MATHEMATICA Policy Research

REPORT

DRAFT REPORT

Guatemala Threshold Education Program: Evaluation Design Report

April 19, 2018
Revised

Edith Felix
Sarah Humpage Liuzzi
Audrey Moore
Seth Morgan
Catalina Torrente
Anthony DeWees (Consultant)

Submitted to:
Millennium Challenge Corporation
1099 14th Street, NW
Washington, DC 20005

Project Officer: Pamela Velez-Vega
Contract Number: MCC-16-CON-0040

Submitted by:
Mathematica Policy Research
1100 First Street, NE, 12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director: Audrey Moore
Reference Number: 50314

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

CONTENTS

TABLES

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

FIGURES

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

I. INTRODUCTION

Guatemala, like many countries in Latin America and the Caribbean (LAC), has made impressive progress in expanding education coverage, particularly at the primary level. However, many youth who could be enrolled in higher levels of education remain out of school, and many of those enrolled in school have no access to high quality education. Although net enrollment in primary school is nearly 100 percent (U.S. Agency for International Development [USAID] 2015), net enrollment in lower- and upper-secondary schools totals only 43 and 24 percent, respectively (Inter-American Development Bank 2015). Students' performance on learning assessments is also low. In 2013, only 18 percent of 191,089 lower-secondary school graduates met the achievement standards in mathematics, and only 15 percent met the standards in language and communications. According to 2015 data, only 9 percent of 149,652 upper-secondary school graduates met the achievement standards in mathematics, and only 26 percent met the standards in language and communications (Dirección General de Evaluación e Investigación Educativa [DIGEDUCA] 2016). As a result, nearly 2 million Guatemalan youth age 15 through 24 lack the foundational skills needed to enter the formal workforce (USAID 2015).

To improve secondary education and youth workforce development, the government of Guatemala has partnered with the Millennium Challenge Corporation (MCC) through a threshold program focusing on secondary education in the country. The Guatemala Threshold Education Project (GTEP) is organized around three activities: (1) the Quality of Education in Support of Student Success (Éxito Escolar) activity, which seeks to strengthen the competencies of educators to promote equitable, high quality secondary education; (2) the Technical and Vocational Education and Training (TVET) activity, which supports the Ministry of Education (MINEDUC) in improving technical and vocational education and training; and (3) the Strengthening of Institutional and Planning Capacity (IPC) activity, which strengthens the institutional capacity of MINEDUC.

MCC has contracted with Mathematica Policy Research to evaluate the GTEP. In this report, we describe the design for the evaluations of the Éxito Escolar and IPC activities (Activities 1 and 3, respectively). The evaluation of the Éxito Escolar activity will include a randomized controlled trial (RCT) and an implementation study. For the IPC activity, we will conduct a performance evaluation to assess the extent to which MINEDUC's strengthened capacity has led to improved planning and budgeting for the enhanced quality of secondary education. We will also conduct a political economy analysis of the decision making that determines how Guatemala allocates resources for education.

In the chapters that follow, we provide context for the evaluations and describe the planned evaluation designs in further detail. In Chapter II, we outline the activities of the GTEP and the program logic and summarize the literature related to the effects of similar technical interventions. In Chapter III, we discuss the research questions that our evaluations seek to answer and describe the evaluation designs and data sources that we will use to conduct the evaluation. We conclude in Chapter IV with a discussion of administrative details related to the evaluation.

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

II. OVERVIEW OF THE PROJECT AND IMPLEMENTATION PLAN

In this chapter, we provide context for the planned evaluations by describing the GTEP activities and the mechanisms through which they are expected to affect outcomes, as set out in the program logic. We also review the existing literature on the impact of similar interventions.

The objective of the GTEP is to support the government of Guatemala's ongoing institutional reforms defined in the Proposal for the Transformation of Secondary Education (Asturias de Barrios 2014) and the Ruta Crítica (MINEDUC 2016). The objective of these reforms is to provide youth in Guatemala with a high quality secondary education that prepares them to succeed in the labor market. The GTEP consists of three complementary activities. In this section, we describe the subactivities, participants, and geographic scope of the first and third activities.

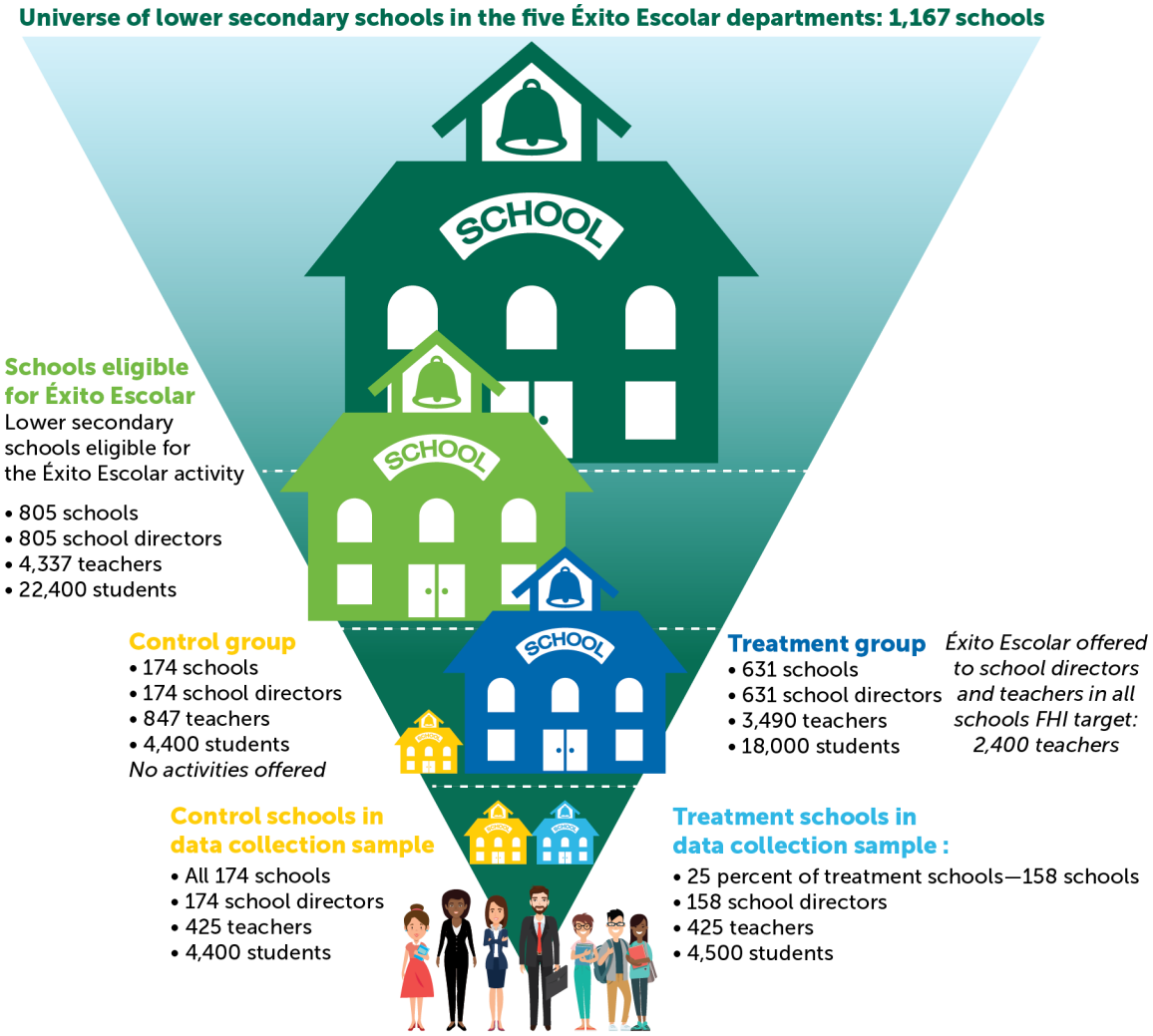
GTEP's Activity 1 is the Éxito Escolar activity, which supports MINEDUC's interventions to improve the quality of lower-secondary schools (grades 7 through 9). Four components make up the Éxito Escolar activity, as follows:

1. **Professional development for teachers.** The GTEP will develop a teacher professional development program with a target of providing 2,400 teachers with an intensive, 20-month training in both content knowledge (that is, language and communication, mathematics, and science) and pedagogic skills. The GTEP will offer the program beginning in the 2018 school year and continuing into the 2019 school year. The program will include over 1,800 hours of in-person instruction and virtual learning, roughly evenly divided between the two. After completing the program, teachers will receive a *profesorado de educación media* (PEM) degree. Participation in the professional development component will be optional.
2. **Pedagogic support and communities of practice.** The GTEP will train about 80 pedagogic advisors and 40 management advisors to help school directors and teachers at lower-secondary schools implement the new pedagogic and management techniques. The advisors will work across all treatment schools. The GTEP will group the schools into communities of practice with neighboring lower-secondary schools that are part of the treatment group. School directors and teachers will participate in pedagogic support activities alongside their counterparts from the neighboring schools. Participation in the pedagogic support and communities of practice activities will be optional.
3. **Establishing school networks.** The GTEP will establish 100 new school networks that link at least five primary schools to one lower-secondary school. Whereas the communities of practice will link lower-secondary schools to other lower-secondary schools, the school networks will link lower-secondary schools to primary schools. One purpose of the school networks is to create a feeder system from primary school to lower-secondary school that helps improve student retention and the transition from primary to lower-secondary education.
4. **Establishing parent councils.** The GTEP will establish parent councils at 400 of the lower-secondary schools participating in Éxito Escolar (some of which may also participate in the newly established school networks). The GTEP will train parents who are council members

to support their respective schools and to serve as liaisons between the schools and families as a way of reducing the number of dropouts in lower-secondary schools.

The impact and implementation evaluations of the Éxito Escolar activity are designed to assess the combined impacts of all four components. The impact evaluation will rely on a randomized controlled trial (RCT), described in detail in Chapter IV. Figure II.1 summarizes how many schools will be placed in a treatment group, how many will be in a control group, and how many schools will be offered each component described above.

Figure II.1. Éxito Escolar activity universe, eligible schools, and data collection sample



Note: The number of teachers selected for data collection is an estimate, which is based on available data, and which may change depending on the actual number of teachers in the schools. Our estimate takes into consideration that some teachers teach multiple subjects and some are also principals.

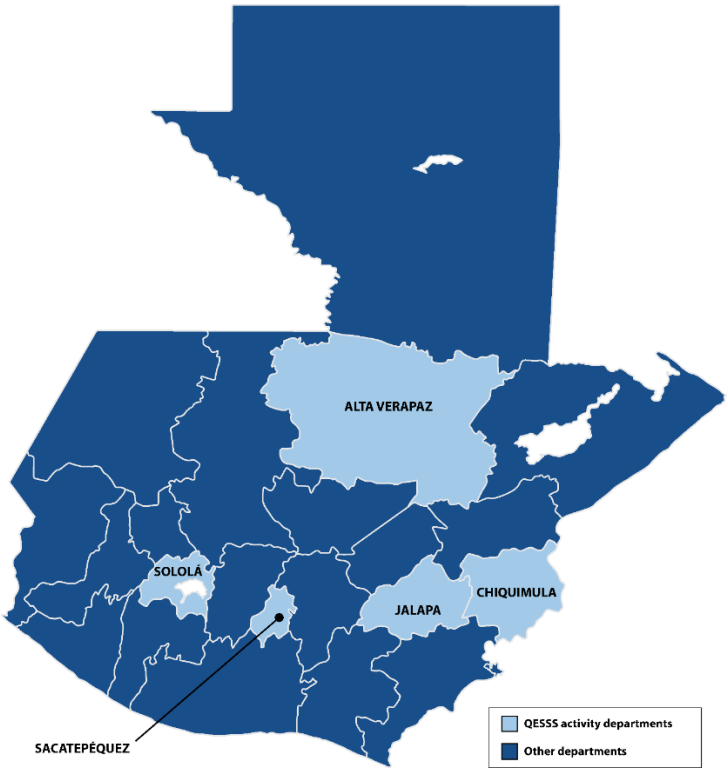
In addition to the Éxito Escolar interventions, DIGEDUCA will develop a series of student assessments (funded separately by MCC) to measure students' outcomes in language and communication (Spanish), mathematics, and science. Mathematica will provide technical support

by participating in the field testing process, providing feedback on the field test, reviewing a final version of the test and providing feedback on that version, and helping to oversee data collection. The test score data generated through these assessments will provide outcome data for the impact evaluation discussed later in this design report.

The main participants in the Éxito Escolar activity are lower-secondary school directors and teachers who participate in activities including professional development, pedagogic support, or communities of practice, and their students, who may benefit from improved school management and teaching. Additional participants include parents who participate in parent councils and primary school directors or teachers who participate in the school network activities (we will learn about the experiences of these groups through our qualitative data collection and analysis, but will not estimate impacts on parents or primary school staff through the impact evaluation).

FHI360 will implement the Éxito Escolar activity in five departments: Alta Verapaz, Chiquimula, Jalapa, Sacatepéquez, and Sololá (Figure II.2).

Figure II.2. Éxito Escolar activity: geographic scope



MINEDUC and MCC are designing the second activity, the Technical Vocational Education and Training (TVET) program. At a later date, we will add to the Evaluation Design Report a description of our approach to evaluate the TVET activity.

The GTEP's third activity is the Strengthening of Institutional Capacity and Planning (IPC for its Spanish acronym) activity. Its objective is to strengthen the institutional capacity of

MINEDUC to improve its planning and budgeting functions to enable it to provide an equitable and high quality secondary education (MCC 2016). The IPC activity will:

1. Assess the effectiveness and efficiency of different models of lower-secondary schools with a sample of about 50 lower-secondary schools. The assessment will identify the minimum inputs needed to provide a high quality education,[1] the inputs currently provided by the government, and the budgetary requirements needed to ensure the provision of minimum inputs in the secondary education system.
2. Strengthen management information systems, support data collection, improve data quality, and promote the use of data as a tool for planning the delivery of secondary education services. This cluster of activities includes management of the training and professional development opportunities targeted to secondary education teachers and the creation of a virtual bank of education resources made available to school directors, teachers, parents, and other community members.
3. Advance the institutionalization of a competitive teacher selection process (including a diagnostic test). The institutionalization process will provide technical and legal assistance to the Jurado Nacional de Oposición[2] and the Dirección de Recursos Humanos in designing and planning a competitive teacher selection process.
4. Develop a geographic analysis of the supply and demand of secondary education as a factor in estimating needed resources for infrastructure, teacher assignments, and materials and in planning and budgeting for a high quality education.

The IPC activity will directly affect a wide range of participants by engaging MINEDUC staff in improving the planning and management of lower-secondary education. The activity will train 50 directors and technicians in MINEDUC's central headquarters as well as department directors responsible for the planning and management of lower-secondary education. The activity will provide for at least two education budget workshops with about 1,000 participants per workshop. Interested civil society and private sector stakeholders will be permitted to participate in some of the institutional capacity-building training workshops.

Given that MINEDUC is responsible for lower-secondary education throughout Guatemala, the IPC will be national in scope. Potential indirect beneficiaries include school directors, teachers, parents, and community members who access the virtual bank of education resources[3] to be established under the IPC activity. These groups would benefit because as the government allocates more resources to the schools, the additional resources could positively affect teacher

[1] Quality is defined as the minimum amount financial and human resources needed to ensure that more primary school students transition into lower-secondary schools, complete secondary education, and find better jobs than they would without secondary education.

[2] The Jurado Nacional de Oposición monitors transparency in the teacher hiring process for both preprimary and primary education. The Dirección de Recursos Humanos is MINEDUC's human resources department.

[3] FHI360 will establish the virtual bank (banco de recursos educativos) as part of the activity to improve planning and decision making.

pay, increased numbers of qualified teachers, and the provision of additional professional development. However, we will not measure or quantify these benefits in the evaluation.

