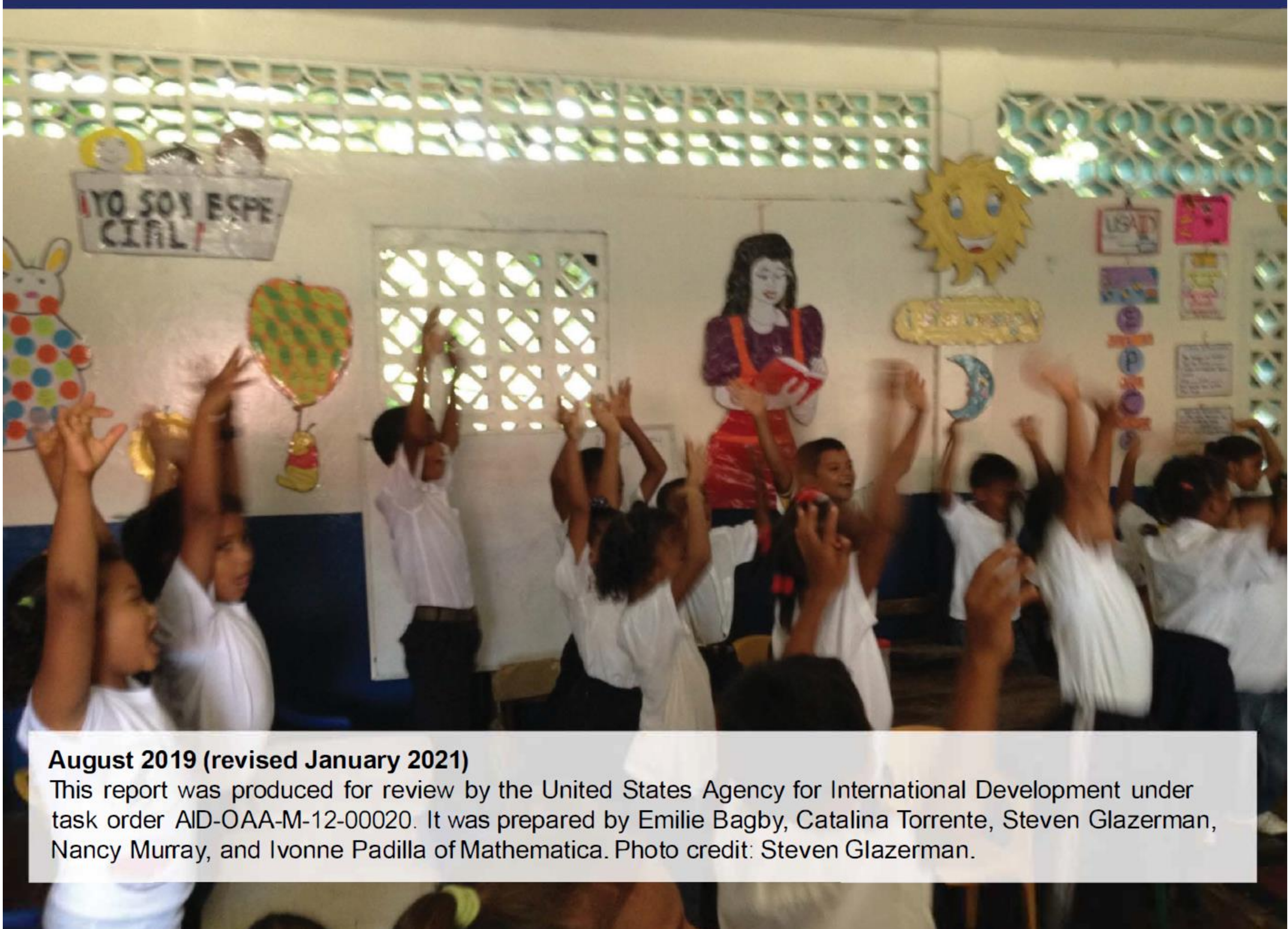
FROM THE AMERICAN PEOPLE

Evaluation of Espacios para Crecer (EpC), an afterschool program, in Nicaragua

Final Report



August 2019 (revised January 2021)

This report was produced for review by the United States Agency for International Development under task order AID-OAA-M-12-00020. It was prepared by Emilie Bagby, Catalina Torrente, Steven Glazerman, Nancy Murray, and Ivonne Padilla of Mathematica. Photo credit: Steven Glazerman.

Evaluation of Espacios para Crecer (EpC), an afterschool program, in Nicaragua

Final Report

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

ABSTRACT

This report presents findings from the impact evaluation of *Espacios para Crecer* (EpC), an afterschool program that involved training and support of educators in the EpC methodology and the delivery of appropriate support materials. EpC facilitators were trained on methods, techniques and activities to facilitate learning and create a positive social environment, to support social development and academic leveling, and to assess and provide differentiated responses to meet children's individual needs. EpC are an important component of a larger project implemented in the Caribbean coast of Nicaragua called Community Action for Reading and Security (CARS).

A two-arm randomized control trial was used to examine impact of the EpC intervention on children's reading skills. In the treatment group, children were exposed to the EpC intervention; and in the control group, children were not exposed to EpC afterschool activities. Treatment and control group participants could have been exposed to other activities offered by CARS, such as education program for parents and community engagement activities aiming to spur community discussions around education. A distinguishing feature of this evaluation is that we randomly assigned—children or communities—to the treatment or control group depending on the size of the community. In larger communities (with more eligible children), we randomly assigned children. In communities with fewer children, it was not possible to form two separate groups, so the communities were the unit of random assignment, with all children in the community assigned together to the treatment or control group. The evaluation followed two cohorts of students for approximately one and a half years of exposure to the EpC program. We collected base-year data to measure children's literacy skills, but could do so only in one of the cohorts and, in some cases, after exposure to the intervention had already begun. We collected follow-up data for each cohort (in 2016 for Cohort 1 and in 2017 for Cohort 2).

Evaluation findings show that that the experimental design worked well in small communities, but not as well in the larger ones, due to lower take-up of and compliance with the program. Findings show that in these small communities, EpC had positive impacts on children's reading outcomes, but not on school attachment or social-emotional outcomes. Impacts were statistically significant for girls and for children who were out of school at intake, but not for boys and children who were already enrolled in school. Cost-effectiveness estimates for the EpC intervention ranged from $45, at steady state, to $358, including startup costs, per 0.1 standard deviations in literacy score improvement.

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

CONTENTS

FIGURES

TABLES

ACKNOWLEDGMENTS

This report is the result of the combined efforts of many people, including Barbara Knox-Seith, on the U.S. Agency for International Development (USAID) Latin America and the Caribbean (LAC) education team. They all provided guidance and support throughout. We are also grateful to Alicia Slate and Angela Cardenas, USAID/Nicaragua, for their helpful input throughout the evaluation.

The report would not have been possible without the contributions of Community Action for Reading and Security (CARS) staff, including Richard Fisher, Megan Gavin, John Helwig, and Solange Casanova, who shared project materials, provided data, responded to emails, and spent hours on the phone helping us to understand their project. We also give special thanks to Miguel Nuñez and Heriberto Hernandez for their tireless patience and support over the course of the evaluation.

Other partners based in Nicaragua also provided important assistance. Mireille Vijil gave insightful comments at various phases of the evaluation and helped ensure the integrity of the evaluation design. Fundación Internacional para el Desafío Económico Global (FIDEG) and Centro de Investigación y Acción Educativa Social (CIASES) provided expert field data collection for this evaluation. We also thank Vanessa Castro for her expertise in developing context-appropriate reading assessments to evaluate the literacy skills of children who participated in this evaluation.

At Mathematica, Virginia Poggio played a major role in all aspects of the evaluation’s fieldwork. Linda Molinari and Cannon Warren provided expert research and programming assistance for this final report. Raquel af Ursin and Sara Woodruff supported measurement development and data collection. In addition, Ken Fortson provided important technical input along the way, and Dan Levy provided valuable advice for this final report.

Last, but not least, we are very appreciative of all the families, school directors, and teachers for allowing us to learn from their perspectives. We are especially grateful to the children who provided us with the important information we are able to share in this report.

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

LIST OF ACRONYMS

CARS	Community Action for Reading and Security
CIASES	Centro de Investigación y Acción Educativa Social
EGRA	Early-grade reading assessment
EpC	Espacios para Crecer (Spaces to Grow)
EPM	Escuelas para Padres y Madres (parent schools)
FIDEG	Fundación Internacional para el Desafío Económico Global
IDEL	Indicadores Dinámicos del Éxito en la Lectura (Dynamic Indicators of Reading Success)
IRR	Interrater reliability
LAC	Latin America and the Caribbean
MDE	Minimum detectable effect
M&E	Monitoring and evaluation
NGO	Nongovernmental organization
RACCN	Región Autónoma de la Costa Caribe Norte (Northern Caribbean Atlantic Autonomous Region)
RACCS	Región Autónoma de la Costa Caribe Sur (Southern Caribbean Atlantic Autonomous Region)
USAID	United States Agency for International Development
USDOL	United States Department of Labor

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

EXECUTIVE SUMMARY

A. Introduction

This report presents the results of the impact evaluation of *Espacios para Crecer* (EpC), an intensive after-school program funded by the U.S. Agency for International Development (USAID) in Nicaragua. The evaluation focuses on Nicaragua's Southern Caribbean Atlantic Autonomous Region, known by the Spanish acronym RACCS, which is the second- poorest of all Nicaragua's regions and departments. DevTech Systems Inc. together with its partners adapted the EpC intervention to the RACCS context and implemented it as a part of the USAID/Nicaragua-funded Community Action for Reading and Security (CARS) project from 2011 through 2017. EpC is a half-day program offered each school day in the afternoon for children who go to school in the morning, or in the morning for children who attend school in the afternoon. EpC provides additional learning time to children who are at risk of poor performance, are having difficulties at school, have dropped out of school before completing grade 3, or have never attended school.

EpC involved training and continued professional development of educators in the EpC methodology and the delivery of appropriate support materials. EpC facilitators were trained on methods, techniques and activities to facilitate learning and create a positive social environment, to support social development and academic leveling, and to assess and provide differentiated responses to meet children's individual needs.

B. Evaluation questions and design

- Was the EpC intervention implemented as intended? What services were delivered at the child, teacher, school, and community levels?
- What impact did the EpC intervention have on children's reading skills, school attachment, and social-emotional skills and attitudes?
- What were the costs of EpC? Were the effects of EpC large enough to justify the costs?

To answer these evaluation questions, we implemented a mixed methods evaluation comprised of an impact evaluation using a randomized-controlled trial design and a qualitative performance evaluation. The analysis draws on EpC intake data collected by the program implementers, household and teacher survey data collected by the evaluation team, child literacy assessments administered by the evaluators in children's homes, and qualitative data collected throughout the evaluation to document implementation.

A distinguishing feature of this evaluation is that we randomly assigned different units—children or educational communities—depending on the number of eligible children in the educational community. An educational community is defined as a school and its immediate catchment area. We categorized those educational communities that had enough children to support an EpC as either *small* or *large*. Small communities had at least 20 and fewer than 50 eligible children. Large educational communities had 50 or more eligible children, which was enough for at least one EpC and one control group. Small *communities* were assigned to either the EpC or the control group, while in large communities we randomly assigned *children* to either group. Thus, in small communities all eligible children belonged to the same experimental group (EpC *or* control), whereas each large community included children assigned to the EpC *as well as children assigned to the* control group.

