participation in training on FA administration	94.7	7.2	1.5	87.5*	(0.000)	5.7	(0.129)
Participation in training on analysis of FA results	97.1	4.4	1.0	92.7*	(0.000)	3.4	(0.255)
Participation in training on adapting lesson plans based on FA results	94.7	7.2	1.5	87.5*	(0.000)	5.7	(0.129)
Schools' teachers report having received training on developing classroom improvement plans based on FAs	96.7	7.0	1.3	89.7*	(0.000)	5.7	(0.107)
Sample size (schools)	60	60	59	120		119	

Source: Principal and Teacher Surveys—Endline 2016. The Group C school that closed during the endline year did not participate in endline data collection.

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.

* $p < 0.05$.

EOG = end-of-grade.

FA = formative assessment.

Table C.4. Access to end-of-grade results

Characteristic	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Difference between FA + EOG and EOG only groups		Difference between EOG only and prevailing practice groups	
				(A) – (B)	p-value	(B) – (C)	p-value
Principals' access to EOG results							
Principal reported that school participated in EOG test the year before	100.0 [†]	99.8	51.9	1.0	(0.864)	47.9*	(0.000)
Principal reported receiving results from past year's EOG test	98.2	100.0 [†]	22.1	-1.8	(0.708)	77.9*	(0.000)
Teachers' access to EOG results							
Teacher received current students' EOG results from the previous year	95.2	98.5	25.0	-3.3	(0.524)	73.5*	(0.000)
Teacher received current students' EOG results from the previous year, at the student level	94.1	93.1	7.9	1.0	(0.818)	85.2*	(0.000)
Sample size (schools)	60	60	59	120		119	

Source: Principal and Teacher Surveys—Endline 2016. The Group C school that closed during the endline year did not participate in endline data collection.

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.

[†] Indicates that the regression-adjusted score was rounded down to 100.

* *p* < 0.05.

EOG = end-of-grade.

FA = formative assessment.

Table C.5. Formative assessment materials and application

Characteristic	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Difference between FA + EOG and EOG only groups		Difference between EOG only and prevailing practice groups	
				(A) – (B)	p-value	(B) – (C)	p-value
Teacher has a copy of the FA instruction manual	96.6	38.7	47.1	58.0*	(0.000)	–8.4	(0.279)
Number of FA copies teacher has per student	1.3	0.4	0.6	1.0*	(0.000)	–0.3*	(0.002)
Number of times teacher has applied FA in reading in the current year	7.2	1.3	3.4	5.9*	(0.000)	–2.1*	(0.000)
Number of times teacher has applied FA in math in the current year	7.2	1.3	3.3	5.9*	(0.000)	–2.1*	(0.000)
Teacher could show enumerator written FA results	68.9	10.7	10.5	58.2*	(0.000)	0.2	(0.982)
Sample size (schools)	60	60	59	120		119	

Source: Teacher Survey—Endline 2016. The Group C school that closed during the endline year did not participate in endline data collection.

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.

* *p* < 0.05.

EOG = end-of-grade.

FA = formative assessment.

Table C.6. Use of end-of-grade results for school action plans

Outcome	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Difference between FA + EOG and EOG only groups		Difference between EOG only and prevailing practice groups	
				(A) – (B)	p-value	(B) – (C)	p-value
Principal survey							
Principal reports that the school has a SAP	99.5	100.0 [†]	16.1	-1.8	(0.667)	85.2*	(0.000)
Principal reports that they are implementing the SAP either completely or partially	99.6	100.0 [†]	16.1	-1.8	(0.667)	85.2*	(0.000)
School received support to develop or implement the SAP	99.9	97.1	1.4	2.8	(0.280)	95.8*	(0.000)
Sample size (schools)	60	60	59	120		119	

Source: Principal Survey—Endline 2016. The prevailing practice school that closed during the endline year did not participate in endline data collection.

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.

[†] Indicates that the regression-adjusted score was rounded down to 100.

* $p < 0.05$.

EOG = end-of-grade.

FA = formative assessment.

SAP = School action plan.

Table C.7. Teachers' use of end-of-grade results for lesson planning

Outcome	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Difference between FA + EOG and EOG only groups		Difference between EOG only and prevailing practice groups	
				(A) – (B)	p-value	(B) – (C)	p-value
Teacher received EOG results from previous year's test	95.2	98.5	25.0	-3.3	(0.524)	73.5*	(0.000)
Use of EOG results on reading							
Teacher adapted lesson plans	90.4	94.7	17.7	-4.3	(0.406)	77.1*	(0.000)
Teacher spent more time on standards that students had low scores on	80.7	75.7	13.0	5.0	(0.451)	62.7*	(0.000)
Teacher spent time outside of class to support struggling students	62.5	60.7	9.5	1.8	(0.823)	51.2*	(0.000)
Identified new teaching methods	62.4	55.6	8.8	6.8	(0.368)	46.8*	(0.000)
Identified new teaching materials	58.4	56.4	3.6	1.9	(0.796)	52.9*	(0.000)
Use of EOG results on math							
Teacher adapted lesson plans	90.5	94.0	17.5	-3.5	(0.506)	76.4*	(0.000)
Teacher spent more time on standards that students had low scores on	83.7	77.9	12.8	5.8	(0.383)	65.1*	(0.000)
Teacher spent time outside of class to support struggling students	63.3	59.8	6.2	3.4	(0.648)	53.7*	(0.000)
Identified new teaching methods	62.9	53.9	7.5	9.0	(0.228)	46.4*	(0.000)
Identified new teaching materials	57.2	54.6	6.7	2.7	(0.742)	47.9*	(0.000)
Sample size (schools)	60	60	59	120		119	

Source: Teacher Surveys—Endline 2016. The prevailing practice school that closed during the endline year did not participate in endline data collection.

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.

* $p < 0.05$.

EOG = end-of-grade.

FA = formative assessment.

Table C.8. Use of formative assessment results to improve teaching and learning

Outcome	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Difference between FA + EOG and EOG only groups		Difference between EOG only and prevailing practice groups	
				(A) – (B)	p-value	(B) – (C)	p-value
Teacher survey							
Schools' teachers report having administered FA	100.0 [†]	30.2	62.9	70.2*	(0.000)	–32.7*	(0.000)
Schools' teachers report modifying lesson plans based on FA results	89.5	24.7	50.8	64.9*	(0.000)	–26.1*	(0.001)
Schools' teachers report offering additional help to students with low scores on FA	96.2	26.5	52.4	69.7*	(0.000)	–25.9	(0.000)
Sample size (schools)	60	60	59	120		119	

Source: Teacher Survey—Endline 2016. The Group C school that closed during the endline year did not participate in endline data collection.

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.

[†] Indicates that the regression-adjusted score was rounded down to 100.

* *p* < 0.05.

EOG = end-of-grade.

FA = formative assessment.

APPENDIX D:

TABLES ON IMPACTS ON TEST SCORES, SUPPLEMENTAL TO CHAPTER V

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

Tables D.1 through D.3 provide supporting detail for results presented in Chapter V. Table D.4 presents a sensitivity check on the main results, by using nonresponse adjustment weights described in Appendix B.

Table D.1. Impacts on reading and math test scores (effects sizes): Overall and subscores

Score	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Impact of FA		Impact of EOG	
				(A) – (B)	<i>p</i> -value	(B) – (C)	<i>p</i> -value
Reading (overall)	0.13	0.00 [†]	–0.15	0.14*	(0.013)	0.15*	(0.017)
Reading comprehension	0.14	–0.02	–0.13	0.16*	(0.004)	0.11	(0.064)
Vocabulary	0.12	0.01	–0.14	0.11*	(0.035)	0.15*	(0.015)
Types of text	0.08	–0.01	–0.08	0.09	(0.090)	0.07	(0.185)
Math (overall)	0.12	–0.02	–0.12	0.14	(0.062)	0.11	(0.166)
Numbers and operations	0.10	–0.04	–0.07	0.14	(0.072)	0.03	(0.693)
Geometry	0.10	0.01	–0.13	0.09	(0.145)	0.14*	(0.042)
Measurement	0.12	–0.01	–0.12	0.14*	(0.028)	0.11	(0.089)
Algebra	0.02	0.02	–0.04	–0.00	(0.985)	0.05	(0.312)
Statistics and probability	0.08	0.04	–0.14	0.04	(0.374)	0.18*	(0.001)
Sample size (students)	2,180	2,273	1,922	4,453		4,195	

Source: EOG reading and math assessments 2016.

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.

[†]Value rounded up from –0.00 to 0.00.

* *p* < 0.05.

EOG = end-of-grade.

FA = formative assessment.

Table D.2. Impacts on reading and math test scores, by urban and rural area (effect sizes)

Score (standard deviations from the sample mean)	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Impact of FA		Impact of EOG	
				(A) – (B)	p-value	(B) – (C)	p-value
Reading							
Urban schools	0.11	0.12	-0.08	-0.01	(0.928)	0.19*	(0.013)
Rural schools	0.15	-0.23	-0.34	0.38*	(0.000)	0.11	(0.209)
Math							
Urban schools	-0.00	-0.01	-0.20	0.01	(0.941)	0.19*	(0.037)
Rural schools	0.37	0.00	0.01	0.37*	(0.003)	-0.01	(0.956)
Number of urban schools	28	28	27	56		55	
Number of rural schools	32	32	32	64		64	
Number of urban students	1,439	1,633	1,297	3,072		2,930	
Number of rural students	741	640	625	1,381		1,265	

Source: EOG reading and math assessments 2016.

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. The p-value on the test for the significance of difference in A–B impacts on reading between urban and rural schools was 0.002. The p-value on the test for the significance of difference in B–C impacts on reading between urban and rural schools was 0.448. The p-value on the test for the significance of difference in A–B impacts on math between urban and rural schools was 0.033. The p-value on the test for the significance of difference in B–C impacts on math between urban and rural schools was 0.264.

* p < 0.05.

EOG = end-of-grade.

FA = formative assessment.

Table D.3. Impacts on reading and math test scores, by gender (effect sizes)

Subject and subgroup (standard deviations from the sample mean)	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Impact of FA		Impact of EOG	
				(A) – (B)	p-value	(B) – (C)	P-value
Reading							
Female students	0.18	0.07	-0.06	0.11	(0.086)	0.13	(0.052)
Male students	0.09	-0.07	-0.23	0.16*	(0.006)	0.16*	(0.022)
Math							
Female students	0.13	-0.02	-0.10	0.15	(0.084)	0.08	(0.301)
Male students	0.12	-0.02	-0.14	0.13	(0.074)	0.13	(0.119)
Number of female students	1,093	1,091	950	2,184		2,041	
Number of male students	1,087	1,182	972	2,269		2,154	

Source: EOG reading assessment 2016.

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. The *p*-value on the test for the significance of difference in A–B impacts between female and male students was 0.426 for reading and 0.819 for math. The *p*-value on the test for the significance of difference in B–C impacts between female and male students was 0.645 for reading and 0.426 for math.

* $p < 0.05$.

EOG = end-of-grade.

FA = formative assessment.

Table D.4. Impacts on reading and math test scores (effect sizes): Overall and subscores with nonresponse weights

Score	EOG and FA (A)	EOG only (B)	Prevailing practice (C)	Impact of FA		Impact of EOG	
				(A) – (B)	<i>p</i> -value	(B) – (C)	<i>p</i> -value
Endline EOG test scores, 2016 (effect sizes)							
Reading (overall, effect sizes)	0.12	0.02	-0.16	0.10	(0.071)	0.18*	(0.003)
Reading comprehension	0.15	-0.03	-0.13	0.17*	(0.002)	0.10	(0.078)
Vocabulary	0.12	0.01	-0.14	0.12*	(0.024)	0.14*	(0.016)
Types of text	0.08	-0.01	-0.08	0.09	(0.066)	0.07	(0.195)
Math (overall)	0.12	-0.01	-0.13	0.13	(0.078)	0.12	(0.113)
Numbers and operations	0.10	-0.04	-0.07	0.14	(0.072)	0.03	(0.693)
Geometry	0.10	0.01	-0.13	0.09	(0.145)	0.14*	(0.042)
Measurement	0.12	-0.01	-0.12	0.14*	(0.028)	0.11	(0.089)
Algebra	0.02	0.02	-0.04	-0.00	(0.985)	0.05	(0.312)
Statistics and probability	0.08	0.04	-0.14	0.04	(0.374)	0.18*	(0.001)
Sample size (students)	2,180	2,273	1,922	4,453		4,195	

Source: EOG reading and math assessments 2016.

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.

* $p < 0.05$.

EOG = end-of-grade.

FA = formative assessment.

APPENDIX E:

SECONDARY QUESTION 1: HOW USEFUL IS NATIONAL CENSUS TESTING COMPARED WITH A SECURELY ADMINISTERED AND SCORED TEST IN A SAMPLE OF SCHOOLS? TEST INTEGRITY IN DIFFERENT TEST-TAKING AND SCORING CONDITIONS

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

To supplement to the study's main evaluation questions, the research team used administrative data to address the following question:

Question

How useful is national census testing compared with a securely administered and scored test? This question addresses the question of test integrity in different test-taking and scoring conditions.

Answer

Results presented below suggest that the two types of test administrations produced very different results in situations where the test and tested population should have been the same, suggesting that the national census testing may not be very useful. We therefore recommend using more secure test administration and scoring procedures for most policy goals.

Background

Honduras is a country with extensive standardized testing. Such testing is carried out in two ways. In recent years, a sample of students has been tested under strict administration and scoring criteria, while all other students have been tested under less secure testing and scoring conditions. The country's Ministry of Education administers national standardized tests to all students in the country in math and Spanish in every year (as funding allows) and in every grade from 1 to 9. The administration and scoring of the national tests, however, are not uniform, and it is not known whether the test results are valid and reliable. Fortunately, a subsample of schools were randomly selected to have the tests administered under more secure conditions (that is, with protections against manipulation of scores) and scored by an external organization. Results from this more secure test administration (STA), given nearly every year since 2007, can be used to gauge the quality of data coming from the census.

Several characteristics of the STA are intended to facilitate accurate measurement of students' knowledge and abilities on the day of the test (summarized in Table E.1). Our first question assesses to what extent these differences in administration may lead to differences in the distribution of scores. Since the STA sample is a random subset of the population of schools targeted by the census, there should be no systematic differences in distributions of scores from the two tests. The only differences should be the result of random sampling error. Any differences in score distributions beyond what we would expect by chance can be attributed to differences in administration and scoring.

Administration of tests by census and STA differs in three ways (Table E.1). The first difference is that schools selected for the STA sample receive the test booklets on the day of the test or one day before, so teachers cannot undermine the test by preparing students for specific questions on the test rather than by preparing them for the tested content more generally. In contrast, the census test typically arrives several days before the test. The second difference is that external observers are present in STA schools to ensure that proper test administration protocols are followed, including moving teachers to observe classes other than their own and verifying that students are sitting at separate desks where they are unable to see their classmates' responses or to work in groups. At census schools, volunteer observers may or may not be

present. The third difference is that at STA schools, the external observers collect the tests and take them to a central location for data entry and scoring. In contrast, at census schools, teachers enter their students’ responses themselves, affording teachers the opportunity to modify responses without being detected.

Table E.1. Characteristics of census testing and secure test administration

Characteristic	Census testing	STA testing
Test booklet security	Booklets arrive days in advance	Booklets sent to school day before or day of administration
Proctoring/monitoring	Local volunteer observers or no observers	Trained external observers present
Scoring	Test results entered by school staff, often by classroom teacher	Tests scored off-site by third party

Methods

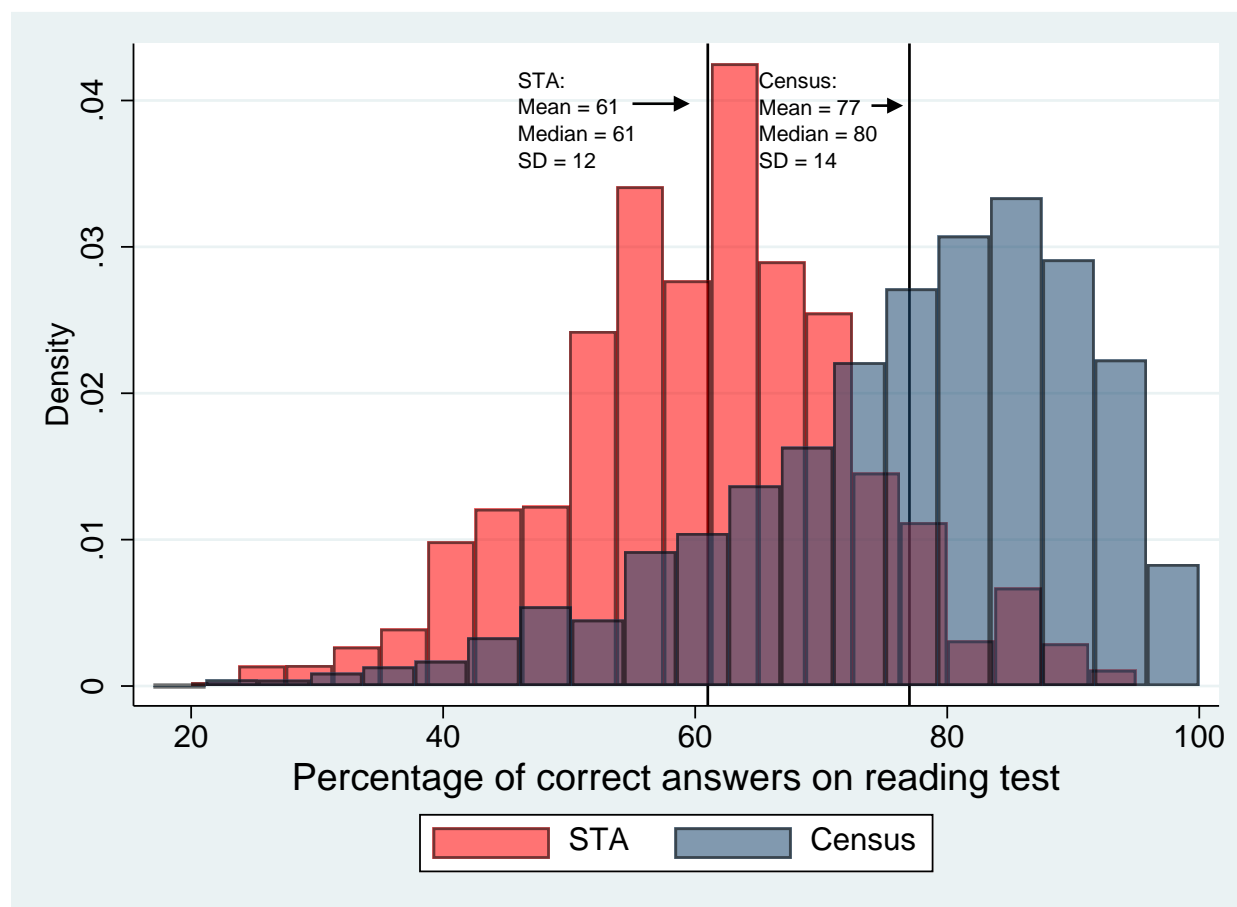
Although testing by census and STA occurred in 2012, 2013, and 2014, we focused our analysis on the testing done in 2013 because the 2013 sample was drawn at random and was comparable to census schools that met the same eligibility criteria. In 2012, a sample of schools that were not selected at random was added to the random sample of schools, making the overall sample not comparable to the population of schools. In 2014, NGOs (nongovernmental organizations) trained external observers for a large number of census schools and did off-site data entry for many of them, making that year’s census results atypical. We focused the analysis on test scores from primary grades 1 through 3, given the focus of our experimental analysis of EducAcción-PRI on those grades.