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

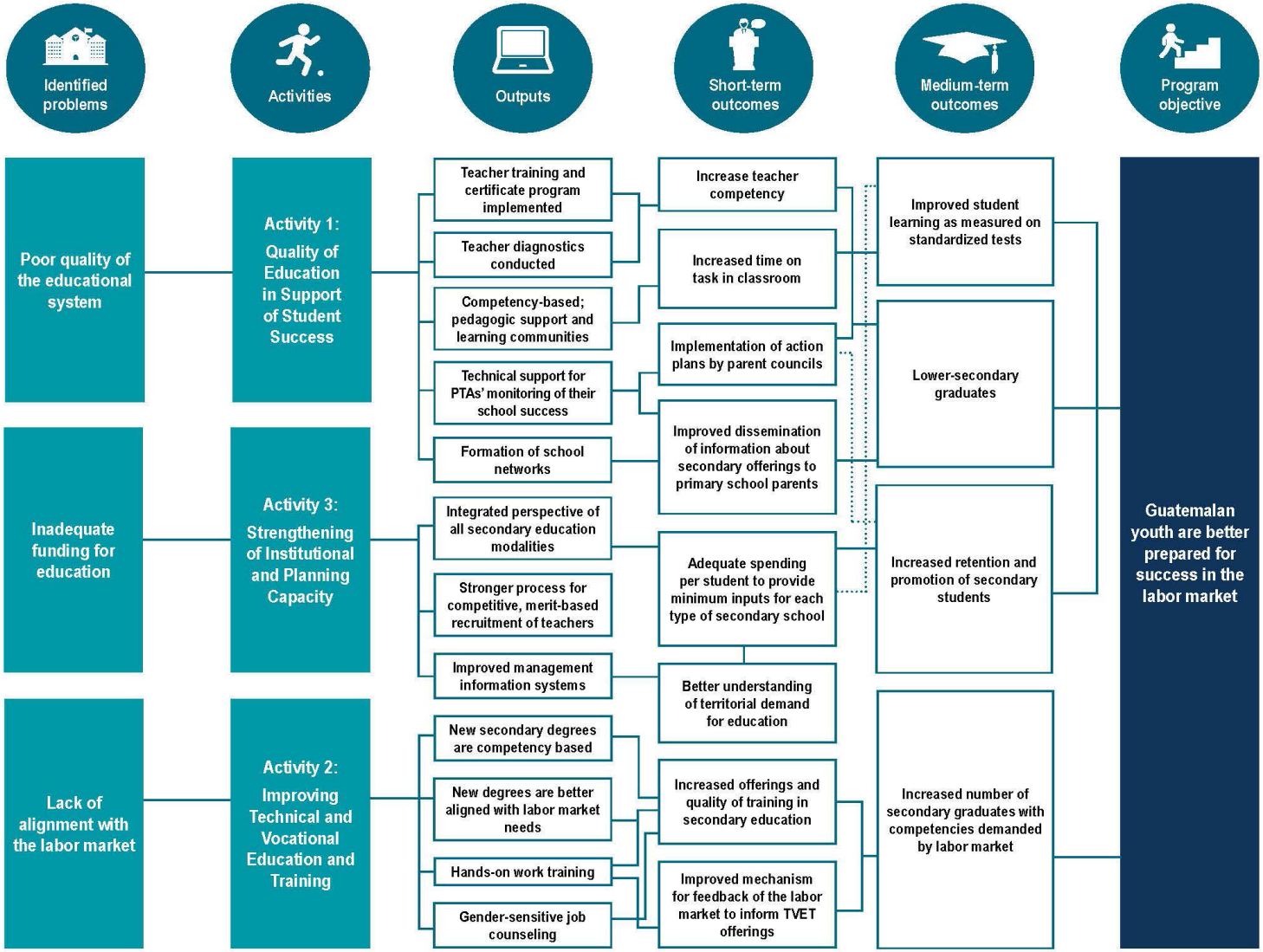
III. THEORY OF CHANGE

The program logic developed by both MCC and MINEDUC staff encompasses a series of hypothesized causal links among program inputs and outputs and short-, medium-, and long-term outcomes that potentially support the program's overarching goal of improved preparation of youth for entry into the workforce (Figure III.1). Each of the links in the program logic represents an assumption by GTEP program designers about how the activities will affect participants—students, teachers, school administrators, and parents and policymakers in relevant government of Guatemala ministries, departments, and offices.

The program logic of Éxito Escolar (also referred to as Activity 1) assumes that the interventions will improve the quality of education in the short term (leading to improved student learning) and increase retention and promotion in the medium term (MCC 2016). The program logic further assumes that improvements in students' outcomes should produce graduates who are better prepared for the workforce over the long term, thus linking the activity to the main project objective of improving the education of Guatemalan youth for success in the labor market.

The IPC's program logic envisions institutional strengthening efforts that will integrate the various secondary education modalities and improve equity and results by developing and implementing policies, systems, and tools to recruit teachers and allocate material and financial resources effectively and equitably.

Figure III.1. The GTEP program logic



Note: Activity 2 is still under development by MCC and MINEDUC.

IV. EVALUATION DESIGN FOR ÉXITO ESCOLAR

In this section, we describe our design for the evaluation of Éxito Escolar. We begin with a review of the literature and describe the contribution we hope to make to the literature through the evaluation. We then list the research questions that the evaluation seeks to address and provide an overview of the evaluation design, which includes several components. Then, we describe in detail each component by activity, the data the evaluation will use for analysis, and how we will draw on the data to answer the research questions. Finally, we discuss some of the limitations and challenges of the evaluation.

A. Literature review

In this section, we present a review of the literature related to the four subactivities included in the Éxito Escolar activity (described in Section II). We then identify the contributions we expect the evaluation of the Éxito Escolar activity to make to the literature.

1. In-service professional development programs for teachers

The literature suggests that support to teachers is critical to the continual development of pedagogic skills. Rigorous evidence demonstrates a positive relationship between teachers' professional development and students' performance. Popova and Evans (2015) looked at 6 reviews (Conn 2014; Glewwe et al. 2014; Kremer et al. 2013; Krishnaratne et al. 2013; McEwan 2015; Murnane and Ganimian 2014) that examine the interventions that improved learning outcomes in low- and middle-income countries. The results of their analyses show that teacher training interventions have the second-largest effects on learning outcomes. McEwan's (2015) meta-analysis of 11 randomized studies with professional development interventions for teachers found a 0.12 standard deviation (SD) improvement in learning (significant at the 1 percent level). However, Popova and Evans (2015) explained that the effectiveness of a professional development program for teachers depends on its specific interventions. For example, Muralidharan and Sundararaman (2010) implemented a randomized experiment in India to examine the impact of low-stakes diagnostic tests[4] and feedback to teachers and of low-stakes monitoring of classroom process; they found no impact on students' learning outcomes. Piper and Korda (2010) rigorously evaluated a program in Liberia that gave teachers scripted lesson plans; they found impacts on seven early-grade literacy outcomes that ranged from 0.39 to 1.23 SDs. The findings suggest that dosage, duration, enabling environment, and type of intervention are factors in program impacts.

Efforts to achieve improved student learning outcomes related to teacher-level interventions require sustained and intensive professional development and support for professional learning (Darling-Hammond et al. 2009). Structured pedagogy programs, which typically address several constraints to learning—including poorly trained teachers—have demonstrated positive effects on students' learning outcomes. However, the programs have not affected school participation

[4] Teachers, school directors, or other school authorities use diagnostic tests to evaluate students' knowledge. These tests are considered low-stakes if the results do not have consequences for the teachers, school directors, or students. In contrast, high-stakes tests might affect students' grades, result in students failing a grade, or affect teachers' or school directors' pay.

and retention, which are more responsive to student-level interventions, such as removing user fees and targeted attention on students (Snilstveit et al. 2016). Cohen and Hill (2001) suggest that teacher professional development can affect teachers' practices when the curriculum is designed to be consistent with the desired reforms. Therefore, continuous professional development that is aligned with curriculum materials and paired with strong pedagogic support is most likely to lead to sustainable change in teachers' behavior and methods—and by extension, in student learning outcomes. Because Éxito Escolar has a similar design, we expect that the changes in teachers' behavior and methods will affect students' outcomes.

2. Time on task and learning

MINEDUC is interested in examining the relationships between teacher professional development and time on task, and how these elements contribute to improved students' learning. Studies in developing countries provide evidence that additional time spent on learning tasks can plausibly improve students' test scores. A randomized evaluation on the effects of short-term tutoring on cognitive and noncognitive skills in Chile found that students from low-performing and poor schools improved their language and communication test scores after participating in a focused three-month program (Cabezas et al. 2011). Bruns and Luque (2015) analyzed teachers' use of instructional time in Latin American classrooms. In Brazil, Colombia, Honduras, and Mexico City, they used students' test data to examine the correlation between teachers' use of time and school-level learning results. Although their study did not establish causality, they found that, in schools with greater academic time on task, students achieved higher learning outcomes. The results were consistent across all four countries. Moore et al. (2012) found similar results when they examined the use of time in schools and classrooms in Ethiopia, Guatemala, Honduras, Mozambique, and Nepal; teachers who were able to use class time effectively and keep students engaged in academics saw their students improve their performance on early-grade reading tests. The teacher professional development program designed by FHI360 is expected to give teachers new skills and resources that improve their use of time in the classroom. As a result, we expect the evaluation of the Éxito Escolar activity to produce data on changes in the use of time in the classroom and contribute to the literature on the types of training and skills that increase time on task and thereby lead to improved student performance.

3. Pedagogic support and school leadership

Although the international literature does not disentangle pedagogic support from teacher professional development programs, the evidence shows that programs with pedagogic support in the form of in-school teacher coaching or long-term teacher mentoring had positive (although not always significant) effects on students' learning. Conn (2014) reviewed four rigorous evaluations of interventions with ongoing teacher professional development elements (Brooker et al. 2013; Lucas et al. 2013; Sailors et al. 2010; Spratt et al. 2013). She found that the interventions led to a pooled effect size of 0.25 SDs on student learning. Rodríguez et al. (2010) identified a correlation between the success of school-based management initiatives and strong leadership (Carr-Hill et al. 2015). Mexico, Reimers and Cardenas (2007) also found that leadership can be a significant enabler of or barrier to the impact of school-based decision-making reforms. However, the evidence on the impacts of school leadership is limited, and few if any studies compare schools with different lead teacher characteristics. Because FHI360's intervention is expected to include pedagogic support (teacher professional development and

communities of practice) and training for school directors, this evaluation can contribute to the literature on pedagogic support and school leadership.

4. Student learning assessments

There is a lack of rigorous evidence on the potential impacts of formative assessments. Perry (2013) reviewed the evidence from formative assessments in Africa and found that it was promising but strikingly limited. Two recent experimental evaluations in Liberia (Piper and Korda 2010) and South Africa (Piper 2009) showed that formative assessments can generate large effects when paired with a detailed curriculum. A growing body of evidence on end-of-grade summative assessments suggests that assessment data can contribute to improved learning outcomes through the dissemination of assessment results. However, teachers need to be trained in how to use the results to support their students. The evidence also shows that summative evaluations can affect learning through an accountability effect. Piper and Korda (2010) found significant growth in test scores simply by reporting test data to families and schools. Andrabi et al. (2015) found that the dissemination of test scores to families in Pakistan through school report cards improved school quality and increased test scores by 0.1 SDs. Even though it is unclear whether or how FHI360 and DIGEDUCA will use the results of various student assessments to support learning, the literature demonstrates potential for such interventions to produce an impact on learning outcomes.

5. Parent councils

A World Bank review of the literature (Bruns et al. 2011) of the impacts of school-based management (SBM) found that, despite the many SBM programs around the globe, there is only a small sample of well-documented rigorous impact evaluations. According to the literature, time is an important consideration because SBM reforms tend to take at least two to three years to achieve their expected results. The first year is usually an adjustment period when changes, such as the creation of a parent council, undergo initial implementation.

A meta-analysis of experimental and quasi-experimental studies in low- and middle-income countries conducted by Carr-Hill et al. (2015) presented the pooled effect sizes of SBM interventions for the following outcomes: dropout, grade repetition, teacher attendance, and test scores. A few individual SBM impact studies of dropout rates appear to demonstrate a beneficial impact when the pooled effect size from 10 studies is small in magnitude and not statistically significant. However, the pooled effect size (−0.09 SDs) from 5 studies indicated that SBM interventions have led to statistically significant reductions in grade repetition. The overall impact of SBM interventions on teacher attendance was positive (0.1 SDs) but not statistically significant when pooled across 7 studies (Carr-Hill et al. 2015).

The evidence of the impact of SBM interventions on test scores is much more mixed. The pooled effect size from five impact studies in the Philippines and Kenya reviewed by Carr-Hill et al. (2015) was 0.21 SDs and was statistically significant. Studies in El Salvador, Mexico, and Nicaragua also found positive effects on test scores (Jimenez and Sawada 2003; Lopez-Calva and Espinosa 2006; King and Özler 2005). In contrast, studies in Brazil and Honduras found no effects of SBM on test scores (Bruns et al. 2011). A meta-analysis of 232 U.S. studies, consisting of more than 1,000 observations of 29 programs, found that, to bring about fundamental school-level changes, SBM interventions require about eight years to affect test scores (Bruns et al.

2011). In the case of the Éxito Escolar activity, parent councils will be responsible for supporting communication within the school networks, and this communication in turn is expected to help increase enrollment, retention, and transition through the years of secondary education. In El Salvador, studies found that SBM committees in the EDUCO program positively affected enrollment, attendance, and retention (Moore et al. 2012; Jimenez and Sawada 2003). We will not be able to determine the impacts of the parent council or SBM activities on these outcome variables because we are only able to assess the package of interventions. We will however, gather qualitative data from participants, which we hope will help us understand how these school-level interventions may have supported students.

6. School networks

For several decades, Latin American countries have employed a variety of school network modalities for pedagogic and institutional management. Guatemala is experienced in implementing several school network modalities in primary schools. One of Guatemala's first experiences with school networks dates back to the late 1940s when the country established school federations, which called for the organization of schools under a federation of grades, each with its own classroom and services. Since then, Guatemala has gained experience with networks of rural schools and networks of regional schools, along with networks run by pedagogic advisors. Guatemala has also attempted to learn from the school networks experiences of Honduras and Peru (MINEDUC 2017).

The evidence on the impacts of schools networks on enrollment, retention, transition, and learning is limited. However, a United Nations Educational, Scientific and Cultural Organization (UNESCO) review of school network case studies found some promising results. A Venezuela school network developed common teaching criteria and methods for different grade levels. Since establishment of the school network, more students have successfully transitioned from grade 6 to grade 7 (Giordano 2008). Another case study in France also found that, when primary students have access to secondary schools within a network, they make an easy transition to the secondary level because they are already familiar with their new school (Duhamel et al. 2003).

Under the Éxito Escolar activity, MINEDUC is planning to create networks that link five primary schools to a ciclo básico school with the hope that the linkage—and support from the parent councils—will help more students transition into the higher levels of education. Our evaluation will not directly measure the impact of the school networks on transition; however, we may be able to add to the literature on networks through our qualitative analysis of school networks' role in the educational transition. We will complement this analysis with analysis of administrative data, although we will not be able to establish causal impacts.

7. The Éxito Escolar evaluation's contribution to the literature

Our evaluation of Éxito Escolar will make several contributions to the existing literature. First, the impact evaluation will generate estimates of the impacts of a package of interventions that includes professional development for teachers and school directors, pedagogic support, and—in some schools—parent councils and the development of school networks. Although our design does not allow us to estimate the effect of any one intervention, the qualitative results will help us understand the factors that contribute to any improvements in students' transition into secondary education or improvements in test scores, thus expanding the body of international

literature. Given the limited evidence on pedagogic support and school leadership, our findings on this package of interventions could make an important contribution to the understanding of how best to train and support content knowledge development among secondary education teachers. We will also use qualitative data on people's perceptions of what has helped improve student learning, and observational data of what practices teachers adopted and how that may have contributed to any changes in student learning.

B. Evaluation overview and research questions for Éxito Escolar

The evaluation of Éxito Escolar seeks to answer a range of research questions, listed in Table IV.1. We have organized the questions into those that are part of the implementation study and those that are part of the impact analysis. This table includes the original research questions proposed by MCC in the request for quotation. Some of these questions are not feasible with the current design. To answer as many questions as possible, we have modified some of the original impact questions. In some cases, we have revised original impact questions that sought to identify the impact of specific components of the Éxito Escolar activity by instead seeking to identify the impact of the Éxito Escolar activity overall. In other cases, we have revised impact questions to be answered as part of the implementation study.

Table IV.1. Research questions for the Éxito Escolar activity: Original and revised

Fidelity of implementation study questions	Original or new question
To what extent were Éxito Escolar's planned activities for teacher professional development, pedagogical support, and development of parent councils and school networks implemented as designed?	
1. What effect do the teacher diagnostics have on teacher motivation?	Original
2. Are the teacher diagnostics useful in training teachers and helping them reach a higher competency level?	Original
3. Are the training instructors and the Ministry able to respond appropriately and with appropriate training when teachers need more support?	Original
4. Did a majority of teachers complete the training?	Original
5. What obstacles did teachers face when completing the training?	Original
6. What kinds of pedagogical support are most important to teachers?	Original
7. Do teachers and their assigned pedagogic advisors meet regularly?	Original
8. Did teacher competency improve after the implementation of Éxito Escolar?	New
9. What were teachers' perceptions of the reasons for changes in student learning outcomes? Did they come as a result of the teacher training program, the parent councils, school networks, or an interaction between all of the components?	New
10. Did the parent councils implement the Action Plans initiated by Éxito Escolar as planned?	Original
11. What were the results of the plans? How effective were they?	Original
12. Are parent councils able to identify and successfully mitigate factors that lead students to drop out of school?	Original
13. Do dropout rates decrease with additional funds from municipalities and capacity building for parent councils?	Original
14. Is the additional support from local government targeted at the right families and students?	Original
15. Are indigenous families represented in the parent councils?	Original

Fidelity of implementation study questions	Original or new question
16. What kind of support do the parent councils provide female students and their families to encourage those students to stay in school?	Original
17. How do school directors, teachers, parents, and students perceive the relative contribution of different subactivities (i.e., teacher professional development, pedagogic support, school networks, and parent councils) to observed changes in students' outcomes?	New
18. How did parents' perceptions of secondary education change as a result of broader dissemination of information to parents of primary school students?	New
19. What were the main facilitators of and barriers to implementing Éxito Escolar activities in terms of reaching hypothesized medium term outcomes including: a. Improvements in student learning b. Higher graduation rates c. Increased retention and promotion of secondary education students	New
Impact evaluation questions	
20. Did teachers adopt new pedagogical approaches as a result of Éxito Escolar? (e.g. active learning, more attention paid to different learning styles of students, adaption of lessons for language minorities, and equal attention paid to both genders and students of all socio-economic backgrounds)	New
21. To what extent did time devoted to learning in the classroom increase as a result of the intervention?	Original
22. How does time devoted to learning in the classroom vary across different teaching subjects?	Original
23. How did Éxito Escolar affect student enrollment in secondary schools (promotion rates, retention rates, dropout rates)?	New
24. How were student learning outcomes affected by Éxito Escolar?	New
25. How do changes in student learning outcomes vary across gender, socio-economic, and language groups?	Original

Note: This list is based on the list of evaluation questions listed under section C.3.6.1 of contract MCC-16-CON-0040. Annex A lists questions that we are unable to answer, and identifies modified research questions added to replace those questions.