C. Summary of findings

The EpC intervention was implemented largely as intended. One-hundred and sixty-six EpC were opened in targeted communities, materials were delivered to those EpC, and enough educators were trained to operate those EpC. EpC facilitators in small and large communities largely engaged in the instructional practices they were trained on for EpC and were much more likely to do so than teachers who were not trained. The challenges CARS confronted in implementing the EpC intervention related to providing sufficient continued support to trained EpC facilitators and to distributing adapted reading materials in a timely fashion to EpC.

Low adherence to the experimental protocol resulted in low contrast between treatment and control groups in large communities. While all planned EpC were delivered, there was considerable non-adherence with the experimental protocol in large communities, where treatment and control children lived in the same community. In those communities, less than half of children assigned to the EpC treatment group reported participating (low take-up) and nearly one-fifth of children in the control group reported participating in EpC (low compliance). In small communities, the non-adherence was not nearly as stark: 73 percent of children assigned to EpC and only about 10 percent of children assigned to the control group reported participating in EpC. Therefore, this report primarily focuses on small communities, though we also present findings for the full sample as laid out in the original analysis plan.

In small communities, EpC had positive impacts on child reading outcomes, but not school attachment or social-emotional outcomes. EpC had a positive impact of 0.12 standard deviations on decoding (2.1 invented words per minute) and of 0.11 standard deviations on oral reading fluency (4.4 words per minute), which are two of the first emerging literacy skills that EpC target. These impacts represent about a 10 percent difference from the control group's mean for these outcomes. EpC also improved reading comprehension, with children in EpC scoring 4 to 5 percentage points better than control group children did.

To the extent we found impacts of EpC, such impacts may have been generated rapidly (by the time base-year data were collected) and did not increase over time. We had found differences between groups in some literacy outcomes during our base year data collection in Cohort 2, which occurred after many EpC had already opened. When adjusted for these base-year differences, the endline impacts on literacy skills were attenuated and became non-significant, or significant at the 10 percent level for reading comprehension.

Learning impacts were statistically significant for girls and for children who were out of school at intake, but not for boys and children who were already enrolled in school. Girls in the treatment group were able to read correctly 3 more invented words and 5.4 more words in one minute, and to answer about half of a question more about the reading passages than girls in the control group. Among children who were not enrolled in school at intake, those in the treatment group correctly read 4.4 more invented words per minute and 11 more words per minute than children in the control group could, and were able to answer correctly nearly one more question out of a total of five questions about the reading passages than children in the control group were. For children who were enrolled in school at intake, the single significant impact, for a reading comprehension outcome, was only significant at the 10 percent level. The differences in impacts by subgroup (between boys and girls, and between out-of-school and enrolled children) were not statistically significant, so we cannot conclude that there were differential impacts across subgroups.

If we consider both small and large communities together, findings suggest that impacts on child reading outcomes were too small to detect, with only one outcome showing a statistically significant difference. Children in EpC were able to decode 1.4 more invented words than children in the control group were, a difference of 0.08 standard deviations.

The lack of impacts on school attachment and child social-emotional outcomes can be due to multiple reasons. EpC might not have affected attendance (our main attachment indicator) because it was already high overall at baseline (over 70 percent in both the treatment and control groups) and EpC was not designed to mitigate common reasons for children's absenteeism in the RACCS (e.g., illness, agricultural or domestic responsibilities, and long commutes). Alternatively, the lack of impacts might be due to characteristics of the intervention. EpC might have been more effective in affecting social-emotional skills if it had focused on fewer social-emotional skills for the full duration of the program.

We estimate cost effectiveness for the EpC intervention to range from $45 to $358 per 0.1 standard deviations in improvement in literacy skills, using our findings in small communities and two different approaches. The first approach uses estimates developed by the CARS project team that aggregated actual costs based on administrative data and were used in planning for new EpC in the region. Based on this method, one EpC is estimated to cost approximately $60 per beneficiary. The second approach uses the original budget from the CARS contract to estimate the costs associated with EpC design and initiation, and with the rollout of Cohorts 1 and 2 that represent about 38 percent of targeted CARS child beneficiaries. Based on this method, we estimate that one EpC costs approximately $850 per beneficiary.

D. Conclusions, limitations, and recommendations

Conclusions

EpC is a promising after-school approach to improving early grade reading outcomes in poor and hard-to-reach communities in Nicaragua. Although the evaluation did not detect impacts of EpC on all hypothesized outcomes for all populations, there was evidence of improvement in word-level decoding, reading fluency, and comprehension skills in small, largely rural communities where there was a meaningful contrast in participation status. These findings were driven by statistically significant impacts for girls and for children who were out of school at intake. Even though the effects are relatively small, our findings indicate that EpC is a promising approach to improve the literacy skills of Nicaraguan children in remote, under-resourced communities. Policymakers should consider the cost, however, because our estimates include the possibility that EpC requires a heavier investment than comparable interventions, which have yet to be tested for cost-effectiveness in the RACCS. However, the lower-bound cost estimate of delivering EpC in the RACCS is less expensive than in the Dominican Republic ($30 compared to $120 per child for 12 months) (ICF Macro 2011). To optimize the cost-effectiveness of future investments, we recommend identifying the core components of EpC that underlie improvements in children's skills and optimal program duration, as our results indicate that impacts may have taken place in the first months of exposure and did not increase over time.

Offering EpC to those that are hardest to reach can be expensive, but when done well can have impact. Children living in harder to reach areas (e.g., remote rural areas) often have lower educational outcomes and more room for improvement than children living in more accessible areas, and their families and communities are more likely to take up the intervention if there are few or no similar services available. In this evaluation, we found positive impacts of EpC on the reading outcomes of children living in hard to reach rural communities. In countries aiming to provide better access to high quality education, large investments may be warranted to improve the academic outcomes of the most disadvantaged populations, including those that are linguistically diverse, poor, and hard to reach.

EpC had challenges reaching older children who were not enrolled in school because they had either dropped out or had never enrolled. This intervention was unable to recruit older children who had dropped out of school. The evaluation is unable to test whether the reason is that EpC did not mitigate the need for older children to work, but prior research in Nicaragua indicates that labor competes with children's participation in school (e.g., Del Carpio & Macours, 2009). Without the right incentives, after-school programs designed to make school attractive might have limited success in overcoming outside constraints to pursuing education.

The findings from this evaluation can help us understand the early stage implementation of EpC in the RACCS. However, because the program continued to evolve, it would be worthwhile to evaluate again while in a steady state. Prior to this evaluation, the EpC intervention had been piloted in other countries and regions in Nicaragua. However, there were challenges specific to the RACCS that the implementing team discovered during roll-out. Therefore, CARS and USAID modified the EpC model in the Caribbean region for later cohorts, as well as for the extension of the project to Northern Caribbean Coast, largely to focus even more strongly on literacy. In addition, the implementing NGOs had no prior experience with EpC, so CARS trained the NGOs in EpC as the intervention was rolled out. Another evaluation of EpC in steady state in the RACCS would be informative.

Limitations

Due to low adherence we are unable to draw conclusions about the effectiveness of EpC in large, primarily urban communities. In large communities, where we implemented within-community random assignment, a sizable proportion of children assigned to the control group reported participating in EpC and there was surprisingly low take-up of EpC among children assigned to participate. These two factors reduced the contrast between the experimental groups and increased the probability of finding null impacts in large communities.

It was not possible to collect baseline data on all children's primary outcomes before the intervention was rolled out. Even though randomized control trials do not require pre-intervention data to draw unbiased conclusions about the impacts of an intervention, such data can increase the statistical power to detect intervention impacts by reducing the amount of unexplained error in the estimation model.

The cost effectiveness analyses conducted in this evaluation are based on imprecise cost estimates due to the complexity of tracking costs across multiple implementing partners and because the program had other objectives aside from what we studied in this evaluation. The analyses provide a broad range for the overall cost effectiveness of EpC in the

RACCS in Nicaragua. They reflect costs ranging from those required to start up a new program to those required to open a new EpC once the program activities are at steady state. Opening EpC in another context, with different languages and social and geographic landscape, would likely have different costs.

Recommendations

A systematic exploration of different packages of materials delivered as a part of EpC and differing exposure periods to EpC would be important to inform future investments in EpC. The EpC model includes many activities that may contribute to the impacts observed, but that also increase program complexity and make it more costly to implement. Identifying the core package of materials and activities, as well as the optimal exposure, that are most cost-effective would be worthwhile, especially if targeting hard to reach populations.

After-school interventions should be targeted to populations that are most in need of the intervention activities. Children in small communities were more disadvantaged, participated in fewer academic and recreational activities outside of school and performed worse in the reading assessments compared to those in large communities. Limited access to extracurricular activities might have increased the perceived value of EpC and contributed to higher take-up among children in small communities. Also, that there was more room for improvement in reading skills possibly made it easier to generate and detect impacts.

Barriers to participation specific to the populations targeted by the intervention should be addressed or the targeted population should be narrowed to those most likely to benefit from EpC activities. Most of the out-of-school children identified as eligible to receive EpC were 5-7 years old and had never been to school, indicating that EpC was not successful at recruiting older out-of-school children who had either dropped out or had never enrolled. Prior research in Nicaragua indicates that labor competes with children's participation in school (e.g., Del Carpio & Macours, 2009). When the EpC model was first tested in Nicaragua, one of the recommendations was to add components such as income generation and technical training for families of working children to compensate for income lost when children leave work to return to school (Macro International Inc. 2009). It is also possible that the EpC activities were not attractive to older children. In Colombia, the EpC model was implemented alongside "Spaces for Entrepreneurship" (Espacios para Emprender, in Spanish) to respond to the needs and interests of different age groups, which targeted youth between the ages of 15 and 18 (ICF Macro 2010).

In randomized-controlled trials with high risk of low take-up in the treatment group and of contamination in the control group, group-level randomization should be preferred over individual-level randomization. Faced with trade-offs between the within-school and the between-school random assignment design (Glazerman 2012), it is tempting to use within-school random assignment because it enables the program to serve as many children as possible while also generating information about impact. We recommend conducting a pilot to estimate the level of contamination in the control group and the level of take-up in the treatment group prior to evaluating. Pilots can also offer an opportunity to detect barriers to take-up and test mitigation strategies. If a pilot is not feasible, and the risk of contamination is deemed high, group-randomization is advisable despite the higher costs that come with using larger sample sizes. When there is uncertainty regarding the risks of contamination and low take-up, hybrid designs, like the one employed in this evaluation, are a useful strategy to balance different risks.