For a primary school to be eligible to participate in STA, it must not have been a weekend or evening school, and it must have had a minimum of 10 students enrolled in 6th grade. We did not have data on enrollment, but we did have data on the number of students that participated in EOG testing at every participating school. We excluded any school that did not have at least 10 6th-grade students participating in EOG census testing or STA and census schools that met in the evenings or on weekends. Excluding schools with fewer than 10 students *participating* in EOG testing is likely to exclude some schools that would have been eligible for STA on the basis of having 10 or more students enrolled. However, without enrollment data, we cannot identify those schools in the census data. We exclude schools with fewer than 10 students in the test score data because we can apply this criterion to both the STA and census groups. This eliminates 73 percent of schools in the census group, which reflects the fact that Honduras has a large number of small schools. This criterion also eliminates 24 percent of the STA sample—these are schools that had at least 10 6th-grade students enrolled, making them eligible for selection in the STA sample, but that had test data on fewer than 10 6th-grade students. We also exclude evening and weekend schools from the census group because evening and weekend schools were ineligible for STA.

Results

The distribution of scores for STA and census schools differed (Figures E.1 and E.2). The distribution of STA scores resembles a normal distribution, whereas the distribution of census scores is skewed, with more students scoring above 75 percent correct on the census test compared to the STA. Furthermore, we find a large difference between the average census and STA schools' test scores in both reading and math. Using the standard deviation from the distribution of census schools, the difference in mean scores is equal to 1.1 standard deviations for reading and 1.4 standard deviations for math.

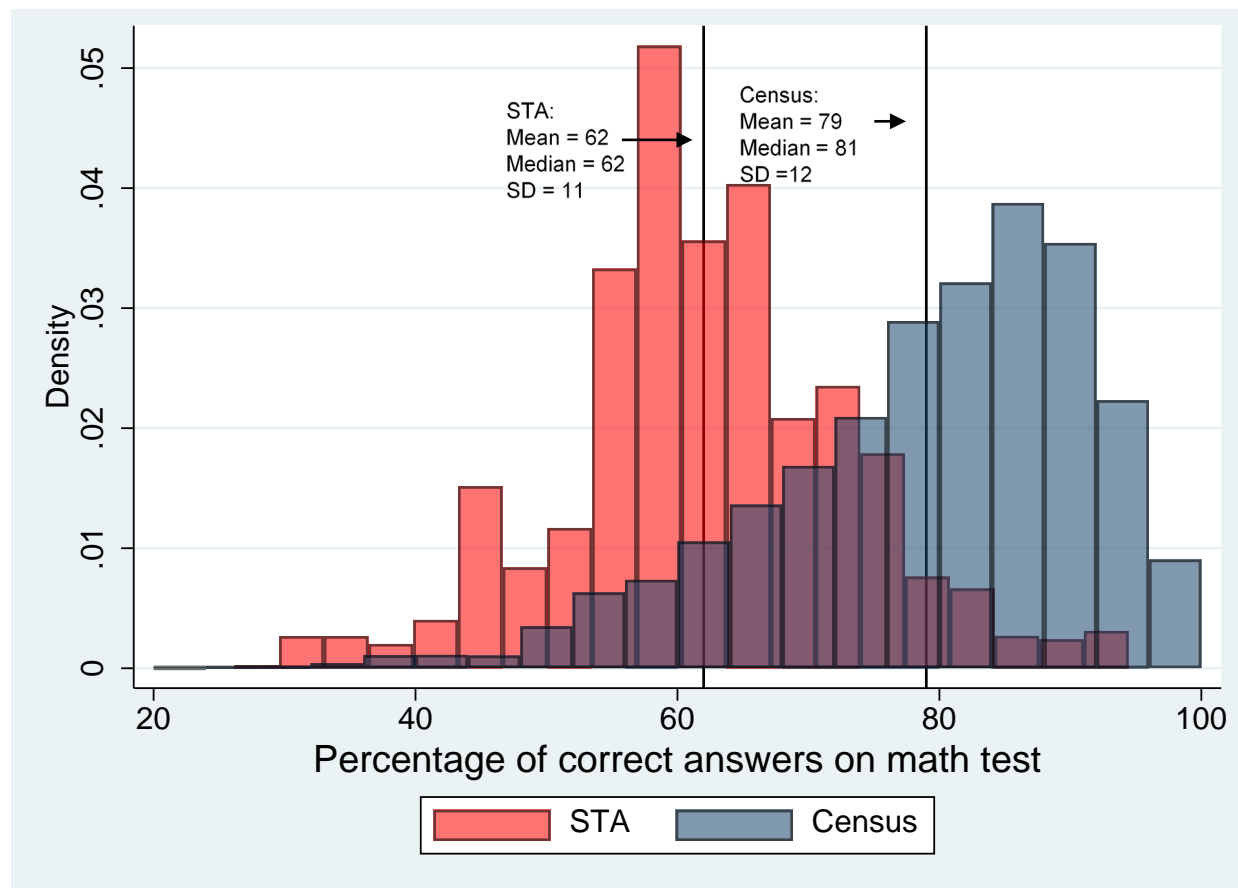
Figure E.1. STA and census score distribution: 2nd grade reading, 2013



Source: 2013 Census, 2013 STA.

Note: Census data for 2013 are available only at the classroom level, and so all results are aggregated to grade-by-school-level before tabulating. Results are weighted by school-grade size (number of students per grade per school) and pertain to grade 2, but they are similar for grades 1 and 3. To replicate the eligibility criteria for STA, we included census or STA data only for schools with test data for at least 10 6th-grade students, and we excluded census schools that had been identified in the 2012 census as being night or weekend schools.

Figure E.2. STA and census score distribution: 2nd grade math, 2013



Source: 2013 census, 2013 STA.

Note: Census data for 2013 are available only at the classroom level, so all results are aggregated to grade-by-school-level before tabulating. Results are weighted by school-grade size (number of students per grade per school) and pertain to grade 2, but they are similar for grades 1 and 3. To replicate the eligibility criteria for STA, we included census or STA data only for schools with test data for at least 10 6th-grade students, and we excluded census schools that had been identified in the 2012 census as being night or weekend schools.

Discussion

What could explain the differences in the distributions? We consider two explanations. First, school staff in census schools where tests were administered less securely may have taken steps to inflate student test scores. In census schools, where schools receive the tests days in advance of the test, teachers could game the tests by preparing their students for the specific questions they knew would be on the test. Furthermore, with less supervision, students in census schools could have received help from teachers inappropriately or copied their classmates' answers. Finally, teachers at census schools had the opportunity to change their students' incorrect responses to artificially improve low scores.

Inflation of census scores due to improper testing environment or manipulation of scores. Under current policy, EOG test results do not have consequences for schools or teachers.

However, in conversations with teachers and Ministry of Education staff in visits in 2014, teachers expressed concern that they might in fact suffer repercussions for poor test results, such as lower salaries. Teachers who feared repercussions for their students' potential poor performance may have been motivated to change student test scores or otherwise influence test scores. Furthermore, even if teachers felt assured that the Ministry of Education would not punish them for their students' poor performance, they may have feared repercussions from their principal or local authorities. In focus groups conducted for our experimental analysis, some coaches reported hearing teachers say that they used to alter student scores, fearing the consequences of low scores, but stopped after they came to appreciate the value of accurate scores as part of EducAcción-PRI and saw that scores were not used to punish teachers.

The second potential explanation is if, even after restricting the analysis to census schools that met the eligibility criteria for sample schools, the two groups are not comparable. We have identified several potential explanations for census schools' higher scores, and one theory that would suggest that census scores should in fact be lower.

- **School selection bias due to unequal internet access.** Census schools submit their test results by uploading them to a Ministry of Education website. Rural schools without electricity or Internet access may be less likely to be able to submit their scores; if these schools also have lower test scores, the omission of these scores could increase census scores on average.
- **Student selection bias due to low-performing student absences.** Lower-performing students may be less likely to attend school on test days. If teachers place less importance on the census test than on the STA test, those students would be more likely to be absent for the census test, increasing the average census test score.
- **Student selection bias due to the census test taking place later in the school year.** Both STA and census testing take place at the end of the academic year, but the census test takes place several weeks later. The student composition may differ in the final days of the school year for two reasons. First, some students stop attending school in the last weeks of school to work in the coffee harvest. If these student workers perform less well than their classmates who stay in school, their absence from census testing will elevate the census scores relative to the STA scores. Second, some schools tell their highest-performing students that they do not need to attend the last few weeks of school. These students are typically asked to return for the census test, but if they are less likely to return for the census test than the STA test, their absence could lower the census scores on average.

Recommendations

This analysis demonstrated a clear trade-off between quality of data and quantity of data. The census testing covers nearly the whole country, but yields a score distribution that looked suspiciously high compared with secure test administration. The STA test, however, was able to cover only a select group of schools. Improving both the quality and quantity—by administering a census under STA conditions—could be very costly. Nevertheless, a reasonable recommendation, given the findings above, would be to administer the census test earlier in each year and provide test booklets only on the day that the test is administered.

Otherwise, the way to resolve the tension between quantity and quality is to consider how the data will be used. Ministries of education have diverse motivations for promoting national EOG testing. Whether low security, census-type testing or higher security testing is more appropriate depends on policy objectives and funding. The recommendations below depend on policy objectives.

- *Monitoring national trends.* We recommend using a nationally representative sample of schools for STA-type testing to gather accurate data on national trends.
- *Monitoring trends at a local level, such as at the municipal or school level.* To monitor individual municipalities or schools, policymakers will require a larger sample and must consider trade-offs between the quality and quantity of data collected.
- *Rewarding or punishing schools through high-stakes testing.* If ministries intend to attach consequences to test results, such as funding for schools or teachers, schools will face incentives to manipulate test results and secure testing will be important.
- *Providing performance data for school improvement.* If the main objective in administering EOG tests is to provide schools with information they can use for school improvement, such as in EducAcción-PRI, test security is less important than participation. Providing test score data for school improvement could be part of efforts to promote a “culture of assessment,” in which assessment becomes part of an ongoing cycle of learning and feedback. If providing test score data for school improvement, ministries should train principals on how to use the test results to improve teaching and learning. If possible, ministries should make test results available by the beginning of the academic year following test administration.

If policymakers have multiple objectives in their testing, they will have to consider the trade-offs in the priorities described above.

APPENDIX F:

SECONDARY QUESTION 2: DOES SECURE ADMINISTRATION OF AN END-OF-
GRADE TEST IMPROVE SUBSEQUENT TEST SCORES?

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

To supplement to the study’s main evaluation questions, the research team used administrative data to address the following question:

Question

Does secure administration of an end-of-grade test improve subsequent test scores? It is possible to answer this question because Honduran authorities randomly selected a subset of schools to have securely administered tests, and subsequently the schools that had been randomized into census testing or secure testing took common tests in later years.

Answer

The results were mixed, but suggest the possibility that schools in which there is secure testing in one year may be less likely to inflate scores in future years, which bolsters the recommendation from Appendix E of relying more heavily on secure test administration than typical census test administration and scoring procedures. This hypothesis is explained in more detail below.

Background

Policymakers deciding between STA testing and less-secure census testing will consider many factors, such as cost and quality of data. Appendix E suggested that STA testing may produce more credible results, providing a possible rationale for the higher cost. However, another consideration is whether actually taking a securely administered and scored test has a positive impact on its own. This could occur if the feedback from the test is more accurate and therefore useful for school improvement planning. It could also focus greater attention on the tests as an outcome. Taking advantage of a natural experiment, the analysis presented in this appendix assesses whether schools that were randomly selected to participate in STA testing in 2013 had higher test scores in 2014 or 2015.

Methods and data

As discussed in Appendix E, eligible schools were randomly selected to participate in STA testing in 2013. Because they were selected at random, we do not expect to find systematic differences between the eligible schools that were selected for STA testing and those that were not. In the framework of a natural experiment, the “treatment” schools are those that participated in STA testing and the “control” schools are those that were eligible but were not chosen, and therefore participated in census testing instead.

To be eligible to participate in STA testing, a primary school must not have been a weekend or evening school, and must have had a minimum of 10 students enrolled in 6th grade. We did not have data on enrollment, but we did have data on the number of students that participated in EOG testing at every participating school. As we describe in Appendix E, we excluded any school that did not have at least 10 6th-grade students participating in EOG census testing or STA testing and census schools that met in evenings or on weekends. Excluding schools with fewer than 10 students participating in EOG testing is likely to exclude schools that would have been eligible for STA testing based on having 10 or more students enrolled. However, without enrollment data, we cannot identify those schools in the census data. We exclude schools with

fewer than 10 students in the test score data because we can apply this criterion to both the STA and census groups. This eliminates 73 percent of schools in the census group, which reflects the fact that Honduras has a large number of small schools. This criterion also eliminates 24 percent of the STA sample—these are schools that had at least 10 6th-grade students enrolled, making them eligible for selection in the STA sample, but had test data on fewer than 10 6th-grade students. We also exclude evening and weekend schools from the census group since evening and weekend schools were ineligible for STA testing.

Random assignment of eligible schools to the STA group allows us to estimate the impact of participating in STA testing on subsequent test scores by comparing outcomes for schools that participated in STA testing with outcomes of those that did not. To estimate the impacts of STA testing in 2013 on performance in 2014, we used student-level census data. For reading and math, we regressed students' 2014 test scores on a binary indicator for whether a school participated in STA testing in 2013, clustering standard errors at the school level. The 2014 STA sample was too small to use for analysis. To estimate the impacts of STA testing in 2013 on learning in 2015, we used classroom-level STA data. Again, we clustered standard errors at the school level. We weighted the analysis by classroom size so the results would represent impacts on the typical student rather than the typical classroom.

Limitations. There are two limitations to this analysis: concerns about (1) external validity and (2) internal validity. The study's external validity—or the extent to which one might expect the results from the study sample to be generalizable to the overall population—is limited by the fact that the analysis sample excludes some types of schools: schools that did not meet the criteria of eligibility for STA testing, schools with fewer than ten 6th-grade students in the 2013 test data (24 percent of STA schools and 73 percent of census schools¹⁶), and evening and weekend schools.

The study's internal validity—or the extent to which the difference in treatment and control group outcomes is attributable to factors beyond being an STA school—is limited by the potential for differential attrition between the two groups. We have identified two ways in which the STA and census samples used in our analysis could differ. The first would raise census scores relative to sample scores and the second could raise or lower census scores.

1. Schools assigned to the census group are excluded from the analysis sample if they fail to upload their census results. Schools without reliable Internet access may be less likely to upload their census results; STA schools do not rely on Internet access, as their tests are scored off-site. If schools without Internet access also have lower test scores than schools that are able to upload their scores, their omission would raise census scores. We do not have enrollment data for 2013, but using a list of 2012 schools, we estimate that as many as 28 percent of schools that were not in the STA data were missing census data.¹⁷

¹⁶ We repeated the analysis and included the small schools in the analysis sample; results were similar.

¹⁷ This is a rough estimate using the number of schools with 2013 test data and a list of all primary schools in Honduras in 2012. The percentage of schools with missing census data would be lower than 28 percent if some types of schools were not expected to participate in STA or census testing.

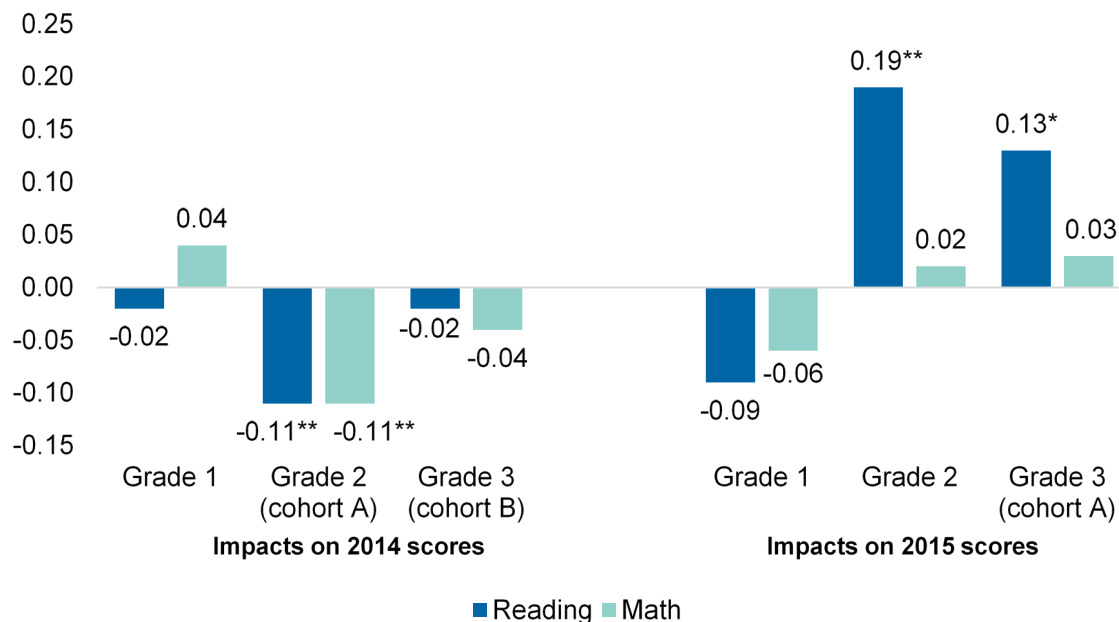
2. The composition of students who participate in EOG testing may differ between the two groups: students who stop attending school before the end of the year would be more likely to miss the census test than the STA test because the census test is administered several weeks after the STA test. The two common reasons for students to stop attending before the end of the school year are (1) some students stop attending to participate in the coffee harvest; and (2) some teachers tell their highest performing students they can stop attending during the final weeks of school, allowing teachers time to focus on lagging students who may be on the cusp of having to repeat a grade. Although students who have stopped attending are asked to return to take the census test, they may not. If their potential scores are different than those of their classmates who do take the test, their absence will affect overall scores. Without school-level enrollment data, we cannot estimate the percentage of missing student-level data.

Results

The estimated impact of a school's participation in 2013 STA testing on 2014 scores varied, by grade, between positive or nonsignificant. STA had no impact on 1st- or 3rd-grade students' test scores in 2014, but significantly decreased 2nd-grade students' test scores in reading and math by 0.11 standard deviations.

The impacts on 2015 scores were positive or nonsignificant. Second- and 3rd-grade students' reading test scores increased by 0.19 standard deviations ($p < 0.05$) and 0.13 standard deviations ($p < 0.10$), respectively, for students at schools that participated in STA testing in 2013 relative to students at schools that participated in census testing in the same year. There was no impact on 1st-grade students' reading test scores or on math scores in any grade. It should be noted that most of those students were not in school for the 2013 STA testing. Those who were in school at that time are labeled with their cohorts in Figure F.1. Cohort A was in 1st grade in 2013 when STA testing occurred. Cohort B was in second grade in 2013. Figure F.1 shows estimated impacts on 2014 census scores and 2015 STA scores for grades 1 through 3. Only two of the cohorts represented in the figure were old enough to have participated in the 2013 EOG testing: those in 2nd or 3rd grade in 2014 (marked in the figure as cohorts A and B, respectively) and those in 3rd grade in 2015 (marked as cohort A).

Figure F.1. Impact of secure test administration on future test scores, by grade, subject, and year



Source: STA and census test score data for 2013, 2014, and 2015.

Note: The analysis sample for estimating impacts on 2014 census scores includes 8,118 to 8,315 students, depending on the cohort, in the treatment group (2013 STA); and 61,194 to 63,404 students in the comparison group (2013 census testing).

The analysis sample for estimating impacts on 2015 STA scores includes 114 schools in the treatment group and 1,002–1,003 schools in the control group, depending on the cohort.

Impacts are expressed as effect sizes. Standard errors (not shown) are clustered at the school level.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

Discussion

The estimated impacts of participating in STA testing were mixed one and two years after testing, with results varying by grade, and then by grade and subject in 2015. Evidence of negative impact in the first year after STA testing turned into positive impacts on reading scores in the following year. Several grade and subject combinations showed no statistically detectable impact.