To answer these questions, we will conduct a mixed-methods evaluation organized into two main components. The first component is an implementation study that will draw on (1) three rounds of key informant interviews with stakeholders and (2) focus groups with school directors, parents, teachers, and students. The study will use purposeful sampling (maximum variation sampling)[5] to select a subset of schools from which to draw focus group participants. The implementation study's qualitative data collection activities will provide the flexibility needed to respond to participants' ideas by, for example, permitting us to explore complex or unexpected ideas. We discuss our implementation study design in greater detail in Section IV.C.

The second component is an impact evaluation with an RCT design. In a public random assignment ceremony, we randomly assigned school districts (geographically based groups of an average of eight schools) to either a treatment or a control group. FHI360 will form communities of practice within treatment districts. Depending on a district's size and the location of the

[5] The maximum variation approach enables us to identify themes that occur consistently across schools and participants regardless of the variation in performance and to identify issues that might be unique to any one group (Patton 1990).

district's schools, communities of practice may comprise all schools in a district; alternatively, more than one community of practice may be formed within a district.

All Éxito Escolar subactivities (professional development for teachers and school directors, pedagogic support, school networks, and parent councils) will take place only in treatment schools. Teachers and school directors at all schools in treatment districts will be invited to participate in the professional development and pedagogic support activities. A subset of 400 treatment schools will be invited to form parent councils, and 100 of the treatment schools will form school networks with neighboring primary schools. Éxito Escolar will not offer any subactivities in control schools during the course of the evaluation.

By randomly assigning districts to the Éxito Escolar subactivities, we will be able to attribute the differences observed between students or teachers in treatment and control districts to the impacts of the subactivities rather than to existing differences between schools. We will use a combination of primary and secondary data to estimate causal impacts on key outcomes of the Éxito Escolar activity over time. We will rely on survey data to gather data systematically from respondents at a sample of study schools, enabling us to make statements about all the schools, teachers, and students in the targeted study area as well as about subgroups of interest (e.g., gender, ethnic identity). The analysis will focus on understanding the differences in outcomes between the treatment and control arms of the RCT (described in detail in the discussion of the impact evaluation design in Section IV.D).

Together, the implementation study and impact evaluation will enable us to answer the types of questions listed in Table IV.1. The impact evaluation will help us determine whether the Éxito Escolar activity had the desired impact on students' and teachers' outcomes, whereas the implementation study will permit us to understand how schools implemented the Éxito Escolar activity, why it was or was not effective, and the mechanisms through which it might have generated impacts.

C. Implementation study

The implementation study will draw on (1) two rounds of qualitative data collected through interviews with key stakeholders, and (2) focus groups with project participants, including school directors, teachers, students, and parents. Before the first round of qualitative data collection, the subcontracted data collection firm, under Mathematica's oversight, will pilot test the interview and focus groups protocols for sequencing, logic, and comprehension. The pilot test will take place during the training workshop for the enumerators. The training will be held in a non-intervention Department (i.e. one of the Departments used for the pilot test of the survey instruments). We will use the results to adapt and finalize the qualitative protocols during the week-long training workshop.

The first round of qualitative data collection (October/November 2018) will capture information about initial facilitators of and barriers to implementation, along with perceptions and attitudes about the implemented activities. The second round will take place in October/November 2020, and will allow us to see whether the school directors or teachers report continued changes in their professional practice one year after the conclusion of the teacher professional development program (the group of school directors and teachers participating in focus groups will likely vary between the first and second rounds). In addition, the

implementation study will benefit from survey data (collected in early 2018 and late 2020) and test score data (collected in early 2018 and late 2020).

The implementation study will complement the impact evaluation in several ways. First, it will enhance the RCT by enabling us to explore *how*, *why*, *where*, and for *whom* the estimated changes in outcomes did or did not take place (related to research questions 4, 5, and 7). Second, the study's qualitative data may shed light on stakeholders' perceptions of the individual contributions of intervention components. It is important to note that the RCT will not permit us to estimate the impacts of individual components because the contrast between the treatment and control groups will allow us to estimate only the combined impact of all Éxito Escolar subactivities. Third, the implementation study will complement the quantitative data gathered from classroom observations by enabling us to understand the findings related to teachers' behavior change and capacity building.

1. Data sources

Assuming that implementation of Éxito Escolar starts at the beginning of the 2018 academic year (which goes from mid-January through October) and that we want the qualitative data to guide program improvements and support the analysis of quantitative data, we will conduct interviews and focus groups with potential program participants and FHI360 during both rounds of qualitative data collection in 2018 and 2020. We will conduct interviews with MCC staff and MINEDUC officials in both rounds.

With the first round of data collection in 2018, the questionnaires for school directors and teachers will yield information about teachers' and school directors' demographic characteristics and any access they have had to education and training in the past. Through classroom observations, we will measure teachers' instruction time and learn about the teaching approaches they use. We will use these results to describe the sample and to check for equivalence between the treatment and control groups at baseline. We expect intervention activities to begin shortly after baseline data collection.

The second and final round of qualitative data collection will take place late in the first year after the Éxito Escolar activity concludes. It will permit us to see to what extent and in what ways schools are continuing to implement the Éxito Escolar components after the conclusion of the intervention period.

We will develop a qualitative data collection protocol customized to each type of data collection method (interview or focus group) and respondent, but the protocols will all cover similar topics related to the research questions.

2. Study sample

We will draw a purposeful sample of school directors, teachers, students, and parents from the treatment and control groups. In Table IV.2, we summarize the data sources we will use and provide illustrative areas of focus provided by each source. Our purposeful approach will use maximum variation sampling, in which we select schools' participants according to the participants' affiliation with a high-, medium-, or low-performing school based on test scores. We will use baseline test score data to identify schools' performance levels at baseline and use

midline test score data to identify performance levels for endline qualitative data collection. The maximum variation approach enables us to identify themes that occur consistently across schools and participants regardless of variation in performance; it also enables us to identify issues that could be unique to any one group, such as high-performing schools or treatment schools (Patton 1990). Mathematica will rely on MINEDUC administrative data (including test performance) to identify thresholds for high-, middle-, and low-performing schools.

The sources include focus group discussions with the following groups (a total of 75 focus groups):

- **Teachers.** We plan to conduct 15 focus groups[6] with teachers at baseline and 21 focus groups at endline. Each focus group will average 8 to 10 participants. We will conduct 9 teacher focus groups with teachers in the treatment group (3 groups per school performance level) and 6 focus groups with teachers in the control group (2 groups per school performance level). By intentionally forming the focus groups by treatment group, we will be able to link results to treatment and control schools. We will try to include in the focus groups teachers who enrolled in the teacher professional development program, teachers who enrolled in the program but only attended infrequently, and teachers who started the program but dropped out. The range in participants' involvement will help us understand the barriers to program completion.
- **Parents.** We plan to conduct nine (six treatment and three control) focus groups with parents who participate in the parent councils in the RCT's treatment schools. We will select parent councils based on the same maximum variation sampling approach discussed above. (that is, we will conduct the parent focus groups at the same schools where we conduct the teacher focus groups). Each focus group will include all members of the parent council who are able to attend the focus group session. The focus groups will primarily explore parents' role in school management and how that role has or has not changed over time.
- **Students.** At baseline, we will conduct 15 focus groups with students in the same schools from which we draw teachers for the teacher focus groups. Each group will include 8 to 10 participants and represent grades 7 through 9. We will randomly select students in each grade from an attendance list. We will also consider separate focus groups for each grade if we determine that joint participation might affect the outcome of cross-grade focus groups. Whatever their composition, the focus groups will help us understand changes in pedagogic delivery, student perceptions of the Éxito Escolar components, and how the interventions have affected the students over time. We will conduct 21 focus groups with students at endline so we can include additional student perspectives (adding 6 focus groups with students from control schools to the 15 focus groups with students from treatment schools conducted at baseline and endline). However, based on the data we gather at baseline, the final number may be adjusted if we expect to reach saturation with fewer focus groups.

[6] Here we describe the largest sample size we are considering. Cost estimates received in the data collection procurement process will allow us to conduct all of these focus groups. However, we will reduce the number if/when we reach a saturation point with each group.

Table IV.2. Plans for qualitative data collection

Data source	Type of data	Approximate number	Illustrative areas of focus
Teachers	Focus groups	15 each at baseline and endline (treatment) and 6 at endline (control)	• Demographics and educational background • Attitudes about teacher professional development program • Facilitators of and barriers to implementation • Pedagogic support • Perceptions of training and capacity building (teacher diagnostics) • Role of parent councils • Implementation of assessments • Perceptions of school networks
Students	Focus groups	15 each at baseline and endline (treatment) and 6 at endline (control)	• Demographics • Experience in education system • Long-term goals • Perceptions and attitudes related to school networks and school management committees • Perceptions of change at school and classroom levels • Support to students (with consideration of gender and ethnic biases)
Parents	Focus groups	9 each at baseline and endline (treatment)	• Role of parent councils • Engagement and activities of parent council (i.e., action plans) • Perceptions of management and change • Role and perceptions of school networks • Perceptions of changes in school quality
School directors	Interviews Questionnaires	15 each at baseline and endline (treatment) and 6 at endline (control)	• Demographics and educational background • Attitudes about teacher professional development program • Facilitators of and barriers to implementation • Pedagogic support • Perceptions of training and capacity building (teacher diagnostics) • Role of parent councils • Implementation of assessments • Perceptions of school networks
MINEDUC officials	Interviews	20 to 25	• Background and experience in MINEDUC • Role in program • Perceptions of reform • Facilitators of and barriers to change • Perceptions of implementation of Éxito Escolar • Planning for sustainability (when relevant)

Data source	Type of data	Approximate number	Illustrative areas of focus
Implementers of teacher professional development program	Interviews	TBD	• Design of program • Teacher performance in program • Participation in program • Facilitators and barriers • Sustainability of program
FHI360	Interview	1 to 5	• Lessons from implementation • Extent to which expected outputs were achieved, and why • Efforts to establish school networks • Efforts to establish parent councils • Role of pedagogic advisors and sustainability of role • Perceived risks to achieving shorter- and longer-term outcomes • Potential for scale-up and sustainability • Successes and challenges
Interview with MCC	Interview	2	• Lessons from implementation • Extent to which expected outputs were achieved, and why • Perceptions of likely effects of project • Perceived risks to achieving shorter- and longer-term outcomes

TBD = to be determined.

In addition to the above focus group sessions, the qualitative data sources will include key informant interviews with the following stakeholders (a total of about 65 structured or semi-structured interviews):

- **School directors.** We will interview at least 15 school directors across the treatment and control groups. We will select the school directors from the same group of schools as the teacher and student focus group participants. For school directors in the treatment group, the interviews will focus on the teacher professional development program, and how they perceive participating teachers benefit from the program. If school directors are also invited to participate in the program, we will also gather data on their experiences and how schools have implemented any relevant changes. For school directors at schools selected for the school network and parent council subactivities, we will also explore their experiences with those subactivities, with attention to the facilitators and barriers to their implementation.
- **MINEDUC officials.** We will interview up to 25 MINEDUC representatives, including staff from the MINEDUC administration (that is, the minister and vice minister); the Dirección de Planificación (DIPLAN); DIGEDUCA; the Dirección General de Currículo; the Sistema Nacional de Acompañamiento Escolar, and DIGECADE: Dirección General de Gestión de Calidad Educativa. We will also interview staff from the departmental offices who work more closely with the schools. The interviews will help us understand the reform implementation process, facilitators, challenges, bottlenecks, and plans for sustainability. The interviews will gather data relevant to both Éxito Escolar and IPC.
- **Implementers of the teacher professional development.** We will interview 10 university professors hired to implement the teacher professional development program. The interviews will enable us to explore various points of view regarding facilitators and barriers to implementing the teacher professional development program, the effects of the teacher professional development on teachers' competencies, and potential barriers associated with the sustainability of the teacher professional development program.
- **FHI360.** We will interview the directors and relevant technical staff from FHI360, the implementing organization. The interviews will help us understand how Éxito Escolar's subactivities have unfolded across departments and schools and help us identify the facilitators and challenges faced by the organization during implementation. We will interview the staff at baseline, midline (including a review of the program logic model), and endline.
- **MCC staff.** We will interview key MCC staff—including those responsible for overseeing GTEP implementation, monitoring and evaluation, and gender issues—about GTEP implementation, their perceptions of successes and challenges, and their expectations of the impact of GTEP on key outcomes. We will conduct the interviews at both baseline and endline.

Finally, we will conduct a document review that focuses primarily on FHI360's reports. Such reports will provide information on aspects of implementation that went smoothly, the main changes made to implementation in response to feedback, and important lessons learned. The document review could also identify topics of possible importance for further exploration through qualitative data collection—for example, ongoing challenges to successful GTEP

implementation. Examples of documents for review include the program logic model, quarterly and annual reports from FHI360, and MINEDUC reports and policies.

3. Qualitative data analysis

We will follow four steps to analyze the data (Creswell 2009):

- **Raw data management.** Raw data management is the process of organizing raw data into formats usable for analysis (that is, from audio files to transcripts). During raw data management, we will review all data and eliminate any that are incomplete or not useful to our analysis.
- **Chunking and initial coding.** Often referred to as data reduction, chunking and initial coding will enable us to read through the interview and focus group transcripts several times and obtain a holistic view of the data. We will develop a detailed initial coding scheme—a set of themes we might encounter in the transcripts, which map to the research questions and logic model. We will also document potential themes, linkages among results, and potential findings.
- **Detailed coding.** Detailed coding will involve refinement of the coding scheme and the recoding of data as we examine the data in greater depth. We will use NVivo software to review and code the transcripts based on the initial codes developed during the chunking process. Use of NVivo software to assign codes to the qualitative data will enable us to access data on a specific topic quickly and organize information in different ways to identify themes and compile evidence supporting the themes. We will expand and refine the codes during the coding exercise and subsequent analysis of the coded transcripts in an iterative process as additional themes emerge. Further, the software will enable us to categorize respondents by gender, age, geographic location, or other salient characteristics to facilitate analysis by subgroup.
- **Data interpretation and writing.** Data interpretation and writing will require the triangulation of the findings across stakeholders to highlight mechanisms, contexts, and similarities and differences in perspectives. The baseline and final reports will use collected qualitative data to explore fully the implementation and results of the Éxito Escolar subactivities.

The qualitative data analysis will explore how Éxito Escolar subactivities were planned, how and why implementation might have varied from the original plan, major barriers and facilitators with regard to implementing the subactivities, and important lessons learned. Mathematica will also triangulate the quantitative and qualitative findings to ensure depth and understanding of the analysis. The qualitative analysis will provide context and meaning to the impact evaluation findings and will help end users understand the roles of the Éxito Escolar subactivities in improving the quality of secondary education.

D. Impact evaluation

In this section, we describe our approach to answering the research questions related to the impacts of Éxito Escolar. We also present the results of our power calculations, describe the

sample for the evaluation, and outline the plan for collecting and using primary and secondary data.

We will use an RCT to answer the research questions on the impacts listed in Table IV.1. Designing a randomized evaluation requires several significant decisions that influence the type of impacts to be detected as well as the components of the Éxito Escolar activity that we can identify a causal impact for. Based on our understanding of the planned implementation of Éxito Escolar's four subactivities, we recommended an RCT with district-level randomization. As described earlier, we randomly assigned districts to a treatment group or a control group. MINEDUC, in collaboration with FHI360, will invite all school directors and teachers in schools in districts assigned to the treatment group to participate in teacher professional development and pedagogic support. MINEDUC and FHI360 will assign a subset of 100 schools in the treatment group to a school network. MINEDUC and FHI360 will invite another (potentially overlapping) group of schools in the treatment group to form parent councils. MINEDUC and FHI360 have yet to determine how they will select the schools for networks and parent councils. We understand that implementation plans could still change; for this reason, we identify our critical assumptions and describe how we will modify our design to adapt to changes in the project design.

1. Approach to random assignment

Random assignment can create treatment groups that are similar to one another in both observable (measurable) and unobservable ways. We can then use the differences in outcomes between these groups to estimate the impact of the intervention that is offered in schools assigned to the treatment group. Observed differences between the two groups can be attributed to the intervention's causal effect because the two groups are similar in all ways other than access to the intervention. However, the success of a randomized evaluation depends critically on other characteristics of random assignment, including the number of treatment groups and the level of random assignment. In this section, we describe our design and its rationale.

Following discussions with MINEDUC and FHI360, we concluded that a two-armed RCT would be the best approach to maximize the precision of our estimates of the overall impact of the Éxito Escolar activity, while allowing flexibility in the implementation of Éxito Escolar. As noted, the Éxito Escolar activity includes four subactivities. We will evaluate the overall impact of the package of subactivities, which will include the offer of teacher and school director training and pedagogic support in all treatment schools, and the invitation to form school networks and parent councils in some treatment schools. We will not estimate the impacts of individual activities, because individual activities will not be randomly assigned. If the implementer later determines that it is feasible to assign treatment schools to parent councils or school networks randomly, we will estimate the marginal effect of offering those subactivities separately.

Based on discussions with MINEDUC and FHI360, we have determined that we should use education districts as the unit of random assignment. Districts[7] are geographically based groups of schools that include an average of eight lower-secondary schools. We recommend randomizing these groups of schools instead individual schools because the Éxito Escolar activity involves the formation of communities of practice, which are groups of schools that will collaborate and share their experiences related to teacher professional development and pedagogic support. Under our current design, communities of practice will be formed within districts. If we randomized at the school level, a control school could be surrounded by treatment schools in a community of practice that would exclude the lone control school. We fear that such an arrangement could result in contamination, or potentially lead to resentment or feelings of exclusion among control schools.