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

I. CONTEXT FOR THE EVALUATION

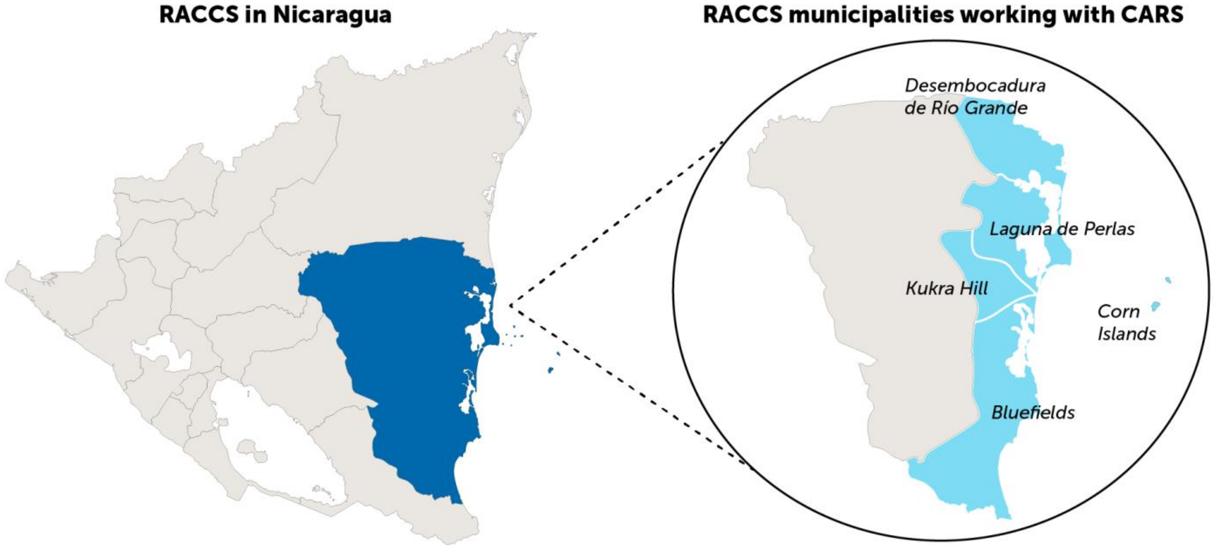
Nicaragua is the poorest nation in Central America and the second poorest in Latin America: 48 percent of Nicaraguans live on less than $1 a day and 76 percent live on less than $2 a day. The average number of years of schooling in Nicaragua is 5.8, which is the second lowest in Central America, and the national primary-level dropout rate is 10 percent (USAID/Nicaragua 2013). Across Nicaragua, approximately half of pre-school-aged children in rural areas are out of the system with the consequential early start of lagging behind and repetition (World Bank 2017). In addition, Nicaragua struggles with poor security outcomes; according to data reported by the World Bank, the nationwide homicide rate in 2012 was 12 per 100,000 inhabitants, a number that increased 72 percent from 1998 to 2009 (DevTech 2013). In 2018, the homicide rate was estimated at 7 per 100,000, a reduction from previous years (U.S. Department of State, 2018).

The Southern Caribbean Atlantic Autonomous Region in Nicaragua, known by the Spanish acronym RACCS (Región Autónoma de la Costa Caribe Sur), is the second-poorest of all Nicaragua's regions and departments; more than half of the population there lives in extreme poverty. Transportation infrastructure is insufficient, and there are many communities in very remote, hard-to-access, rural areas. The region is multilingual; Spanish and English are the dominant languages of instruction in schools but Kriol/English, Ulwa, and Miskitu are also spoken.

The RACCS has poorer education outcomes than the rest of the country. The average number of years of schooling in the RACCS is 2.9 years. Twelve and a half percent of first graders on the Caribbean Coast met international standards for reading in Spanish, whereas 25 percent of first graders in other parts of the country did (USAID 2012). In addition, the primary-level dropout rate is 17 percent in the RACCS, nearly twice as high as the national rate. Children attend primary school in the RACCS for five hours per day, and many schools have two shifts serving different groups of students.

To address these issues, the U.S. Agency for International Development (USAID) Nicaragua funded the Community Action for Reading and Security (CARS) program in five municipalities in the RACCS from 2011 through 2017. CARS consists of several activities across four programmatic components—formal and non-formal reading programs, community engagement, local capacity development, and knowledge generation and management—which are intended to strengthen educational outcomes and community security. The activities include an after-school program that is the focus of this evaluation, as well as a formal school program in private and subsidized schools that includes preschool and changes to the reading curriculum in grades 1 to 3. In addition, CARS also works to engage parents and communities in investing in reading and school safety. As part of the project, CARS provides technical assistance to local nongovernmental organizations (NGOs) to strengthen their capacity to launch early grade reading and parent engagement activities. CARS also aims to generate knowledge and inform local stakeholders and policymakers about reading and security. DevTech Systems Inc. and its local partners in Nicaragua implemented the $13.1 million dollar CARS program from 2013 to 2019. Figure I.1 shows the location of the RACCS in Nicaragua and the RACCS municipalities in which CARS was working during this evaluation.

Figure I.1. Location of the RACCS in Nicaragua and its municipalities



A core component of CARS is Spaces to Learn, known by its Spanish name *Espacios para Crecer* (EpC), which serves children who are having or are at risk of having difficulties at school, have dropped out of school, or have never attended school. EpC was expected to serve about two thirds of CARS beneficiaries. The CARS project adapted the EpC model to the RACCS in Nicaragua—where Spanish is the primary language of instruction but the region is multilingual and has poor reading outcomes—broadening it to incorporate new early-grade reading teaching methodologies in local languages.

DevTech Systems originally designed the EpC after-school approach with funding from the U.S. Department of Labor (USDOL) International Labor Affairs Bureau, to eradicate child labor and protect children's right to education. The approach was based on the Quantum Learning method, which uses such methods as accelerated learning, neurolinguistic programming, and experimental and cooperative learning to create a dynamic learning environment. In Nicaragua, DevTech also integrated the I Learn, Practice and Apply approach (APA, for its name in Spanish), which encompasses activities such as a welcome circle, recreation, and activities to link lesson content to the family and community contexts (DevTech 2013). EpC consists of three main components that aim to complement the formal school curriculum: (1) academic performance, (2) personal growth, and (3) recreation. The first component reinforces areas in the national primary education curriculum, and includes homework support and academic reinforcement or "leveling" in math, reading, and writing. The second component, personal growth, emphasizes the development of life skills. The third component includes games, sports, and artistic activities aimed at making learning enjoyable (DevTech 2013). The EpC model was first implemented in the Dominican Republic with promising results (ICF Macro 2011) and was piloted on the Caribbean coast of Nicaragua as part of Primero Aprendo, a project the USDOL funded to

> **"[The EpC] is where the child can be exposed to an environment where teaching is focused on games, activities, music, etc. EpC give [children] an opportunity to fall in love with reading and education. . . . It's a space to share their experiences, and do work in a fun way."**
>
> ***— DevTech representative***
>
> (Blair et al. 2017)

identify best practices to eradicate child labor from 2004 to 2008. A qualitative evaluation of Primero Aprendo indicated that EpC was a promising model to reduce child labor and increase school access and retention in Nicaragua (Macro International Inc. 2009). In 2010, the Department of Jinotega in Nicaragua replicated the model. Monitoring data collected by the implementing partner supported the conclusions from Primero Aprendo, indicating high rates of retention in EpC and reductions in the number of hours children dedicated to work (CARE 2010). The model was also implemented in Colombia from 2007 to 2010, with promising results (ICF Macro 2010). We discuss in greater detail in Chapter II the EpC model as originally adapted by CARS.

Evidence is inconclusive regarding the effectiveness of after-school programs[1] such as EpC on educational and non-academic outcomes for children with varying backgrounds. A meta-analysis of 68 studies (of which only 35 percent used a randomized design) in high-income countries found that high quality after-school programs that foster personal and social skills lead to improvements in children's attachment to school, positive school behaviors, school grades, and academic achievement. The mean effect size on achievement test scores was 0.31 standard deviations, which is comparable to results from other school-based educational interventions (Durlak et al. 2010). In contrast, a systematic review and meta-analysis that centered on the effects of after-school programs on at-risk primary and secondary students in the United States found small and nonsignificant effects on behaviors and school attendance, but did not examine effects on academic performance. The meta-analysis included 24 studies that used experimental or quasi-experimental designs only (Kremer et al. 2015).

Evidence is scarce on the effectiveness of after-school programs on educational and non-academic outcomes of children in low- and middle-income countries. Most of the existing studies on after-school programs rely on research designs that do not permit causal conclusions about program impacts (for example, studies use monitoring data collected by the implementer, qualitative designs, or pre-post designs without a comparison group[2]). Two exceptions include studies of the School Dropout Prevention Pilot in Tajikistan and Timor-Leste and Glasswing's After-School Clubs in El Salvador. The School Dropout Prevention Pilot found that an early warning system to help teachers and school directors identify and support students at risk of dropout, combined with an after-school intervention to facilitate engagement in school, had a positive impact on at-risk students' attendance and emotional attitudes toward school in Tajikistan, and on behavioral attitudes toward school in Timor-Leste (Creative and Mathematica 2015). A small experimental evaluation of Glasswing's After-School Clubs found that the program, implemented in five public schools in El Salvador, had small but significant effects on misbehavior (reduction of 0.17 standard deviations), school absenteeism (23 percent decrease), and school grades (increases of 0.11 and 0.13 standard deviations for math and science, respectively) (Dinarte 2017).

Given the inconclusive and scarce evidence discussed earlier, as well as the overall CARS investment, the fact that the EpC component was going to serve two thirds of the 12,500

[1] After-school programs are defined as offering one or more activities outside of normal school hours, during at least part of the school year, under the supervision of adults (Durlak et al., 2010).

[2] See, for example, the study conducted in Colombia in 2010 by ICF Macro (ICF Macro 2010).

beneficiaries targeted by the overall program, and that the EpC approach was increasingly being used in LAC, and in Nicaragua in particular, it was important to assess whether EpC, as adapted for the RACCS, was having the intended effects. This evaluation is one of a suite of evaluations conducted under the Latin America and the Caribbean Reads (LAC Reads) evaluation contract, which aims to address the need for high quality evidence on interventions that improve student outcomes, particularly in reading. USAID contracted with Mathematica Policy Research as its independent evaluator to design and conduct impact evaluations and cost-effectiveness analyses of promising reading interventions and education-access interventions in the LAC region. Through LAC Reads, Mathematica has conducted evaluations in El Salvador, Guatemala, Honduras, and Peru, as well as this evaluation in Nicaragua. In this report, we present the final results of our impact evaluation of EpC as well as a discussion of the program's cost-effectiveness. This is the first evaluation to estimate the causal impacts of EpC on children's outcomes, specifically children's attachment to school, literacy skills, and social-emotional outcomes.

II. INTERVENTION: AFTER-SCHOOL ENRICHMENT PROGRAM FOR CHILDREN IN EARLY GRADES

A. EpC

The CARS project adapted the EpC model to support children's acquisition of literacy skills, strengthen children's attachment to school, improve overall academic performance, and support social-emotional development, through which it can act as a bridge for out of school children to enter the schooling system and a safety net for those in the schooling system. To do this, EpC offers opportunities for children to participate in recreational and learning activities meant to make learning more enjoyable. The intervention is a half-day program offered each school day in the afternoon for children who go to school in the morning, or in the morning for children who attend school in the afternoon.[3] Because the program is implemented mostly in communities where other similar programs do not exist, it provides additional learning time (three hours per day, five days per week, for approximately 18 months) to children who are considered at risk. The project defines children at risk as those who are having or are at risk of having difficulties at school, have dropped out of school, or have never attended school.

Figure II.1 presents a theory of change of the intervention that shows how EpC activities are expected to lead to improved academic and non-academic skills. The intervention provides learning materials—reading materials for children and instruction materials for EpC facilitators — that meet the linguistic and other needs of children in the RACCS.[4] EpC facilitators, who are often teachers in the school, receive training and support to implement: (1) methods, techniques and activities that aim to facilitate learning and create a positive social environment; (2) social development and academic leveling activities that include grouping children by grade level ability (a best practice; McEwan, 2015) and having children share with their peers what they learned during the EpC and what they can use it for; and (3) methods to assess and provide differentiated responses to meet children's individual needs. Implicit in the theory of change is the assumption that training and coaching of educators in EpC methodologies will lead to changes in their behavior (best practices in teaching in general, and teaching reading in particular) and in the classroom environment. Those modifications, in turn, are expected to lead to improvements in child outcomes. For this reason, we consider changes in instructional practices to be intermediate outcomes or antecedents to changes in children's outcomes.