We can only speculate on what might explain these findings. It seems unlikely that participating in STA testing would have made students learn less in the following year. A more plausible explanation is that the experience of participating in STA testing in 2013 might have lowered test scores in 2014 by reducing manipulation (inflation) of test scores, as evidence presented in Appendix E suggested. For example, after participating in STA testing, school staff may have incorporated practices such as separating students’ desks and refraining from helping students during the test. Furthermore, after receiving written reports on STA test results, school staff may have developed an appreciation for the value of accurate test results, reducing any incentive to take steps to inflate students’ scores on the 2014 census test. This is consistent with

reports in focus groups (see Chapter IV) that teachers were less likely to manipulate students' scores after learning to use EOG test results through EducAcción-PRI and developing an appreciation for their value.

Test scores at schools that participated in STA testing were significantly higher for reading, with no impact on math scores, two years after their participation. This could be explained by a lagged impact on reading learning, such as through making better use of data from prior years to improve instruction. Alternatively, if school staff did in fact develop an appreciation for the value of STA testing, or EOG testing in general, they may have been more motivated to prepare their students for EOG testing, demonstrating an accountability effect.

Recommendations

These findings suggest that participating in STA testing may lead to lower rates of manipulation on the census test after one year, and higher rates of learning, as measured by STA testing, after two years. The regression results are not conclusive, however, given that impacts are not found in all grades or in both subjects, and given that the regression estimates indicate impacts but do not indicate what led to the impacts. Nonetheless, in combination with the findings of the impact evaluation of EducAcción-PRI, these results support more widespread adoption of STA. The results on the impact of STA testing in subsequent years, however, are not enough by themselves to support that conclusion.

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

APPENDIX G:

SECONDARY QUESTION 3: WHAT ARE THE IMPACTS OF ASSESSMENT-RELATED TRAINING ON TEST SCORES?

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

To supplement to the study's main evaluation questions, the research team used administrative data to address the following question:

Question

What are the impacts of assessment-related training on test scores? The question can be addressed using administrative data because USAID had been funding precursor programs to EducAcción-PRI in the years preceding that project, a period over which considerable test data are available. The study team took advantage of the fact that the program activities were rolled out over time, so there are some municipalities that received a related intervention later than others, and could serve as a comparison group for those that received services earlier.

Answer

The best available non-experimental evidence suggests that the impact of assessment-related training on test scores was negative. This surprising finding was overturned by the subsequent randomized controlled trial presented in the body of this report. Two key differences between the non-experimental evidence presented in this appendix and the experimental evidence presented in Chapter V are the methods (use of random assignment) and the maturity of the program. The more rigorous impact evaluation using random assignment was conducted on a second-generation implementation of the intervention, which was more intensive and presumably more targeted than the intervention described below. This finding motivates the recommendation, presented in Chapter VII, for policymakers to ensure that evaluations are rolled out in a manner that allows for rigorous impact evaluation of a mature program.

Background

In 2011, USAID contracted with American Institutes for Research to help schools in Honduras design, collect, and make better use of student assessment data through two projects called MIDEH and EducAcción. MIDEH supported the development of a new curriculum and a national testing system, including summative end-of-grade (EOG) testing and formative assessments linked to that curriculum. Both MIDEH and EducAcción trained municipal education authorities, schools, and teachers in targeted municipalities. MIDEH's school-level activities were short-term, whereas EducAcción's were a combination of short-term and ongoing. Each project targeted specific municipalities, as discussed below.

To complement our impact evaluation, which uses a random assignment design to estimate the impacts of the EducAcción-PRI intervention on student achievement, the analysis in this appendix presents quasi-experimental evidence on a similar question with a less intensive set of interventions. The EducAcción-PRI intervention, which is the focus of this report, included materials and structured bimonthly support for principals and teachers. The EducAcción and MIDEH activities that preceded EducAcción-PRI included a diverse set of interventions. Much of the work was with municipal authorities. Of the school-level work, many of the trainings lasted one to two days, in contrast to the long-term approach taken with EducAcción-PRI. This evaluation of MIDEH and EducAcción activities that preceded EducAcción-PRI takes advantage

of a natural experiment—specifically, from 2012 to 2015¹⁸—when MIDEH and EducAcción projects were already operating and longitudinal (school-level) test score data were available for that period.

Methods

The natural experiment arises from the fact that the programs were rolled out gradually and did not reach all municipalities in the country. To prioritize municipalities' need for receiving support services, the two projects each used an index that captured each municipality's education, human development, and security indicators. The indices were continuous measures with a sharp cutoff, such that municipalities above the cutoff were selected as target municipalities and municipalities with scores below the cutoff were not. We used an instrumental variables approach to estimate impacts, in which the index variable is used to predict participation, and then the predicted participation variable is effectively the treatment variable, not the participation indicator itself.

EducAcción's index did not significantly predict selection as a target municipality, meaning that we were unable to use the instrumental variables method to estimate the impacts of being selected as one of EducAcción's target municipalities. This may be partly because EducAcción selected some target municipalities before developing the index. Therefore, some of their target municipalities have scores below the cutoff. Regardless of the reason, the inability of the index to predict a municipality's intervention status means it was not a valid instrument for EducAcción participation status.

The index was able to predict whether a municipality received services through the MIDEH project. We conducted instrumental variables analysis for MIDEH to estimate the project's impacts on EOG reading and math test scores from 2013 (census and STA), 2014 (census and STA), and 2015 (STA only)—one, two, and three years after beginning implementation in target municipalities. The analysis included grades 1, 2, and 3. We used MIDEH's continuous index for identifying target municipalities as the instrumental variable. As a robustness check, we repeated the analysis using a binary version of the index as the instrument.

Results

We conducted a specification test to verify that MIDEH's index for selecting targeted municipalities was a valid instrument; results confirmed that it was. Table G.1 shows the results of the first-stage regressions, in which we regress a binary indicator for being a MIDEH target municipality on the index variable that MIDEH used to select target municipalities. This table shows results of first-stage regressions for 3rd-grade reading (results for other grades and for math were similar). We find a significant relationship between the MIDEH index and selection as a target municipality; the coefficient estimate on the MIDEH index was significantly different from zero at the 1 percent level in all regressions. The F-statistics were over 15 in all cases.

¹⁸ There was a short period of overlap in the second half of 2015 when MIDEH and EducAcción activities were ongoing and EducAcción-PRI had begun.

Table G.1. First-stage regressions: Relationship between MIDEH index and selection as a MIDEH target municipality

Statistic	2013		2014		2015
	Census	STA	Census	STA	STA
Coefficient on MIDEH index	0.095***	0.094***	0.091***	0.100***	0.099***
Coefficient standard error	(0.017)	(0.016)	(0.017)	(0.017)	(0.010)
F-statistic	19.47	15.16	27.32	17.62	52.22

Sources: EOG census test score data (2013 and 2014) and EOG STA data (2013, 2014, and 2015). MIDEH index data and MIDEH target municipality data. Results are shown for grade 3. Results for grades 1 and 2 and for math for all three grades were similar.

*** Significantly different from zero at the .01 level, two-tailed test.

In the second stage, we estimated the impacts of being a MIDEH target municipality on reading and math test scores. We found significant negative impacts on census EOG scores in 2013 and 2014, and on STA scores in 2014 and 2015. We found smaller impacts, both positive and negative, on the 2013 STA test scores. Table G.2 summarizes these results. Results for reading were similar to results for math.

Table G.2. Second-stage regressions: Estimated impact of selection as a MIDEH target municipality on 2013, 2014, and 2015 end-of-grade test scores for reading and math

Outcomes	Statistic	Reading			Math		
		Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3
2013 Census	Coefficient on MIDEH municipality ⁺	-0.23**	-0.27***	-0.11	-0.25*	-0.37***	-0.14
	Standard error	0.093	0.097	0.115	0.133	0.101	0.129
	Sample size (municipalities)	292	293	291	292	293	291
2013 STA	Coefficient on MIDEH municipality	0.19	0.08	0.30	0.13	-0.04	-0.03
	Standard error	0.157	0.315	0.291	0.242	0.202	0.248
	Sample size (municipalities)	155	155	155	155	155	155
2014 Census	Coefficient on MIDEH municipality	-0.21***	-0.31**	-0.22*	-0.20**	-0.30**	-0.18
	Standard error	0.082	0.126	0.124	0.087	0.133	0.131
	Sample size (municipalities)	297	297	297	297	297	297
2014 STA	Coefficient on MIDEH municipality	-0.40***	-0.22	-0.29	-0.22	-0.37***	-0.06
	Standard error	0.144	0.164	0.186	0.189	0.140	0.172
	Sample size (municipalities)	164	163	164	164	163	164
2015 STA	Coefficient on MIDEH municipality	-0.48***	-0.28***	-0.33***	-0.29*	-0.41***	-0.35***
	Standard error	0.158	0.101	0.113	0.146	0.111	0.127
	Sample size (municipalities)	298	298	298	298	298	298

Sources: EOG census test score data (2013 and 2014) and EOG STA data (2013, 2014, and 2015). MIDEH index data and MIDEH target municipality data.

*The coefficient is estimated for the predicted status as a targeted MIDEH municipality, based on MIDEH's continuous municipal index, rather than on a municipality's actual status as a targeted MIDEH municipality.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

These findings are sensitive to the specification of the instrumental variables regression. When we repeated our analysis of the impacts of MIDEH using a binary version of the municipal index, impact estimates were smaller in absolute value and were not significant for some years. Of 30 tests run, 8 were significantly negative at the 5 percent level with the binary instrument compared to 14 for the continuous instrument.

Discussion

The data suggest that being selected as one of MIDEH's target municipalities led to lower EOG test scores in four of the five tests we included, but the results were mixed and inconclusive for the fifth test. Results were not significant when we used an alternative specification for the instrumental variables regression. Negative impacts on census testing could be attributable to school staff learning about appropriate testing methods, and a reduction in the manipulation of test results. Negative impacts on STA test results are harder to explain, as these explanations are not as applicable.

MIDEH has made important contributions to Honduras's education system, among them developing its current curriculum and testing system and working with national institutions to ensure that national staff have the capacity to continue MIDEH's work when the MIDEH project ends. The impacts of this work, which was national in scope rather than targeted at certain municipalities, are not captured in the impact estimates presented here. The impact estimates, drawn from the quasi-experimental design used for this analysis, suggest that MIDEH's training efforts in schools did not lead to measurable learning growth.

This analysis differs in two important ways from our experimental analysis of EducAcción-PRI. First, the interventions were different. MIDEH's activities were diverse, serving municipal authorities and a selection of schools in target municipalities. The support that schools received was often in the form of one-time trainings. This is in contrast to the EducAcción-PRI intervention, which involved a long-term presence in a concentrated sample of schools. The second difference is that we used a randomized controlled trial to estimate EducAcción-PRI's impacts in study schools, in contrast to the quasi-experimental design using existing administrative data to estimate the impacts of MIDEH's work in targeted municipalities. In the EducAcción-PRI study, schools received the treatment they were assigned, resulting in clean contrasts among the three treatment groups, increasing our ability to detect impacts. In contrast, our evaluation of MIDEH's work in targeted municipalities compares municipalities in which schools either received MIDEH's intervention indirectly through training municipal authorities, or in which only some schools received training directly. This contrast is weaker, making it more difficult to detect impacts.

Recommendations

Absent any other information, and when available resources allow, policymakers may want to consider focusing efforts on more intensive training efforts like those used in EducAcción-PRI rather than the one- to two-day trainings offered through MIDEH. However, to answer the question about the impacts of a project like MIDEH, it would be important to conduct a more rigorous study, ideally a controlled experiment, to draw more firm conclusions.

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

APPENDIX H:

TEACHER AND PRINCIPAL SURVEY INSTRUMENTS: BASELINE, MIDLINE,
AND ENDLINE YEARS

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

The baseline, midline, and endline survey instruments will be available online upon request at USAID's Development Data Library website: <https://data.usaid.gov/>.

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

APPENDIX I:

DISCUSSION GUIDES FOR FOCUS GROUPS AND INTERVIEWS:
MIDLINE AND ENDLINE YEARS

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

This appendix includes the guides used for interviews and focus groups conducted at midline and endline. The interviews and focus groups were all conducted in Spanish and the guides included here are in the original Spanish.

For endline, we have included both an interview guide and a focus group discussion guide for coaches because one of the teaching coaches did a phone interview because of being unable to travel to participate in the focus group.

Guides used at midline (2015):

- | | |
|--|------|
| 1. Interview guide for teaching coaches (<i>asesores</i>) | I.5 |
| 2. Interview guide for lead teaching coaches (<i>asesores principales</i>) | I.9 |
| 3. Interview guide for key EducAcción staff | I.13 |
| 4. Focus group discussion guide for principals | I.17 |
| 5. Focus group discussion guide for teachers | I.23 |

Guides used at endline (2016):

- | | |
|--|------|
| 1. Interview guide for teaching coaches (<i>asesores</i>) | I.29 |
| 2. Interview guide for lead teaching coaches (<i>asesores principales</i>) | I.33 |
| 3. Interview guide for key EducAcción staff | I.37 |
| 4. Focus group discussion guide for teaching coaches (<i>asesores</i>) | I.41 |
| 5. Focus group discussion guide for principals | I.45 |
| 6. Focus group discussion guide for teachers | I.53 |

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

ESTUDIO CUALITATIVO DE IMPLEMENTACIÓN

GUÍA DE DISCUSIÓN PARA ENTREVISTAS CON ASESORES

Hola, soy César Valenzuela de Espirállica. Estamos llevando a cabo un estudio sobre pruebas formativas (PF) y pruebas de fin de grado (PFG). El estudio de investigación es financiado por USAID, la Agencia de Estados Unidos para el Desarrollo Internacional, y está siendo llevado a cabo por Mathematica Policy Research. Usted ha sido seleccionado para la discusión porque es un asesor.

La entrevista incluirá preguntas acerca de su trabajo como asesor. Esta información servirá para comprender mejor el uso de pruebas formativas (PF) (pruebas mensuales) y de fin de grado (PFG) (pruebas anuales). Toda la información que obtenemos se mantendrá estrictamente confidencial y esta información no se dará a conocer en ninguna forma que permita identificarle a usted o sus respuestas. La información que proporcione puede ser vista por USAID. Esta información se utilizará solamente para fines de evaluación, y se removerá toda información identificable como nombres o información de contacto. Una vez que se complete el estudio, los datos del estudio que no le identifican personalmente pueden estar a disposición del público para permitir análisis adicionales.

Esta entrevista llevará alrededor de una hora. Su participación es voluntaria y puede optar por no contestar alguna o todas las preguntas por cualquier motivo. En otras palabras, tiene la alternativa de no participar. No hay sanción o cambio en su situación laboral si no participa en este estudio. No hay riesgos ni beneficios directos para usted por participar en este estudio.

¿Está de acuerdo en participar en este estudio?

Debido a que no podemos anotar toda la conversación, nos gustaría grabar la entrevista. La grabación se usará únicamente para transcribir la conversación y toda información que permita identificarlo será removida.

¿Está de acuerdo en que grabemos la entrevista?

GUÍA TEMÁTICA

A. Perfil del Asesor

- Cuénteme un poco sobre usted.
 - ¿Cuánto tiempo ha estado con el proyecto?
 - ¿Qué es lo que entiende que es su trabajo?
 - ¿Qué tan bien preparado se siente para hacerlo?
 - ¿Recibió alguna capacitación? De ser así, ¿qué temas cubrió?
 - ¿En qué grupo de evaluación ha estado trabajando (grupo A o grupo B)?
 - ¿Cuántas escuelas le fueron asignadas?

- ¿Qué tipo de trabajo hizo antes de empezar a trabajar con EducAcción PRI?
 - Si un maestro: ¿Qué grados enseñó?
- Hábleme de cómo trabaja con directores y maestros en las escuelas que cubre. *Sondear para entender:*
 - ¿Trabaja más con directores o maestros? ¿Qué discute con directores y maestros?
 - ¿Trabaja con maestros en grupos o individualmente?
 - ¿Trabaja más con maestros de 1er a 3er grado (*primer ciclo*) o maestros de 4to a 6to grado?
 - ¿Con qué frecuencia visita? ¿Visita algunas escuelas con más frecuencia que otras? ¿Cuánto tiempo pasa en cada visita?
 - ¿Qué tipo de trabajo hace para prepararse para sus visitas?
- ¿Qué conceptos han sido más difíciles de transmitir a los maestros? ¿Tiene algún ejemplo?
- ¿Qué conceptos han parecido más útiles para los maestros? ¿Tiene algún ejemplo?

B. Para Asesores trabajando en Grupo A (PFG + PF) solamente

[ESTA SECCIÓN ES PARA ASESORES TRABAJANDO EN GRUPO A (PFG + PF) SOLAMENTE.]

Hablemos de su trabajo en el grupo de intervención A.

- [PREGUNTA CLAVE]¿Cómo proporciona apoyo a los maestros? ¿Puede dar ejemplos? (*Sondeo:* ¿Qué cambios está sugiriendo? ¿Cómo sugiere estos cambios?)
- [PREGUNTA CLAVE]¿Cómo han reaccionado los maestros a su apoyo?
- ¿A qué maestros ha podido dar copias de PF? (*Sondeo:* ¿Qué grados? ¿A todos los maestros en esos grados?)
- ¿Cuántas copias ha podido dar a cada maestro? ¿Es esto suficiente para todos sus estudiantes?
- ¿Le dio a todos los maestros una copia del manual de instrucciones para los/el grado(s) que enseña?
- ¿Percibe que el proyecto está teniendo un impacto? En caso afirmativo, ¿qué impacto está teniendo? Si no, ¿por qué cree que no está teniendo ningún impacto?

C. Para Asesores trabajando en Grupo B (PFG) solamente

[ESTA SECCIÓN ES PARA ASESORES TRABAJANDO EN GRUPO B SOLAMENTE]

Hablemos de su trabajo en el grupo de intervención B.

- [PREGUNTA CLAVE] ¿Cómo proporciona apoyo a los maestros? ¿Puede dar ejemplos? (Sondeo: ¿Qué cambios está sugiriendo? ¿Cómo sugiere estos cambios?)
- [PREGUNTA CLAVE]¿Cómo han reaccionado los maestros a su apoyo?
- ¿Han sugerido los maestros usar evaluaciones formativas afín de mejorar el aprendizaje?
- ¿Ha utilizado alguna estrategia para disminuir la probabilidad de que los maestros usen evaluaciones formativas?
- ¿Ha enfrentado algún desafío de maestros queriendo ampliar el uso de evaluaciones formativas en sus escuelas (del grupo B)?
- ¿Percibe que el proyecto está teniendo un impacto? En caso afirmativo, ¿qué impacto está teniendo? Si no, ¿por qué cree que no está teniendo ningún impacto?

D. Para todos los Asesores

Esta sección es para todos los asesores.

- ¿Ha notado maestros que reciben copias de PF de alguna otra fuente que no sea EducAcción PRI? ¿De quién? ¿Cuántos y con qué frecuencia?
- ¿Sabe de maestros que reciben apoyo técnico o pedagógico para PFG o PF de alguien que no sea EducAcción PRI?
- ¿Cuáles fueron los aspectos más difíciles en sus tareas como *asesor*?
- ¿Qué cree que se puede hacer para mejorar el uso de PFG y PF?

E. Capacitación

Hablemos sobre capacitación.

- ¿Utiliza la capacitación recibida de EducAcción de otras maneras además del apoyo a maestros en las escuelas a que fue asignado? [TRATE DE AVERIGUAR SI TRABAJA CON OTROS MAESTROS EN LAS MISMAS ESCUELAS, OTRAS ESCUELAS, POSIBLEMENTE DEL GRUPO C, O SI HAY BENEFICIOS POTENCIALMENTE NO MEDIDOS]

F. Conclusión

- En general, ¿cree que las cosas cambiaron para las escuelas desde que empezó la intervención (por ejemplo, con respecto a la dotación de personal, plan de estudios o actividades)? Si es así, ¿cómo han cambiado las cosas?
- En su opinión, ¿cuáles cree que son los mayores desafíos que enfrentan las escuelas donde trabaja?