In contrast, by randomizing districts, we can be assured that treatment and control schools will also be located near neighboring schools with the same treatment status, even though control districts will border treatment districts. We recognize that randomization by district instead of by individual school will make our estimates less precise (that is, the standard errors of our estimates will increase with randomization of clusters [districts] of schools instead of individual schools, but this approach does not introduce bias); however, we consider randomization at the district level necessary to minimize the risk of contamination and feelings of exclusion and to facilitate implementation.

Across 103 districts, 816 schools were eligible to participate in the randomized evaluation. The implementer noted the need to include roughly 80 percent of teachers in the treatment group in order to reach its target number of teachers for the subactivity of teacher professional development. To accommodate this, we assigned 80 percent of districts (82 districts) to the treatment group. The treatment group includes 80.5 percent of study school teachers. We conducted random assignment at a public lottery in Guatemala. We describe the public lottery and the outcome of the random assignment in detail in Annex B. We expect that parent councils will be formed in roughly 400 treatment schools, and that 100 of the treatment schools will form a school network with neighboring primary schools. We present our random assignment design in Figure IV.1.

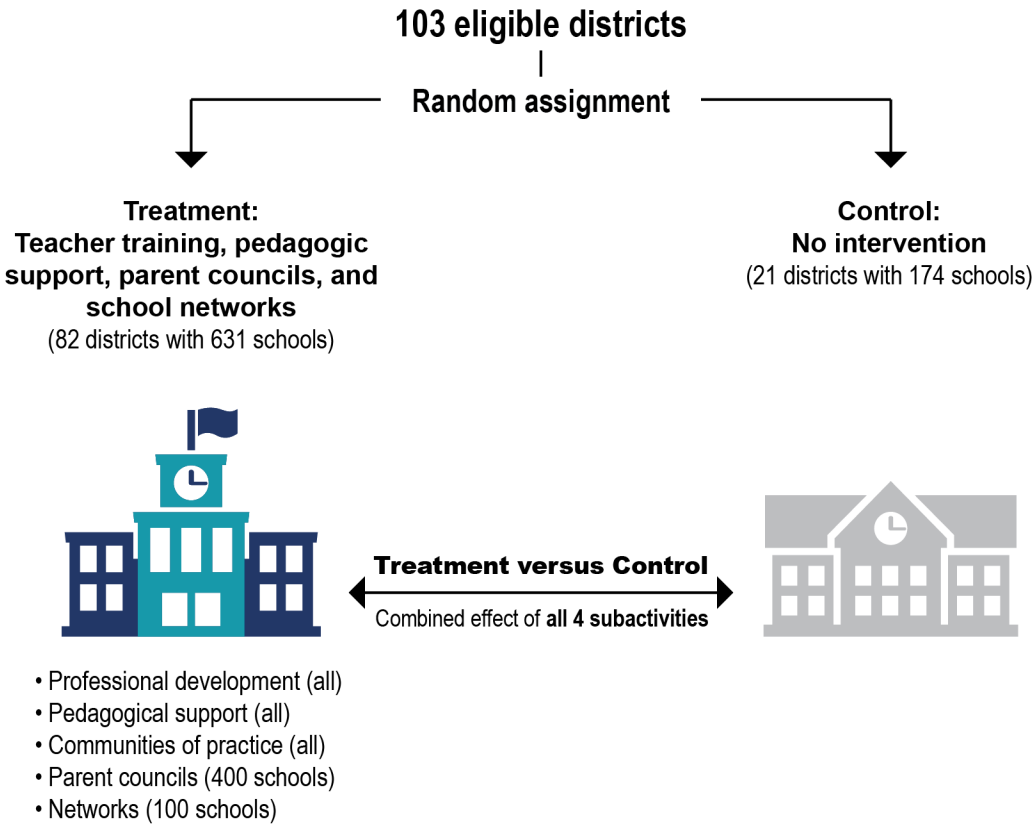
2. Stratification

We conducted stratified random assignment, stratifying districts first by department and then by 2013 standardized language and communications and mathematics scores to ensure balance on those key characteristics and to improve the precision of our estimates of impacts on learning. Within each department, we sorted schools by their 2013 test scores. For schools with no 2013 test score, we imputed scores using their municipality's mean score. To reduce the odds of assigning a disproportionate number of large districts (districts with many teachers) to the control group, or of failing to assign the implementer's minimum of 80 percent of teachers assigned to the treatment group, we reassigned some districts to different strata. These reassignments involved clustering the largest districts in strata with other large districts. This

[7] MINEDUC, FHI360, and Mathematica agree that the designation of districts as the unit of random assignment makes the most sense to avoid forming new groups of schools for the communities of practice. The districts, which MINEDUC formed, are intended to be lasting units of organization.

way, if a large district is assigned to the control group, other large districts are guaranteed to be assigned to the treatment group. We describe this in detail in Annex B.

Figure IV.1. Random assignment design



Note: All eligible schools in treatment districts will be invited to participate in teacher professional development and pedagogic support activities. Only some schools in treatment districts will be invited to join a school network or to form a parent council. Treatment effect estimates will reflect the impact of the offer of teacher professional development and pedagogic support activities and the chance of the offer to join a school network or form a parent council.

3. Quantitative data analysis

We could analyze the results of the randomized evaluation in any of several ways. We recommend estimating the intent to treat (ITT) estimate and the local average treatment effect (LATE), both of which we describe here. We recommend this combination of approaches because it includes an estimate that is generalizable to the population of public schools[8] in the study area (ITT) as well as another estimate that is closer to impacts experienced in schools that actually participate in Éxito Escolar (LATE).

[8] All lower-secondary schools except private schools, municipal schools, and a small number of specialized schools, such as agricultural training schools, are eligible. We refer to this group of schools as public schools.

- **ITT analysis.** The ITT estimate captures the impact of *being offered* a treatment, regardless of participation. It provides estimates that generalize to the full population that was invited to participate. The results are policy-relevant because they include the null effect for anyone who chooses not to participate. This allows the Government of Guatemala (and others) to understand the average impact of the program, taking into consideration the fact that some will choose not to participate. In analytic terms, we identify the estimate as the coefficient on the treatment variable in a standard regression that includes the full sample of those invited to participate in the Éxito Escolar activity as well as those in the control group regardless of their participation.
- **LATE analysis.** The LATE estimate approximates the treatment effect on the treated by capturing the impact of the treatment for those induced to participate because of their treatment assignment.[9] To estimate LATE, we will use the same sample used for the ITT analysis to estimate the impact of participation in *treatment* rather than the impact of *being invited* to participate in treatment, as we do with ITT. However, given that participation is not randomly determined, we cannot directly estimate the impact of participation. Instead, we will rely on an instrumental variables regression, using the random assignment as the instrumental variable.

We propose to estimate both the ITT, which is most policy-relevant, and the LATE, which approximates the actual impact among participants.

4. Impact estimates

Impact estimates on outcomes observed at the student or teacher level will be clustered at the district level. We will estimate the ITT effect by using regressions that include a treatment group indicator, baseline covariates, and strata fixed effects. The inclusion of baseline covariates and strata fixed effects should improve the precision of our estimates. To calculate LATE, we will include the same covariates but will rely on an instrumental variables regression, using the random assignment as an instrument for participation.

5. Precision of our estimates

We conducted power calculations to estimate the size of effects that we will estimate in our design. In Table IV.3, we present the minimum detectable effect sizes (MDES) for our design. We also outline several assumptions related to the MDES estimates. The MDES is the smallest effect size that we would be able to detect with the design. If the Éxito Escolar activity's true effect is smaller than the MDES, we do not expect to be able to detect that effect with our proposed design and sample size.

We focus on two key outcomes: (1) impacts on students' test scores, measured through the annual tests of mathematics, language and communications, and science to be administered by MINEDUC; and (2) impacts on teachers' time on task, which we will measure with the Stallings classroom observation tool. To estimate MDES for test score outcomes, we used data from

[9] LATE does not approximate the treatment effect for people whose participation is not affected by their treatment assignment: never-takers and always-takers. We identify the proportion of never-takers as those in the treatment group who do not participate and identify the proportion of always-takers as those in the control group who do participate.

MINEDUC's tests of grade 9 lower-secondary students in 2009 and 2013 and estimated intracluster correlations of test scores within and between districts and the proportion of individual- and group-level variance explained by the covariates that we expect to use in our impact regressions. We did not have data on time on task to use for estimating ICCs and variance reduction for impacts on teachers' time on task. Instead, we used ICCs based on data reported from Honduras (Bruns and Luque 2015), along with conservative estimates of variance reduction based on our experience in conducting power calculations.

In Table IV.3, in which we present two MDES for mathematics and language and communications test scores—(1) for intent to treat analysis and (2) for estimates of the local average treatment effect—the results of the power analysis show that we should be able to detect impacts as small as 0.12 to 0.13 SDs with ITT analysis, which is not adjusted for contamination or noncompliance. We assume that all students at all study schools are invited to take the endline test and that we have endline data for 75 percent of students in the sample frame. We estimate that, for the LATE analysis, which in cases of noncompliance (for example, teachers from control schools participating in intervention activities, or teachers from treatment schools not participating) inflates the impact estimate to approximate the impact for *participants*, we would be able to detect impacts as small as 0.20 to 0.22 SDs.

Table IV.3. MDES for test score outcomes (effect sizes)

	ITT estimates	LATE
Mathematics	0.12	0.20
Language and communications	0.13	0.22

Note: We calculated ICCs and variance reduction by using 2009 and 2013 student test score data for public lower-secondary schools in the five study departments. To calculate the MDES, we assume 80 percent power, a two-tailed test, and a 5 percent significance level. For mathematics (language and communications), we estimate that the covariates we plan to include in the impact regression model will explain 4 (6) percent of the individual-level variance in the outcome variable and 73 (84) percent of the group-level variance in the outcome variable. The ICC for students within districts is 0.08 (0.19). The results pertain to students from one grade. We assume that we will have test score data for 75 percent of students in sample schools. The MDES shown in the ITT column is the MDES for analysis that does not adjust for contamination. The MDES shown in the LATE column is the MDES for analysis that does adjust for contamination, such as the LATE analysis. For this, we assume that the schools of 70 percent of students in treatment schools receive their intended treatment, while the schools of 10 percent of students in control schools access treatment.

In Table IV.4, we present MDEs for impacts on teachers' time on task. Rather than presenting effect sizes, we present changes in the percent of time on task. We will base the estimates on classroom observation data, which we will collect in the study schools.[10] Our estimates are grounded in the assumption that we collect data in all schools or in subsets of schools of two different sizes. If we collect data in a subset of schools, we will ensure representation of all districts. We find that, with the ITT analysis, we would be able to detect effects as small as a 10 to 11 percentage point change in class time on task. With the LATE analysis, we would be able to detect effects as small as a 17 to 19 percentage point change in

[10] The final number of schools for the classroom observation sample will be based on the cost estimates received from the data collection firm. We expect to receive these costs in December 2018. This includes traditional public schools (INEB), telesecundaria, PEMEM.

class time on task. According to these calculations, the change in MDE is small even for a large change in sample size; we found almost no difference in MDE between observing teachers at all 800 study schools and observing teachers at a sample of 480 schools, and the MDE grew by only 1 to 2 percentage points with a sample of 160 schools. With all three sample sizes shown, we assume representation from all 101 districts. The small change in MDE after a large change in sample size reflects the fact that, with clustered randomization, statistical precision is more a function of the number of clusters (in this case, districts) than of total sample size (in this case, teachers). The calculations suggest no need to observe teachers at all 800 study schools to improve the statistical precision of our estimates.

Table IV.4. MDEs for teacher time on task (change in percentage of teaching time on task)

Number of schools	Total number of teachers observed	ITT	LATE
Three teachers each at 800 schools	2,100	10	17
Three teachers each at 480 schools	1,260	10	17
Three teachers each at 160 schools	420	11	19

Note: To estimate ICCs, we used data reported in Bruns and Luque (2015) on variance properties of time-on-task data collected in Honduras (the country most similar to Guatemala among those reported). We assumed an ICC of 0.4 of teachers within districts. To calculate the MDEs, we assume 80 percent power, a two-tailed test, and a 5 percent significance level. The MDEs shown in the ITT column are the MDEs for analysis that does not adjust for contamination. The MDE shown in the LATE column is the MDE for analysis that does adjust for contamination. For this, we assume that 70 percent of treatment schools receive their intended treatment, while 10 percent of control schools access treatment.

The detectable effect sizes are in the range of the effect sizes that we expect for the types of subactivities composing Éxito Escolar. In Table IV.5, we summarize similar effect sizes from the literature.

Table IV.5. Effect size comparison

Literature review of similar interventions	Effect sizes
Carr-Hill et al. (2015) review of 5 SBM studies' impacts on test scores	0.21 SDs
McEwan (2015) meta-analysis of 5 experimental evaluations of school management interventions' impacts on student learning	0.06 SDs
McEwan (2015) meta-analysis of 11 randomized studies of teacher professional development interventions' impacts on student learning	0.12 SDs
Piper and Korda (2010) rigorous evaluation of teachers scripted lesson plans' impacts on early-grade literacy outcomes	0.39 to 1.23 SDs
Conn (2014) review of 4 rigorous evaluations of teacher professional development interventions' impacts on student learning	0.25 SDs
Bruns and Luque (2015) review of time on task in Mexico (controls for prior test score and teacher characteristic, but not a causal effect)	0.14 to 0.16 (Spanish) 0.13 to 0.16 (mathematics)
Hanushek and Rivkin (2010) evaluation of teachers' relative effectiveness on student learning	0.11 SDs (mathematics)
Rothstein (2010) evaluation of teachers' relative effectiveness on student learning in North Carolina.	0.11 SDs (reading) 0.15 SDs (mathematics)

Literature review of similar interventions	Effect sizes
Kane, Rockoff, and Staiger (2008) evaluation of teachers' relative effectiveness on student learning in New York City	0.08 SDs (reading) 0.11 SDs (mathematics)

6. Gender and social inclusion analyses

Guatemala has made important progress in improving access to education for traditionally disadvantaged populations, such as indigenous Guatemalans, Guatemalans living in rural areas, and girls. However, these groups still face challenges in obtaining high quality education, particularly at the secondary level (Asturias de Barrios 2014). The subactivities' effects are likely to vary by group. For example, the teacher professional development program could have greater impacts in traditionally disadvantaged areas and populations if the program's teachers received additional training. Parent councils' action plans might have greater impacts on girls than on boys if the councils can identify previously unresolved challenges for female students. In our evaluations of all three GTEP activities, we will review outcomes for subgroups by students' ethnic identification and gender. For our quantitative analysis, we will analyze how impacts differ among these subgroups. For subgroups that are large enough, we will estimate impacts by subgroup.

Our ability to interpret results for specific subgroups clearly depends on how schools are selected for parent councils and networks. If these subactivities, which will be offered to only some treatment schools, are correlated with subgroup status, like urban or rural location, we will not be able to differentiate between the variation in impacts by location and the increased likelihood of receiving additional subactivities in some locations.

7. Time frame of exposure

Éxito Escolar's subactivities rely on knowledge transfer—to school directors, teachers, and parents—to improve student learning. The transfer of knowledge to school directors, teachers, and parents will take time. Similarly, when school directors, teachers, and parents have absorbed their new knowledge and skills, additional time will have to elapse before any impacts on student learning become evident. School directors and teachers will begin their training in early 2018 and continue to participate in training through the 2019 school year. We will repeat data collection in 2020 to gather data after two full years of implementation. We will evaluate impacts on student learning at the end of the second and third years of the intervention, providing additional time for school directors, teachers, and parents to apply what they have learned through Éxito Escolar's subactivities.

8. Study sample

FHI360 will implement the Éxito Escolar's subactivities in approximately 640 public lower-secondary schools in the five study departments. We will use MINEDUC's administrative data on the public lower-secondary schools in the five study departments as our sample frame. Until we determine data collection costs, we cannot specify the number of schools from which we will collect survey and classroom observation data. Once we determine costs, we will finalize our sample size. We have asked bidders to submit estimates for collecting data in 800, 480, or 160 schools, evenly distributed across the five departments.

E. Impact evaluation outcomes

In this section, we describe the outcomes we will use to measure the impact of Éxito Escolar and the data sources we will use to carry out the analysis. For the analysis, we will draw on both primary and secondary data sources from all five departments targeted for Éxito Escolar. The sources include monitoring data gathered by FHI360 in schools participating in Éxito Escolar (treatment schools), questionnaires that Mathematica will administer with our data collection partner, teacher and student assessments, classroom observations, and administrative data gathered by MINEDUC in treatment and control schools.

We will use teacher surveys and classroom observations to measure impacts on teachers' knowledge and behavior. We will also review the results of tests administered to teachers before and after participating in teacher professional development for the treatment group to determine knowledge gains among teachers participating in teacher professional development. We will rely on MINEDUC's administrative data on enrollment to measure impacts on students' promotion, retention, and dropout rates. We will use scores from national tests administered by MINEDUC to measure students' academic performance. We will gather data to estimate impacts on student learning and teacher behavior at the end of the 2020 school year. This timing allows one full year after the professional development subactivity has ended, and is the latest point in time we can gather endline data within our contract. Table IV.6, we summarize the data sources that we will use to answer each of the impact evaluation research questions.