Through increased learning opportunities and the enjoyable atmosphere in EpC, children are expected to become more engaged in and attached to school. A major focus of EpC is on developing children's literacy skills, in particular the reading comprehension skills fundamental for children to do well in school. The program also targets the development of decoding and fluency skills, which are requisites for reading comprehension (Gove, 2009). Children are expected to develop their social-emotional skills through visual arts, dance and movement, and collaborative and teamwork activities that embed instruction on values, positive habits, and that

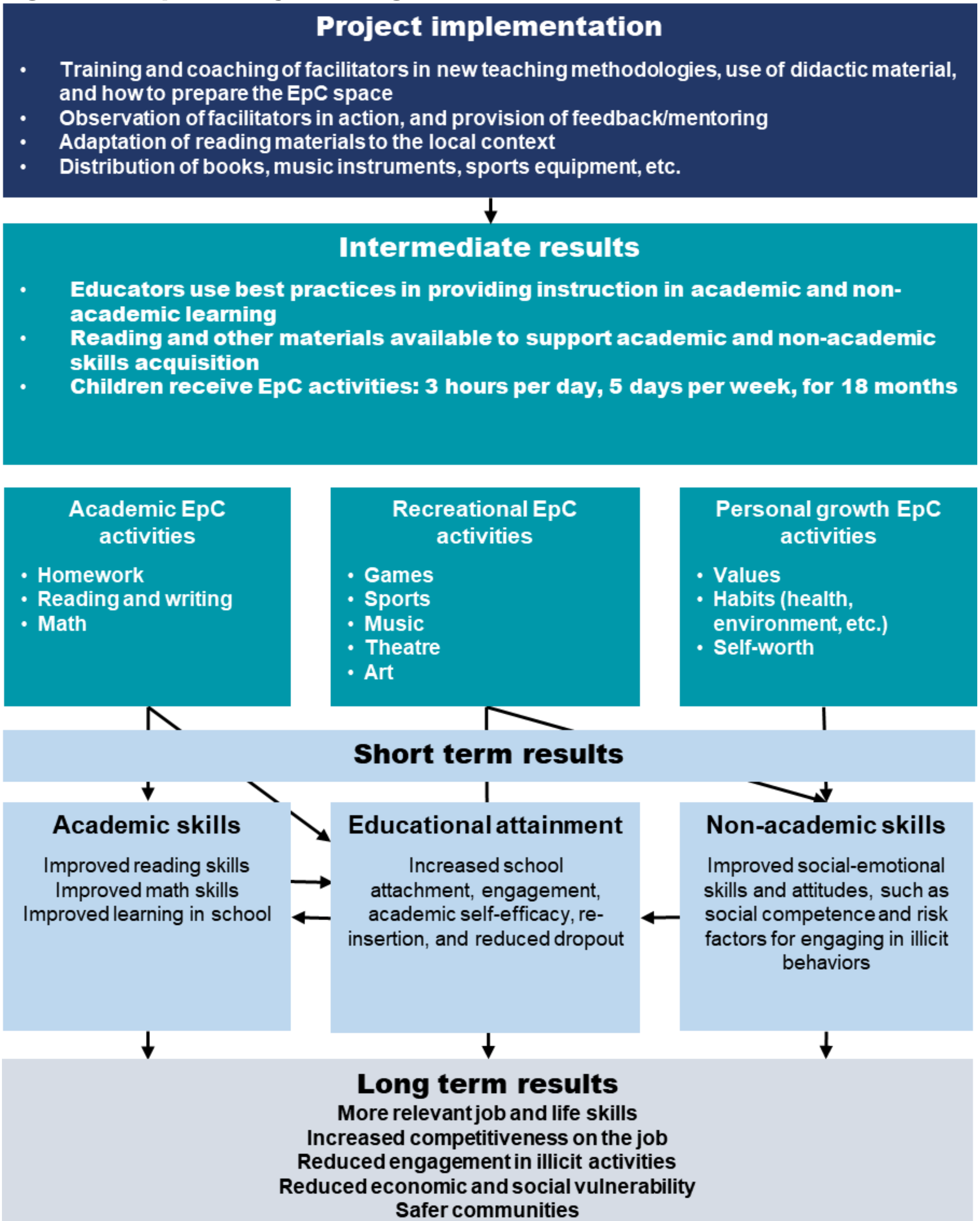
[3] Children who are out of school were required to enroll in school in order to attend EpC.

[4] The intervention design called for adaptation of materials to the local context, and that work began at the same time as roll-out of EpC. Therefore, the EpC included in the impact evaluation did not receive reading materials for children adapted to the context of the RACCS. The cohorts of children in this evaluation received reading materials in Spanish, Miskitu, or English that were already available for purchase.

seek to promote self-worth. For example, EpC modules cover topics such as "identifying emotions," "expressing emotions and feelings of affection," and "dialogue," which seek to promote self-regulation and communication skills, both of which are essential to developing and maintaining good relationships with others, and achieve other positive life outcomes. Effective after-school programs that foster social-emotional skills (such as relationship skills) have shown positive results in children's attachment to school, reduced problem behaviors, and school performance (Durlak et al., 2010), all of which are areas that EpC seeks to influence.

By spending time in a safe learning environment after or before school, children have the opportunity to develop these desirable skills rather than engaging in other (potentially less productive or even harmful) activities, such as child labor, or being exposed to violence or abuse. The EpC model posits that in the long run, the increased attachment to school, accelerated learning in reading and academics in early grades, and improved social-emotional skills should result in improvements in educational achievement and employment, and reductions in children's participation in illicit or unsafe activities. This should reduce the economic and social vulnerability of children and of the region as a whole.

Figure II.1. EpC Theory of change



Source: Authors, based on interviews with CARS project staff and consultation of research literature on after-school programs.

B. Eligibility for EpC

The team implementing CARS, in consultation with local education authorities, identified communities in five municipalities in the RACCS (shown in Figure I.1) to be eligible for the project, as well as children within those communities eligible to participate in project activities.

To be eligible to participate in the EpC intervention, communities had to be accessible (that is, the implementing team had to be able to travel to the community within a reasonable time frame and at a reasonable cost[5]), have schools that offered first to third grades, and have at least 20 children who were eligible to participate in the EpC intervention. The CARS team specified the number of children that an EpC could serve (20 to 35), balancing the need to be small enough to be manageable but large enough to be cost-effective.

Children had to meet the following requirements, specified by CARS, to be eligible to participate in EpC:

- School age (6 to 16 years old; 5-year-olds in first grade were also eligible).[6]
- Out of school with the equivalent of a 3rd-grade education or less, or enrolled in first, second, or third grades. At the outset of the project, CARS staff expected that many eligible children in the region would be on the upper end of the age range and would have dropped out of school, and they worked to identify and recruit such children.
- If enrolled in school, classified by school staff as being "at risk," meaning having one of the following risk factors:
  - Failed a grade or had poor academic performance (that is, low grades)
  - Showed high rates of absenteeism
  - Reported a mother tongue different from the language of instruction
- In some particularly impoverished communities, all children in the community were eligible.[7]

C. Counterfactual

This evaluation aimed to compare the outcomes of children who were offered the opportunity to participate in EpC, to the outcomes of the same children had they not been offered this opportunity. *This counterfactual* cannot be observed, but as discussed in the next chapter, is approximated in the evaluation design by using random assignment to form a control group that is equivalent to the treatment group, except for the opportunity to participate in EpC. Here, the

[5] In practice, in order to reach the target number of beneficiaries and meet the evaluation design requirements, the CARS team recruited communities that were harder to reach than initially anticipated.

[6] Five-year-olds enrolled in 1st grade accounted for only 1.3 percent of all children recruited to participate.

[7] In the RACCS, many rural communities lack the services covering basic needs such as easy access to potable water, lighting or roads/pathways to navigate the community, and have poor human development indicators with extremely low levels of literacy and high levels of child mortality. In addition, many schools have only one teacher covering more than one grade and do not have basic infrastructure. CARS considered all children in such communities to be eligible for EpC.

counterfactual (what children would have experienced had they not been offered the chance to participate in EpC) is access to the full set of programs offered by CARS and any other educational opportunities available in the community, with the exception of the EpC.

The CARS program included several other activities that took place within both the treatment and control groups and that could have affected the implementation and estimated impacts of EpC. Two activities are of particular importance, but there were other activities occurring in the regions, including new preschools being opened and private subsidized schools being supported. First, parents of school-age children in communities receiving CARS activities, including EpC, were invited to participate in an adult education program for parents, known as *Escuelas para Padres y Madres* or EPM. EPM are conducted in public and private schools regularly, and the CARS project supported these activities, providing content and encouraging participation in communities where the project was implemented. Some of the content of EPM could reinforce themes related to EpC, such as the importance of attending school and reading. Second, as a part of the broader community engagement activity, there were communication campaigns in the entire region. CARS developed and transmitted radio spots and programs, bulletins, and other means of communication directed at parents, teachers, children, and leaders. The objective of these communications was to spur community discussions around education and the importance of reading, security, empowerment, participation, inclusion, and sustained community development, all of which could have an impact on children's education outcomes (DevTech, 2013).

In addition to direct or indirect exposure to EPM and communication campaigns, the assumption was that children in eligible communities would have limited access to after-school academic activities other than EpC. Instead, they would spend their time in recreational activities, such as playing games or watching TV, helping with domestic chores, or working for pay.

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

III. EVALUATION DESIGN

The main purpose of this evaluation is to generate rigorous evidence on the effectiveness of EpC at improving children outcomes. We examine the impact of EpC relative to prevailing practice, on children's literacy skills, attachment to school, and social-emotional outcomes.

A. Evaluation questions

The main questions this evaluation addressed are:

- Was the EpC intervention implemented as intended? What services were delivered at the child, school, teacher, and community levels?
- What impact does the EpC intervention have on reading skills, school attachment, and social-emotional skills and attitudes?
- What are the costs of EpC? Are the effects of EpC large enough to justify the costs?

To answer the first evaluation question we use household surveys, school visits, focus groups with parents and with teachers, and in-depth interviews with school directors from a performance evaluation of the package of services delivered by CARS as a whole (see Blair et al. 2018). We answer the last two evaluation questions with a randomized controlled trial (RCT) which is described in greater detail in the evaluation plan (Bagby et al. 2016) and summarized in the next section.

B. Impact evaluation design: RCT

The research design for the impact evaluation of EpC consists of randomly creating two experimental groups, one in which children receive the EpC intervention (the "treatment" or "EpC" group) and another in which children do not (the "control" group). A distinguishing feature of this evaluation is that we used a flexible random assignment protocol, adapted to local conditions and the needs of the program implementation. The flexible protocol consisted of randomly assigning different units—children or educational communities—depending on the size of the educational community (an educational community is defined as a school and its immediate catchment area.) We refer to this approach as a hybrid design, because we use two different units of assignment

1. Hybrid randomization design: child- and community-level randomization

Educational communities vary in size from just a few children to hundreds of children. This is important because EpC required a minimum number of 20 children to be cost-efficient, and they would become unwieldy if they had more than 34 children[8]. We categorized those

[8] CARS had to evaluate each prospective community's ability to support an EpC without having to exclude a small minority of children (for example, if there were exactly 36 or 40 eligible children) and its ability to form a control group of eligible children within the same community. During the process of identifying and recruiting children eligible to participate in EpC and the evaluation—which took place over approximately 18 months —recruiting children became more challenging, so we adjusted the minimum thresholds. Initially, the CARS program planned to open 270 EpC, and each EpC would have not fewer than 30 or more than 35 children. However, due to the number of eligible children and communities in the region, as well as program beneficiary targets, CARS decided to be flexible in these goals, so some EpC served as few as 17 children and others as many as 37.