- Que ustedes sepan, ¿de qué otra manera supervisan las escuelas el aprendizaje de los estudiantes?
- ¿Hay algo más que le gustaría compartir con nosotros?

ESTUDIO CUALITATIVO DE IMPLEMENTACIÓN

GUÍA DE DISCUSIÓN PARA ENTREVISTAS CON ASESORES PRINCIPALES

Hola, soy Chantal Toledo de Mathematica Policy Research. Estamos llevando a cabo un estudio sobre la implementación PRI. El estudio de investigación es financiado por USAID, la Agencia de Estados Unidos para el Desarrollo Internacional, y está siendo llevado a cabo por Mathematica Policy Research. Usted ha sido seleccionado(a) para una entrevista porque es un asesor principal.

La entrevista incluirá preguntas acerca de su trabajo como asesor principal. Esta información servirá para comprender mejor la implementación PRI. Toda la información que obtenemos se mantendrá estrictamente confidencial y esta información no se dará a conocer en ninguna forma que permita identificarle a usted o sus respuestas. La información que proporcione puede ser vista por USAID. Esta información se utilizará solamente para fines de evaluación, y toda información identificable como nombre o información de contacto será removida. Una vez que se complete el estudio, los datos del estudio que no le identifican personalmente pueden ser puestos a disposición del público para permitir análisis adicionales.

Esta entrevista llevará alrededor de una hora. Su participación es voluntaria y puede optar por no contestar alguna o todas las preguntas por cualquier motivo. En otras palabras, tiene la alternativa de no participar. No hay sanción o cambio en su situación laboral si no participa en este estudio. No hay riesgos ni beneficios directos para usted por participar en este estudio.

¿Está de acuerdo en participar en este estudio investigativo?

Debido a que no podemos anotar toda la conversación, nos gustaría grabar la entrevista. La grabación se usará únicamente para transcribir la conversación y toda información que permita identificarlo será removida.

¿Está de acuerdo en que grabemos la entrevista?

GUÍA TEMÁTICA

A. Funciones y responsabilidades

- Cuénteme un poco sobre usted.
 - ¿En qué grupo de evaluación ha estado trabajando?
 - ¿Cuántas escuelas le fueron asignadas?
 - ¿Hace cuánto que es maestro?
 - ¿Qué grados enseña/ enseñaba?

Me gustaría preguntarle acerca de su función en la implementación PRI.

- ¿Cuál es su función y cuáles son sus responsabilidades en la implementación PRI?
- ¿Qué tipo de capacitación y apoyo recibió durante la implementación del programa?
¿Quién proporcionó la capacitación y apoyo?
- ¿Cómo está organizado el equipo de implementación? ¿Con quién se comunica para organizar las tareas y discutir las responsabilidades y actividades de campo?
- ¿De qué forma siente que ha contribuido a la implementación (tiempo, habilidades, experiencia)?

B. Fidelidad

- ¿Qué tan útil fue la formación de asesores de Mayo 2015? Cómo fue diseñada la formación? ¿Cuánto aporte del Ministerio de Educación tuvieron?
- ¿Cuáles han sido los principales retos en la capacitación y entrenamiento de maestros?
¿Cómo fueron resueltos? (*SONDEO: EXPLORAR DESAFÍOS ESPECÍFICOS POR GRUPO EXPERIMENTAL: GRUPO A Y GRUPO B*).
- ¿Cómo se implementó el componente prueba formativa (PF) de la evaluación? (*Sondeo: ¿Quién implementó el componente? ¿Cuáles fueron los diferentes aspectos? ¿Cuándo y dónde se realizó? ¿Cuáles fueron los desafíos para obtener y usar datos de PF? ¿Se usaron los datos de PF? ¿Fueron los datos de PF comprendidos por los maestros? (Sondeo: ¿Pudieron los maestros usar los datos de PF para fundamentar la instrucción?)*)
- ¿Cómo se implementó el componente prueba de fin de grado (PFG) de la evaluación? (*Sondeo: ¿Quién implementó el componente? ¿Cuáles fueron los diferentes aspectos? ¿Cuándo y dónde se realizó? ¿Cuáles fueron los desafíos para obtener y usar datos de PFG? ¿Se usaron los datos de PFG? ¿Fueron los datos de PFG comprendidos por los maestros? (Sondeo: ¿Pudieron los maestros usar los datos de PFG para fundamentar (informar/ajustar) la instrucción?)*)
- ¿Qué (aspectos de la implementación) no se realizó de acuerdo al plan o directrices?
- ¿Han habido retos derivados de maestros queriendo ampliar su uso de PF en las escuelas del grupo B?
- ¿Hubo alguna diferencia en servicio (implementación) en las escuelas debido a su tamaño, ubicación geográfica o cualquier otra característica relevante?
- ¿Hubo alguna escuela con problemas de seguridad que afectaron su trabajo?
- En general, ¿cómo describiría la experiencia global de los asesores?

- ¿Cree que la diferencia en la carga de trabajo de los asesores en los grupos A y B puede haber afectado la *intensidad* del componente PFG de la intervención? (*Sondeo*: ¿Perciben los asesores una diferencia en la carga de trabajo? ¿Hay también una diferencia en la intensidad de la ayuda establecida a las escuelas?)

C. Percepciones del asesor principal

- ¿Cómo han respondido los maestros, padres y estudiantes a la implementación?
- ¿Cómo se están implementando las pruebas formativas (PF) en las escuelas? ¿Cómo se están implementando las pruebas de fin de grado (PFG) en las escuelas?
- ¿Cómo han recibido las actividades generales de PRI los directores, maestros, padres y niños? ¿Cómo han recibido las actividades PF? ¿Cómo han recibido las actividades PFG?
- ¿Hubo algún reto surgido de la participación (de padres o niños) en las actividades de implementación? ¿Cómo se abordaron esos retos?
- ¿Cómo describiría su interacción con las autoridades locales referente a la implementación de PF? ¿Y con respecto a la implementación de PFG?
- ¿De qué forma piensa que las pruebas PF y PFG pueden ser mejoradas para servir mejor a maestros y estudiantes?
- ¿De qué forma piensa que se puede mejorar el apoyo a maestros y escuelas en cómo usar los resultados de PF y PFG para servir mejor a maestros y estudiantes? ¿Qué cambios serían necesarios?

D. Metas del programa e impactos percibidos

- ¿Cómo hace seguimiento y mide el logro de las metas del programa? (*Sondeo*: ¿Cómo se establecen puntos de referencia y seguimiento? ¿Cómo se ha ajustado para cumplir las metas de implementación?).
- ¿Cree que los estudiantes en el programa en realidad están mejorando sus niveles de lectura y matemáticas como un resultado de la intervención? ¿Ha beneficiado a niños y sus familias el programa? En caso afirmativo, ¿puede dar un ejemplo?

E. Conclusión

- ¿Ha habido cambios importantes en el currículo, el personal de la escuela, o de apoyo externo (nuevos programas) este año? Si es así, ¿en qué han consistido estos cambios?
- ¿Hay algo más que le gustaría compartir con nosotros?

ESTUDIO CUALITATIVO DE IMPLEMENTACIÓN

GUÍA DE DISCUSIÓN PARA ENTREVISTAS CON PERSONAL CLAVE DE EDUCACIÓN

Hola, soy Chantal Toledo de Mathematica Policy Research. Estamos llevando a cabo un estudio sobre la implementación PRI. El estudio de investigación es financiado por USAID, la Agencia de Estados Unidos para el Desarrollo Internacional, y está siendo llevado a cabo por Mathematica Policy Research. Usted ha sido seleccionado(a) para una entrevista porque es un miembro clave del personal de EducAcción.

La entrevista incluirá preguntas acerca de su trabajo como personal clave de EducAcción. Esta información servirá para comprender mejor la implementación PRI. Toda la información que obtenemos se mantendrá estrictamente confidencial y esta información no se dará a conocer en ninguna forma que permita identificarle a usted o sus respuestas. La información que proporcione puede ser vista por USAID. Esta información se utilizará solamente para fines de evaluación, y toda información identificable como nombre o información de contacto será removida. Una vez que se complete el estudio, los datos del estudio que no le identifican personalmente pueden ser puestos a disposición del público para permitir análisis adicionales.

Esta entrevista llevará alrededor de una hora. Su participación es voluntaria y puede optar por no contestar alguna o todas las preguntas por cualquier motivo. En otras palabras, tiene la alternativa de no participar. No hay sanción o cambio en su situación laboral si no participa en este estudio. No hay riesgos ni beneficios directos para usted por participar en este estudio.

¿Está de acuerdo en participar en este estudio investigativo?

Debido a que no podemos anotar toda la conversación, nos gustaría grabar la entrevista. La grabación se usará únicamente para transcribir la conversación y toda información que permita identificarlo será removida.

¿Está de acuerdo en que grabemos la entrevista?

GUÍA TEMÁTICA

A. Funciones y responsabilidades

- Cuénteme un poco sobre usted.
 - ¿Cuánto tiempo ha estado trabajando con EducAcción?
 - ¿Cuál es su función y cuáles son sus responsabilidades en la implementación PRI?
- ¿Cómo/cuando se enteró del diseño y de los objetivos de la intervención PRI? ¿Qué tipo de apoyo recibió para entender los objetivos de la intervención?
- ¿Cómo está organizado el equipo de implementación PRI? (*Sondeo*: línea de comunicación, toma de decisiones, cambios en la organización durante el proyecto).
- ¿Qué retos de implementación PRI ha encontrado en su función? ¿Cómo los ha enfrentado?
- ¿Nuestra evaluación ha influenciado su trabajo y cómo ha implementado la intervención? Es decir, ¿ha implementado la intervención de forma diferente debido a la evaluación de Mathematica?

B. Fidelidad

- ¿Qué tan útil fue la formación de asesores de Mayo 2015? Cómo fue diseñada la formación? ¿Cuánto aporte del Ministerio de Educación tuvieron?
- ¿Cómo decidieron cuántas escuelas asignar a cada asesor en los grupos A y B? [NOTA: SABEMOS QUE NO ESTABAN TRATANDO DE ASIGNAR EL MISMO NÚMERO DE ESCUELAS PARA ASESORES DE LOS GRUPOS A Y B, ASÍ QUE ¿CUÁL FUE LA REGLA DE DECISIÓN?]
- ¿Cuáles han sido los principales retos en la capacitación y entrenamiento de maestros? ¿Cómo fueron enfrentados? (*SONDEO*: EXPLORAR DESAFÍOS ESPECÍFICOS POR GRUPO EXPERIMENTAL: GRUPO A Y GRUPO B)
- ¿Cómo está siendo implementando el componente prueba formativa (PF) de la evaluación? [*Sondeo*: ¿Cuáles son las principales actividades de los asesores en la implementación de este componente?]
 - Por lo que usted sabe, ¿qué desafíos han enfrentado los maestros de grupo A en la implementación de PF en el aula?
 - ¿Qué desafíos enfrentaron los maestros al usar los resultados de PF para ajustar su enseñanza?
 - ¿Pudieron los maestros usar los datos de PF para informar o ajustar la instrucción? ¿Cómo usaron los datos de PF para informar o ajustar la instrucción? ¿Tiene ejemplos?

- Para las escuelas en el grupo A, ¿considera que existen sinergias entre los componentes PF y PFG? Es decir, ¿cree que el componente PF fortalece el componente PFG o que el componente PFG podría hacer más fuerte al componente PF?
 - ¿Han habido aspectos del componente PF que no han sido implementados como esperaba?
- ¿Cómo se implementó el componente prueba de fin de grado (PFG) de la evaluación?
 - ¿Cuáles fueron los retos en la obtención o uso de datos PFG?
 - Por lo que usted sabe, ¿entendieron los directores y maestros los datos PFG?
 - ¿Cómo usaron los datos los directores?
- ¿Usaron los maestros los datos para fundamentar (informar o ajustar) su instrucción? ¿Tiene ejemplos concretos? ¿Hubo algún aspecto de la implementación del componente PFG que no se produjo de acuerdo al plan o directrices?
- ¿Tiene conocimiento de algún reto derivado de maestros que quieran ampliar su uso de PF en las escuelas del grupo B?

C. Percepciones del personal clave de EducAcción en la implementación e impacto del programa

- ¿Cómo recibieron las actividades generales del PRI los directores, maestros, padres de familia y los niños? ¿Cómo recibieron las actividades PF? ¿Cómo recibieron las actividades PFG?
- ¿Hubo algún reto surgido de la participación de niños en las actividades de implementación? ¿Cómo se abordaron esos retos?
- ¿Cree que los maestros han cambiado la forma en que utilizan su tiempo como resultado de la intervención? Si es así, ¿cómo? (*Sondeo*: están trabajando más horas o están organizando su tiempo de otra manera como consecuencia de la intervención?)
- ¿Cree usted que los directores han cambiado la forma en que utilizan su tiempo como resultado de la intervención? Si es así, ¿cómo? (*Sondeo*: están trabajando más horas o están organizando su tiempo de otra manera como consecuencia de la intervención?)
- ¿Cómo hace seguimiento y mide el logro de las metas del programa? (*Sondeo*: ¿Cómo se establecen puntos de referencia y seguimiento? ¿Cómo se ha ajustado para cumplir las metas de implementación?).

- ¿Cree que los estudiantes en el programa están realmente mejorando sus niveles de lectura y matemáticas como resultado de la intervención? ¿Ha beneficiado a niños el programa? En caso afirmativo, ¿puede dar un ejemplo?
- ¿Cómo describiría sus interacciones con las autoridades locales con respecto a la aplicación de la PF? ¿Con respecto a la aplicación de la PFG?
 - Si hubiera problemas: ¿cómo piensa que las pruebas PF y PFG podrían ser mejoradas para servir mejor a maestros y estudiantes?
- ¿Qué cree que se puede hacer para mejorar el uso de PF y PFG (si esto se hiciera a nivel nacional)?

D. Conclusión

- En general, ¿cree que la implementación de cualquiera de los componentes cambió desde que empezó la intervención? Si es así, ¿cómo han cambiado las cosas?
- ¿Anticipa modificar de alguna forma la implementación de los componentes PF o PFG el año próximo?
- ¿Hay algo más que le gustaría compartir con nosotros?

ESTUDIO CUALITATIVO DE IMPLEMENTACIÓN

GUÍA DE DISCUSIÓN PARA GRUPOS FOCALES CON DIRECTORES

Hola, soy César Valenzuela de Espirállica. Estamos llevando a cabo un estudio sobre las pruebas formativas (PF) y pruebas de fin de grado (PFG). El estudio de investigación es financiado por USAID, la Agencia de Estados Unidos para el Desarrollo Internacional, y está siendo llevado a cabo por Mathematica Policy Research. Ustedes han sido seleccionados para la discusión porque son directores.

La discusión incluirá preguntas acerca de su trabajo como director. Esta información servirá para comprender mejor el uso de pruebas formativas (PF) (pruebas mensuales) y de fin de grado (PFG) (pruebas anuales). Toda la información que obtenemos se mantendrá estrictamente confidencial y esta información no se dará a conocer en ninguna forma que permita identificarles a ustedes o sus respuestas. La información que proporcionen puede ser vista por USAID. Esta información se utilizará solamente para fines de evaluación, y se removerá toda información identificable como nombres o información de contacto. Una vez que se complete el estudio, los datos del estudio que no les identifican personalmente pueden estar a disposición del público para permitir análisis adicionales.

Esta discusión llevará alrededor de una hora y media. Su participación es voluntaria y pueden optar por no contestar alguna o todas las preguntas por cualquier motivo. En otras palabras, tienen la alternativa de no participar. No hay sanción o cambio en su situación laboral si no participan en este estudio. No hay riesgos ni beneficios directos para ustedes por participar en este estudio.

¿Están de acuerdo en participar en este estudio?

Debido a que no podemos anotar toda la conversación, nos gustaría grabar la discusión. La grabación se usará únicamente para transcribir la conversación y toda información que permita identificarlos será removida.

¿Están de acuerdo en que grabemos la discusión?

GUÍA TEMÁTICA

A. Perfil del director

Para empezar, por favor preséntese al grupo. Por favor díganos su nombre, el nombre de su escuela, el tiempo que ha sido director de su escuela, y cuántos estudiantes hay en su escuela y si también trabaja en otras escuelas.

B. Resultados de fin de grado (PFG)

Hablemos de las pruebas y los resultados de fin de grado.

- ¿Han visto los resultados de fin de grado para la prueba de 2014? ¿Qué tipo de información recibieron?
 - EN CASO **NEGATIVO**, NO HAN VISTO RESULTADOS (PFG) DE SU ESCUELA:
 - ¿Saben cómo es posible acceder a los resultados de fin de grado de su escuela? ¿Cómo?
 - ¿Sienten que sería útil para ustedes ver los resultados de fin de grado de su escuela?
 - ¿Cómo les gustaría recibir la información? Por maestro, por estudiante, por grado, a nivel de escuela?
 - **PASE A LA SECCIÓN PF**
 - EN CASO **AFIRMATIVO**, CONTINUAR:
 - ¿De quién recibieron los resultados?
 - ¿En qué momento del año recibieron los resultados?
 - ¿Cómo le gustaría recibir la información? a nivel de escuela, maestro, grado, sección o estudiante?
 - ¿Recibió información a nivel de escuela, maestro, grado, sección o estudiante?
- ¿Encontraron útiles los resultados de fin de grado? (*Sondeo: ¿Cómo fueron útiles? ¿Los utilizaron para su propia planificación? ¿Qué hicieron diferente en la forma en que planearon basado en los resultados de la PFG? ¿Los utilizaron sus maestros para su planificación? ¿Cómo cambió la planificación/instrucción de los maestros el uso de PFG? Si los resultados no son útiles, ¿por qué no?, ¿Qué les gustó más acerca de los resultados de fin de grado?*)
- ¿Creen que las pruebas de fin de grado proporcionan a los maestros información que ellos no podrían tener trabajando con sus alumnos? Si es así, ¿qué tipo de información?
- Que ustedes sepan, ¿han ustedes o alguien de su escuela participado en eventos de capacitación o actividades relacionadas con los resultados de fin de grado en general? De ser así, ¿fueron útiles? ¿Ha recibido su escuela apoyo de alguna organización por *usar* los

resultados de la prueba de fin de grado? [TRATE DE AVERIGUAR SI HAN RECIBIDO EL APOYO DE EducAcción PRI, PERO SEA CONSCIENTE DE QUE PUEDEN NO SABER EL NOMBRE DEL PROGRAMA. LUEGO TRATE DE SONDEAR SI HAN RECIBIDO OTRO TIPO DE APOYO DE OTROS GRUPOS.]

- ¿Creen que sus maestros enfrentan retos en el uso de los resultados de fin de grado para ajustar su enseñanza? Si es así, ¿qué tipo de retos?
- ¿Pueden pensar en alguna manera en que los resultados de fin de grado podrían ser más útiles para su escuela? ¿Necesitan resultados diferentes? ¿Los necesita antes? ¿Están presentados en una manera que es difícil de entender? ¿Qué les gustaría mejorar en los resultados de fin de grado?

C. Resultados de pruebas formativas (PF)

Hablemos de las pruebas formativas.

- ¿Ha recibido su escuela alguna copia de los materiales de las pruebas formativas este año? ¿De quién?
- ¿Ha recibido su escuela asistencia técnica o pedagógica para el uso de pruebas formativas este año? Por lo que saben, ¿alguien de su escuela ha participado en eventos o actividades relacionadas con pruebas formativas? Si es así, ¿fueron útiles?
- Por lo que saben, ¿usan sus maestros pruebas formativas? Si no usan, ¿por qué no? Si usan, ¿por qué?
- ¿Qué piensan de las pruebas formativas? ¿Son útiles? ¿Por qué o por qué no? (*Sondeo:* ¿En qué ayudan a la escuela / al maestro? ¿Cómo ayudan a los estudiantes las pruebas formativas?)
- ¿Qué les gusta más sobre las pruebas formativas?
- ¿Qué les gustaría mejorar en las pruebas formativas? ¿Pueden pensar en alguna forma en que las pruebas formativas podrían ser más útiles para sus maestros?