Table IV.6. Impact evaluation research questions, outcomes, and data sources

Impact evaluation questions	Outcomes	Data sources
20. Did teachers adopt new pedagogical approaches as a result of their training? This could include more active learning, more attention paid to different learning styles of students, adaption of lessons for language minorities, and equal attention paid to both genders and students of all socio-economic backgrounds	Teacher pedagogic approaches	Classroom observations
21. Did time devoted to learning in the classroom increase as a result of the intervention?	Teacher time on task	Classroom observations
22. How does time devoted to learning in the classroom vary across different teaching subjects?	Teacher time on task	Classroom observations
23. How did Exito Escolar affect student enrollment in secondary schools (promotion rates, retention rates, dropout rates)?	Student enrollment status by year, unique ID	MINEDUC enrollment data
24. How were student learning outcomes affected by Exito Escolar?	Student academic performance	DIGEDUCA test score data
25. How do changes in student learning outcomes vary across gender, socio-economic, and language groups?	Student academic performance, sex, SES, and language group	DIGEDUCA test score data and associated factors data

Note: Question numbers align with numbering from Table IV.1.

The table lists separate questions about specific components of Éxito Escolar and therefore includes more research questions than those listed in Table IV.1.

- **Teachers' knowledge.** FHI360 will administer the teachers' knowledge tests at the beginning and end of the teacher professional development program. We do not consider it feasible to administer the same tests to teachers (in treatment or control schools) who are not participating in the teacher professional development program. Nonetheless, we will evaluate the feasibility of incorporating questions to gauge teachers' knowledge in the teacher questionnaire, which we will administer to teachers in treatment and control schools. We will also analyze the results of the pre-post tests for teachers in the teacher professional development program to determine the effect on teachers who participated for the duration of the program.
- **Time on task.** We will use classroom observations to gather data on how teachers spend their class time. Among other outcomes, the classroom observations will focus on time spent on task (including active instruction time, time for group work, or time for individual work). We will ensure uniformity across observations by conducting all observations for the same amount of time.
- **Pedagogic approaches.** We will use teacher surveys and classroom observations to gather information on teachers' professional background and pedagogic approaches. The teacher surveys will include general and specific questions about the pedagogic approaches used by teachers. We will code classroom observations to identify specific teaching techniques.
- **Students' academic performance.** We will use students' scores from tests administered by MINEDUC to determine academic performance. At a minimum, we will use scores from mathematics, language and communications, and science tests. We will also evaluate impacts on other subjects if MINEDUC tests on other subjects during the evaluation time period.
- **Students' enrollment, promotion, retention, and dropout rates.** MINEDUC gathers data on student enrollment at the beginning and end of each academic year, including data on whether each student is repeating a grade. MINEDUC uses the data to calculate students' promotion, retention, and dropout rates. Below, we describe how we plan to calculate each of these indicators. We will review our approaches with MINEDUC and MCC to be sure that our approaches capture the information of interest.
    - **Enrollment.** We will review MINEDUC's data on enrollment in each year of lower-secondary school and evaluate the annual changes in enrollment at the school level. We expect enrollment to increase as the quality of education at the lower-secondary level improves. Specifically, enrollment may increase for two reasons: first, primary school graduates may be more likely to go on to secondary school if their primary school is part of a network; and second, improved quality in lower-secondary school may decrease dropout rates, thereby increasing enrollment. We will take care to examine enrollment of new students to avoid reporting reductions in grade repetition as reductions in enrollment. We will also gather enrollment data from schools as a way to verify MINEDUC's administrative data.
    - **Promotion.** We will calculate promotion as the percentage of students who began a grade and are promoted to the next grade at the end of the academic year.

- **Retention.** We will calculate retention as the percentage of students enrolled at the beginning of one school year who are enrolled the following school year, including students repeating a grade.
- **Dropout rates.** We will calculate dropout rates in two ways. The first approach measures the percentage of students who drop out of school during the school year by dividing the number of students who no longer attend school at the end of the school year by the total number of students enrolled in school at the beginning of the school year. The second approach includes students who may drop out from one year to the next by dividing the number of students who are no longer enrolled at the beginning of one school year by the number of students who were enrolled at the beginning of the previous school year.

F. Data collection approach

We will gather the data needed to conduct our analysis through a combination of primary data collection, reliance on the monitoring and evaluation data collected by FHI360, and use of the administrative data collected by MINEDUC on an ongoing basis. We summarize our approach in Table IV.7.

Table IV.7. Éxito Escolar: data collection approach

	2018				2019				2020			
	January–March	April–June	July–September	October–December	January–March	April–June	July–September	October–December	January–March	April–June	July–September	October–December
Éxito Escolar Implementation												
Treatment schools	x	x	x	x	x	x	x	x	x	x	x	x
Teacher training		x	x	x	x	x	x	x				
Data collection												
School director survey (Mathematica)	BL[S]	BL[S]										FU2[S]
Teacher survey (Mathematica)	BL[S]	BL[S]										FU2[S]
Teacher knowledge test (FHI360)	BL[T]						FU1[T]					
Classroom observations (Mathematica)	BL[S]	BL[S]										FU2[S]
Qualitative data collection (Mathematica)				QD[Q]								QD[Q]
Student national examinations (MINEDUC)		BL[A]										FU[A]
Monitoring data (FHI360)	x[T]	x[T]	x[T]	x[T]	x[T]	x[T]	x[T]	x[T]	x[T]	x[T]	x[T]	x[T]

Note: Parties responsible for data collection shown in parentheses.

[A]All students in study schools.

[Q]Subset of schools selected for qualitative data collection, including treatment and control schools.

[S]Survey sample. Number of schools included in survey sample will depend on data collection cost estimates, which we have not yet received. Regardless of the number of schools, the survey sample will encompass each included school's school director and three teachers (or fewer if the school has only one or two teachers).

[T]All treatment schools.

BL = Baseline; FU1 = Follow-up 1; FU2 = Follow-up 2; QD = Qualitative data collection.

1. Primary data collection

Primary data collection will begin with the process of field testing the quantitative instruments in February 2018. Following the pilot, we will revise the instruments and plan to collect baseline data in late March to early April 2018. We will use a combination of questionnaires, classroom observations, standardized assessments, and monitoring and evaluation indicators to measure key outcomes for the impact and implementation evaluations. In addition, we will measure selected characteristics of participants, and their contexts, any of which could influence responsiveness to Éxito Escolar, fidelity of implementation, and impacts on teachers and students.

In this section, we first focus on the quantitative instruments that Mathematica will develop and administer and then address the timing of data collection, data collection staff qualifications, and quality assurance protocols. Next, we provide a brief overview of data that the FHI360, DIGEDUCA, and DIPLAN will collect and that we will use as secondary data sources to answer key research questions.

- **School director and teacher questionnaires.** We will administer school director and teacher surveys at baseline and endline. Before it administers the baseline surveys, the subcontracted data collection firm, with Mathematica's oversight, will pilot the school director and teacher questionnaires to test their internal consistency: sequencing, flow, and comprehension of questions. The pilot will be conducted in 25 schools from five departments that neighbor the study departments. The pilot departments and schools are the same ones that DIGEDUCA used to pilot-test the student assessment instruments. We will use the results from the pilot test to adapt and improve the internal consistency of the questionnaires before we field them. We will administer the baseline survey in late March to early April of 2018, before the program implementation activities begin. The endline survey will take place about 11 months after the conclusion of the Éxito Escolar activities in 2020. Baseline surveys will be designed to gauge readiness for change, access to support services, and other characteristics that could influence school directors' and teachers' take-up of Éxito Escolar's subactivities and thus key outcomes. Endline data collection will enable us to evaluate lasting impacts after the conclusion of the teacher professional development program. Baseline and endline surveys will include the following modules:
    - **Demographic characteristics and qualifications.** School directors and teachers will report on their gender, age, and ethnic group; languages spoken; and level of education and years of experience in the education field.
    - **Professional development.** We will ask school directors and teachers about their recent professional development training activities on topics covered by Éxito Escolar, such as language and communications, mathematics, natural sciences, leadership, and administration.
    - **Work conditions.** We will ask school directors and teachers about their work load, including the number of hours spent on administrative, teaching, and professional development activities. Teachers will report on the grades and subjects they teach and the number of students per classroom. Work conditions are likely to affect school directors' and teachers' willingness and ability to complete the training program and implement the new pedagogic techniques.

- **Interest and self-efficacy.** At baseline, we will ask school directors and teachers about their interest in participating in Éxito Escolar's subactivities and their perceived self-efficacy for completing them. Self-efficacy has been shown to predict the implementation of new practices as well as persistence in the face of difficulties (Bandura 2006).
- **Pedagogic practices.** We will ask teachers about their classroom practices and beliefs about best practices in teaching.
- **Pedagogic support.** School directors and teachers will report on the level of pedagogic support received in the past 12 months.

At endline, we will include questions to gauge the fidelity of implementation. The questions will address participation in teacher professional development activities, the frequency of meetings with pedagogic advisors, meetings of the parent councils, and activities related to school networks. We will keep these questions relatively simple because teachers may not be able to remember specific details about activities from the previous academic year. Monitoring and evaluation data collected by FHI360 will complement information about the frequency of meetings (or dosage). In addition, we will ask questions to identify potential barriers to implementation. In-depth focus groups with a select sample of school directors and teachers will round out the information collected from teachers.

- **Classroom observations.** During both rounds of data collection, independent observers will complete structured classroom observations by using an adaptation of the Stallings Classroom Snapshot (SCS) (Stallings and Mohlman 1988). The SCS measures teachers' use of instructional time and materials, core pedagogic abilities, and the ability to engage students. It uses categories that are easy to understand and that have been shown to produce high inter-rater reliability among observers with limited training, making the tool well suited for data collection with large-scale samples in developing countries (Jukes et al. 2006). In fact, more than seven LAC countries have recently used the SCS, and some countries have decided to mainstream periodic SCS observations into their regular school supervision systems (Bruns and Luque 2015). We will make needed adaptations to the observation protocol in consultation with local education experts and before finalizing it, we will pilot the observation protocol in the same 25 schools located outside of the study departments.
- **Rounds and timing.** We will conduct two rounds of data collection in the treatment and control groups. We will conduct baseline data collection in March/April 2018, the first year of implementation. We expect that baseline data collection will take place before implementation begins, or early enough in the implementation period that implementation will not have affected outcomes.

  We will conduct follow-up data collection in the 2020 school year. In particular, we plan to conduct follow-up survey data collection and qualitative data collection in the final months of the 2020 school year. Students' tests will be administered at about the same time as the survey and qualitative data collection in each year.
- **Respondents.** We will survey all consenting school directors at participating schools in both rounds. We will sample teachers at each school, ensuring representation of all three grades and coverage of mathematics, language and communications, and science instruction.

2. Secondary data

We will take advantage of Guatemala's rich administrative data sources and the valuable data that FHI360 will collect as part of the implementation.

- **Administrative data.** We will use data collected by DIPLAN to measure key school characteristics and student enrollment data. Important school-level characteristics include schools' modality (that is, national basic education institutes, cooperative, telesecondary, and so on); sector (public or municipal); number of teachers (total and disaggregated by grade and gender); teacher and student mobility; student enrollment (total and disaggregated by grade and gender); teachers' and students' absenteeism; and presence of an active parent council. In addition, information on infrastructure conditions—including the number of classrooms in use, working toilets, access to potable water, and availability of facilities such as libraries, computers, and laboratories—would be useful for characterizing students' learning environments. These data could be collected as part of school director surveys in the event that the data are not available from DIPLAN.

  We will use student data collected by DIPLAN to assess the following key outcomes of the evaluation: students' transition from primary to lower-secondary school, transition from lower- to upper-secondary school, and retention and dropout rates in lower-secondary school. We will also use students' demographic data to examine variation in these outcomes as a function of students' grade, gender, race, and ethnicity.

  To assess the quality of the administrative data, Mathematica will apply internal data quality review processes to all its administrative data sets. We will obtain copies of the data collection forms to improve our understanding of the data sets provided and check the data for consistency, including checking for outliers and missing values. We will ask MINEDUC to verify suspicious data points. We may not be able to correct inaccurate data, but we will attempt to identify them, and will describe any limitations that result from data issues when we report on our results.

- **Student national assessments.** We will use scores from the standardized national assessments administered by DIGEDUCA in seventh grade at baseline, and seventh and ninth grades at follow-up, to measure the main outcomes of the evaluation: students' mathematics, language and communications, and science performance.

- **Teachers' knowledge test.** At the beginning and end of the intervention, FHI360 will design and administer a knowledge test to all school directors and teachers in treatment schools who express interest in enrolling in the teacher professional development program. The knowledge test will assess performance in mathematics, language and communication, and science; knowledge and implementation of the national basic curriculum (Curriculo Nacional Base, in Spanish); and pedagogic competencies. The test is intended to identify the areas where teachers need the most support. We will use teachers' scores on the knowledge test to analyze changes over time in teachers' knowledge, although without data from the control group, we cannot attribute any changes directly to the intervention. As described above, Mathematica will apply its internal data quality assurance process to check and verify the reliability, accuracy, and validity of the data provided by FHI360. If we discover deficiencies in the data, we will not be able to correct them, but we will describe limitations to the analysis that may result from them when reporting our results.

- **Monitoring and evaluation data.** The implementation evaluation will take advantage of the indicators collected by FHI360 for each of Éxito Escolar's subactivities: the number of school networks, communities of practice, parent councils, and action plans established; the number of school days completed; and so on. The full list of indicators appears in Annex C.

G. Challenges

1. Limitations of interpretations of results

The main limitation of a randomized evaluation is that the quantitative estimates do not explain how or why an impact is or is not achieved. Two strategies, however, will minimize the impact of this limitation. First, we will explore the results of our qualitative data collection to understand the nuances of stakeholders' implementation experiences. Second, our surveys will ask questions about implementation, enabling us to understand implementation across all study schools and to investigate whether issues identified in focus groups and interviews persist across the sample.

Nevertheless, the interpretation of our impact estimates will pose a challenge. Only some treatment schools will be invited to join a school network or to form a parent council, and these schools will not be randomly selected. Therefore, we will interpret impact estimates as the impact of both the offer of teacher professional development and pedagogic support and of the opportunity to join a school network or form a parent council. We recognize that the unique combination of interventions and the fact that only some treatment schools have the opportunity to join a network or form a parent council may limit the value of the research to others interested in the results of the study.

2. Risks to the study

Low take-up. The success of any randomized evaluation depends on project participants' adherence to their treatment assignment. Low take-up, or low rates of participation in the treatment group, will diminish our ability to detect impacts because the expected average impacts will be reduced to the extent that members of the treatment group do not participate in the treatment. In particular, the teacher professional development program is time-intensive, requiring 1,800 hours of teachers' time over two years. Despite the incentive of the PEM credential offered through the teacher professional development subactivity, teachers may find it difficult to remain in the training program in the face of competing professional and personal commitments. To the extent that teachers stop participating in the teacher professional development program, potential impacts will decline. In our power calculations, we presented estimates of the precision of ITT estimates, which are not adjusted for non-compliance to treatment assignment (which includes low take-up), and estimates of the precision of LATE estimates, which are adjusted for non-compliance. For the LATE estimates, we assume that 70 percent of treatment schools participate in Éxito Escolar's subactivities, and 10 percent of control schools gain access to them. LATE estimates will be sensitive to how we define whether a student's school is participating or not, because each student is likely to have some teachers who participate and others who do not. We will estimate and report conservative and liberal definitions of participation.

Crossover and contamination. A related risk occurs if school directors, teachers, or parents gain access to intervention activities that do not correspond to their treatment assignment. As with low take-up, crossover or contamination diminishes the contrast between groups and decreases our ability to detect impacts. If school directors and teachers from control schools access the teacher professional development or pedagogic support program, they will compromise our ability to measure the impacts of both teacher professional development and pedagogic support. Such a possibility makes it essential for the implementer to establish a system that restricts access to Éxito Escolar exclusively to treatment schools. Maintaining control over which schools gain access to treatment is more challenging to the extent that there are larger numbers of gatekeepers—individuals determining who gains access—and to the extent that those individuals do not understand the reasons for maintaining treatment assignments. Access control is complicated further if the gatekeepers are personal friends, neighbors, or coworkers of those meant to be excluded from treatment. We will encourage the implementer to identify which individuals will support the evaluation and maintain its integrity. We will coordinate with the implementer to invite as many stakeholders as possible to a half-day session on evaluation design that we will host to ensure stakeholders understand the design. We will also work with the implementer to monitor access to Éxito Escolar subactivities.

It seems unlikely that school directors or teachers would organize school networks or parent councils on their own; however, local government offices might be inspired by Éxito Escolar to offer similar activities in control schools. The implementer should communicate regularly with local government authorities to be sure that local education authorities understand the evaluation's experimental design and do not sponsor activities that might jeopardize it.

During the life of the project, some school directors and teachers will likely change schools. We expect that, in some cases, teachers will move from one treatment group to another (either from treatment to control, or vice versa). Given that 80 percent of districts were assigned to the treatment group, school directors or teachers leaving control schools will be more likely to transfer to a treatment school than to leave a treatment schools and transfer to a control school. We cannot control school director or teacher transfers. However, we will analyze MINEDUC's school director and teacher data to report how many school directors or teachers move from one treatment group to another, and we will consider the rate when interpreting our analysis.

Spillovers. In addition to contamination and crossover, spillover effects could occur, diminishing our ability to detect impacts if school directors, teachers, or parents who participate in treatment activities share information or materials from those activities with school directors, teachers, or parents in control schools. However, we have found such spillover effects to be small. Although teachers might tell other teachers about what they are learning, we do not expect the impact of such conversations to be comparable to the impact of direct participation. These conversations could, however, inspire others to try to gain access to Éxito Escolar's subactivities. If they do, we will rely on effective gatekeepers to ensure that participants follow their treatment assignment.