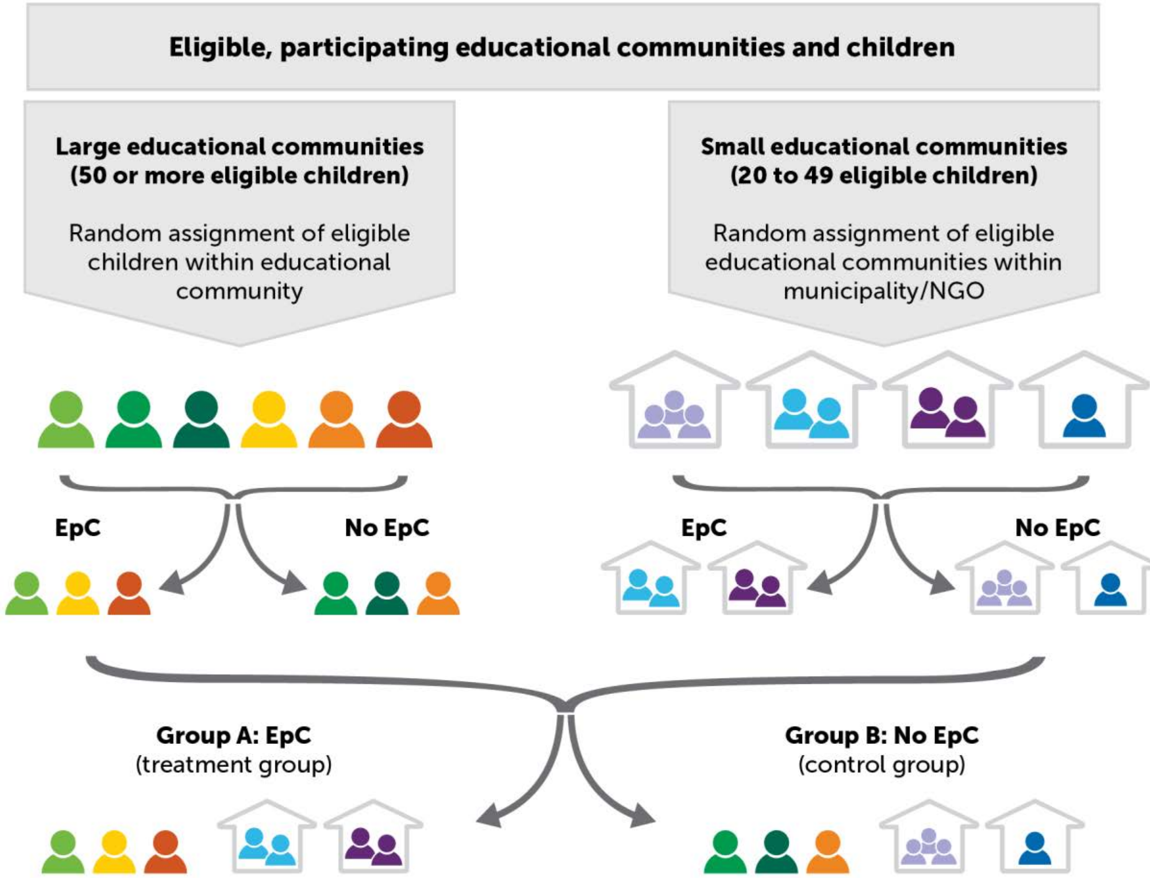
educational communities that had enough children to support an EpC (at least 20 children), as either *small* or *large*.[9] Small communities had at least 20 and fewer than 50 children. In practice, this meant that they were large enough for one EpC, but not large enough to assign half of the children to a control group as well. Large educational communities were those with 50 or more children, which was enough for at least one EpC and one control group. Thus, in small communities we assigned *communities* as a whole to either the EpC or the control group, while in large communities we randomly assigned *children* to either group. As a result, in small communities all eligible children belonged to the same experimental group (EpC *or* control) and the risk of contamination was low, whereas each large community included children assigned to the EpC *as well as children assigned to the* control group and the risk of contamination was higher than in small communities, because children in the treatment and control groups shared the same community[10].

Figure III.1 illustrates how we conducted random assignment depending on the size of the community. After random assignment across all communities in the evaluation, we have two groups of children: those in Group A are assigned to receive the EpC intervention and those in Group B (control) are assigned not to receive the EpC intervention. Of 9,223 recruited children across 219 educational communities, 4,542 are in the control group, and 4,681 are in the EpC group.

[9] This terminology has been updated from the baseline report, where we refer to large, medium and small communities. At baseline, small communities referred to those that were too small to be eligible to participate in EpC, while large and medium sized communities could participate. For simplicity, in this report we refer to large and small communities instead of large and medium.

[10] Within the small communities with 35 to 49 eligible children, there was a wait list from which program staff could fill slots as they became open using child random assignment. Wait-listed children were not a part of the evaluation sample frame.

Figure III.1. Random assignment[11]



For the impact analysis, we classified children according to the group to which they were originally assigned, regardless of their actual participation. This approach is known as intent-to-treat (ITT) because the analysis ignores non-compliance and therefore maintains the balance generated from the original random treatment allocation. ITT produces a conservative impact estimate and minimizes the risk of bias that may be introduced by comparing groups that differ in variables that may be associated with non-adherence (take-up and compliance) to the experiment. For example, among children offered the opportunity to participate in EpC, those who actually participate may have more motivation to learn than those who do not. Therefore, excluding children who do not adhere to their randomly assigned group may result in an unbalanced sample and result in biased impact estimates. Excluding children from the EpC group who did not take-up the EpC intervention from the analysis would bias the impact estimates upwards since highly motivated children who are more likely to reap the benefits of the program

[11] Following the original analysis plan, this report presents estimates of the difference between Groups A and B (at the bottom of the figure), using the full sample of large and small educational communities. However, we emphasize the comparison between "EpC" and "No EpC" groups in small communities (right side in the figure) because non-compliance in large educational communities reduced the contrast between the experimental groups, making it unlikely to detect impacts of EpC in those communities, as well as in the full sample of large and small communities. Results from large communities are included in the appendices.

would be overrepresented in the treatment group. Excluding children from the control group who did take-up the EpC intervention from the analysis would likewise bias the impact estimates.

2. Timing of randomization with phased rollout of EpC

The CARS team followed a phased approach to implement the EpC intervention, beginning in 2014, and the evaluation followed suit. We identified two cohorts of children to include in the evaluation on the basis of the year in which they entered the EpC program: Cohort 1 began in 2014; Cohort 2 began in 2015.[12] Within each cohort, there was a further degree of phase-in: one group started early in the year and another group started later in the year. We refer to these groups as Cohorts 1A, 1B, 2A, and 2B. The CARS team hired three different local NGOs to be responsible for rollout of EpC as part of the project's effort to build local institutional capacity and promote sustainability of activities.[13] The timing of the recruitment and intake processes of eligible beneficiaries varied slightly depending upon the NGO and location. Implementation started in the first half of 2014 for Cohort 1A, at the end of 2014 for Cohort 1B, from April to August 2015 for Cohort 2A, and from September to November 2015 for Cohort 2B.[14,15]

The two implementation cohorts have different representations of small and large communities. Cohort 1 consists of 46 large communities and only 8 small communities; Cohort 2 consists of only 6 large communities and 139 small communities.

Once the CARS team completed the recruitment and intake process for each cohort, Mathematica used the intake data for the evaluation population to conduct random assignment. This means we conducted random assignment in batches, as children and communities were identified, following the phased rollout of the intervention.[16] CARS or Mathematica notified participants of the results. The CARS team notified Cohort 1A communities of the results of random assignment in March of 2014 and Cohort 1B communities in October and November of 2014. Similarly, Mathematica notified Cohort 2A communities of results from March through July of 2015 and Cohort 2B communities from August through November of 2015. The date of notification of stakeholders is important because when the program is rolled out determines the earliest point at which treatment effects could emerge.

[12] The CARS team implemented EpC in additional cohorts, but these were implemented too late to permit following the students as part of the impact evaluation.

[13] The organizations were Fundación Zamora Terán, Fundación Hermanamiento Rama, and Universidad de las Regiones Autónomas de la Costa Caribe Nicaragüense.

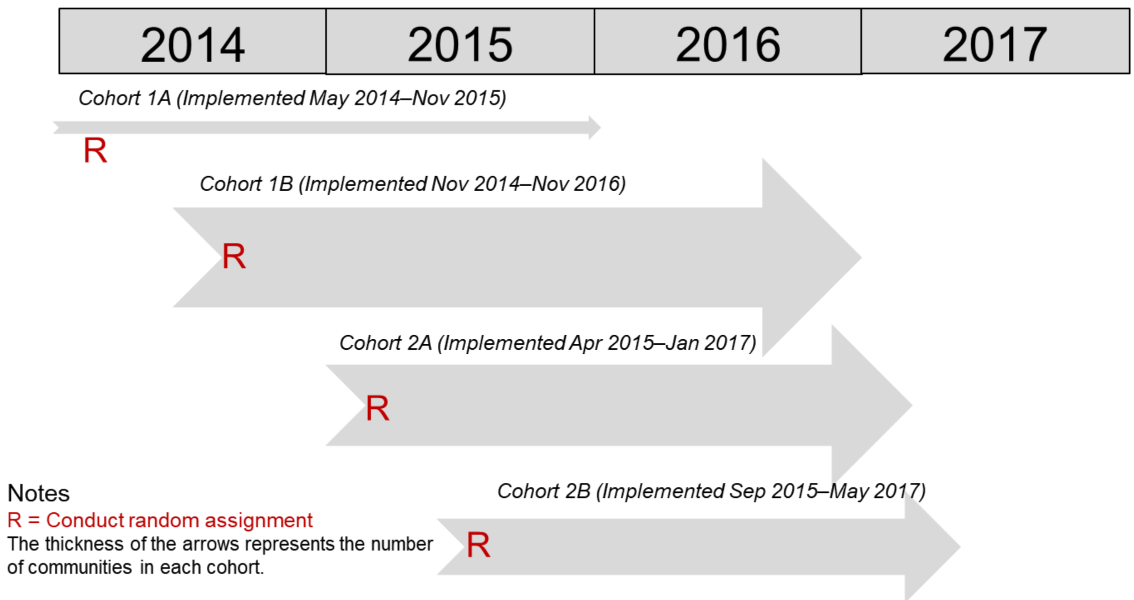
[14] Implementation for Cohort 2 was to start in the first half of 2015, but identification and recruitment of eligible educational communities were delayed. As a result, implementation in 27 of the Cohort 2A communities started in April 2015 and in June and August of 2015 in the remaining 37 and 57 communities, respectively. Similarly, implementation for Cohort 2B started in September 2015 in 26 of these communities, and in November of the same year in the remaining 11 communities.

[15] Of the 9,223 children in the evaluation, 586 were in Cohort 1A; 3,520 were in Cohort 1B; 4,154 were in Cohort 2A; and 963 were in Cohort 2B.

[16] We used the best available data at each round of random assignment to stratify the sample. We stratified on municipality and other community characteristics in small educational communities and on child characteristics, such as age, gender, and risk status, in large educational communities. At each round of random assignment, approximately half of children were assigned to the EpC group and half to the control group.

The goal of CARS was to provide EpC for each cohort for 18 months. However, due to the variation in the timing of random assignment and the school year, the length of time that EpC were open varied by cohort, ranging from 15 to 20 months (Figure III.2). As a result, the actual exposure to EpC can vary across children. Appendix A provides additional detail by cohort, on the number and size of educational communities, the number of EpC, the municipalities, the implementing NGOs, the notification date of the results of random assignment, the length of exposure at the time of data collection, and the start and end date of EpC activities.