D. Apoyo externo

Hablemos de apoyo externo.

- ¿Utilizan sus escuelas libros, cuadernos, hojas de trabajo, o papel escrito proveniente de organizaciones, agencias gubernamentales, o gente afuera de la escuela? Si es así, ¿quién proporciona estos materiales? [*SONDEE POR NOMBRES DE ONGS, MOE / DDE / ALCALDE, PADRES*]

- ¿Proporciona la escuela sus propios materiales? ¿Qué hay de los padres, pagan ellos o proporcionan papel en blanco, cuadernos, materiales fotocopiados? ¿Es difícil para ustedes / los padres acceder a una fotocopidora o impresora?
- En la preparación de clases y pruebas, ¿recibe la escuela ayuda de organizaciones externas [además de EducAcción si están en los grupos A o B]? En caso afirmativo, *sondear*: ¿Quién? ¿Cuánto tiempo pasan en la escuela? ¿Proporcionan materiales escritos?
- [SI ESTÁN EN LOS GRUPOS A O B, HAGA LAS SIGUIENTES PREGUNTAS. DE LO CONTRARIO, VAYA A LA SECCIÓN E.] Cuando la escuela está recibiendo apoyo de EducAcción: ¿qué parte del día es? Si es durante el horario de clase, ¿hay alguien cubriendo el aula o qué están haciendo sus estudiantes? Si es fuera del horario de clase, ¿es durante horario regular de trabajo o terminan usted y los maestros quedándose más horas de lo normal en la escuela? ¿Cómo ha cambiado el apoyo la manera en que directores/maestros trabajan en la escuela?
- [PREGUNTA CLAVE] Pensando específicamente en EducAcción, ¿han recibido:
 - capacitación en PFG (grupos A y B),
 - capacitación en PF (grupo A),
 - capacitación en cómo interpretar resultados de PFG (grupos A y B),
 - capacitación en cómo interpretar resultados de PF (grupos A),
 - capacitación en cómo usar resultados de PFG para adaptar la enseñanza (grupos A y B),
 - capacitación en cómo usar resultados de PF (grupos A),
 - fotocopias de PF (grupo A)?

E. Otras pruebas

- En la escuela, ¿hacen otras cosas para evaluar el progreso de los estudiantes? (*Sondeo*: ¿Aplican los maestros la prueba diagnóstica al comienzo del año? ¿Desarrollan los maestros sus propias pruebas para evaluar el progreso? ¿Se basan los maestros en la interacción con los niños en el aula en lugar de en algo escrito?)

F. Conclusión

- En general, ¿cuáles consideran son sus mayores retos como director? (*Sondeo*: ¿Cuáles son los desafíos a la enseñanza y el aprendizaje del primero al tercer grado? Cuáles son los retos de la escuela en general?)
- En general, ¿qué consideran son los desafíos más grandes como maestro?
- En general, ¿qué consideran son los mayores retos para los estudiantes en las escuelas?

- [Para grupos A y B solamente:] En general, ¿han cambiado las cosas desde que comenzó la intervención? Si es así, ¿cómo han cambiado las cosas? (*Sondeo*: en términos de actividades, dotación de personal, plan de estudios)
- ¿Hay algo más que les gustaría compartir con nosotros?

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

ESTUDIO CUALITATIVO DE IMPLEMENTACIÓN

GUÍA DE DISCUSIÓN PARA GRUPOS FOCALES CON MAESTROS

Hola, soy César Valenzuela de Espirállica. Estamos llevando a cabo un estudio sobre pruebas formativas (PF) y pruebas de fin de grado (PFG). El estudio de investigación es financiado por USAID, la Agencia de Estados Unidos para el Desarrollo Internacional, y está siendo llevado a cabo por Mathematica Policy Research. Ustedes han sido seleccionados para la discusión porque son maestros.

La discusión incluirá preguntas acerca de su trabajo como maestro. Esta información servirá para comprender mejor el uso de pruebas formativas (PF) (pruebas mensuales) y de fin de grado (PFG) (pruebas anuales). Toda la información que obtenemos se mantendrá estrictamente confidencial y esta información no se dará a conocer en ninguna forma que permita identificarles a ustedes o sus respuestas. La información que proporcionen puede ser vista por USAID. Esta información se utilizará solamente para fines de evaluación, y se removerá toda información identificable como nombres o información de contacto. Una vez que se complete el estudio, los datos del estudio que no les identifican personalmente pueden estar a disposición del público para permitir análisis adicionales.

Esta discusión llevará alrededor de una hora y media. Su participación es voluntaria y pueden optar por no contestar alguna o todas las preguntas por cualquier motivo. En otras palabras, tienen la alternativa de no participar. No hay sanción o cambio en su situación laboral si no participan en este estudio. No hay riesgos ni beneficios directos para ustedes por participar en este estudio.

¿Están de acuerdo en participar en este estudio?

Debido a que no podemos anotar toda la conversación, nos gustaría grabar la discusión. La grabación se usará únicamente para transcribir la conversación y toda información que permita identificarlos será removida.

¿Están de acuerdo en que grabemos la discusión?

GUÍA TEMÁTICA

A. Perfil del maestro

Para empezar, por favor preséntese al grupo. Por favor díganos su nombre, el nombre de su escuela, el tiempo que ha sido maestro en total y en su escuela, y el o los grados y turnos que enseña.

B. Resultados de pruebas fin de grado (PFG)

Hablemos de las pruebas y los resultados de fin de grado.

- ¿Han recibido resultados de fin de grado de la prueba del año pasado para su escuela? ¿Para el grado(s) que ustedes enseñan y para sus alumnos?
 - EN CASO **NEGATIVO**, NO HAN VISTO RESULTADOS PFG DE SU ESCUELA:
 - ¿Saben cómo es posible acceder a los resultados de fin de grado de su escuela? ¿Cómo?
 - ¿Sienten que sería útil para ustedes ver los resultados de fin de grado de su clase? ¿Por qué o por qué no?
 - ¿Sienten que sería útil para ustedes ver los resultados de fin de grado de sus estudiantes? ¿Por qué o por qué no?
 - PASE A LA SECCIÓN PF
 - EN CASO **AFIRMATIVO**, HAN VISTO LOS RESULTADOS PFG DE SUS ESCUELAS:
 - ¿De quién recibieron los resultados? ¿En qué momento del año escolar recibieron la información? ¿Qué tipo de información había en esos informes? (*Sondeo*: ¿cómo se presentó la información: gráficamente, numéricamente, por estudiante, clase o escuela?)
 - ¿Aprendieron algo de la información que recibieron? En caso afirmativo, ¿qué aprendieron? ¿Fueron útiles los resultados de las pruebas de fin de grado? ¿Cómo o cómo no? ¿Qué les gustó más acerca de los resultados de fin de grado?
 - ¿Qué otra información les gustaría tener sobre los resultados de sus alumnos en la prueba de fin de grado o acerca de lo que sus estudiantes han aprendido? ¿Qué le gustaría mejorar en la presentación de resultados de fin de grado?
 - ¿Alguien les ayudó a comprender los informes recibidos? ¿Alguien les ayudó a entender qué hacer con la información?
 - ¿Participaron ustedes o alguien de su escuela en eventos o actividades relacionadas con los resultados de fin de grado (capacitaciones o reuniones)? De ser así, ¿fueron útiles?
 - ¿Alguna vez han utilizado los resultados de fin de grado para ayudarles a planear su instrucción? Por ejemplo, ¿se enteraron de áreas en las que sus estudiantes estaban luchando y decidieron dedicar más tiempo a esas áreas?
 - **En caso afirmativo**: ¿Cómo los han utilizado? ¿Pueden dar algunos ejemplos?

- **En caso afirmativo:** ¿Han encontrado algún desafío utilizando los informes de fin de grado para planificar sus clases? Si es así, ¿qué retos han encontrado?
- **En caso afirmativo:** ¿Había algo que fue especialmente útil para ustedes en el uso de los resultados de la prueba de fin de grado para planificar sus clases?
- ¿Les han ayudado los resultados de la prueba de fin de grado a identificar estudiantes específicos que podrían necesitar apoyo adicional?
 - **En caso afirmativo:** ¿Qué están haciendo con esos estudiantes?
 - **En caso negativo:** *Sondeo:* ¿Por qué no?

C. Resultados de pruebas formativas (PF)

Hablemos de las pruebas formativas.

- ¿Están familiarizados con las pruebas formativas? [EN CASO CONTRARIO MOSTRAR UN EJEMPLO. SI NO LAS CONOCEN, OMITA ESTA SECCIÓN.]
- ¿Utilizan las pruebas formativas mensuales en sus clases?
 - **En caso negativo:** ¿Por qué no las usan?
 - ¿Las han usado en el pasado?
 - De ser así, ¿fueron útiles? ¿Cómo?
 - Si no fueron útiles, ¿por qué no? ¿Recibieron los materiales o capacitación para usarlas?
- ¿Han recibido copias de las pruebas formativas este año?
 - i) De ser así, ¿de quién?
 - ii) ¿Cuántas copias?
 - iii) ¿Cómo se relaciona este número con la cantidad de estudiantes que enseñan?
 - iv) ¿Recibieron copias del manual de instrucciones?
- [PASE A LA SIGUIENTE SECCIÓN SI NO HAN UTILIZADO PF ESTE AÑO]
¿Cómo deciden cuándo administrar las pruebas formativas? [NOTA: ESTAMOS INTERESADOS EN SABER CÓMO DECIDEN LA SINCRONIZACIÓN CON

RESPECTO A CUÁNDO ESTÁN CUBRIENDO QUÉ MATERIALES CON SUS ALUMNOS. POR EJEMPLO, ¿ESTÁN ESPERANDO HASTA DESPUÉS DE HABER CUBIERTO EL MATERIAL EN LA PF ANTES DE SU ADMINISTRACIÓN, O SIMPLEMENTE SIGUIENDO EL MES CALENDARIO?]

- Cuéntenos cómo las administran. (Por ejemplo, ¿trabajan por su cuenta los estudiantes? ¿Tienen una cantidad fija de tiempo? ¿Escriben en sus propios cuadernos o en el papel que dan al maestro?)
- Que ustedes sepan, ¿han ustedes, o alguien de su escuela, participado en eventos de capacitación o actividades relacionadas con pruebas formativas? Si es así, ¿fueron útiles?
- ¿Han encontrado algún desafío utilizando las pruebas formativas para planificar sus clases? De ser así, ¿qué retos han encontrado? ¿Qué le gustaría mejorar en las pruebas formativas?
- ¿Hay elementos que fueron particularmente exitosos durante el uso de las pruebas formativas? De ser así, ¿cuáles fueron los éxitos? ¿Qué les gustó más acerca de las pruebas formativas?
- [PREGUNTA CLAVE]: ¿Están enfocando la instrucción de acuerdo a los resultados?
 - **En caso afirmativo:** ¿Qué están haciendo?
 - ¿Han cambiado su forma de enseñar a nivel estudiantil? ¿Han cambiado su forma de enseñar a nivel de clase?
 - ¿Pueden dar ejemplos?

D. Apoyo externo

Hablemos de apoyo externo.

- ¿Utilizan libros, cuadernos, hojas de trabajo, o papel escrito que vienen de afuera de la escuela? Si es así, ¿quién proporciona estos materiales? [SONDEE POR NOMBRES DE ONGS, MOE / DDE / ALCALDE, PADRES]
- ¿Proporcionan sus propios materiales? ¿Qué pasa con los padres, pagan o proporcionan papel en blanco, cuadernos, materiales fotocopiados? ¿Es difícil para ustedes / los padres acceder a una fotocopidora o impresora?
- En la preparación de clases y pruebas, ¿reciben ayuda de organizaciones externas [además de EducAcción si están en los grupos A o B]? En caso afirmativo, *sondear*: ¿Quién? ¿Cuánto tiempo pasan con ustedes? ¿Proporcionan materiales escritos?

[PREGUNTAS CLAVE] [SI ESTÁN EN LOS GRUPOS A O B, HAGA LAS SIGUIENTES PREGUNTAS. DE LO CONTRARIO, VAYA A LA SECCIÓN E.]

- Cuando están recibiendo apoyo de EducAcción, ¿qué parte de su día es? Si es durante el horario de clase, ¿hay alguien cubriendo su clase o qué están haciendo sus estudiantes? Si

es fuera del horario de clase, ¿es durante su horario de trabajo normal o terminan quedándose más horas de lo normal en la escuela?

- ¿Cuánto tiempo pasan trabajando en pruebas formativas o de fin de grado?
- Pensando específicamente en EducAcción, ¿han recibido:
 - capacitación en PFG (grupos A y B),
 - capacitación en PF (grupo A),
 - capacitación en cómo interpretar resultados de PFG (grupos A y B),
 - capacitación en cómo interpretar resultados de PF (grupo A),
 - capacitación en cómo usar resultados de PFG para adaptar la enseñanza (grupos A y B),
 - capacitación en cómo usar resultados de PF para adaptar la enseñanza (grupo A),
 - fotocopias de PF (grupo A)?

E. Otras pruebas

- ¿Hacen otras cosas para evaluar el progreso de los estudiantes? (*Sondeo*: ¿Aplican la prueba diagnóstica al comienzo del año? ¿Desarrollan sus propias pruebas para evaluar el progreso? ¿Se basan en la interacción con los niños en el aula en lugar de en algo escrito?)

F. Conclusión

- En general, ¿cuáles consideran son sus mayores retos como maestros?
- En general, ¿qué consideran son los desafíos más grandes para los estudiantes en las escuelas?
- [Para grupos A y B solamente:] En general, ¿han cambiado las cosas desde que comenzó la intervención? Si es así, ¿cómo han cambiado las cosas?
- ¿Hay algo más que les gustaría compartir con nosotros?

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

ESTUDIO CUALITATIVO DE IMPLEMENTACIÓN

GUÍA DE DISCUSIÓN PARA ENTREVISTAS CON ASESORES

Línea Final de Evaluación: Agosto 2016

Hola, soy César Valenzuela de Espirállica. Estamos llevando a cabo un estudio sobre PF (pruebas formativas) y PFG (pruebas de fin de grado). El estudio es financiado por USAID, la Agencia de Estados Unidos para el Desarrollo Internacional, y está siendo llevado a cabo por Mathematica Policy Research. Usted ha sido invitado a participar en una entrevista porque es un asesor de EducAcción.

La entrevista incluirá preguntas acerca de su trabajo como asesor y llevará alrededor de una hora. Esta información servirá para comprender mejor el uso de PF y de PFG. Toda la información que obtengamos se mantendrá en estricta privacidad y esta información no se dará a conocer en ninguna forma que permita identificarle a usted o sus respuestas. La información que proporcione estará a disposición de USAID. La información recogida durante la entrevista se utilizará solamente para fines de evaluación, y toda información que lo identifique a usted individualmente como nombres o información de contacto será removida. Una vez que se complete el estudio, los datos del estudio pueden ser puestos a disposición del público para permitir análisis adicionales, excluyendo la información que lo identifique personalmente.

Su participación en la entrevista es voluntaria y usted puede optar por no participar. Si decide participar, usted puede dejar de contestar cualquiera de las preguntas durante la entrevista si así lo prefiere. No hay sanción o cambio en su situación laboral si no participa en este estudio. No hay riesgos ni beneficios directos para usted por participar en este estudio.

¿Está de acuerdo en participar en este estudio?

Debido a que no podemos anotar toda la conversación, nos gustaría grabar la entrevista. La grabación se usará únicamente para transcribir la conversación y toda información que permita identificarlo será removida.

¿Está de acuerdo en que grabemos la entrevista?

GUÍA TEMÁTICA

A. Perfil del Asesor (para todos los asesores)

Cuénteme un poco sobre usted (su formación, experiencia, características de su trabajo).

- ¿Cuánto tiempo ha estado con el proyecto?
- ¿Qué es lo que entiende que es su trabajo?

- ¿Qué tan bien preparado se siente para hacerlo?
- ¿Recibió alguna capacitación este año? De ser así, ¿qué temas cubrió?
- Si es asesor nuevo(a): ¿Cómo ha sido su integración con el resto del equipo de trabajo?

- ¿En qué grupo de evaluación ha estado trabajando (grupo A o grupo B)?
- ¿Cuántas escuelas le fueron asignadas?
- Si es asesor(a) desde el año pasado: ¿hubo algún cambio en el número de escuelas que le fueron asignadas?
- ¿Qué tipo de trabajo hizo antes de empezar a trabajar con EducAcción PRI?
 - Si fue maestro: ¿Qué grados enseñó?
- Cuénteme cómo trabaja con directores y maestros en las escuelas que cubre este año.
Sondear para entender:
 - ¿Trabaja más con directores o maestros? ¿Qué discute con directores y maestros?
 - ¿Qué cambios está sugiriendo? ¿Cómo sugiere estos cambios?
 - ¿Trabaja con maestros en grupos o individualmente?
 - ¿Trabaja más con maestros de 1er a 3er grado (*primer ciclo*) o maestros de 4to a 6to grado?
 - ¿Con qué frecuencia visita? ¿Visita algunas escuelas con más frecuencia que otras? ¿Cuánto tiempo pasa en cada visita?
 - ¿Qué tipo de trabajo hace para prepararse para sus visitas?
 - ¿Cómo han reaccionado los directores a su apoyo? ¿Cómo han reaccionado los maestros a su apoyo?
- ¿Qué conceptos han parecido más útiles para los maestros/directores este año? ¿Tiene algún ejemplo?
- ¿Qué conceptos han sido más difíciles de transmitir a los maestros/directores este año? ¿Tiene algún ejemplo?
- ¿Utiliza la capacitación recibida de EducAcción de otras maneras además del apoyo a maestros en las escuelas a que fue asignado? [TRATE DE AVERIGUAR SI TRABAJA CON OTROS MAESTROS EN LAS MISMAS ESCUELAS, OTRAS ESCUELAS, POSIBLEMENTE DEL GRUPO C, O SI HAY BENEFICIOS POTENCIALMENTE NO MEDIDOS]
- ¿Percibe que el proyecto está teniendo un impacto? En caso afirmativo, ¿qué impacto está teniendo? En caso negativo, ¿por qué cree que no está teniendo ningún impacto?

B. Cambios este año (para todos los asesores)

- Si es asesor(a) desde el año pasado: ¿Han habido cambios en la forma en que trabaja con los directores y maestros en las escuelas con respecto al año pasado? ¿Qué cambios ha habido?
- ¿Cómo se desarrollaron las siguientes actividades y que tan útiles fueron (en su opinión):
 - ¿la reunión de cierre del 2015? (si es asesor(a) desde el año pasado)
 - ¿la evaluación anual participativa?
 - ¿la inclusión de una especialista en temas de género?
 - ¿la contratación de más asesores?
 - ¿la revisión de las PF y de sus manuales de instrucción?

C. Para Asesores trabajando en Grupo A (PFG + PF) solamente

[ESTA SECCIÓN ES PARA ASESORES TRABAJANDO EN GRUPO A (PFG + PF) SOLAMENTE.]

Pensando en este año solamente, hablemos de su trabajo en el grupo de intervención A.

- ¿A qué maestros ha podido dar copias de PF? (*Sondeo: ¿Qué grados? ¿A todos los maestros en esos grados?*)
- ¿Cuántas copias ha podido dar a cada maestro? ¿Es esto suficiente para todos sus estudiantes?
- ¿Le dió a todos los maestros una copia del manual de instrucción para los/el grado(s) que enseña?
- ¿Pudieron usar las PF actualizadas (corregidas)?

D. Para Asesores trabajando en Grupo B (PFG) solamente

[ESTA SECCIÓN ES PARA ASESORES TRABAJANDO EN GRUPO B (PFG) SOLAMENTE]

Pensando en este año solamente, hablemos de su trabajo en el grupo de intervención B.