Resentment over treatment assignment. School directors and teachers may learn that other schools in the same district are being offered different services. If school directors and teachers are accustomed to the offer of uniform services, the offer of different services could become a source of tension. We will encourage the implementer to be forthright in describing the

Éxito Escolar activities to school directors and teachers, stating that capacity constraints limit the participation of all schools. If MINEDUC plans to expand Éxito Escolar's subactivities to all schools in the study departments at the conclusion of the study, and is confident that those plans will not change, the implementer might communicate such information. We anticipate that the randomization of districts instead of individual schools may alleviate any potential conflict between treatment and control schools.

Delayed implementation. The implementer is charged with designing the details of the Éxito Escolar's subactivities and hiring staff in five departments to implement them; both tasks are significant undertakings that may be time-consuming. If the implementer is unable to begin intervention activities on time, we might have to modify our evaluation timeline. If it is not feasible to modify the evaluation timeline, we will evaluate final impacts after schools have had less exposure to the intervention activities, potentially attenuating program impacts.

V. EVALUATION DESIGN FOR ACTIVITY 3: STRENGTHENING INSTITUTIONAL AND PLANNING CAPACITY (IPC)

Activity 3 (IPC) of the GTEP focuses on improving MINEDUC's institutional capacity to plan and budget so that it can provide an equitable and high quality secondary education (MCC 2016). Section V of the EDR presents the design for the performance evaluation of IPC. This performance evaluation will assess how well the project effectively and efficiently delivered the planned activities. It will also generate lessons learned on how the project's strategies and operational decisions improved or did not improve MINEDUC's institutional capacity (MCC 2016). Section V begins with a review of the literature on institutional change and the components that often make up education reform. Subsection B presents an overview of the design and evaluation questions for IPC. Subsection C discusses our methodology and plans for data analysis.

A. Literature review

1. Institutional change and scaling up reform efforts

Well-targeted and well-timed technical and policy support are critical to the implementation of major reforms and require a sharp focus on a limited set of core system functions (Gillies 2010). Indeed, such a focus is what leads to the system-level reforms that improve student learning (DeStefano and Crouch 2006). However, when programs either extend the reach of education interventions to increasing numbers of communities or scale up, what may be considered reasonable gains? What interventions will help close the learning gap, with what intensity, for how long, and in what types of environments? What institutional changes are needed to support the changes over the long term?

Designing and implementing reforms that create change in thousands or even millions of classrooms in which teachers might lack proper skills is an enormous challenge (Bold et al. 2013; Glennan et al. 2004; Management Systems International [MSI] 2012; Thompson and Wiliam 2007). The difficulties of moving from pilot educational interventions to scale include the sheer number of classrooms; the complex system in which these classrooms often function (that is, different environments even within the same country); the "black box" or closed nature of classrooms; and the type of pedagogic support that teachers receive on a regular basis (Bold et al. 2013; Glennan et al. 2004; MSI 2012; Thompson and Wiliam 2007). Thompson and Wiliam (2007) argue that in theory, for interventions to demonstrate impacts on student learning, every teacher must understand the reform and carry out all aspects of the intervention as prescribed. If teachers cannot deliver the interventions at a high level of quality at the appropriate intensity, the implementation effort will fail (Ibid). Moreover, interventions may work well and be based on solid theories of action, but the enabling environment might undermine an intervention. In addition, interventions may be difficult to sustain among large numbers of teachers and students (Bold et al. 2013; Glennan et al. 2004; MSI 2012; Thompson and Wiliam 2007).

Institutional change as it relates to education reform is also a call for governments to come to grips with the reality of their resource constraints. If governments are committed to changes at scale, they will have to find the resources to support institutional change. For example, in the GTEP, MCC will support the addition of 40 pedagogic advisors; however, if MINEDUC cannot absorb and sustain the support provided by these advisors after the GTEP's conclusion, teachers'

performance might decline and thus affect student learning. Even if the political will for scale-up is great, it does not guarantee that the same extends to securing adequate budgets.

An *enabling environment* is a set of interrelated conditions (political, institutional, technical, and cultural) that affect the capacity of actors (teachers, parents, communities, donors, and governments) to engage in development processes in a sustained and effective manner. Frequently, the challenge in implementing education reform is that a country's situation is not accidental. Instead, well-entrenched interest groups often want to ensure that the institutional arrangements from which they benefit directly do not change (Conn 2014; DeStefano and Crouch 2006). Interest groups typically see proposed school reforms or pilot programs as affecting only a small part of the system and therefore do not feel threatened by them. However, when changes (such as policies or the scale-up of interventions) begin to force a shift in resource allocations, interest groups that might not benefit from the change respond, often aggressively, and erect barriers or bottlenecks to change (DeStefano and Crouch 2006). Successful reform requires strong leaders skilled and experienced in conducting political battles and negotiating needed trade-offs. In some cases, a favorable enabling environment may already exist, in which case a project will immediately benefit. In other cases, however, a project must create its own enabling environment to ensure that an intervention achieves its desired impact on its target population.

The enabling environment extends to institutional capacity. The people who are trained to support institutional changes must remain in the system. However, in developing countries, we often find that strong leaders and skilled managers leave government service to seek higher-paid employment elsewhere as they acquire marketable skills (Gillies 2010). The success of the IPC activity under the GTEP will hinge, in part, on the ability of the education system to retain both teachers and MINEDUC personnel.

2. Teacher recruitment strategies

Bruns and Luque (2015) present the latest evidence on teacher recruitment within and outside LAC countries. Ample evidence shows that the recruitment of top talent into teaching is a critical factor in improving the quality of education systems. In addition, efforts to attract top talent not only require the offer of competitive salaries but also depend on a selective teacher recruitment system. Bruns and Luque (2015) draw on global evidence to identify three important approaches for increasing selectivity in recruitment. One approach is to raise the entry standards for teacher education. Education ministries in Latin America, which are limited by the principle of university autonomy, can raise entry standards by adopting the following strategies: (1) closing low quality teacher preparation schools overseen by the education ministry; (2) establishing a national teacher university that is controlled by the education ministry; (3) offering special scholarships to top students; and/or (4) requiring autonomous tertiary institutions to meet higher accreditation standards. To raise the quality of teacher education schools, education ministries can adopt competitive funding programs to incentivize schools to reform and thus raise the quality of teacher education. Finally, ministries can raise teacher hiring standards through the establishment of national teacher standards, pre-employment examinations of teachers' skills and competencies, and/or the recruitment of teachers trained in other disciplines (alternative certification). For example, rigorous evidence from Mexico supports the use of certification examinations; students taught by a test-hired teacher scored 0.78 SDs higher in

language and 0.66 SDs higher in mathematics than students taught by a traditionally hired teacher (Estrada 2013).

Guarino et al. (2006) reviewed the U.S. literature on policies that promote teacher recruitment. They looked at a large number of studies on compensation policies, which consistently find that higher salaries are associated with lower teacher attrition. In addition, the in-service policy studies reviewed by Guarino et al. (2006) show that schools with mentoring and induction programs exhibit lower turnover rates among beginning teachers than their counterpart schools that do not offer such programs.

3. Strategies to increase funds for secondary education

Developing countries may turn to a variety of strategies to increase funds for education. Lewin and Caillods (2001) examined several case studies in African, Asian, and Latin American countries and suggested some strategies for increasing funds for education. The first strategy calls for looking for additional funding for education among existing public funds: expanding the education budget as a percentage of government expenditure, increasing spending on secondary education as a percentage of government educational expenditure, and potentially requiring decentralized authorities to finance some of the costs of education. A second strategy stretches existing budgets by reducing costs: unit costs of secondary schools may be reduced by increasing the proportion of teaching assistants, reducing non-salary costs, or adopting a core curriculum with limited options at the lower-secondary level. In a third strategy, Lewin and Caillods (2001) suggest reducing costs by increasing efficiency in secondary schools by reducing dropout and repetition, eliminating ghost workers on the salary payroll, and providing incentives to increase the efficiency of school management.

4. Increases in education budget and student learning

Financial resources represent an important influence on the level of student learning outcomes. Bruneforth et al. (2004) looked at the average amount of money that countries (including five Latin American countries) spend per student from the beginning of primary education until age 15 and compared that amount with mean student performance in language and communications, mathematics, and science literacy. They found that, as per student expenditures on education increase, the national mean performance in literacy achievement also increases. However, they also reported that spending alone was not sufficient to reach high levels of achievement.

B. Evaluation overview and research questions for IPC

IPC comprises interrelated activities intended to strengthen MINEDUC's institutional capacity to provide a high quality and equitable lower-secondary education (MCC 2016), including support to the broader institutional strengthening process envisioned in the Plan for the Transformation of Secondary Education.

We will use a mixed-methods approach to conduct the performance evaluation, which has two main components. The first is a **trend analysis**, for which we will use secondary data to examine changes over time in key outcomes related to the project (for example, changes in budget allocations to secondary education; changes in teacher hiring and retention data). We will draw on data from all departments in Guatemala (IPC is national in scope) and will rely on

project monitoring data from the MCC monitoring and evaluation (M&E) plan. The second component is a **qualitative component**, for which we will draw on two rounds of key informant interviews with stakeholders and two rounds of focus groups with program participants such as teachers, school directors, and university faculty. The interviewees will be national staff, departmental staff, and FHI360 staff. The focus group participants will be teachers, school directors, and parents from the five target departments where Éxito Escolar will take place. We will include questions relevant to IPC as part of the baseline and endline data collection process discussed under Éxito Escolar. We will also conduct a midline data collection round of key informant interviews and focus groups so we can understand whether institutional changes are taking place and provide that feedback to MINEDUC and FHI360.

For the evaluation of IPC, we will document each planned component, highlighting completed activities and any trends that can be identified through administrative data. We will also compare products developed under IPC against the GTEP implementation commitments and will assess to what extent IPC's overall performance in strengthening MINEDUC's capacity to improve effectiveness, efficiency, and equity (including gender, ethnic, and socioeconomic concerns) in lower-secondary education may contribute to improvements in learning. To understand the contributions of the institutional strengthening components to changes in student learning, we will employ a Drivers of Change (DOC) framework during the data analysis phase. The DOC framework is an analytic framework for applying Political Economy Analysis (PEA) that enables evaluators to systematically assess how project design and implementation decisions addressed contextual factors that may affect the achievement of project goals. The learning outcomes from the evaluation of IPC will provide lessons learned to MCC and to Guatemala stakeholders. In Table V.1, we summarize the research questions for the IPC evaluation, along with the research dimensions to be analyzed and data sources.

Table V.1. Research questions, research dimensions, and data sources for IPC

Research questions
1. To what extent was the implementer able to complete activities in accordance with his or her work plan? a. Were activities as implemented consistent with the objectives of the GTEP and IPC? b. What factors facilitated implementation of the activities? c. What factors were barriers to implementation? d. What steps did the implementer take to address any barriers faced during the implementation process? 2. To what extent did IPC help improve decision making and resource utilization in MINEDUC? a. What factors contributed to or constrained translation of the investment in activities into improved quality, efficiency, and equity in decision making and resource utilization in lower-secondary education?
Research dimensions (Drivers of Change)
• Structural challenges (political, demographic, and macroeconomic contexts) • Institutional challenges (legal, policy, and administrative practices) • Agents (individual and organizational interests or incentives and capacity)

Data sources
• EMIS[b] data
• Other administrative data
• Budget documents
• Key informant interviews (national and in sample of departments/districts/schools)

[b]EMIS = Education Management Information System.

C. Trend analysis

In this section, we describe the data sources and analysis approach for the first component of the IPC performance evaluation—a trend analysis that will examine changes in key institutional outcomes over time. This analysis will draw on secondary data sources and the MCC M&E monitoring indicators for the GTEP project, including administrative data (for example, annual budgets, teacher databases, and student repetition, dropout, and completion rates) and documents that describe the budgeting process, teacher recruitment strategy, planned changes to the system, and completion of key deliverables. We will draw on other secondary data from MINEDUC's national databases, departmental offices, and FHI360.

1. Key outcomes and data sources

The outcomes we will analyze using secondary data are linked to the research questions and draw on discussions with MCC, MINEDUC, and FHI360 in 2017. Many of the outcomes are project monitoring indicators that MCC and FHI360 are tracking as part of their reporting requirements. These proposed outcomes should be considered preliminary and may be modified in accordance with what the early data collection reveals. Information about the outcomes is typically available at an aggregated level—national or departmental—on an annual basis. To the extent possible, we plan to capture the information for a period covering several years before the project and throughout the life of the project in order to enable us to examine trends.

We describe the key outcomes we plan to examine here and summarize them in Table V.2.

- **Student transition rates.** We will use MINEDUC administrative data to examine any changes in student transition rates between grades 6 through 10. We will look at data beginning in 2013 (three years before the beginning of the threshold program) through the end of the project in 2020.
- **Student dropout rates.** We will use MINEDUC administrative data to examine any changes in student dropout rates in grades 6 through 10. We will look at data beginning in 2013 (three years before the beginning of the threshold program) through the end of the project in 2020.
- **Student repetition rates.** We will use MINEDUC administrative data to examine any changes in student repetition rates in grades 6 through10. We will look at data beginning in 2013 (three years before the beginning of the threshold program) through the end of the project in 2020.
- **Changes in annual budget allocations for lower-secondary education.** We will collect annual budgetary data from MINEDUC to look at any changes in budget allocations for lower-secondary education. Although changes in annual budget allocations may reflect a

variety of external factors, we will attempt to triangulate any changes with data gathered from interviews with MINEDUC and ministry of finance officials and FHI360 representatives to understand how project support contributes to any identified changes.

- **Changes in the distribution of budget allocations for lower-secondary education.** We will attempt to collect secondary data from the ministry of education on the distribution of budget allocations within lower-secondary education and understand (through qualitative sources) whether any identified changes are related to GTEP support.
- **Legal, financial, or policy changes adopted (teacher recruitment and lower-secondary education budget).** We will review all documents related to any legal, financial, or policy changes that take place during the life of the project. Although a wide range of factors is likely to influence policy changes, we will rely on interviews with key decision makers to understand whether and how the GTEP contributed to new policies governing teacher recruitment and budgetary allocations for lower-secondary education.

Table V.2. Outcomes from secondary data

Research question	Outcomes	Source
1. To what extent was the implementer able to complete activities in accordance with his or her work plan? a. Were activities as implemented consistent with the objectives of the GTEP and IPC? b. What factors facilitated implementation of the activities? c. What factors were barriers to implementation? d. What steps did the implementer take to address any barriers faced during the implementation process?	• Increased quantity and quality of secondary education teachers • Improved retention and transition of secondary education students • Increased budget for secondary education	• FHI360 monitoring data • MINEDUC EMIS • Key informant interviews
2. To what extent did IPC help improve decision making and resource utilization in MINEDUC?	• New teacher hiring and recruitment strategy adopted by MINEDUC • Increased budget allocations for secondary education • Online “one stop shop” database provides just in time data for MINEDUC.	• Key informant interviews • Focus groups • MINEDUC policy documents
3. What factors contributed to or constrained translation of the investment in activities into improved quality, efficiency, and equity in decision making and resource utilization for lower-secondary education?	• Increased quantity and quality of secondary education teachers • Improved retention and transition of secondary education students • Increased budget for secondary education	• Key informant interviews • Focus groups

2. Analysis approach

The analysis of the secondary data will be largely descriptive and focus on an illustration of trends. Wherever possible, we will conduct the analysis separately for the five targeted

departments and other departments because the changes associated with the GTEP's IPC might affect the target departments differently, given that schools are receiving additional support through Éxito Escolar. We will necessarily exercise caution in interpreting the findings from the analysis and take care not to unduly attribute any observed changes in trends to the impacts of the project that we are measuring in Éxito Escolar. Specifically, factors unrelated to the project could drive any observed changes in trends; without a valid counterfactual, we might identify changes that we cannot fully attribute to the project. Nevertheless, our approach will still yield useful evidence about changes over time, and the qualitative study discussed below will help us understand the project's possible influence.

D. Qualitative study

In this section, we describe the data sources and analytic approach for the second component of the performance evaluation of IPC—a qualitative study. The study will draw on three rounds of interviews with key stakeholders and two rounds of focus groups with school directors, teachers, and parents. The first round of interviews (in the first two quarters of 2018) will capture information about the current status of the system (from an institutional perspective) and perceptions of the teacher hiring and recruitment strategy, budget allocations for secondary education, and secondary education system needs. The second round of data collection (in the first two quarters of 2019) will capture mid-term results and help guide the trend analysis. The final data collection process will take place at the end of 2020 alongside data collection for Éxito Escolar.

Questions related to IPC outcomes will be included as part of the qualitative protocols discussed under Éxito Escolar for certain stakeholders at endline. We will develop separate protocols to use for baseline and follow-up 1. For the endline, we will group the questions into a separate module that focuses on the secondary education system's institutional framework and capacity.

1. Key data sources

The data sources include key informant interviews with the following IPC– specific stakeholders (a total of 32 structured or semi-structured interviews in the first round and 26 in the second round):

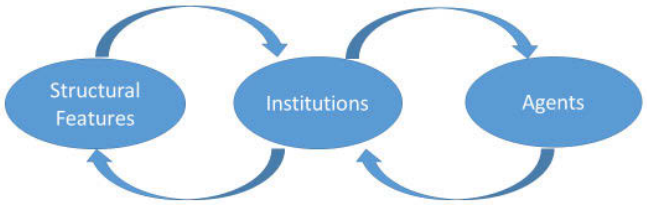
- **Key MINEDUC officials.** We will interview key stakeholders who are likely to play a role in changes to both the teacher recruitment strategy and the budget allocation process. Key officials may include the minister of education, vice minister of education, minister of finance, representatives of the Consejo Nacional de Educación, representatives of the Dirección de Planificación Educativa (DIPLAN), representatives of the Dirección General de Gestión de Calidad Educativa (DIGECADE), representatives of the Dirección de Recursos Humanos, representatives of the Dirección de Desarrollo y Fortalecimiento Institucional, representatives from the Jurado Nacional de Oposicion, representatives from DIGEDUCA and DIGEMOCA, and representatives of the departments. The interviews, which will yield data relevant to both Éxito Escolar and IPC, will help us understand the reform implementation process; the facilitators, challenges, and bottlenecks associated with change; and plans for sustainability.