Figure III.2. EpC phased implementation and random assignment



3. Impact estimation

In a simple RCT, impacts are defined as the difference between the two experimental groups in average outcomes at end line. The analysis is typically conducted using a regression that also incorporates additional information including data on baseline differences that may have arisen due to chance to increase the precision of the impact estimates. This enhances the researcher's ability to draw conclusions. Regression analysis also enabled us to estimate average impacts across both parts of our "hybrid" design, which is what we had originally planned, as we noted in the evaluation plan (see Bagby et al. 2016; Bagby et al. 2017a).

Focus on community-level randomization in small communities. Following the original evaluation design, we estimated impacts for the full sample and separately for small and large communities. This report primarily focuses on small communities because, as we explain below, compliance with the random assignment protocol differed considerably in large and small communities, in ways that threatened the validity of the experiment for the large communities.

The tradeoffs between community- and child-level randomization are well documented (Glazerman 2012). When whole communities are assigned to experimental conditions (as small communities in this evaluation were), lower statistical power is offset by lower risks of contamination and spillover. Community-level randomization bears a minimal risk that children in control communities would attend an EpC in a different community, that their literacy and/or social-emotional skills would change from interacting with children assigned to EpC, through peer influence, for example, or that they would be exposed to educators trained in EpC methodologies. All of these risks are greater when children in the same communities are assigned to different conditions (as were children in large communities in this evaluation), but statistical power to detect impacts is typically much greater too.

In this evaluation, the risk of low statistical power associated with community-level randomization in small communities was not realized. In Table III.1, we present the minimum detectable effect (MDE) for each of the designs. The MDE is defined as the smallest impact that the evaluation will be able to detect with conventional standards of significance and statistical power. Smaller MDEs are desirable, as they indicate more statistical power to detect intervention impacts. The MDEs achieved are nearly identical for small communities and for large communities. For analysis by out-of-school status, the MDEs are smaller for small communities relative to large communities, which is somewhat unexpected. The reason is that the sample of out-of-school children in large communities is much smaller than in small communities (7 percent of children were out-of-school at intake in large communities, compared to 29 percent in small communities).

Table III.1. Minimum detectable effects, in standard deviations (sd)

Sample description	Full sample (sd)	Gender (sd)	Out of school (sd)	Children in sample (n)	Communities in sample (n)
Estimated based on full sample, 80 percent response rate, and a set of assumptions	0.13	0.16	0.33	2,480	219
Achieved based on full sample	0.14	0.19	0.28	2,371	199
Achieved for small communities	0.20	0.25	0.30	1,182	147
Achieved for large communities	0.21	0.29	0.79	1,189	52

Source: Follow-up household survey.

Notes: We use a binary outcome measure (for example, whether or not a child is enrolled in school) with a mean of 0.5 and a standard deviation of 0.25. For the estimated values, we assumed an equal distribution of communities across EpC and the comparison group. We also assumed that our sample will comprise 10 children per eligible small community and 20 children per eligible large community, on average, half of which will be female and the other half, male. We assumed that 25 and 5 percent of children from small and large communities, respectively, will be out of school. We also assumed that 10 percent of the variation in the outcomes of interest occurs at the community level and that covariates explain 20 percent of the variation in each outcome. For the achieved values, we used actual follow-up data and sample sizes, including an R-squared of 0.2.

sd = standard deviation, n = number.

End-of-year data on children's participation in EpC, as reported by caregivers, suggest that there was lower than expected adherence to the evaluation design, especially in large communities. That is, take-up of the intervention was relatively low by those assigned to receive it, and contamination was relatively high (in that a non-trivial number of children assigned to not receive EpC did participate).[17] In large communities, children assigned to the control group had

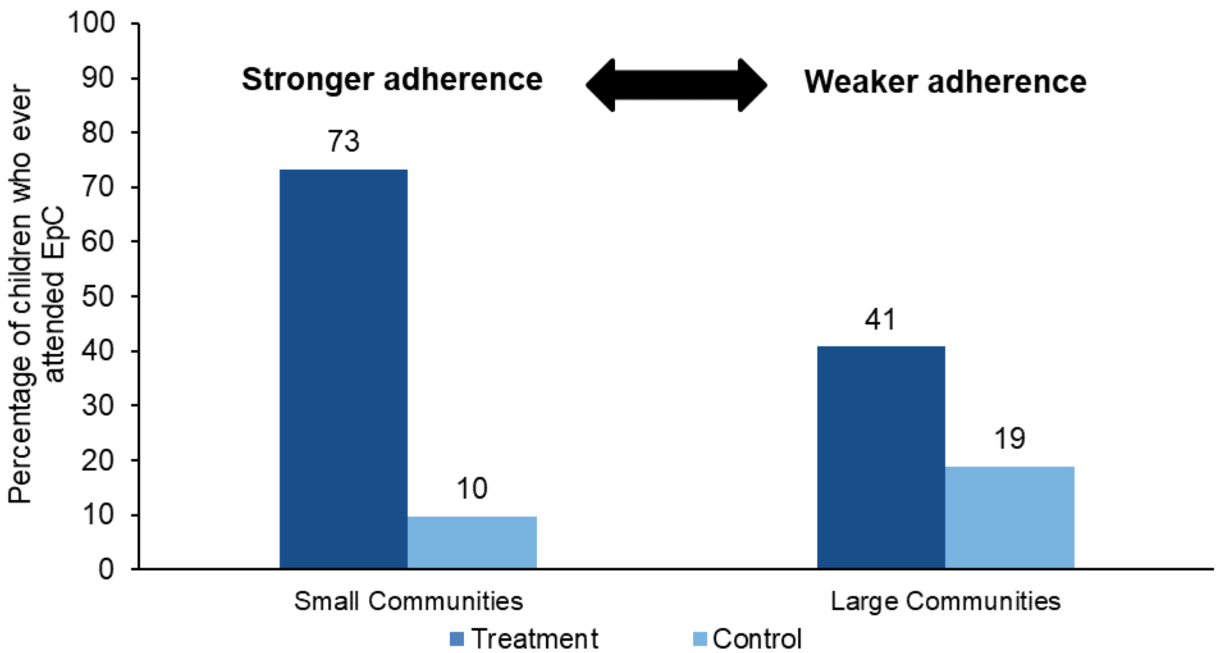
[17] We also examined individual-level attendance records collected in 5 small and 2 large EpC in 2016. Average attendance in small communities was about 50 percent; in large communities, the average attendance was about 40 percent in the records we reviewed. These numbers may include children who never attended EpC but they do not

higher rates of participation in EpC than children in small control communities, indicating that contamination was more acute in large communities (where children in the same community were assigned to the EpC or control conditions) (Figure III.3). These children's exposure to treatment reduces the differences between the treatment and control groups and therefore decreases the likelihood of our ability to detect impacts. In addition, unmeasured contact between the groups can further decrease our ability to identify impacts.

Additionally, compared to small communities, a smaller percentage of children assigned to the treatment group actually participated in EpC in large communities (Figure III.3). With any intervention, there are multiple potential reasons for low take-up by potential beneficiaries. Project implementers must both understand potential barriers to take-up and engage in actions to encourage take-up by potential beneficiaries that are the hardest to reach. In addition, interventions rarely, if ever, can overcome all barriers to take-up. Therefore, understanding barriers to engage the intervention's target population is important to setting realistic expectations about take-up. We discuss potential explanations for the low take-up in large communities in chapter VII.

Our findings indicate that the experiment worked well in the small communities but not as well in the large communities. Contamination and low take-up decreased the contrast between experimental conditions and our ability to detect the true impacts of EpC in large communities.

Figure III.3. EpC participation rates, by community size



Source: Follow-up household survey.

include all children assigned to receive EpC. Nonetheless, these data lend support to findings from the impact evaluation suggesting that EpC participation was much lower than expected.

For the reasons above (sufficient statistical power to estimate impacts in small communities only and low adherence to the experimental protocol in large communities), we estimate and describe intervention impacts for small communities while controlling for the specific cohort in our analyses. Following our original analysis plan, we also describe impacts for the full sample of small and large communities. We anticipate, however, that impact estimates for the full sample are diluted due to the contamination and low take-up in large communities. Details about our empirical specification are in Appendix C. We present impact analyses for the sample of large communities separately in Appendix J.

C. Impact analysis

We report findings for all key outcomes regardless of statistical significance, but we focus our discussion of impacts on results that are significant at least at a 5 percent level. It is important to recall that we conducted many tests, and when doing so there is a higher chance of false positives (falsely concluding that an impact is statistically significant). There are several methods available for adjusting the statistical analyses to account for the higher chance of false positives.[18] However, we focused our analysis and testing on the outcomes of primary interest that were defined before we collected data for the evaluation, based upon the program's theory of change. These outcomes are: (1) literacy skills (measured by decoding, oral reading fluency, and reading comprehension); (2) attachment to school (measured by attendance); and (3) social-emotional skills and attitudes (measured by social competence and beliefs that justify wrongdoing).[19] We considered findings for additional outcomes of interest as being useful for understanding the findings from the main outcomes and conclude that an impact estimate is meaningful only if it is part of a pattern. Thus, we report hypothesis test results based on conventional p-values, but have structured our discussion to give most weight to our primary outcomes, as defined in our analysis plan.

Child impacts. We estimated the impact on child education and social-emotional outcomes by comparing average outcomes of children in the treatment group to those of children in the control group. In our analysis, we used dummy variables to control for such child characteristics as age, gender, and school enrollment status at the time of intake to the evaluation. We also accounted for stratification used for random assignment using dummies for each stratum, the clustering of children in educational communities, as well as the oversampling of children who were out of school at intake using weights.

Educator analyses. CARS worked with the regional education authorities to identify educators in the region to train in the EpC methodology and assign them to communities selected to receive an EpC. In some cases, teachers in schools in EpC-receiving communities were trained and then worked as both regular teachers and EpC facilitators (teacher-facilitators). In other

[18] To mitigate against finding statistically significant impacts for a particular outcome are not simply statistically significant by chance because of the number of comparisons being made in that domain (known as the multiple comparisons problem; see Schochet [2009]), one can adjust the statistical significance level or select primary outcomes before beginning the analyses.

[19] In the literature, this is known as "moral disengagement", "neutralization" or "delinquent beliefs that neutralize guilt for offending" (Hennigan et al. 2014).

cases, individuals who were not currently teachers were selected and trained in EpC methodologies and then sent to communities to act as EpC facilitators (facilitators-only).

Because educators were not randomly chosen to participate in EpC training, and there is a high likelihood that those selected to participate were different from those that were not, we compare EpC facilitators' instructional practices to those of traditional teachers only to estimate non-causal impacts of EpC training on facilitator behaviors. In small communities, we compared EpC facilitators (including those who were both a teacher and a facilitator, as well as those who were only a facilitator) in the EpC group to teachers in the control group. The goal was to determine if EpC facilitators received training and acted on that training and used different instructional practices in EpC than typical teachers use in the classroom. If they did so, we might expect to see an impact on child outcomes. If they did not, we would be less likely to expect to see an impact on child outcomes (though we might see an impact from children spending the extra time in a safe learning environment). We conducted similar analyses in large communities, but those comparisons should be interpreted more cautiously since EpC facilitators and those who are only teachers were working in the same schools, increasing the risk of spillover from facilitators to teachers.[20]

EpC implementation analyses. We also present findings on the implementation of EpC in evaluation communities. For instance, we observed the availability of materials that were to have been delivered as a part of EpC and observed any differences between EpC and control communities. In such comparisons, we incorporated characteristics such as evaluation cohort. We compared small treatment and control communities in these analyses. In large communities it was not possible to compare the availability of school resources in EpC and control communities because all large communities received an EpC. However, we compared EpC facilitators to teachers in the same schools who were not trained on EpC to explore whether teaching practices, such as the use of didactic materials, differed between the two groups of educators. Even though we cannot attribute differences between the two groups to EpC, the findings can help us understand the larger pattern of results.