- ¿Han sugerido los maestros usar pruebas formativas para mejorar el aprendizaje?
- ¿Ha utilizado alguna estrategia para disminuir la probabilidad de que los maestros usen pruebas formativas?
- ¿Ha enfrentado algún desafío de maestros queriendo ampliar el uso de pruebas formativas en sus escuelas (del grupo B)?

E. Para todos los asesores

[ESTA SECCIÓN ES PARA TODOS LOS ASESORES]

- ¿Ha notado maestros que reciben copias de PF de alguna otra fuente que no sea EducAcción PRI? ¿De quién? ¿Cuántos y con qué frecuencia?

- En caso positivo:
 - ¿Pudieron los maestros usar las copias?
 - ¿Su trabajo ha sido afectado por la distribución de PF? ¿Cómo si o cómo no?
- ¿Sabe de maestros que reciben apoyo técnico o pedagógico para PFG o PF de alguien que no sea EducAcción PRI (por ejemplo, MIDEH o alguna otra entidad)?
 - En caso positivo: ¿Cómo se compara este apoyo con el de EducAcción? (*sondeo*: ¿es parecido? ¿en qué se diferencian? ¿menos intenso? ¿más intenso?)
- En su opinión, ¿cuáles son las posibilidades de que otras escuelas (que no están en los grupos A o B) se beneficien indirectamente de la intervención? Por ejemplo, pueden escuelas fuera del programa aprender a través de interacciones con maestros de los grupos A o B a aumentar su uso de PFG o de PF, a tabular datos o a ajustar su instrucción utilizando los resultados de las pruebas?
- ¿Cree que los maestros han cambiado la forma en que utilizan su tiempo como resultado de la intervención este año? Si es así, ¿cómo? (*Sondeo*: están trabajando más horas o están organizando su tiempo de otra manera como consecuencia de la intervención?)
- ¿Cree usted que los directores han cambiado la forma en que utilizan su tiempo como resultado de la intervención este año? Si es así, ¿cómo? (*Sondeo*: están trabajando más horas o están organizando su tiempo de otra manera como consecuencia de la intervención?)
- Que ustedes sepan, ¿de qué otra(s) manera(s) supervisan las escuelas el aprendizaje de los estudiantes?
- En general, ¿Cómo describiría su experiencia como asesor?
- ¿Cuáles fueron los aspectos más difíciles en su tarea como asesor?
- ¿Qué cree que se puede hacer para mejorar el uso de PFG y PF?

F. Conclusión (para todos los asesores)

- En general, ¿cree que las cosas cambiaron para las escuelas este año (por ejemplo, con respecto a la actitud del docente o director, plan de estudios o actividades)? Si es así, ¿cómo han cambiado las cosas?
- ¿Qué aspectos fueron particularmente exitosos en la implementación de la intervención?
- En su opinión, ¿cuáles cree que son los mayores desafíos que enfrentan las escuelas donde trabaja?
- ¿Hay algo más que le gustaría compartir con nosotros?

ESTUDIO CUALITATIVO DE IMPLEMENTACIÓN

GUÍA DE DISCUSIÓN PARA ENTREVISTAS CON ASESORES PRINCIPALES

Línea Final de Evaluación: Agosto 2016

Hola, soy Chantal Toledo de Mathematica Policy Research. Estamos llevando a cabo un estudio sobre la implementación PRI. El estudio es financiado por USAID, la Agencia de Estados Unidos para el Desarrollo Internacional, y está siendo llevado a cabo por Mathematica Policy Research. Usted ha sido invitado a participar en una entrevista por su trabajo como asesor principal en EducAcción.

La entrevista incluirá preguntas acerca de su trabajo con EducAcción y llevará alrededor de una hora. Esta información servirá para comprender mejor la implementación PRI. Toda la información que obtengamos se mantendrá en estricta privacidad y esta información no se dará a conocer en ninguna forma que permita identificarle a usted o sus respuestas. La información que proporcione estará a disposición de USAID. La información recogida durante la entrevista se utilizará solamente para fines de evaluación, y toda información que lo identifique a usted individualmente como nombre o información de contacto será removida. Una vez que se complete el estudio, los datos del estudio pueden ser puestos a disposición del público para permitir análisis adicionales, excluyendo la información que lo identifique personalmente.

Su participación en la entrevista es voluntaria y usted puede optar por no participar. Si decide participar, usted puede dejar de contestar cualquiera de las preguntas durante la entrevista si así lo prefiere. No hay sanción o cambio en su situación laboral si no participa en este estudio. No hay riesgos ni beneficios directos para usted por participar en este estudio.

¿Está de acuerdo en participar en este estudio ?

Debido a que no podemos anotar toda la conversación, nos gustaría grabar la entrevista. La grabación se usará únicamente para transcribir la conversación y toda información que permita identificarlo será removida.

¿Está de acuerdo en que grabemos la entrevista?

GUÍA TEMÁTICA

A. Funciones y responsabilidades

Cuénteme un poco sobre usted (su formación, experiencia, características de su puesto) y los cambios en su trabajo con EducAcción PRI con respecto al año pasado.

- ¿Ha cambiado en algo su función y sus responsabilidades en la implementación PRI desde el año pasado?
- ¿Qué retos ha encontrado en la implementación del PRI este año? ¿Cómo los ha enfrentado desde su función?

B. Fidelidad

- ¿Cómo se desarrollaron las siguientes actividades y que tan útiles fueron (en su opinión)?:
 - ¿la reunión de cierre del 2015?
 - ¿la evaluación anual participativa?
 - ¿la inclusión de una especialista en temas de género?
 - ¿la contratación de más asesores?
 - ¿la revisión de las pruebas formativas (PF) y de sus manuales de instrucción? ¿qué revisiones concretas realizaron? ¿fueron útiles? ¿pudieron los maestros utilizar las nuevas versiones?
- Hablemos sobre las pruebas de fin de grado (PFG). ¿Cómo se implementó el componente PFG de la evaluación este año? [*Sondeo*: ¿Cuáles son las principales actividades de los asesores en la implementación de este componente?]
 - ¿Qué tipo de retos enfrentó este año en la obtención o uso de datos PFG?
 - ¿Cuáles han sido los principales retos en la capacitación y entrenamiento de maestros con las PFG este año? ¿Cómo fueron enfrentados?
 - Por lo que usted sabe, ¿en qué medida entendieron los directores y maestros los datos PFG este año?
 - ¿Cómo usaron los datos los directores este año?
 - ¿En qué medida usaron los maestros los datos para fundamentar (informar o ajustar) su instrucción este año? ¿Me puede dar ejemplos concretos de cómo los maestros usaron los datos?
 - ¿Hubo algún aspecto de la implementación del componente PFG que no se dió de acuerdo al plan o directrices este año? ¿De qué manera se desvió de los planes originales?

- Hablemos sobre la PF. ¿cómo está siendo implementando el componente PF de la evaluación este año? [*Sondeo*: ¿Cuáles son las principales actividades de los asesores en la implementación de este componente?]
 - ¿Pudieron los maestros usar los datos de PF para informar o ajustar la instrucción este año? ¿Cómo usaron los datos de PF para informar o ajustar la instrucción? ¿Me puede dar algunos ejemplos de cómo se usaron los datos de PF?
 - Por lo que usted sabe, ¿qué desafíos han enfrentado los maestros del grupo A en la implementación de PF en el aula este año? *Sondeo*: ¿Qué tipo de desafíos han enfrentado al usar los resultados de PF para ajustar su enseñanza?
 - Para las escuelas del grupo A y pensando en este año, ¿considera que existen sinergias entre los componentes PF y PFG? Es decir, ¿cree que el componente PF afecta al componente PFG o que el componente PFG afecta al componente PF?
 - ¿Cuáles han sido los principales retos en la capacitación y entrenamiento de maestros con las PF este año? ¿Cómo fueron enfrentados?
 - ¿Han habido aspectos del componente PF que no han sido implementados como se esperaba este año? ¿Cuáles componentes no se implementaron como se esperaba y por qué no fue posible implementarlos como se esperaba?
 - ¿Qué diferencias han habido con respecto a la capacitación o uso de PF entre el año pasado y este año? ¿Cómo describiría estas diferencias (positivas, negativas, iguales)?
- ¿Hubo alguna diferencia en la implementación en las escuelas debido a su tamaño, ubicación geográfica o cualquier otra característica de la escuela relevante este año?
- ¿Se dieron problemas de seguridad en las escuelas que posiblemente afectaran la implementación este año?
- ¿Se dieron diferencias en la carga de trabajo de los asesores en los grupos A y B este año? ¿Se dieron diferencias en la intensidad de la ayuda provista a las escuelas en los grupos A y B este año? En qué medida las diferencias en la carga de trabajo de los asesores en los grupos A y B puede haber afectado la *intensidad* del componente PFG de la intervención este año?

C. Percepciones del asesor principal sobre la influencia de organizaciones externas y autoridades locales

- ¿En su opinión, afecta a la evaluación la distribución de PF por otras organizaciones en Lempira, Santa Bárbara, Tegucigalpa y La Ceiba? En caso positivo: ¿Cuánto afecta a la evaluación la distribución de PF por otras organizaciones en Lempira, Santa Bárbara, Tegucigalpa y La Ceiba?

- ¿Tiene conocimiento de algún reto derivado de maestros que quieran ampliar el uso de PF en las escuelas del grupo B?
- ¿Ha tenido interacciones con autoridades locales (departamentales y/o distritales)? ¿Cómo describiría sus interacciones con las autoridades locales con respecto a la aplicación de la PFG? ¿con respecto a la aplicación de la PF?
- ¿De qué manera se podrían mejorar las pruebas PFG y PF a favor de los maestros y estudiantes? ¿Qué cree que se puede hacer para mejorar el uso de PFG y PF (si esto se hiciera a nivel nacional)?
- Si pudiera escoger solamente una prueba, ¿preferiría tener PFG o PF? ¿Por qué?

D. Percepciones del asesor principal sobre la aceptación y el impacto del programa

- ¿Cómo recibieron las actividades de PFG los directores, maestros, padres de familia y los niños este año? ¿Cómo recibieron las actividades de PF los directores, maestros, padres de familia y los niños este año?
- ¿Cree que los maestros han cambiado la forma en que planifican su instrucción como resultado de la intervención este año? Si es así, ¿cómo?
- ¿Cree usted que los directores han cambiado la forma en que utilizan su tiempo como resultado de la intervención este año (por ejemplo, han aumentado su tiempo de trabajo o lo han reorganizado)? Si es así, ¿cómo?
- ¿Cree que los estudiantes en el programa están realmente mejorando sus niveles de lectura y matemáticas como resultado de la intervención este año? ¿Cómo ha impactado el programa el aprendizaje de los niños?
- ¿En su opinión, las escuelas del grupo C se podrían beneficiar indirectamente de la intervención en PFG o PF que se está haciendo en las escuelas de los grupos A y B? En caso positivo: ¿De qué manera?
- ¿Hay algo más que quisiera agregar acerca de la experiencia de los asesores con el programa?

E. Conclusión

- ¿Hay algún otro componente de la implementación que cambió desde el año pasado? Si es así, ¿Qué cambió?
- ¿Qué aspectos fueron particularmente exitosos en la implementación de la intervención?

¿HAY ALGO MÁS QUE LE GUSTARÍA COMPARTIR CON NOSOTROS?

ESTUDIO CUALITATIVO DE IMPLEMENTACIÓN

GUÍA DE DISCUSIÓN PARA ENTREVISTAS CON PERSONAL CLAVE DE EDUCACIÓN

Línea Final de Evaluación: Agosto 2016

Hola, soy Chantal Toledo de Mathematica Policy Research. Estamos llevando a cabo un estudio sobre la implementación PRI. El estudio de investigación es financiado por USAID, la Agencia de Estados Unidos para el Desarrollo Internacional, y está siendo llevado a cabo por Mathematica Policy Research. Usted ha sido invitado a participar en una entrevista porque es un miembro clave del personal de EducAcción.

La entrevista incluirá preguntas acerca de su trabajo como personal clave de EducAcción y llevará alrededor de una hora y media. Esta información servirá para comprender mejor la implementación PRI. Toda la información que obtengamos se mantendrá en estricta privacidad y esta información no se dará a conocer en ninguna forma que permita identificarle a usted o sus respuestas. La información que proporcione estará a disposición de USAID. La información recogida durante la entrevista se utilizará solamente para fines de evaluación, y toda información que lo identifique a usted individualmente como nombre o información de contacto será removida. Una vez que se complete el estudio, los datos del estudio pueden ser puestos a disposición del público para permitir análisis adicionales, excluyendo la información que lo identifique personalmente.

Su participación en la entrevista es voluntaria y usted puede optar por no participar. Si decide participar, usted puede dejar de contestar cualquiera de las preguntas durante la entrevista si así lo prefiere. No hay sanción o cambio en su situación laboral si no participa en este estudio. No hay riesgos ni beneficios directos para usted por participar en este estudio.

¿Está de acuerdo en participar en este estudio?

Debido a que no podemos anotar toda la conversación, nos gustaría grabar la entrevista. La grabación se usará únicamente para transcribir la conversación y toda información que permita identificarlo será removida.

¿Está de acuerdo en que grabemos la entrevista?

GUÍA TEMÁTICA

A. Funciones y responsabilidades

Cuénteme un poco sobre usted (su formación, experiencia, características de su puesto) y los cambios en su trabajo con EducAcción PRI con respecto al año pasado.

- ¿Ha cambiado en algo su función y sus responsabilidades en la implementación PRI desde el año pasado?
- ¿Qué retos ha encontrado en la implementación del PRI este año? ¿Cómo los ha enfrentado desde su función?

B. Fidelidad

- ¿Cómo se desarrollaron las siguientes actividades y qué tan útiles fueron (en su opinión)?:
 - ¿la reunión de cierre del 2015?
 - ¿la evaluación anual participativa?
 - ¿la inclusión de una especialista en temas de género?
 - ¿la contratación de más asesores?
 - ¿la revisión de las pruebas formativas (PF) y de sus manuales de instrucción?
- ¿Cuáles fueron los principales hallazgos del grupo de monitoreo? ¿Se pudo usar la información obtenida? En caso positivo: ¿Cómo se usó? En caso negativo: ¿Por qué no?
- Hablemos sobre las pruebas de fin de grado (PFG). ¿Cómo se implementó el componente PFG de la evaluación este año? [*Sondeo*: ¿Cuáles son las principales actividades de los asesores en la implementación de este componente?]
 - ¿Ha notado algún cambio en la facilidad o dificultad con la que obtuvo datos de PFG este año, comparado al año pasado? Por ejemplo, ¿ha notado algún cambio en la actitud de directores y/o docentes?
 - Por lo que usted sabe, ¿en qué medida entendieron los directores y maestros los informes con los resultados de las PFG este año?
 - ¿Cómo usaron los datos los directores este año?
 - ¿En qué medida usaron los maestros los datos para fundamentar (informar o ajustar) su instrucción este año? ¿Me puede dar algunos ejemplos concretos de cómo los maestros usaron los datos? ¿Hubo algún aspecto de la implementación del componente PFG que no se produjo de acuerdo al plan o directrices este año?
 - ¿Cuáles han sido los principales retos en la capacitación y entrenamiento de maestros con las PFG este año? ¿Cómo fueron enfrentados?

- Hablemos sobre la PF. ¿Cómo está siendo implementando el componente PF de la evaluación este año? [*Sondeo*: ¿Cuáles son las principales actividades de los asesores en la implementación de este componente?]
 - ¿Pudieron los maestros usar los datos de PF para informar o ajustar la instrucción este año? ¿Cómo usaron los datos de PF para informar o ajustar la instrucción? ¿Me puede dar algunos ejemplos de cómo se usaron los datos de PF?
 - Por lo que usted sabe, ¿qué desafíos han enfrentado los maestros del grupo A en la implementación de PF en el aula este año? *Sondeo*: ¿Qué tipo de desafíos han enfrentado al usar los resultados de PF para ajustar su enseñanza?
 - Para las escuelas del grupo A y pensando en este año, ¿considera que existen sinergias entre los componentes PF y PFG? Es decir, ¿cree que el componente PF afecta al componente PFG o que el componente PFG afecta al componente PF?
 - ¿Cuáles han sido los principales retos en la capacitación y entrenamiento de maestros con las PF este año? ¿Cómo fueron enfrentados?
 - ¿Han habido aspectos del componente PF que no han sido implementados como esperaba este año? ¿Cuáles componentes no se implementaron como se esperaba y por qué no fue posible implementarlos como se esperaba?
 - ¿Qué diferencias han habido con respecto a la capacitación o uso de PF entre el año pasado y este año? ¿Cómo describiría estas diferencias (positivas, negativas, iguales)?

C. Percepciones del personal clave de EducAcción sobre la influencia de organizaciones externas y autoridades locales

- ¿En su opinión, afecta a la evaluación la distribución de PF por otras organizaciones en Lempira, Santa Bárbara, Tegucigalpa y La Ceiba? En caso positivo: ¿Cuánto afecta a la evaluación la distribución de PF por otras organizaciones en Lempira, Santa Bárbara, Tegucigalpa y La Ceiba?
- ¿Tiene conocimiento de algún reto derivado de maestros que quieran ampliar el uso de PF en las escuelas del grupo B?
- ¿Cómo describiría sus interacciones con las autoridades locales (departamentales y distritales) con respecto a la aplicación de la PFG? ¿Con respecto a la aplicación de la PF?
- ¿De qué manera se podrían mejorar las pruebas PFG y PF a favor de los maestros y estudiantes? ¿Qué cree que se puede hacer para mejorar el uso de PFG y PF (si esto se hiciera a nivel nacional)?
- Si pudiera escoger solamente una prueba, ¿preferiría tener PFG o PF? ¿Por qué?

D. Percepciones del personal clave de EducAcción sobre la aceptación y el impacto del programa

- ¿Cómo recibieron las actividades de PFG los directores, maestros, padres de familia y los niños este año? ¿Cómo recibieron las actividades de PF los directores, maestros, padres de familia y los niños este año?
- ¿Cree que los maestros han cambiado la forma en que planifican su instrucción como resultado de la intervención este año? Si es así, ¿cómo?
- ¿Cree usted que los directores han cambiado la forma en que utilizan su tiempo como resultado de la intervención este año (por ejemplo, han aumentado su tiempo de trabajo o lo han reorganizado)? Si es así, ¿cómo?
- ¿Cree que los estudiantes en el programa están realmente mejorando sus niveles de lectura y matemáticas como resultado de la intervención este año? ¿Cómo ha impactado el programa el aprendizaje de los niños?
- ¿En su opinión, las escuelas del grupo C se podrían beneficiar indirectamente de la intervención en PFG o PF que se está haciendo en las escuelas de los grupos A y B? En caso positivo: ¿De qué manera?
- ¿Hay algo más que quisiera agregar acerca de la experiencia de los asesores con el programa?

E. Conclusión

- ¿Hay algún otro componente de la implementación que cambió desde el año pasado? Si es así, ¿Qué cambió?
- ¿Qué aspectos fueron particularmente exitosos en la implementación de la intervención?
- ¿Hay algo más que le gustaría compartir con nosotros?

ESTUDIO CUALITATIVO DE IMPLEMENTACIÓN

GUÍA DE DISCUSIÓN PARA GRUPOS FOCALES CON ASESORES

Línea Final de Evaluación: Agosto 2016

Hola, soy César Valenzuela de Espirállica. Estamos llevando a cabo un estudio sobre PF (pruebas formativas) y PFG (pruebas de fin de grado). El estudio es financiado por USAID, la Agencia de Estados Unidos para el Desarrollo Internacional, y está siendo llevado a cabo por Mathematica Policy Research. Ustedes han sido invitados a participar en la discusión porque son asesores(as) de EducAcción.