- **FHI360.** We will interview the directors and relevant technical staff from FHI360, the implementing organization, who work specifically on IPC. The interviews will help us understand how IPC's subactivities unfolded within MINEDUC and enable us to identify the facilitators and challenges associated with implementation. We will interview the staff at baseline and endline for IPC.
- **School directors.** We will gather data related to institutional changes from our school director interviews, as discussed in Section C.2 under Éxito Escolar.
- **Teachers.** We will gather data related to institutional changes in the recruitment system from our teacher focus groups, as discussed in Section C.2 under Éxito Escolar.
- **Parents.** We will gather data related to institutional changes from our parent focus groups, as discussed in Section C.2 under Éxito Escolar.

2. Data analysis

Data analysis for IPC will use the DOC framework to code and map the qualitative data. In applying the DOC (Warrener 2004), we will assess project performance with respect to how project design and implementation addressed the contextual factors of structure, institutions, and agents. In Figure V.1, we demonstrate the interrelated nature of the DOC analysis framework, as discussed in Warrener (2004).

Figure V.1. Conceptual framework for understanding DOCs



Source: Warrener 2004.

We have applied Warrener's (2004) three conceptual areas to the Guatemalan context as follows:

1. **Structure.** The political structure/history of education in Guatemala, trajectory of social and economic development (internal and external), and demographic trends.
2. **Institutions.** The relevant legal framework, government policies (from the education and other sectors), formal administrative and financial processes, and informal rules that influence the behavior of agents.
3. **Agents.** Organizations and individuals who pursue their interests. In the given evaluation, agents include politicians and political appointees, public service staff employed by MINEDUC (at all levels), and teachers' unions, school directors, teachers, and parents.

The institutional modules that we will add to the qualitative protocols will include questions related to the three conceptual areas (Annex D). We will code the results and then map them to show the changing political and economic relationships that occur over time. The mapping process will allow us to document how different institutions and agents can influence the decision-making process and hence the political economy within MINEDUC. DOC analysis provides insight into what, how, and why change takes root in a given sector and examines the change process through interviews and document reviews over several periods. Although the project evaluation design does not call for an impact evaluation for IPC, an assessment of project performance associated with structures, institutions, and agents requires a concise description and baseline assessment of current (before the project) processes for managing the education sector. Mathematica will follow the same data analysis procedures as described for the Éxito Escolar implementation study (Section IV.C).

In Table V.3, we show the timeline for the IPC performance evaluation and the content for the trend analysis, key informant interviews, and focus groups.

Table V.3. Timeline of analysis and data collection for IPC

Timeline for data collection	Topics and themes	Data source
Baseline (March/April 2018)	Current distribution of resources in lower-secondary education: learning materials, infrastructure, and teachers	Administrative data (EMIS)
	Current distribution of budgetary resources in lower-secondary education institutions (budget review and estimation of monetary equivalent of physical and human resources at school level)	Administrative data (EMIS)
	Description of current systems for allocating resources: learning materials, infrastructure, teachers, and financial resources Description of current systems for allocating resources: learning materials, infrastructure, teachers, and financial resources	Document review; key informant interviews with relevant entities at national level, in two participating regions, and, possibly, in subregions, depending on decision-making locus
Midline (March/April 2019)	Annual reporting on performance against criteria/ standards of performance and content knowledge developed by MINEDUC	Document review and key informant interviews
	Progress and constraints with respect to structure, institutions, and agents	Document review and key informant interviews, focus groups
	Teacher resource management system (projection of requirements, recruitment and deployment, and professional development)	Document review and key informant interviews, focus groups
	Infrastructure planning system (projection of requirements and asset management database)	Document review and key informant interviews
	Planning and budgeting for lower-secondary education	Document review and key informant interviews

Timeline for data collection	Topics and themes	Data source
Follow-up (October/November 2020)	Comparison of distribution of resources in lower- secondary education: learning materials, infrastructure, and teachers at baseline	MINEDUC administrative data (MIS)
	Comparison of distribution of budgetary resources in lower-secondary education institutions (budget review and estimation of monetary equivalent of physical and human resources at school level) at baseline	MINEDUC administrative data (MIS)
	Annual reporting on performance against criteria/ standards of performance and content knowledge developed by MINEDUC	Document review and key informant interviews
	Progress and constraints with respect to structure, institutions, and agents	Document review and key informant interviews, focus groups
	Teacher resource management system (projection of requirements, recruitment and deployment, and professional development)	Document review and key informant interviews, focus groups
	Infrastructure planning system (projection of requirements and asset management database)	Document review and key informant interviews
	Planning and budgeting for lower-secondary education	Document review and key informant interviews

VI. ADMINISTRATIVE

In this section, we discuss several administrative issues relevant to managing the evaluation and then present a timeline of the evaluation activities.

A. Summary of institutional review board requirements and clearances

Mathematica is committed to protecting the rights and welfare of human subjects and will prepare and submit an application for approval of the research and data collection plans to an institutional review board (IRB) registered with the Office for Human Research Protections, U.S. Department of Health and Human Services. We intend to use Health Media Lab as our IRB because of our positive experience with it on other MCC projects. For each IRB application, we will submit a set of required documents, including a research protocol providing details of the study and data collection activity, copies of all data collection instruments, and a completed IRB questionnaire summarizing the key elements of the research protocol and plans for protecting participants' confidentiality. The data collection instruments (both the quantitative survey and qualitative protocols) that we will prepare and submit to the IRB will include consent statements approved by MCC that guarantee the confidentiality of respondents to the extent possible. We will apply for an expedited IRB review, although we might have to undergo a full IRB review because the research involves minors, a vulnerable population.

We will provide evidence of IRB approval to MCC. IRB approval is valid for one year, and we will submit annual renewals for subsequent approvals as data collection proceeds through follow-up collection processes. We expect the annual renewals to require only minimal updates to the core application materials because we will be collecting similar data from year to year. If data collection instruments change substantially from those approved by the IRB, we will reapply for review. Small changes to the instruments (such as rewording of questions, reordering of questions, or editing changes) do not require reapplication, but the finalized instruments must be submitted to the IRB for documentation. We will submit the instruments for review in both English and Spanish.

Local authorities in Guatemala have told us we will need to apply for a local IRB as well. Mathematica will submit the research protocols and instruments to our U.S.-based IRB, and the local survey firm hired by Mathematica will obtain permits or clearances from the relevant national or local government offices before initiating field work. The survey firm, MCC, and Mathematica will work together to accommodate any changes the IRB board recommends to the final protocol before the start of data collection.

B. Data protection

Mathematica and the local survey firm will ensure confidentiality of all survey respondents, including confidentiality of survey participation, confidentiality of personally identifiable information, and confidentiality of other sensitive data. The data collection instruments (both the quantitative survey and qualitative protocols) will include consent statements approved by MCC that guarantee the confidentiality of respondents to the extent possible. When it collects data, the local survey firm will ensure the safe handling and transport of questionnaires from the field to the main office for data entry where the questionnaires will be stored in lock-and-key cabinets. The electronic data files will be stored on a secure Mathematica server and will be accessible

only to project team members who use the data. All project team members have signed a nondisclosure agreement pertaining to confidential information. Electronic data files will be shared with Mathematica via a secure file transfer system, such as a file transfer protocol (FTP), file exchange website (FX site), or a SharePoint site. For internal control and audit purposes, the local survey firm will retain the data files, both in paper and electronic form, for the entire duration of the project, which includes the base contract and the subsequent option contracts. All of the collected data and databases are the property of Mathematica and will be delivered to Mathematica at the end of the contract.

C. Preparing data files for access, privacy, and documentation

Public-use data will enable any stakeholder, researcher, or agency to understand the source data and analysis behind MCC evaluations and might inspire a wide range of new policy-relevant research, maximizing the benefits of MCC's investments in large-scale data collection efforts in developing countries. Before the analysis, we will attempt to de-identify the data to identify any limitations to analysis that may result from working with a de-identified file. We will alert MCC to discuss the appropriate solution if we find we are unable to conduct analysis using the de-identified file. In addition to de-identified data files, we will provide users' manuals and codebooks according to the most recent guidelines set forth by MCC. Public-use data files will be free of personal or geographic identifiers that would enable unassisted identification of individual respondents or their households, and we will remove or adjust variables that introduce reasonable risks of deductive disclosure of the identity of individual participants. We will also recode unique and rare data by using top and bottom coding or replacing affected observations with missing values. If necessary, we will also collapse into less easily identifiable categories any variables that make an individual highly visible because of geographic or other factors.

We will also produce de-identified transcripts of focus groups and interviews—with the exception of interviews with MINEDUC, PRONACOM, or FHI360 officials; we will not provide publicly available transcripts of interviews with these officials because those positions are unique and transcripts of their interviews cannot be de-identified.

D. Dissemination plan

The Mathematica team will present the final evaluation findings in person to MCC at MINEDUC. We will also participate in any other MCC-financed dissemination and training events related to the findings from the baseline and endline reports. To ensure that the results and lessons from the evaluation reach a wide audience, we will work with MCC to increase the visibility of the evaluation and findings within the education sector, particularly for policymakers and practitioners. We expect the broader research community to have a strong interest in the evaluation findings. To facilitate wider dissemination of findings and lessons, we will collaborate with MCC and other stakeholders to identify additional forums—conferences, workshops, and publications—for disseminating the results and will encourage other donors and implementers to integrate the findings into their programming.

E. Evaluation team's roles and responsibilities

Our team has extensive experience and expertise in evaluation design, data collection, and analysis and therefore will be able to meet MCC's evaluation needs. **Dr. Audrey Moore**

oversees the project team and provides technical leadership. She is responsible for managing the evaluation team, leading the design and implementation of the evaluations, and overseeing quantitative and qualitative data collection. Dr. Moore also monitors the project's budget and schedule and manages communication with MCC, local partners, and other stakeholders. **Drs. Sarah Liuzzi** and **Catalina Torrente** assist Dr. Moore as senior analysts, overseeing the quality of data collection in the field and working on the design of the impact and performance evaluations and analysis. **Mr. Seth Morgan**, junior analyst on the team, assist with conducting the project's evaluability assessment and support other evaluation activities, particularly data collection in the field and data analysis. **Ms. Galina Lapadatova**, also a junior analyst on the team, manages the project internally for Mathematica and will support the training, data collection, and analysis tasks.

Our three consultants also play an important role in the project. **Dr. Anthony DeWees** serves as the political economist on the team and provides technical analysis, advice, and guidance in political science and political economy. **Dr. Elizabeth Katz** is the gender and social inclusion specialist on the team and is responsible for technical and methodological leadership of the social and gender-related aspects of the evaluation. **Dr. Larissa Campuzano** provides quality assurance reviews for all of the project's key deliverables.

F. Evaluation timeline and reporting schedule

In Table VI.1 we present the evaluation timeline.

Table VI.1. Evaluation timeline

Round	Data collection	Data cleaning and analysis	First draft report expected	Final draft report expected
Baseline quantitative, Éxito Escolar	March/April 2018	September 2018	November 2018	January 2019
Baseline qualitative, IPC	March/April 2018	September 2018	November 2018	January 2019
Baseline qualitative, Éxito Escolar	October 2018	March 2019	May 2019	July 2019
First follow-up, IPC	March/April 2019	September 2020	November 2020	January 2021
Endline quantitative, Éxito Escolar and endline qualitative, Éxito Escolar and IPC	October 2020	May 2021	July 2021	September 2021

Note: We assume that we will receive data one month after the data collection month listed.

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

REFERENCES

Andrabi, Tahir, Jishnu Das, and Asim I. Khwaja. "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets." 2015. Available at http://documents.worldbank.org/curated/en/789721468191341184/text/WPS7226.txt. Accessed October 30, 2017.

Asturias de Barrios, Linda. "Propuesta de transformación del nivel de educación media: versión revisada y validada. Documento para la discusión." Guatemala City, Guatemala: USAID, January 2014.

Bandura, A. "Guide for Constructing Self-Efficacy Scales." Greenwich, CT: Information Age Publishing, 2006.

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur. "Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education." 2013. Available at https://www.cgdev.org/publication/scaling-what-works-experimental-evidence-external-validity-kenyan-education-working. Accessed October 30, 2017.

Brooker, Simon, Catherine Goodman, Caroline Jones, Benson Estambale, Matthew Jukes, Hellen Inyega, Elizabeth Juma, and Kiambo Njagi. "Impact of Malaria Control and Enhanced Literacy Instruction on Educational Outcomes Among Kenyan School Children: A Multi-Sectoral, Prospective, Randomised Evaluation." 2013. Available at http://www.3ieimpact.org/media/filer_public/2013/10/18/malaria_kenya_consolidated_dfr.pdf. Accessed October 30, 2017.

Bruneforth, Michael, Albert Motivans, and Yanhong Zhang. "Investing in the Future: Financing the Expansion of Educational Opportunity in Latin America and the Caribbean." Montreal, Quebec: UNESCO Institute for Statistics, 2004.

Bruns, Barbara, Deon Filmer, and Harry A. Patrinos. *Making Schools Work: New Evidence on Accountability Reforms.* Washington, DC: World Bank Publications, 2011.

Bruns, Barbara, and Javier Luque. *Great Teachers: How to Raise Student Learning in Latin America and the Caribbean.* Washington, DC: World Bank Publications, 2015.

Cabezas, Verónica, José I. Cuesta, and Francisco A. Gallego. "Effects of Short-Term Tutoring on Cognitive and Non-Cognitive Skills: Evidence from a Randomized Evaluation in Chile." Santiago, Chile: PUC-Chile and J-PAL, 2011.

Carr-Hill, Roy, Caine Rolleston, Tejendra Pherali, and Rebecca Schendel. "The Effects of School-Based Decision Making on Educational Outcomes in Low-and Middle-Income Contexts: A Systematic Review." 3ie Grantee Final Review. London: International Initiative for Impact Evaluation (3ie), 2015. Available at http://www.3ieimpact.org/media/filer_public/2015/11/20/dfid-funded-decentralisation-review.pdf. Accessed October 30, 2017.

Cohen, David K., and Heather C. Hill. *Learning Policy: When State Education Reform Works.* Binghamton, NY: Vail-Ballou, 2001.

Conn, Katharine M. "Identifying Effective Education Interventions in Sub-Saharan Africa: A Meta-Analysis of Rigorous Impact Evaluations." 2014. Available at https://search.proquest.com/openview/32e4aee063ff3a8c0e5729167a1e792d/1?pq-origsite=gscholar&cbl=18750&diss=y. Accessed October 31, 2017.

Creswell, J.W. *Research Design: Qualitative, Quantitative, and Mixed-Methods Approaches*. Washington, DC: SAGE Publications, 2009.

Darling-Hammond, Linda, Ruth C. Wei, Alethea Andree, Nikole Richardson, and Stelios Orphanos. "Professional Learning in the Learning Profession." Washington, DC: National Staff Development Council, 2009.

DeStefano, Joseph, and Luis Crouch. *Education Reform Support Today*. Washington, DC: USAID, 2006.

Dirección General de Evaluación e Investigación Educativa (DIGEDUCA). "Resultados generales de evaluación educativa." Guatemala: Ministerio de Educación, 2016.

Duhamel, M., A. Houchot, H.G. Richon, J.F. Cuby, Y. Moulin, O. Roze, and C. Symankiewicz. "L'évolution Du Réseau Des Écoles Primaires: Rapport d'étape." Paris: Ministere De La Jeunesse, De l'Education Nationale Et De La Recherche, 2003.

Estrada, Ricardo. "Rules Rather Than Discretion: Teacher Hiring and Rent Extraction." Paris: Paris School of Economics, 2013.

Evans, David K., and Anna Popova. "What Really Works to Improve Learning in Developing Countries?" 2015. Available at https://openknowledge.worldbank.org/bitstream/handle/10986/21642/WPS7203.pdf?sequence=1. Accessed October 31, 2017.

Gillies, John. "The Power of Persistence: Education System Reform and Aid Effectiveness." 2010. Available at http://www.equip123.net/docs/E2-Power_of_Persistence.pdf. Accessed March 9, 2017.

Giordano, Elizabeth, A. "School Clusters and Teacher Resource Centres." Paris: UNESCO, International Institute for Educational Planning, 2008.

Glennan, Thomas K., Susan J. Bodilly, Jolene Galegher, and Kerri A. Kerr. *Expanding the Reach of Education Reforms: Perspectives from Leaders in the Scale-Up of Educational Interventions*. Rand Corporation, 2004. Available at http://www.rand.org/pubs/monographs/MG248.html. Accessed March 9, 2017.

Glewwe, Paul W., Eric A. Hanushek, Sarah D. Humpage, and Renato Ravina. *School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010.* 2011. Available at http://www.nber.org/papers/w17554. Accessed October 31, 2017.

Guarino, Cassandra M., Lucrecia Santibanez, and Glenn A. Daley. "Teacher Recruitment and Retention: A Review of the Recent Empirical Literature." *Review of Educational Research,* vol. 76, no. 2, 2006, pp. 173-208.

Hanushek, Eric, and Steven Rivkin. "Using Value-Added Measures of Teacher Quality." Policy Brief 9. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research, 2010.

Inter-American Development Bank. "Guatemala: Education Quality and Coverage Improvement Program." GU-L1087 loan proposal. 2015. Available at http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=39997432. Accessed March 9, 2017.