Additional analyses. We present results for several additional analyses that are important to understanding the impact of the EpC approach in Nicaragua. These include findings by children's gender and enrollment status at intake, two categories that policymakers identified as being important before data were collected for the evaluation. Gender effects of any program are a strategic focus for USAID across programs it supports. The EpC intervention aims to re-enroll older children who had dropped out of school, so CARS identified this group as being of special importance and emphasized their recruitment.

We also present exploratory findings for additional outcomes that are of interest to the evaluation, but are not primary outcomes. These include alternative outcomes within each primary outcome domain, as well as on additional domains such as security related outcomes.

[20] In addition to contact among teachers in the same school, spillover could have also taken place more broadly through Ministry of Education staff who were invited to participate in EpC trainings, as well as through professional development workshops held by the Ministry of Education (for example "Talleres de Evaluación, Programación, y Capacitación" (TEPCE)), where teachers from different communities may have exchanged experiences and techniques.

These exploratory findings are included in the body of the report for small communities, with some exceptions, such as alternative calculations of some variables, which are included in Appendix H. Exploratory findings for the full sample are described in Appendix I. Finally, we conducted several additional analyses to determine if our results are robust, including analysis adjusting for child base-year literacy skills.

D. Implementation and cost-effectiveness analysis

We conducted a descriptive analysis of the implementation of the EpC intervention using qualitative and quantitative data. In addition to the base-year and follow-up surveys for this impact evaluation, we collected quantitative data that included CARS monitoring and evaluation (M&E) indicators and data from educator interviews. We also collected qualitative data including programmatic reports, semi-structured interviews, and focus groups with stakeholders. These data provided information on intervention rollout, take-up, and exposure that helped us to understand our impact evaluation findings.

Although implementation analysis is important for interpreting the impact findings, because it describes the intervention that generated the impacts, a cost-effectiveness analysis is needed to interpret whether the impacts are large enough to justify the investment. This evaluation includes a simple cost-effectiveness analysis that estimates the average cost of implementing EpC and compares that cost to the impacts it generated, with both costs and impacts reported on a per-child basis. Average costs incorporate fixed costs that would be incurred regardless of how many EpC were opened and how many children were served. For example, they include the cost of developing and implementing facilitator training and of staffing an organization to recruit communities. For this reason, we describe the scale of the program as implemented for the evaluation and caution that average costs per child may be higher when the program is rolled out on a smaller scale, or lower when it is rolled out on a larger scale or for a shorter time. Other costs vary by EpC or by child. EpC-specific costs, such as paying for a facilitator's training and stipend, will be higher (per child) if EpC serve fewer children. It is also important to recognize that input prices, such as for labor or transportation, may vary from context to context, so the cost for EpC on Nicaragua's Atlantic Coast may be different from the cost of a similar program that might be implemented in another regions in Nicaragua, or in other countries.

IV. DATA

The data for this evaluation come primarily from households and schools. The CARS team collected basic information from teachers or school leaders on the age, grade, and academic risk factors of most children in the evaluation as part of the program intake process, and shared the data with us for the evaluation. Mathematica hired the Fundación Internacional para el Desafío Económico Global (FIDEG) to conduct base-year and follow-up household visits, including interviews with children and their primary caregivers, as well as a child literacy assessment. FIDEG also conducted school visits at follow-up, which included interviews with school directors, teachers, and EpC facilitators, as well as school and classroom infrastructure observations. Our evaluation team also worked with FIDEG to document implementation progress and challenges through site visits (during the performance evaluation mentioned earlier), including in-depth interviews and focus groups with school directors, educators, and parents, in addition to interviews with project implementers and USAID/Nicaragua staff at the midpoint of the intervention.

A. Timing of data collection relative to random assignment

We developed the evaluation design—including the data collection schedule—around the timing of the rollout of the intervention (see Figure III.1) to minimize delays in program implementation related to the evaluation and to gather the data efficiently. The intake data used to identify the population of children and communities eligible to receive the EpC intervention represent the baseline data for the evaluation. Intake data consisted of the child's name, age, gender, school enrollment status, grade, and in some cases, his/her maternal language, but did not include baseline literacy skills. We collected base-year data to measure children's literacy skills, but could do so only in Cohort 2 communities and, in some cases, after exposure to the intervention had already begun.[21] We collected follow-up data after approximately 18 months of exposure to the EpC intervention for each cohort (in 2016 for Cohort 1 and in 2017 for Cohort 2). See Appendix A for additional information.

B. Indicators and data collection sources

This impact analysis focuses on a set of primary implementation activities, intermediate outcomes, and child outcomes (shown in Table IV.1). At the school level, we focused on implementation outputs, specifically the availability of materials distributed by the program (for example, music instruments, sports equipment, art supplies, and reading materials); the percentage of educators trained in topics central to the EpC model (for example, how to teach

[21] To avoid delaying rollout of EpC, given the time required to conduct fieldwork for data collection, the timing of base-year data collection varied across communities from within 0 to 5 months of program rollout. This means that base-year (or pre-test) scores were collected after children were exposed, even if for a short period, to the EpC intervention. Collecting *late* pre-test data to improve the precision of estimations is typical in many randomized controlled trials in the education field (Schochet, 2010). In this evaluation, this occurred in approximately 94 percent of small communities in Cohort 2 (11 communities had not been exposed, 120 had between 1 and 3 months of exposure, and 27 had about 5 months of exposure). In large communities, base-year data were collected in only 7 of 53 communities (those in Cohort 2). Exposure to EpC ranged from 1 to 5 months. We were unable to collect similar data for Cohort 1 due to the timing of implementation activities and random assignment (Bagby et al., 2017b). We discuss the potential implications of collecting late pre-test data when we present the analyses of these data in chapter VIII of this report.

reading and writing, and the use of standardized assessments); and whether educators received continuous support from CARS.

Table IV.1. Primary child and educator outcome indicators

Domain	Construct	Description	Data source
Educator outcome indicators			
Instructional practices promoted by EpC	Strategies to promote student engagement	Educators answer questions about their use of music, traditions and routines, didactic materials, visual aids, and other strategies that facilitate student engagement.	Educator survey
	Strategies to promote concept development	Educators answer questions about their use of methods to promote the development of higher-order skills, including discussion about the motivations of characters in stories, allowing opportunities for students to ask questions and to work in pairs and small groups.	Educator survey
	Use of assessments	Educators answer questions about their use of student reading assessments	Educator survey
Child outcome indicators			
Attachment to school	Attendance at school	Caregivers reported the number of days the child attended school in the last month.	Household survey
Literacy – decoding	Pseudo-word (or invented word) decoding	Children read out loud as many pseudo-words as they could in 60 seconds. Pseudo-words are vowel-consonant combinations that follow the target language's phonological and spelling rules but are not actually meaningful words.	Child literacy assessment
Literacy – reading fluency	Oral reading fluency (correct words per minute)	Children were assessed on their ability to read out loud grade-level passages with speed and accuracy.	Child literacy assessment
Literacy – reading comprehens ion	Reading comprehension (passage 1, explicit questions)	Children were assessed on their ability to answer explicit questions about a grade-level passage they read silently to themselves.	Child literacy assessment
Social-emotional skills	Relationship skills – social competence	Social competence encompasses abilities to establish and maintain healthy and rewarding relationships with others (CASEL 2013). Caregivers rated the frequency of positive and negative interactions with their child to assess the child's ability to establish and maintain positive relationships with adults using a four-point scale (0 = "almost always" and 3 = "never," for example). Higher scores denote a higher level of social competence, and therefore are more desirable.	Household survey
Social-emotional attitudes	Risk factors for engaging in illicit activities (beliefs that justify wrongdoing)	Beliefs that justify engaging in unlawful or unethical behaviors are known as "moral disengagement" or "neutralization" and have been linked with involvement in illicit activities (Esbensen and Osgood 1999). Children rated their level of agreement with statements that justify wrongdoing using a four-point scale (0 = "disagree a lot" and 3 = "agree a lot," for example). Higher scores denote a higher level of risk, and therefore are less desirable.	Child survey

To examine any changes the educators might have made due to the EpC intervention, we focused on three instructional practices: educators' use of EpC strategies that promote student engagement, use of strategies to promote concept development, and use of reading assessments to monitor students' progress. Strategies to promote student engagement include the use of music and traditions and routines, arranging children in a flexible way, and using a diversity of materials to facilitate children's learning. Strategies to promote concept development refer to the degree to which academic activities promote higher-order thinking versus rote or fact-based learning. We chose to measure these strategies because reading comprehension, a primary outcome, requires students to go beyond fact-based learning and memorization to interpret and make inferences about what they read. Teachers also received training in the use of a simple reading assessment in the classroom. We measured whether or not teachers applied the assessment, but we did not attempt to understand how they used the result of the findings in a quantitative way. We measured these three outcomes using educators' self-reports.

The primary child outcomes on which we focus our findings (shown in Table IV.1) are targeted by EpC, namely school attendance (an indicator of school attachment), invented word decoding, reading fluency, reading comprehension (indicators of literacy skills), and social competence and moral disengagement (indicators of social-emotional skills and attitudes). We measured a dimension of social competence called relationship skills that encompass skills to establish and maintain positive relationships, such as listening, expressing ideas clearly, and emotion regulation (CASEL 2013). Moral disengagement refers to beliefs that justify wrongdoing. Even though EpC does not directly target such beliefs, it aims to promote positive values (for example, solidarity, honesty, and responsibility) that could impact children's beliefs and reduce the risk of their engaging in illicit behavior.

We measured additional outcomes and outputs at the school, educator, and household and child level. A detailed description of all indicators/constructs we used in our impact analysis, including definitions and measurement, are in Appendices F and G.

We used primarily survey data including child and caregiver interviews and literacy assessments applied in households, as well as intake data that had been collected for the evaluation population, in our impact analysis. We also used base-year data for sensitivity analyses such as the ones presented in Chapter VIII. In Table IV.2, we describe these data sources in more detail. In addition, we use qualitative data we collected to better understand implementation and contextualize findings. The survey instruments, data documentation, and data files used for this evaluation are available online upon request as restricted use data files at USAID's Development Data Library website (http://data.usaid.gov).