La discusión incluirá preguntas acerca de su trabajo como asesores(as) y llevará alrededor de una hora y media. Esta información servirá para comprender mejor el uso de PF y de PFG. Toda la información que obtengamos se mantendrá en estricta privacidad y esta información no se dará a conocer en ninguna forma que permita identificarle a usted o sus respuestas. La información que proporcione estará a disposición de USAID. La información recogida durante la discusión se utilizará solamente para fines de evaluación, y toda información que los identifiquen a ustedes individualmente como nombres o información de contacto será removida. Una vez que se complete el estudio, los datos del estudio pueden ser puestos a disposición del público para permitir análisis adicionales, excluyendo la información que los identifiquen personalmente.

Su participación en la discusión es voluntaria y pueden optar por no participar. Si deciden participar, pueden dejar de contestar cualquiera de las preguntas durante la discusión si así lo prefieren. No hay sanción o cambio en su situación laboral si no participan en este estudio. No hay riesgos ni beneficios directos para ustedes por participar en este estudio.

¿Están de acuerdo en participar en este estudio?

Debido a que no podemos anotar toda la conversación, nos gustaría grabar la discusión. La grabación se usará únicamente para transcribir la discusión y toda información que permita identificarlos será removida.

¿Están de acuerdo en que grabemos la discusión?

GUÍA TEMÁTICA

A. Perfil del Asesor (para todos los asesores)

Para empezar, por favor preséntese al grupo. Por favor díganos su nombre, cuanto tiempo han estado con el proyecto, en que grupo y zona geográfica trabaja, cuantas escuelas le fueron asignadas, y cuál ha sido su formación y experiencia previa al proyecto (si fue maestro, ¿Qué grados enseñó?).

- ¿Qué es lo que entienden que es su trabajo?

- ¿Qué tan bien preparados se sienten para hacerlo?
- ¿Recibieron alguna capacitación este año? De ser así, ¿qué temas cubrió?
- Si es asesor nuevo(a): ¿Cómo ha sido su integración con el resto del equipo de trabajo?
- Si es asesor(a) desde el año pasado: ¿hubo algún cambio en el número de escuelas que le fueron asignadas?
- Cuéntenme cómo trabajan con directores y maestros en las escuelas que cubren este año.
Sondear para entender:
 - ¿Trabajan más con directores o maestros? ¿Qué discuten con directores y maestros?
 - ¿Qué cambios están sugiriendo? ¿Cómo sugieren estos cambios?
 - ¿Trabajan con maestros en grupos o individualmente?
 - ¿Trabajan más con maestros de 1er a 3er grado (*primer ciclo*) o maestros de 4to a 6to grado?
 - ¿Con qué frecuencia visitan? ¿Visitan algunas escuelas con más frecuencia que otras? ¿Cuánto tiempo pasan en cada visita?
 - ¿Qué tipo de trabajo hacen para prepararse para sus visitas?
 - ¿Cómo han reaccionado los directores a su apoyo? ¿Cómo han reaccionado los maestros a su apoyo?
- ¿Qué conceptos han parecido más útiles para los maestros/directores este año? ¿Tienen algún ejemplo?
- ¿Qué conceptos han sido más difíciles de transmitir a los maestros/directores este año? ¿Tienen algún ejemplo?
- ¿Utilizan la capacitación recibida de EducAcción de otras maneras además del apoyo a maestros en las escuelas a que fue asignado? [TRATE DE AVERIGUAR SI TRABAJA CON OTROS MAESTROS EN LAS MISMAS ESCUELAS, OTRAS ESCUELAS, POSIBLEMENTE DEL GRUPO C, O SI HAY BENEFICIOS POTENCIALMENTE NO MEDIDOS]
- ¿Perciben que el proyecto está teniendo un impacto? En caso afirmativo, ¿qué impacto está teniendo? En caso negativo, ¿por qué creen que no está teniendo ningún impacto?

B. Cambios este año (para todos los asesores)

- Si es asesor(a) desde el año pasado: ¿Han habido cambios en la forma en que trabajan con los directores y maestros en las escuelas con respecto al año pasado? ¿Qué cambios ha habido?
- ¿Cómo se desarrollaron las siguientes actividades y que tan útiles fueron (en su opinión):

- ¿la reunión de cierre del 2015? (si es asesor(a) desde el año pasado)
- ¿la evaluación anual participativa?
- ¿la inclusión de una especialista en temas de género?
- ¿la contratación de más asesores?
- ¿la revisión de las PF y de sus manuales de instrucción?

C. Para Asesores trabajando en Grupo A (PFG + PF) solamente

[ESTA SECCIÓN ES PARA ASESORES TRABAJANDO EN GRUPO A (PFG + PF) SOLAMENTE.]

Pensando en este año solamente, hablemos de su trabajo en el grupo de intervención A.

- ¿A qué maestros han podido dar copias de PF? (*Sondeo*: ¿Qué grados? ¿A todos los maestros en esos grados?)
- ¿Cuántas copias han podido dar a cada maestro? ¿Es esto suficiente para todos sus estudiantes?
- ¿Le dieron a todos los maestros una copia del manual de instrucción para los/el grado(s) que enseña?
- ¿Pudieron usar las PF actualizadas (corregidas)?

D. Para Asesores trabajando en Grupo B (PFG) solamente

[ESTA SECCIÓN ES PARA ASESORES TRABAJANDO EN GRUPO B (PFG) SOLAMENTE]

Pensando en este año solamente, hablemos de su trabajo en el grupo de intervención B.

- ¿Han sugerido los maestros usar pruebas formativas para mejorar el aprendizaje?
- ¿Han utilizado alguna estrategia para disminuir la probabilidad de que los maestros usen pruebas formativas?
- ¿Han enfrentado algún desafío de maestros queriendo ampliar el uso de pruebas formativas en sus escuelas (del grupo B)?

E. Para todos los Asesores

[ESTA SECCIÓN ES PARA TODOS LOS ASESORES]

- ¿Han notado maestros que reciben copias de PF de alguna otra fuente que no sea EducAcción PRI? ¿De quién? ¿Cuántos y con qué frecuencia?
 - En caso positivo:
 - ¿Qué hicieron con las copias que recibieron?
 - ¿Su trabajo ha sido afectado por la distribución de PF? ¿Cómo si o cómo no?

- ¿Sabían de maestros que reciben apoyo técnico o pedagógico para PFG o PF de alguien que no sea EducAcción PRI (por ejemplo, MIDEH o alguna otra entidad)?
 - En caso positivo: ¿Cómo se compara este apoyo con el de EducAcción? (*sondeo*: ¿es parecido? ¿en qué se diferencian? ¿menos intenso? ¿más intenso?)
- En su opinión, ¿cuáles son las posibilidades de que otras escuelas (que no están en los grupos A o B) se beneficien indirectamente de la intervención? Por ejemplo, pueden escuelas fuera del programa aprender a través de interacciones con maestros de los grupos A o B a aumentar su uso de PFG o de PF, a tabular datos o a ajustar su instrucción utilizando los resultados de las pruebas?
- ¿Creen que los maestros han cambiado la forma en que utilizan su tiempo como resultado de la intervención este año? Si es así, ¿cómo? (*Sondeo*: están trabajando más horas o están organizando su tiempo de otra manera como consecuencia de la intervención?)
- ¿Creen que los directores han cambiado la forma en que utilizan su tiempo como resultado de la intervención este año? Si es así, ¿cómo? (*Sondeo*: están trabajando más horas o están organizando su tiempo de otra manera como consecuencia de la intervención?)
- Que ustedes sepan, ¿de qué otra(s) manera(s) supervisan las escuelas el aprendizaje de los estudiantes?
- En general, ¿Cómo describirían su experiencia como asesores(as)?
- ¿Cuáles fueron los aspectos más difíciles en su tarea como asesores?
- ¿Qué creen que se puede hacer para mejorar el uso de PFG y PF?

F. Conclusión (para todos los asesores)

- En general, ¿creen que las cosas cambiaron para las escuelas este año (por ejemplo, con respecto a la actitud del docente o director, plan de estudios o actividades)? Si es así, ¿cómo han cambiado las cosas?
- ¿Qué aspectos fueron particularmente exitosos en la implementación de la intervención?
- En su opinión, ¿cuáles creen que son los mayores desafíos que enfrentan las escuelas donde trabaja?
- ¿Hay algo más que les gustaría compartir con nosotros?

ESTUDIO CUALITATIVO DE IMPLEMENTACIÓN

GUÍA DE DISCUSIÓN PARA GRUPOS FOCALES CON DIRECTORES

Línea Final de Evaluación: Agosto 2016

Hola, soy César Valenzuela de Espirállica. Estamos llevando a cabo un estudio sobre las PF (pruebas formativas) y PFG (pruebas de fin de grado). El estudio es financiado por USAID, la Agencia de Estados Unidos para el Desarrollo Internacional, y está siendo llevado a cabo por Mathematica Policy Research. Ustedes han sido invitados a participar en la discusión porque son directores.

La discusión incluirá preguntas acerca de su trabajo como director y llevará alrededor de una hora y media. Esta información servirá para comprender mejor el uso de PF y de PFG. Toda la información que obtengamos se mantendrá en estricta privacidad y esta información no se dará a conocer en ninguna forma que permita identificarles a ustedes o sus respuestas. La información que proporcionen estará a disposición de USAID. La información recogida durante la discusión se utilizará solamente para fines de evaluación, y toda la información que los identifiquen a ustedes individualmente como nombres o información de contacto será removida. Una vez que se complete el estudio, los datos del estudio pueden estar a disposición del público para permitir análisis adicionales, excluyendo la información que los identifiquen personalmente.

Su participación en la discusión es voluntaria y pueden optar por no participar. Si deciden participar, pueden dejar de contestar cualquiera de las preguntas durante la conversación por cualquier motivo si así lo prefieren. No hay sanción o cambio en su situación laboral si no participan en este estudio. No hay riesgos ni beneficios directos para ustedes por participar en este estudio.

¿Están de acuerdo en participar en este estudio?

Debido a que no podemos anotar toda la conversación, nos gustaría grabar la discusión. La grabación se usará únicamente para transcribir la conversación y toda información que permita identificarlos será removida.

¿Están de acuerdo en que grabemos la discusión?

GUÍA TEMÁTICA

A. Perfil del director

Para empezar, por favor preséntese al grupo. Por favor díganos su nombre, el nombre de su escuela, el tiempo de experiencia que lleva como director, el tiempo que lleva en su escuela actual, cuántos estudiantes hay en su escuela, y si también trabaja en otras escuelas.

B. Resultados de fin de grado (PFG) (para todos los grupos, A, B y C)

Hablemos de las pruebas y los resultados de fin de grado del 2015.

- ¿Han visto los resultados de fin de grado para la prueba de 2015?
 - EN CASO **NEGATIVO**, NO HAN VISTO RESULTADOS PFG DE 2015 DE SU ESCUELA:
 - ¿Saben cómo es posible acceder a los resultados de fin de grado de su escuela? ¿Cómo?
 - ¿Piensan que sería útil para ustedes ver los resultados de fin de grado de su escuela? ¿Por qué o por qué no?
 - ¿Cómo les gustaría recibir la información? ¿A nivel de escuela, grado, sección, maestro, o estudiante?
 - ¿Creen que las PFG proporcionan a los maestros información única que no se podría obtener trabajando directamente con sus alumnos? Si es así, ¿qué tipo de información única proporcionan las PFG?
 - [PREGUNTA CLAVE] Que ustedes sepan:

[TRATE DE AVERIGUAR SI HAN RECIBIDO EL APOYO DE EDUCACIÓN PRI Y/O MIDEH O SUS SOCIOS, PERO SEPA QUE PUEDEN NO SABER EL NOMBRE DEL PROGRAMA. LUEGO TRATE DE SONDEAR SI HAN RECIBIDO OTRO TIPO DE APOYO DE OTROS GRUPOS.]
 - ❖ ¿Han ustedes o alguien de su escuela participado en eventos de capacitación o actividades relacionadas con los resultados de fin de grado en general este año? De ser así, ¿Quién proporcionó la capacitación o evento? ¿Qué aspectos de la capacitación fueron útiles?
 - ❖ ¿Ha recibido su escuela apoyo de alguna organización para *usar* los resultados de la prueba de fin de grado este año? De ser así, ¿Quién proporcionó el apoyo? ¿Qué aspectos de ese apoyo les han servido?
 - ❖ En caso mencionen MIDEH, EROC, CADERH, CIDEH, Fundaempresa Unitec, o Asociación CESAL, y para los grupos A y B solamente: ¿Cómo se comparan la capacitación y/o actividades relacionadas con las PFG con la capacitación de EducAcción? ¿son parecidas? ¿En qué se diferencian? ¿más intensas? ¿menos intensas?

▪ **PASE A LA SECCIÓN PF (SECCION C)**

○ EN CASO **AFIRMATIVO**, HAN VISTO LOS RESULTADOS PFG DEL 2015 DE SU ESCUELA:

- ¿Qué tipo de información recibieron?
- ¿Recibieron información a nivel de escuela, grado, sección o estudiante? ¿A qué nivel les gustaría recibir la información?
- ¿De quién recibieron los resultados? ¿En qué momento del año escolar recibieron la información? ¿Qué tipo de información había en esos informes? (*Sondeo*: ¿cómo se presentó la información: gráficamente, numéricamente, por estudiante, clase o escuela?)
- ¿Aprendieron algo de la información que recibieron? En caso afirmativo, ¿qué aprendieron? ¿Fueron útiles los resultados de las PFG? ¿Cómo o cómo no? ¿Qué les gustó más acerca de los resultados de fin de grado?
- [PREGUNTA CLAVE] Que ustedes sepan:

[TRATE DE AVERIGUAR SI HAN RECIBIDO EL APOYO DE EDUCACIÓN PRI Y/O MIDEH O SUS SOCIOS, PERO SEPA QUE PUEDEN NO SABER EL NOMBRE DEL PROGRAMA. LUEGO TRATE DE SONDEAR SI HAN RECIBIDO OTRO TIPO DE APOYO DE OTROS GRUPOS.]

- ❖ ¿Han ustedes o alguien de su escuela participado en eventos de capacitación o actividades relacionadas con los resultados de fin de grado en general este año? De ser así, ¿Quién proporcionó la capacitación o evento? ¿Qué aspectos de la capacitación o actividades fueron útiles?
- ❖ ¿Ha recibido su escuela apoyo de alguna organización para *usar* los resultados de la prueba de fin de grado este año? De ser así, ¿Quién proporcionó el apoyo? ¿fue útil?
- ❖ En caso mencionen MIDEH, EROC, CADERH, CIDEH, Fundaempresa Unitec, o Asociación CESAL, y para los grupos A y B solamente: ¿Cómo se comparan la capacitación y/o actividades relacionadas con las PFG con la capacitación de EducAcción? ¿En qué se diferencian? ¿son parecidas? ¿más intensas? ¿menos intensas?
- ¿Encontraron útiles los resultados de fin de grado? (*Sondeo*: ¿Cómo fueron útiles? ¿Los utilizaron para su propia planificación? ¿Qué hicieron diferente en la forma en que planearon basado en los resultados de la PFG? ¿Los

utilizaron sus maestros para su planificación? ¿Cómo cambió la planificación/instrucción de los maestros el uso de PFG?) Si los resultados no son útiles, ¿por qué no?

- ¿Creen que las PFG proporcionan a los maestros información única que no se podría obtener trabajando con sus alumnos? Si es así, ¿qué tipo de información única proporcionan las PFG?
- ¿Creen que sus maestros enfrentan retos en el uso de los resultados de fin de grado para ajustar su enseñanza este año? Si es así, ¿qué tipo de retos?
- ¿De qué manera podrían ser más útiles los resultados de fin de grado para su escuela? ¿Necesitan resultados diferentes? ¿Los necesita antes? ¿Están presentados en una manera que es difícil de entender?
- ¿Perciben algún cambio entre el año pasado y este año con respecto a su conocimiento y/o uso de PFG?

C. Resultados de Pruebas Formativas (PF) (para todos los grupos, A, B y C)

Hablemos de las pruebas formativas.

- ¿Han recibido copias de las PF este año?
 - **En caso positivo:**
 - ¿de quién? (*Sondeo:* EducAcción, MIDEH, EROC, CADERH, CIDEH, Fundaempresa Unitec, o Asociación CESAL, Secretaria de Educacion, BID (Banco Interamericano de Desarrollo), otro)
 - ❖ [PREGUNTA CLAVE] Si mencionan alguien que no sea EducAcción: ¿guardaron las copias de PF?
 - ¿Cuántas copias?
 - Para directores del grupo A solamente: ¿Tienen conocimiento de que las PF fueron corregidas este año? ¿Pudieron usar las PF corregidas? (*Para el entrevistador:* nos estamos refiriendo a los errores en las PF, no a nuevas versiones de PF)
 - **En caso negativo:**
 - ¿Caso no hayan recibido copias este año, tienen acceso a copias de las PF por otros medios? Por ejemplo, a través de copias del año pasado u otros años?

- Que ustedes sepan, ¿han ustedes, o alguien de su escuela, participado en eventos de capacitación o actividades relacionadas con PF? Si es así, ¿fueron útiles?
 - **En caso positivo:** ¿Qué entidad(es) proporcionó(aron) la capacitación u organización de actividades relacionadas con PF?
 - Para el grupo A solamente: ¿como se compara la capacitación u organización de actividades relacionadas con PF que recibieron de esta(s) entidad(es) con las que reciben de EducAcción? ¿En qué se diferencian? son parecidas? más intensas? menos intensas?

- Por lo que saben, ¿usan sus maestros PF este año? Si no usan, ¿por qué no? Si usan, ¿por qué?
- ¿Qué piensan de las PF? ¿Son útiles? ¿Por qué o por qué no? (*Sondeo:* ¿En qué ayudan a la escuela / al maestro? ¿Cómo ayudan a los estudiantes las PF?)
- ¿Qué les gusta más sobre las PF?
- ¿Qué les gustaría mejorar en las PF? ¿Pueden pensar en alguna forma en que las PF podrían ser más útiles para sus maestros?

D. Apoyo externo

(las 3 primeras preguntas son para todos los grupos, A, B y C, y las que siguen son para los grupos A y B solamente)

Hablemos de apoyo externo este año.

- ¿Utilizan libros, cuadernos, hojas de trabajo, o papel escrito que vienen de afuera de la escuela? Si es así, ¿quién proporciona estos materiales? [SONDEE POR NOMBRES DE ONGS, MOE / DDE / ALCALDE, BID, PADRES]
- ¿Proporcionan sus propios materiales? ¿Los padres pagan o proporcionan papel en blanco, cuadernos, o materiales fotocopiados? ¿Es difícil para ustedes / los padres acceder a una fotocopidora o impresora?
- En la preparación de clases y pruebas, ¿reciben ayuda de organizaciones externas [además de EducAcción si están en los grupos A o B]?
 - En caso afirmativo, *sondear:*
 - ¿De quién reciben la ayuda?
 - ¿Cuánto tiempo pasan con ustedes?
 - ¿Qué tipos de materiales proporcionan ?
 - Para los grupos A y B solamente: ¿Cómo se compara esta ayuda con la que reciben de EducAcción? ¿En qué se diferencia?

[PREGUNTAS CLAVE] [SI ESTÁN EN LOS GRUPOS A O B, HAGA LAS SIGUIENTES PREGUNTAS. DE LO CONTRARIO, VAYA A LA SECCIÓN E.]

- ¿Tienen retos relacionados al apoyo que proporciona EducAcción? Por ejemplo, ¿pierden tiempo dedicado a otras actividades de la escuela o tienen un aumento en su tiempo de trabajo? ¿Cómo han enfrentado esos retos?
- [PREGUNTA CLAVE] Para el grupo A solamente: ¿Trabajan en PF también o solo en PFG? En caso trabajen con las PF: ¿Cuánto tiempo le dedican a trabajar con las PF? La intervención de PF ha aumentado su tiempo de trabajo o han reorganizado su tiempo (sin aumentar su tiempo de trabajo)?
- [PREGUNTA CLAVE] ¿Cuánto tiempo le dedican a trabajar con las PFG? ¿La intervención de PFG ha aumentado su tiempo de trabajo o han reorganizado su tiempo (sin aumentar su tiempo de trabajo)?
- ¿Ha cambiado el apoyo que reciben la manera en que los directores/maestros trabajan en la escuela? En caso positivo, ¿Cómo?
- Pensando específicamente en el apoyo que proporciona EducAcción, han recibido:
 - ¿capacitación en PFG?
 - ¿capacitación en PF?
 - ¿capacitación en cómo interpretar resultados de PFG?
 - ¿capacitación en cómo interpretar resultados de PF?
 - ¿capacitación en cómo usar resultados de PFG para adaptar la enseñanza?
 - ¿capacitación en cómo usar resultados de PF para adaptar la enseñanza?
 - ¿fotocopias o copias originales de PF?
- ¿Ha cambiado su conocimiento o su uso de PF y/o PFG entre el año pasado y este año?
En caso positivo: ¿Cómo?