Jimenez, Emmanuel, and Yasuyuki Sawada. "Does Community Management Help Keep Kids in Schools? Evidence Using Panel Data from El Salvador's EDUCO Program." Washington, DC: World Bank, 2003.

Jukes, M.C.H., S.B. Vagh, and Y. Kim. "Development of Assessments of Reading Ability and Classroom Behavior." Washington, DC: World Bank, 2006.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." *Economics of Education Review*, vol. 27, no. 6, 2008, pp. 615-631.

King, Elizabeth M., and Berk Ozler. "What's Decentralization Got to Do with Learning? School Autonomy and Student Performance." *Kyoto University: Interfaces for Advanced Economic Analysis,* vol. 54, 2005, pp. 51-60.

Kremer, M., C. Brannen, and R. Glennerster. "The Challenge of Education and Learning in the Developing World." *Science (New York, NY),* vol. 340, no. 6130, 2013, pp. 297-300.

Krishnaratne, Shari, Howard White, and Ella Carpenter. "Quality Education for All Children? What Works in Education in Developing Countries." Working Paper No. 20. New Delhi: International Initiative for Impact Evaluation (3ie), 2013.

Lewin, Keith, and Françoise Caillods. "Financing Secondary Education in Developing Countries: Strategies for Sustainable Growth." Paris: UNESCO, International Institute for Educational Planning, 2001.

Lopez-Calva, Luis F., and Luis D. Espinosa. "Efectos Diferenciales De Los Programas Compensatorios Del CONAFE En El Aprovechamiento Escolar." In *Efectos Del Impulso a La Participación De Los Padres De Familia En La Escuela.* Mexico: CONAFE, 2006.

Lucas, Adrienne M., Patrick J. McEwan, Moses Ngware, and Moses Oketch. "Improving Early-Grade Literacy in East Africa: Experimental Evidence from Kenya and Uganda." *Journal of Policy Analysis and Management*, vol. 33, 2013, pp. 950–976.

McEwan, Patrick J. "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments." *Review of Educational Research,* vol. 85, no. 3, 2015, pp. 353-394.

Millennium Challenge Corporation. "MCC-16_RFQ-0131 (Guatemala Education Project)." Washington, DC: May 24, 2016.

MINEDUC. "Redes Educativas Rurales: Estrategia territorial para una atención educativa relevante y pertinente." Guatemala: Ministerio de Educación, January 2017.

MINEDUC. "Ruta Crítica: Avances y Retos. Prioridades del Plan de Implementación Estratégica de Educación 2012-2016." Guatemala: Ministerio de Educación, July 2014.

Moore, Audrey-Marie Schuh, Joseph DeStefano, and Elizabeth Adelman. "Opportunity to Learn: A High-Impact Strategy for Improving Educational Outcomes in Developing Countries." Education Quality Improvement Program 2. Washington, DC: FHI 360, 2012.

MSI. "Scaling Up—From Vision to Large-Scale Change: A Management Framework for Practitioners." Washington, DC: MSI, 2012. Available at http://www.msiworldwide.com/wp-content/uploads/MSI-Scaling-Up-Framework-2nd-Edition.pdf. Accessed March 9, 2017.

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, 2010, pp. F187-F203.

Murnane, Richard J., and Alejandro J. Ganimian. *Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Evaluations.* National Bureau of Economic Research, 2014.

Patton, Michael Q. *Qualitative Evaluation and Research Methods.* Washington, DC: SAGE Publications, 1990.

Perry, Lindsey. "Review of Formative Assessment Use and Training in Africa." *International Journal of School & 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 International, 2009.

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

Reimers, Fernando, and Sergio Cárdenas. "Who Benefits from School-Based Management in Mexico?" *Prospects,* vol. 37, no. 1, 2007, pp. 37-56.

Rodríguez, Catherine, Fabio Sánchez, and Armando Armenta. "Do Interventions at School Level Improve Educational Outcomes? Evidence from a Rural Program in Colombia." *World Development,* vol. 38, no. 3, 2010, pp. 415-428.

Rothstein, Jesse. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *The Quarterly Journal of Economics*, vol. 125, no. 1, 2010, pp. 175-214.

Sailors, M., J.V. Hoffman, P.D. Pearson, S.N. Beretvas, and B. Matthee. "The Effects of First- and Second-Language Instruction in Rural South African Schools." *Bilingual Research Journal,* vol. 33, no. 1, 2010, pp. 21-41.

Snilstveit, Birte, Jennifer Stevenson, Radhika Menon, Daniel Phillips, Emma Gallagher, Maisie Geleen, Hannah Jobse, Tanja Schmidt, and Emmanuel Jimenez. "The Impact of Education Programmes on Learning and School Participation in Low- and Middle-Income Countries." London, 3ie: 2016.

Spratt, Jennifer, Simon King, and Jennae Bulat. "Independent Evaluation of the Effectiveness of Institut Pour l'Education Populaire's 'Read-Learn-Lead' (RLL) Program in Mali." Unpublished manuscript. Research Triangle Park, NC: RTI International, 2013.

Stallings, J.A., and G.G. Mohlman. "Classroom Observation Techniques." In *Educational Research, Methodology, and Measurement: An International Handbook,* edited by J.P. Keeves. West Yorkshire, England: Emerald Group, 1988.

Thompson, Marnie, and Dylan Wiliam. "Tight but Loose: A Conceptual Framework for Scaling Up School Reforms." Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, 2007.

USAID. "Guatemala Education Situation Analysis." June 2015. Available at https://www.usaid.gov/guatemala/education. Accessed March 9, 2017.

Warrener, Debbie. "Synthesis Paper 3: The Drivers of Change Approach." London: Overseas Development Institute, 2004.

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

ANNEX A

RESEARCH QUESTIONS: ORIGINAL, NEW, AND DOCUMENTATION OF CHANGES

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

Table A.1. Research questions for the Éxito Escolar activity

Fidelity of implementation study questions	Original or new question
To what extent were Éxito Escolar's planned activities for teacher professional development, pedagogical support, and development of parent councils and school networks implemented as designed?	
4. What effect do the teacher diagnostics have on teacher motivation?	Original
5. Are the teacher diagnostics useful in training teachers and helping them reach a higher competency level?	Original
6. Are the training instructors and the Ministry able to respond appropriately and with appropriate training when teachers need more support?	Original
7. Did a majority of teachers complete the training?	Original
8. What obstacles did teachers face when completing the training?	Original
9. What kinds of pedagogical support are most important to teachers?	Original
10. Do teachers and their assigned pedagogic advisors meet regularly?	Original
11. Did teacher competency improve after the implementation of Éxito Escolar?	New—replaces impact question 26
12. What were teachers' perceptions of the reasons for changes in student learning outcomes? Did they come as a result of the teacher training program, the parent councils, school networks, or an interaction between all of the components?	New—replaces impact question 27
13. Did the parent councils implement the Action Plans initiated by Éxito Escolar as planned?	Original
14. What were the results of the plans? How effective were they?	Original
15. Are parent councils able to identify and successfully mitigate factors that lead students to drop out of school?	Original
16. Do dropout rates decrease with additional funds from municipalities and capacity building for parent councils?	Original
17. Is the additional support from local government targeted at the right families and students?	Original
18. Are indigenous families represented in the parent councils?	Original
19. What kind of support do the parent councils provide female students and their families to encourage those students to stay in school?	Original
20. How do school directors, teachers, parents, and students perceive the relative contribution of different subactivities (i.e., teacher professional development, pedagogic support, school networks, and parent councils) to observed changes in students' outcomes?	New
21. How did parents' perceptions of secondary education change as a result of broader dissemination of information to parents of primary school students?	New—replaces impact question 32
22. What were the main facilitators of and barriers to implementing Éxito Escolar activities in terms of reaching hypothesized medium term outcomes including: a. Improvements in student learning b. Higher graduation rates c. Increased retention and promotion of secondary education students	New

Impact evaluation questions	
23. Did teachers adopt new pedagogical approaches as a result of Éxito Escolar? (e.g. active learning, more attention paid to different learning styles of students, adaption of lessons for language minorities, and equal attention paid to both genders and students of all socio-economic backgrounds)	New—replaces impact question 28
24. To what extent did time devoted to learning in the classroom increase as a result of the intervention?	Original
25. How does time devoted to learning in the classroom vary across different teaching subjects?	Original
26. How did Éxito Escolar affect student enrollment in secondary schools (promotion rates, retention rates, dropout rates)?	New—replaces impact questions 29 and 31
27. How were student learning outcomes affected by Éxito Escolar?	New—replaces impact question 30
28. How do changes in student learning outcomes vary across gender, socio-economic, and language groups?	Original
Impact evaluation questions that are not feasible to answer	
29. Has teacher competency improved as a result of the teacher training and certificate program?	Replaced by implementation question 8
30. What were the reasons for these changes in student learning outcomes? Did they come as a result of the teacher training program, the parent councils, school networks, or an interaction between all of the components?	Replaced by implementation question 9
31. Did teachers adopt new pedagogical approaches as a result of their training? This could include more active learning, more attention paid to different learning styles of students, adaption of lessons for language minorities, and equal attention paid to both genders and students of all socio-economic backgrounds	Replaced by impact question 20
32. How did teacher trainings' improvement in teacher performance affect student enrollment in secondary schools (promotion rates, retention rates, dropout rates)?	Replaced by impact question 23
33. How were student learning outcomes affected by teacher training' improvement in teacher performance?	Replaced by impact question 24
34. Does the establishment of parent councils have an impact on decreasing the dropout rate?	Replaced by impact question 23
35. How did parents' perceptions of secondary education change as a result of broader dissemination of information to the parents of primary school students?	Replaced by implementation question 18

Note: This list includes all of the evaluation questions listed under section C.3.6.1 of contract MCC-16-CON-0040.

ANNEX B

RANDOM ASSIGNMENT: PROCEDURE AND RESULTS

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

Random assignment plan

Overview. In a public lottery, we conducted stratified random assignment at the district level, assigning 80 percent of districts to treatment and 20 percent to control. Districts are clusters of schools formed by MINEDUC, ranging in size from 2 to 18 schools, with an average of 8 schools per district. Most strata had five districts, and within each stratum, we assigned four districts to treatment and one to control (three strata had six districts). We formed strata ahead of time, and the random assignment took place in a public lottery in Guatemala. With our approach, we attempted to balance two goals: (1) creating equivalent study groups, and (2) ensuring that FHI360 has an adequate number of teachers to train (our target was to assign at least 76 percent of all teachers to districts that were assigned to treatment).

Stratification. First, we stratified on department. Within each department, we ranked districts by their average test score on the most recent standardized test, which was done in 2013, and grouped districts with similar test scores to form strata. However, we reassigned some districts to ensure that districts of roughly similar size (in terms of number of teachers) were grouped together. To avoid randomly assigning so many of the largest districts to the control group that we would fail to assign a minimum of 76 percent of teachers to the treatment group, we reassigned some of the largest districts to be in the same strata as other large districts. This way, if one large district was assigned to control, other large districts within the same stratum would be assigned to treatment.

Results of stratification. Our strata have the following characteristics:

- **Likely range of treatment group sizes**: We ran 15 simulations, and the percentage of teachers assigned to the treatment group ranged from 78 to 84 percent. The smallest possible treatment group, which would result if the largest district in every stratum were assigned to the control group, would have 70.3 percent of teachers in the treatment group. This outcome was extremely unlikely.
- **Similarity on test scores within strata:** We used the district-level mean test score, averaged across math and language and communication, to group districts on the similarity of their test scores. The scores in the test data file that we downloaded from the MINEDUC website are labeled as having a mean of zero and a standard deviation of one, and the distribution of scores observed in the data is similar to that. The schools in our sample are below the national average in terms of their scores. The district means, averaged across language and communication and math, range from 1.11 to 0.07 standard deviations below the national mean. Within strata, the largest range of scores from the lowest to highest-scoring district is 0.31, and the average within-stratum range is 0.15 standard deviations. The average range sorting just on test score is 0.11 standard deviations. The ranges in our strata are a bit larger because we reassigned some districts out of strata with more similar test scores in order to cluster the larger districts together.

Strategies to create strata with five units of randomization (districts). The sample includes 103 districts. We wanted to form strata of five districts to ensure that every district would have an 80 percent chance of treatment regardless of its stratum. Because 103 is not a multiple of 5, in three strata we paired two districts and treated them as one district or unit of random assignment—that is, they were randomized together. These pairings have no

implications for implementation. When we conduct regression analysis, we will cluster standard errors at the district level. For the paired districts, each pair of districts will be considered one district. Four of the five departments have a number of districts that is not a multiple of five. In two of these departments (Sacatepéquez and Jalapa), we formed two paired districts, which allowed us to form strata without combining districts from different departments. We formed one stratum with districts from different departments: this stratum included two districts from Sololá and three from Chiquimula. Table B.1 shows the number of districts by department.

Table B.1. Districts by department

Department	Number of districts	Number of "leftover" districts beyond the largest multiple of 5	Number of paired districts	Number of districts in combined stratum
Alta Verapaz	45	0	Already a multiple of 5; none needed	
Chiquimula	18	3	None	3
Jalapa	12	2	2	None
Sololá	17	2	None	2
Sacatepéquez	11	1	1	None
Total	103	3	3	5

Public lottery. We randomized within strata in a public lottery ceremony in Guatemala. The lottery was attended by departmental and central representatives from the five departments in the Threshold Program; representatives of DIGEDUCA, DIGEMOCA, DIGEBI, DIGECADE, DIGECUR, DIDEFI; the vice minister of education; advisors to the vice minister of education, and staff from MCC, PRONACOM, and FHI360.

We brought 20 cloth bags and 100 wooden blocks. Each bag represented a stratum. The blocks represented districts. Blocks were labeled from one through 100. In the case of paired districts, one block represented both districts. In the lottery, attendees took turns drawing blocks from each bag to select treatment districts. We gave attendees printed sheets showing the districts assigned to each stratum so they could follow along and mark which districts were assigned to treatment and which were assigned to control. At the same time, PRONACOM personnel recorded the results on PowerPoint slides that were displayed on a projector screen. Once all districts had been assigned to treatment or control, the results were read aloud so attendees could confirm their accuracy.

Outcome of the random assignment. The lottery event was well received. The random assignment resulted in a treatment group of 82 districts and a control group of 21 districts. We estimate that the districts in the sample have a total of 4,337 teachers. The districts assigned to the treatment group have a total of 3,490 teachers, which represent 80.5 percent of teachers in the sample. Treatment and control districts were also similar on student test scores; the mean test score for districts assigned to the treatment group was −0.498, and the mean for districts assigned to the control group was −0.500.

ANNEX C

IMPLEMENTATION INDICATOR LIST

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

Table C.1. Proposed MCC monitoring indicators

Subactivity	Indicators
Teacher professional development	Number of teacher professional development instructors trained
Teacher professional development	Number of educators (school directors and teachers) enrolled in teacher professional development, by school, position, gender, and grade level
Teacher professional development	Number of educators at risk of dropping out of the teacher professional development program, by school, position, gender, and grade level
Teacher professional development	Number of educators at risk of dropping out of the teacher professional development program who received support to stay in the program, by school, position, and gender
Teacher professional development	Number of educators (school directors and teachers) who complete the teacher professional development, by school, position, gender, and grade level
Teacher professional development	Knowledge as measured by teacher professional development entrance and exit examinations, by individual educator, school, position, gender, and grade level
Pedagogic support	Number of school visits by management advisors (time frame to be determined), by school
Pedagogic support	Number of teachers benefiting from pedagogic support, by school, gender, and grade level
Pedagogic support	Number of school visits by pedagogic advisors (time frame to be determined), by school, gender, and grade level
Parent councils	Parent councils established, by school
Parent councils	Action plans established, by school
School networks	Number of teachers participating in any learning community activities, by school, gender, and grade level
School networks	Number of school directors participating in any learning community activities, by school and gender
School networks	Number of school networks established by municipality
School networks	Number of primary teachers involved in networks, by school, gender, and grade level
School networks	Number of lower-secondary teachers involved in networks, by school, gender, and grade level
School networks	Number of school directors involved in networks, by school and gender
School networks	Number of parents involved in networks, by school, gender, and students' grade level
School networks	Number of educational network activities (time frame to be determined), by network
School networks	Number of participants attending educational network activities, by network and type of participant (e.g., school directors, teachers, parents, students; by gender and ethnicity and so on)

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

ANNEX D

POLITICAL ECONOMY ANALYSIS PROTOCOL: TOPICS AND METHODS

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

Table D.1. Political economy analysis protocol: topics and methods

Area of political analysis	Proposed topics for data collection and analysis using drivers of change, bottleneck analysis, and power analysis	Method
`Enabling environment (crosses structure, institutions, and agents)	1. **Social norms:** Expectations of young people and households and how the expectations differ with respect to gender, ethnicity, and wealth.	Focus group discussions
	2. **Policy/legal framework**: Development policy and priorities, Education and other social sector policies, norms for public service.	Document review
	3. **Budget/expenditure**: Adequacy of financial resources to meet policy goals and the effectiveness and efficiency of spending.	Key informant interviews and document review
	4. Management/coordination: Locus of decision making (for key quality and equity determinants); organizational structure; norms of reporting and accountability; and explicit and implicit incentives for decision makers.	Key informant interviews and document review
Supply of secondary education (institutions)	1. Physical accessibility to secondary education.	Focus group discussions
	2. Equity in distribution of key resources (financial, material, human).	Key informant interviews
Demand for secondary education (institutions, agents)	1. Financial accessibility. 2. Culture and expectations. 3. Labor market returns.	Labor market surveys, Key informant interviews and focus group discussions on the perceived labor market returns
Quality (institution, agents)	1. Adherence to national goals/standards and international norms.	Key informant interviews, document review

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

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research