E. Otras pruebas (para todos los grupos, A, B y C)

- ¿Tienen otras formas de evaluar el progreso de los estudiantes? (*Sondeo*: ¿Aplican los maestros la prueba diagnóstica al comienzo del año? ¿Desarrollan los maestros sus propias pruebas? ¿Se basan los maestros en la interacción con los niños en el aula?)
 - Caso apliquen la prueba diagnóstica: ¿Qué tan útil es la prueba diagnóstica comparada con la PFG? ¿Necesitan recibir un informe de los resultados de la prueba diagnóstica como para la PFG o pueden analizar los resultados sin necesidad de tener un informe?
- Si pudiera escoger solamente una prueba ¿qué prueba les parece más útil entre la prueba diagnóstica, la PF y la PFG? ¿Por qué?
- Si pudiera escoger solamente una prueba ¿qué prueba les parece más útil entre la prueba la PF y la PFG? ¿Por qué?

F. Conclusión (para todos los grupos, A, B y C)

- En general, ¿cuáles consideran son sus mayores limitaciones/desafíos como director? (*Sondeo*: ¿Cuáles son los desafíos con respecto a la enseñanza y el aprendizaje del primero al tercer grado? Cuáles son los retos de la escuela en general?)
- En general, ¿qué consideran son las limitaciones/desafíos más grandes como maestro?
- En general, ¿qué consideran son las limitaciones/desafíos más grandes para los estudiantes en las escuelas?
- Para el grupo C solamente: En general, ¿ha cambiado su conocimiento o uso de las PFG y/o PF desde el año pasado? Si es así, ¿cómo ha cambiado?
- ¿Qué aspectos fueron particularmente exitosos este año con respecto a las PFG y/o PF?
- ¿Hay algo más que les gustaría compartir con nosotros?

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

ESTUDIO CUALITATIVO DE IMPLEMENTACIÓN

GUÍA DE DISCUSIÓN PARA GRUPOS FOCALES CON MAESTROS

Línea Final de Evaluación: Agosto 2016

Hola, soy César Valenzuela de Espirállica. Estamos llevando a cabo un estudio sobre PF (pruebas formativas) y PFG (pruebas de fin de grado). El estudio es financiado por USAID, la Agencia de Estados Unidos para el Desarrollo Internacional, y está siendo llevado a cabo por Mathematica Policy Research. Ustedes han sido invitados a participar en la discusión porque son maestros de primer ciclo.

La discusión incluirá preguntas acerca de su trabajo como maestro y llevará alrededor de una hora y media. Esta información servirá para comprender mejor el uso de PF y de PFG. Toda la información que obtengamos se mantendrá en estricta privacidad y esta información no se dará a conocer en ninguna forma que permita identificarles a ustedes o sus respuestas. La información que proporcionen estará a disposición de USAID. La información recogida durante la discusión se utilizará solamente para fines de evaluación, y toda la información que los identifiquen a ustedes individualmente como nombres o información de contacto será removida. Una vez que se complete el estudio, los datos del estudio pueden estar a disposición del público para permitir análisis adicionales, excluyendo la información que los identifiquen personalmente.

Su participación en la discusión es voluntaria y pueden optar por no participar. Si deciden participar, pueden dejar de contestar cualquiera de las preguntas durante la conversación si así lo prefieren. No hay sanción o cambio en su situación laboral si no participan en este estudio. No hay riesgos ni beneficios directos para ustedes por participar en este estudio.

¿Están de acuerdo en participar en este estudio?

Debido a que no podemos anotar toda la conversación, nos gustaría grabar la discusión. La grabación se usará únicamente para transcribir la conversación y toda información que permita identificarlos será removida.

¿Están de acuerdo en que grabemos la discusión?

GUÍA TEMÁTICA

A. Perfil del maestro

Para empezar, por favor preséntese al grupo. Por favor díganos su nombre, el nombre de su escuela, el tiempo de experiencia que lleva como maestro, el tiempo que lleva en su escuela actual, y el o los grados y jornadas que enseña.

B. Resultados de pruebas fin de grado (PFG) (para todos los grupos, A, B y C)

Hablemos de las pruebas y los resultados de fin de grado del 2015.

- ¿Han recibido resultados de fin de grado de la prueba del año pasado (2015) para su escuela? ¿Recibieron resultados de sus alumnos para el grado(s) que ustedes enseñan?
 - EN CASO **NEGATIVO**, NO HAN VISTO RESULTADOS PFG DEL 2015 DE SU ESCUELA:
 - ¿Sabes cómo es posible acceder a los resultados de fin de grado de su escuela? ¿Cómo?
 - ¿Piensan que sería útil para ustedes ver los resultados de fin de grado? ¿Por qué o por qué no?
 - ¿Preferirían ver los resultados de PFG a nivel de sección o de alumno? ¿Por qué?
 - ¿Creen que las PFG proporcionan a los maestros información única que no se podría obtener trabajando directamente con sus alumnos? Si es así, ¿qué tipo de información única proporcionan las PFG?
 - [PREGUNTA CLAVE] Que ustedes sepan:

[TRATE DE AVERIGUAR SI HAN RECIBIDO EL APOYO DE EDUCACIÓN PRI Y/O MIDEH O SUS SOCIOS, PERO SEPA QUE PUEDEN NO SABER EL NOMBRE DEL PROGRAMA. LUEGO TRATE DE SONDEAR SI HAN RECIBIDO OTRO TIPO DE APOYO DE OTROS GRUPOS.]

- ❖ ¿Han ustedes o alguien de su escuela participado en eventos de capacitación o actividades relacionadas con los resultados de PFG en general este año? De ser así, ¿Quién proporcionó la capacitación o evento? ¿Qué aspectos de la capacitación fueron útiles?
 - ❖ ¿Ha recibido su escuela apoyo de alguna organización para *usar* los resultados de PFG este año? De ser así, ¿Quién proporcionó el apoyo? ¿Qué aspectos de ese apoyo les han servido?
 - ❖ En caso mencionen MIDEH, EROC, CADERH, CIDEH, Fundaempresa Unitec, o Asociación CESAL, y para los grupos A y B solamente: ¿Cómo se comparan la capacitación y/o actividades relacionadas con las PFG con la capacitación de EducAcción? ¿son parecidas? ¿En qué se diferencian? ¿más intensas? ¿menos intensas?
- **PASE A LA SECCIÓN PF (SECCION C)**

○ EN CASO **POSITIVO**, HAN VISTO LOS RESULTADOS PFG DEL 2015 DE SU ESCUELA:

- ¿Qué tipo de información recibieron?
- ¿Recibieron información a nivel de escuela, grado, sección o estudiante? ¿A qué nivel les gustaría recibir la información?
- ¿De quién recibieron los resultados? ¿En qué momento del año escolar recibieron la información? ¿Qué tipo de información había en esos informes? (*Sondeo*: ¿cómo se presentó la información: gráficamente, numéricamente, por escuela, grado, sección o estudiante?)
- ¿Aprendieron algo de la información que recibieron? En caso positivo, ¿qué aprendieron? ¿Fueron útiles los resultados de las PFG? ¿Cómo o cómo no? ¿Qué les gustó más acerca de los resultados de fin de grado?
- [PREGUNTA CLAVE] Que ustedes sepan:

[TRATE DE AVERIGUAR SI HAN RECIBIDO EL APOYO DE EDUCACIÓN PRI Y/O MIDEH O SUS SOCIOS, PERO SEPA QUE PUEDEN NO SABER EL NOMBRE DEL PROGRAMA. LUEGO TRATE DE SONDEAR SI HAN RECIBIDO OTRO TIPO DE APOYO DE OTROS GRUPOS.]

- ❖ ¿Han ustedes o alguien de su escuela participado en eventos de capacitación o actividades relacionadas con los resultados de fin de grado en general este año? De ser así, ¿Quién proporcionó la capacitación o evento? ¿Qué aspectos de la capacitación o actividades fueron útiles?
- ❖ ¿Ha recibido su escuela apoyo de alguna organización para *usar* los resultados de la prueba de fin de grado este año? De ser así, ¿Quién proporcionó el apoyo? ¿fue útil?
- ❖ En caso mencionen MIDEH, EROC, CADERH, CIDEH, Fundaempresa Unitec, o Asociación CESAL, y para los grupos A y B solamente: ¿Cómo se comparan la capacitación y/o actividades relacionadas con las PFG con la capacitación de EducAcción? ¿En qué se diferencian? ¿son parecidas? ¿más intensas? ¿menos intensas?
- ¿Alguna vez han utilizado los resultados de fin de grado para ayudarles a planear su instrucción? Por ejemplo, ¿se enteraron de áreas en las que sus estudiantes estaban teniendo dificultades y decidieron dedicar más tiempo a esas áreas?
 - ❖ En caso positivo: ¿Cómo los han utilizado? ¿Pueden dar algunos ejemplos?

- ❖ **En caso positivo:** ¿Han encontrado algún desafío utilizando los informes de fin de grado para planificar sus clases? Si es así, ¿qué retos han encontrado?
- ❖ **En caso positivo:** ¿Hubo algo que fue especialmente útil para ustedes en los resultados de la prueba de fin de grado para planificar sus clases?
- ❖ Su director(a) de centro ha trabajado con ustedes para usar los resultados de fin de grado?
 - ¿Les han ayudado los resultados de la prueba de fin de grado a identificar estudiantes específicos que podrían necesitar apoyo adicional?
 - ❖ **En caso positivo:** ¿Qué están haciendo con esos estudiantes?
 - ❖ **En caso negativo:** *Sondeo:* ¿Por qué no?
- ¿De qué manera podrían ser más útiles los resultados de fin de grado para ustedes? ¿Necesitan resultados diferentes? ¿Los necesitan antes? ¿Están presentados en una manera que es difícil de entender?
- ¿Perciben algún cambio entre el año pasado y este año con respecto a su conocimiento y/o uso de PFG?

C. Resultados de Pruebas Formativas (PF) (para todos los grupos, A, B y C)

Hablemos de las pruebas formativas.

- ¿Están familiarizados con las PF? [EN CASO CONTRARIO MOSTRAR UN EJEMPLO. SI NO LAS CONOCEN, OMITA ESTA SECCIÓN.]
- ¿Han recibido copias de las PF este año?
 - **En caso positivo:**
 - ¿de quién? (*Sondeo:* EducAcción, MIDEH, EROC, CADERH, CIDEH, Fundaempresa Unitec, o Asociación CESAL, Secretaría de Educación, BID (Banco Interamericano de Desarrollo), otro)
 - ❖ [PREGUNTA CLAVE] Si mencionan alguien que no sea EducAcción: ¿guardaron las copias de PF?
 - ¿Cuántas copias?
 - ¿Recibieron suficientes copias para la cantidad de estudiantes que enseñan?

- ¿Recibieron copias del manual de instrucción?
- **En caso negativo:**
 - ¿Caso no hayan recibido copias este año, tienen acceso a copias de las PF por otros medios? Por ejemplo, a través de copias del año pasado u otros años?
 - Que ustedes sepan, ¿han ustedes, o alguien de su escuela, participado en eventos de capacitación o actividades relacionadas con PF? Si es así, ¿Qué aspectos fueron útiles?
 - ❖ En caso positivo: ¿Qué entidad(es) proporcionó(aron) la capacitación u organización de actividades relacionadas con PF?

[PASE A LA SECCIÓN D SI NO TIENEN COPIAS DE LAS PF (O SI DEVOLVIERON TODAS)]

- ¿Utilizan las PF mensuales en sus clases?
 - **En caso negativo:**
 - ¿Por qué no las usan?
 - ¿Las han usado en el pasado?
 - ❖ En caso positivo: ¿fueron útiles? ¿Cómo?
 - ❖ En caso negativo: ¿por qué no? ¿Recibieron los materiales o capacitación para usarlas? ¿Qué aprendieron sobre el uso de las PF mensuales?

[PASE A LA SECCIÓN D SI NO HAN UTILIZADO PF ESTE AÑO]

- **En caso positivo: (Para todos los que utilizan PF este año, independiente de como las consiguieron)**
- ¿Cómo deciden cuándo administrar las PF? ¿las aplican en el mes programado o cuando se terminó de cubrir el material? [NOTA: ESTAMOS INTERESADOS EN SABER CÓMO DECIDEN LA SINCRONIZACIÓN CON RESPECTO A CUÁNDO ESTÁN CUBRIENDO QUÉ MATERIALES CON SUS ALUMNOS. POR EJEMPLO, ¿ESTÁN ESPERANDO HASTA DESPUÉS DE HABER CUBIERTO EL MATERIAL EN LA PF ANTES DE SU ADMINISTRACIÓN, O SIMPLEMENTE SIGUIENDO EL MES CALENDARIO?]

- Cuéntenos cómo las administran este año. (Por ejemplo, ¿trabajan por su cuenta los estudiantes? ¿Tienen una cantidad fija de tiempo? ¿Escriben en sus propios cuadernos o en el papel que dan al maestro?)
- Para docentes del grupo A solamente: ¿Tienen conocimiento de que las PF fueron corregidas este año? ¿Pudieron usar las PF corregidas? (*Para el entrevistador:* nos estamos refiriendo a los errores en las PF, no a nuevas versiones de PF)
- Que ustedes sepan, ¿han ustedes, o alguien de su escuela, participado en eventos de capacitación o actividades relacionadas con PF? Si es así, ¿fueron útiles? ¿de qué manera?
 - **En caso positivo:** ¿Qué entidad(es) proporcionó(aron) la capacitación u organización de actividades relacionadas con PF?
 - Para el grupo A solamente: ¿como se compara la capacitación u organización de actividades relacionadas con PF que recibieron de esta(s) entidad(es) con las que reciben de EducAcción? son parecidas? En qué se diferencian? más intensas? menos intensas?
- ¿Hay elementos que fueron particularmente exitosos durante el uso de las PF este año? De ser así, ¿cuáles fueron los éxitos? ¿Qué les gustó más acerca de las PF?
- ¿Han encontrado algún desafío utilizando las PF para planificar sus clases este año? De ser así, ¿qué retos han encontrado? ¿Qué le gustaría mejorar en las PF?
- [PREGUNTA CLAVE]: ¿Están enfocando la instrucción de acuerdo a los resultados este año?
 - **En caso positivo:**
 - ¿Qué están haciendo?
 - ¿Han cambiado su forma de enseñar a los estudiantes de su grado o la forma de diferenciar la enseñanza según el nivel cada estudiante individualmente?
 - ❖ En caso positivo: ¿Qué aspectos de su enseñanza han cambiado?
 - ¿Han cambiado su forma de enseñar a nivel grupal?
 - ❖ En caso positivo: ¿Qué aspectos de su enseñanza han cambiado?
 - ¿Pueden dar ejemplos de las maneras cómo han ajustado su enseñanza este año?

D. Apoyo externo

(las 3 primeras preguntas son para todos los grupos, A, B y C, y las que siguen son para los grupos A y B solamente)

Hablemos de apoyo externo este año.

- ¿Utilizan libros, cuadernos, hojas de trabajo, o papel escrito que vienen de afuera de la escuela? Si es así, ¿quién proporciona estos materiales? [SONDEE POR NOMBRES DE ONGS, MOE / DDE / ALCALDE, BID, PADRES]
- ¿Proporcionan sus propios materiales? ¿Los padres pagan o proporcionan papel en blanco, cuadernos, o materiales fotocopiados? ¿Es difícil para ustedes / los padres acceder a una fotocopidora o impresora?
- En la preparación de clases y pruebas, ¿reciben ayuda de organizaciones externas [además de EducAcción si están en los grupos A o B]?
 - En caso positivo, *sondear*:
 - ¿De quién reciben la ayuda?
 - ¿Cuánto tiempo pasan con ustedes?
 - ¿Qué tipo de materiales proporcionan?
 - Para los grupos A y B solamente: ¿Cómo se compara esta ayuda con la que reciben de EducAcción? ¿En qué se diferencia?

[PREGUNTAS CLAVE] [SI ESTÁN EN LOS GRUPOS A O B, HAGA LAS SIGUIENTES PREGUNTAS. DE LO CONTRARIO, VAYA A LA SECCIÓN E.]

- ¿Tienen retos relacionados al apoyo que proporciona EducAcción? Por ejemplo, ¿pierden tiempo de clases o tienen un aumento en su tiempo de trabajo? ¿Cómo han enfrentado esos retos?
- [PREGUNTA CLAVE] Para el grupo A solamente: ¿Cuánto tiempo le dedican a trabajar con las PF? La intervención de PF ha aumentado su tiempo de trabajo o han reorganizado su tiempo (sin aumentar su tiempo de trabajo)?
- [PREGUNTA CLAVE] ¿Cuánto tiempo le dedican a trabajar con las PFG? ¿La intervención de PFG ha aumentado su tiempo de trabajo o han reorganizado su tiempo (sin aumentar su tiempo de trabajo)?
- ¿Ha cambiado el apoyo que reciben la manera en que los directores/maestros trabajan en la escuela? En caso positivo, ¿Cómo?
- Pensando específicamente en el apoyo que proporciona EducAcción, han recibido:
 - ¿capacitación en PFG?
 - ¿capacitación en PF?
 - ¿capacitación en cómo interpretar resultados de PFG?
 - ¿capacitación en cómo interpretar resultados de PF?
 - ¿capacitación en cómo usar resultados de PFG para adaptar la enseñanza?
 - ¿capacitación en cómo usar resultados de PF para adaptar la enseñanza?
 - ¿fotocopias o copias originales de PF?

- ¿Ha cambiado su conocimiento o su uso de PF y/o PFG entre el año pasado y este año?
En caso positivo: ¿Cómo?

E. Otras pruebas (para todos los grupos, A, B y C)

- ¿Tienen otras formas de evaluar el progreso de los estudiantes? (*Sondeo*: ¿Aplican la prueba diagnóstica al comienzo del año? ¿Desarrollan sus propias pruebas? ¿Se basan en la interacción con los niños en el aula?)
 - Caso apliquen la prueba diagnóstica: ¿Qué tan útil es la prueba diagnóstica comparada con la PFG? ¿Necesitan recibir un informe de los resultados de la prueba diagnóstica como para la PFG o pueden analizar los resultados sin necesidad de tener un informe?
- Si pudieran escoger solamente una prueba ¿qué prueba les parece más útil entre la prueba diagnóstica, la PF y la PFG? ¿Por qué?
- Si pudieran escoger solamente una prueba ¿qué prueba les parece más útil entre la prueba la PF y la PFG? ¿Por qué?

F. Conclusión (para todos los grupos, A, B y C)

- En general, ¿cuáles consideran son sus mayores limitaciones/desafíos como maestros?
- En general, ¿qué consideran son las limitaciones/desafíos más grandes para los estudiantes en las escuelas?
- Para el grupo C solamente: En general, ¿ha cambiado su conocimiento o uso de las PFG y/o PF desde el año pasado? Si es así, ¿cómo ha cambiado?
- ¿Qué aspectos fueron particularmente exitosos este año con respecto a las PFG y/o PF?
- ¿Hay algo más que les gustaría compartir con nosotros?

U.S. Agency for International Development

1300 Pennsylvania Avenue, NW

Washington, DC 20523

Tel: (202) 712-0000

Fax: (202) 216-3524

www.usaid.gov