Table IV.2. Data collection sources

Data source	Survey respondent / informant	Purpose	Type of data collected	Additional notes
Baseline data				
Intake data	School director, households	Determine child and community eligibility and describe the evaluation population at baseline	Community characteristics, such as the number of eligible children and the language of instruction Children's age, enrollment status, current grade if enrolled, and information about their mother tongue	Cohort 2 includes additional school information, such as measures of school infrastructure quality and number of teachers
Base-year survey data				
Household survey	Primary caregiver	Characterize child's home environment	Household composition, adult's education level and occupation, household assets and migration, transportation, and quality of familial relationships	Collected only for Cohort 2
Child survey	Child	Capture child characteristics	Child characteristics, such as mother tongue, reading habits at home, and social-emotional skills	Collected only for Cohort 2
Child literacy assessment	Child	Measure child's early literacy skills in the base year	Base-year early literacy skills in the language of instruction in schools in the community (Spanish)	Collected only for Cohort 2
Follow-up survey data				
Household survey	Primary caregiver	Characterize child's home environment and child outcome data	Same as base year	
Child survey	Child	Capture child characteristics and child outcome data	Same as base year	
Child literacy assessment	Child	Measure child's early literacy skills at follow-up	Follow-up early literacy skills in the language of instruction in schools in the community (Spanish, English, and Miskitu)	
School director survey	School director	Capture school-level conditions	School characteristics including instructional resources, teacher characteristics, and teacher participation in professional development	

Data source	Survey respondent / informant	Purpose	Type of data collected	Additional notes
School register	School director, administrative data, EpC facilitator	Capture teacher characteristics and school and EpC enrollment and attendance	School and EpC enrollment and attendance data Teacher information, including grades taught and EpC facilitator status[a]	
Educator survey	EpC facilitator and teacher	Capture context and outcome data on teachers and EpC facilitators	Educator background, participation in professional development, and instructional practices	
School observations	Enumerators	Capture information about school infrastructure, EpC, and classrooms	Number of classrooms, adequacy of characteristics of classroom for teaching, and condition of school facilities	

Notes: CARS collected intake data, and FIDEG collected survey data. See Appendix E for additional information. Literacy assessments were adapted from Early Grade Reading Assessment (Dubeck & Gove, 2015) and the Dynamic Indicators of Reading Success (Baker et al. 2006)

[a] We had to shorten data collection for Cohort 2 in 2017 (compared to the duration for Cohort 1 in 2016), due to insecurity in the region, the amount of time necessary to travel between remote communities, and the need for CARS to begin opening Cohort 3 EpC (some in evaluation communities). To do so, we captured the total number of enrolled students for each school for Cohort 2; in Cohort 1 we had captured enrollment data disaggregated by grade and gender.

C. Characteristics of the analysis samples

In this section, we describe the samples of children and communities we use in the analyses, and test whether the EpC and control groups are balanced in a set of characteristics measured before the rollout of EpC and in the base year (a few months after implementation started). Intake (baseline) data were unavailable for 8.8 percent of children. In such cases, we used characteristics from follow-up survey data, such as child's maternal language, that were unlikely to have changed as a result of the EpC intervention. We present the measures that are most important for illustrating the context of the evaluation in this chapter, and provide additional information in Appendix D.

Drawing the analysis sample. The evaluation team selected a random sample of children from the evaluation population for primary data collection and analysis. We oversampled children who were reported to be out of school at intake so as to ensure adequate representation of this important group—which is smaller than the school-going population but of interest to policymakers and stakeholders. We drew a proportional number of boys and girls among those currently enrolled in school. We drew a replacement list to use in the event that a sampled child was actually ineligible for the EpC intervention. See Appendix B for details on our sampling strategy.

The analysis samples are representative of the evaluation population. The evaluation population consists of the universe of children determined eligible to participate in the evaluation. The analysis sample represents those who were sampled and successfully interviewed (response rates were approximately 81 percent in small communities and 87 percent in large communities and were comparable across treatment and control groups [see Appendix B]). As we show in Table IV.3, the evaluation population and analysis samples are similar in a range of characteristics in the full sample, as well as in the samples of small and large communities. Those characteristics include children's gender, age, enrollment status, and whether or not the child walks to school with another child. In addition, there is no differential attrition between experimental groups in either small or large communities (see Appendix B for details).

Table IV.3. Comparison of intake data between evaluation population and analysis sample, by community size

Characteristic	Evaluation population (A)	Analysis sample (B)	Difference (A-B)	Sample size (A)	Sample size (B)
Panel A: All communities					
Female (percentage)	46.8	48.4	-1.6	9,223	2,371
Age (years)	8.7	8.6	0.1	9,223	2,371
Caregiver attended community assembly during intake process (percentage)	64.9	66.8	-1.9	6,808	1,766
Grade level (percentages)					
Out of school	16.2	14.7	1.6	9,223	2,371
Grade 1	38.0	38.7	-0.7	9,213	2,368
Grade 2	24.9	25.8	-1.0	9,213	2,368
Grade 3	20.0	20.7	-0.7	9,213	2,368
Walks to school with another child (percentage)	28.3	25.7	2.6	9,223	2,371
Panel B: Small communities					
Female (percentage)	47.4	47.4	0.0	4,876	1,182
Age (years)	9.2	9.1	0.1	4,876	1,182
Caregiver attended community assembly during intake process (percentage)	68.1	71.2	-3.1*	4,410	1,057
Grade level (percentages)					
Out of school	25.7	23.7	2.0	4,876	1,182
Grade 1	35.8	36.6	-0.8	4,866	1,179
Grade 2	21.2	22.4	-1.2	4,866	1,179
Grade 3	16.8	17.2	-0.4	4,866	1,179
Walks to school with another child (percentage)	44.9	44.2	0.7	4,876	1,182
Panel C: Large communities					
Female (percentage)	46.3	49.4	-3.1*	4,347	1,189
Age (years)	8.1	8.1	0.0	4,347	1,189
Caregiver attended community assembly during intake process (percentage)	58.9	59.9	-1.0	2,398	709
Grade level (percentage)					

Characteristic	Evaluation population (A)	Analysis sample (B)	Difference (A-B)	Sample size (A)	Sample size (B)
Out of school	5.6	5.6	0.0	4,347	1,189
Grade 1	40.5	40.9	-0.4	4,347	1,189
Grade 2	29.0	29.3	-0.2	4,347	1,189
Grade 3	23.5	24.2	-0.8	4,347	1,189
Walks to school with another child (percentage)	9.7	7.2	2.5***	4,347	1,189
Number of all communities	**219**	**199**			
Number of small communities	**166**	**147**			
Number of large communities	**53**	**52**			

Source: CARS intake data.

Notes: Column A shows means for the universe of children determined eligible to participate in the evaluation. Column B shows means for the subset of the eligible population we successfully interviewed at follow-up. T-tests were used to examine differences between groups. Age at intake was filled in with base-year or follow-up survey data, if missing. Attendance at community assemblies could not be collected in all communities.

*Difference in group means is statistically significant at the .10 level.

*** Difference in group means is statistically significant at the .01 level.

Educational community and school characteristics. The team of staff implementing CARS, in consultation with local education authorities, identified 219 educational communities in the RACCS as eligible for the evaluation, based on the criteria described in Section III.B. Of the 219 educational communities, 166 were small. All communities agreed to participate in the evaluation, and 147 small communities and 52 large communities were surveyed at follow-up and are in the overall impact analysis sample. Due to security challenges in the region related to common crime (such as theft) and armed activity in opposition to an infrastructure project that caused displacement of communities in the path of a planned canal, 20 small communities had no follow-up visit for data collection. These communities are approximately balanced across treatment and control groups (11 treatment and 9 control), and we do not believe that their attrition introduces bias to our impact estimates. (See Appendix B for detailed analyses.) They are, however, communities that are somewhat harder to reach than the other eligible communities, which may affect the generalizability of the evaluation's findings.

Community size is closely related to cohort, and eligible communities were distributed across the five targeted municipalities (Table IV.4). Most large communities were in Cohort 1, and were in urban areas across all five municipalities. Most small communities were in Cohort 2, and were in rural areas in three municipalities. Spanish was the primary language of instruction in all small communities and in 80 percent of large communities. This reflects the linguistic diversity of the sample, where children in 25 percent of large communities reported that Spanish was not their maternal language. However, children in all small communities except one reported Spanish as their maternal language.

Table IV.4. Geographic distribution and primary languages of communities in the analysis sample, by community size and cohort

Community characteristics	All communities		Small communities		Large communities	
	Cohort 1	Cohort 2	Cohort 1	Cohort 2	Cohort 1	Cohort 2
Municipality						
Bluefields	24	84	2	82	22	2
Kukra Hill	12	29	6	27	6	2
Laguna de Perlas	10	32	0	30	10	2
Desembocadura de Rio Grande	4	0	0	0	4	0
Corn Island	4	0	0	0	4	0
Primary language of instruction						
Spanish	43	145	8	139	35	6
English/Kriol	5	0	0	0	5	0
Miskito, Ulwa, or Garifuna	6	0	0	0	6	0
Children’s maternal language						
Spanish	42	143	8	138	34	5
Kriol/Rama-Kriol	6	2	0	1	6	1
Miskito	6	0	0	0	6	0
Total	**54**	**145**	**8**	**139**	**46**	**6**

Source: Follow-up school director survey; follow-up household survey; CARS M&E data.

Balance in child characteristics at intake. Children in the EpC and control groups had similar characteristics at intake, before EpC rollout had begun, including gender, age, parental community assembly participation, and enrollment status (Table IV.5). Although there are a few small differences in the proportion of children who were in 1st and 3rd grades at intake in small communities, they are significant only at the 10 percent level. There are no statistically significant differences in child intake characteristics in the full sample or in large communities.

Table IV.5. Child characteristics at intake, by treatment group and community size

	All communities			Small communities			Large communities		
Child characteristics	EpC (A)	Control (B)	Difference (A-B)	EpC (C)	Control (D)	Difference (C-D)	EpC (E)	Control (F)	Difference (E-F)
Female (percentage)	49.5	48.0	1.6	47.9	47.6	0.3	51.1	48.1	3.1
Age (years)	8.5	8.6	0.0	9.1	9.1	0.0	8.0	8.1	0.0
Parent attended community assembly at intake (percentage)	65.4	67.9	-2.5	70.9	71.6	-0.6	57.2	62.7	-5.4
Grade level (percentage)									
Out of school	15.9	14.9	1.0	25.1	23.3	1.8	6.7	6.5	0.2
Grade 1	36.5	39.7	-3.3	32.8	38.7	-6.0*	40.1	40.3	-0.2
Grade 2	25.8	26.2	-0.3	22.9	22.7	0.2	28.9	29.7	-0.8
Grade 3	21.8	19.1	2.6	19.1	15.1	4.1*	24.3	23.5	0.7
Walks to school with another child (percentage)	24.3	27.6	-3.2	42.8	47.0	-4.2	5.9	8.2	-2.3
Number of children	**1290**	**1081**		**598**	**584**		**692**	**497**	
Number of communities	**126**	**125**		**74**	**73**		**52**	**52**	

Source: 2014 Intake data. Base-year and follow-up data were used when intake age and grade were missing.

Notes: Columns A and B, C and D, and E and F present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level. Column A–B, C-D, and E- F present differences in the regression-adjusted group means. All regressions include weights to account for oversampling of out-of-school children. The CARS teams identified poor performance during the intake process by talking with teachers about their students' performance; they did not collect such information for all students. Parent attendance to community assemblies could not be collected in all cases. Sample sizes vary due to missing data.

*Difference in group means is statistically significant at the .10 level.

Balance in child characteristics in the base year and follow-up. We found that other characteristics of children were balanced across EpC and control groups in the full sample and in small and large communities. Using survey data for the respondent sample (base-year data for Cohort 2 and follow-up data for Cohort 1), the groups were balanced in 23 out of 27 characteristics that are unlikely to have changed as a result of the EpC program (Table D.1). These include the child's maternal language, agreement between the child's home language and the school's language of instruction, whether both parents live with the child, the child's means of transportation and time to get to school, the school shift the child attended, whether the child reports the school offers a meal, the household’s poverty level and assets, and migration.

The groups were also balanced on seven outcomes that were measured for Cohort 2 children during the base year in the full sample and in both small and large communities (Table IV.6). Specifically, the groups were balanced on whether the child reported knowing how to read, was enrolled in school, repeated a grade, attended preschool, the percentage of days the child attended school, and whether the child had worked with or without compensation in the week before data collection.

Table IV.6. Base-year child outcomes, by treatment group and community size (Cohort 2 only)

	All communities			Small communities			Large communities		
Child outcomes	**EpC (A)**	**Control (B)**	**Difference (A–B)**	**EpC (C)**	**Control (D)**	**Difference (C–D)**	**EpC (E)**	**Control (F)**	**Difference (E–F)**
Child can read (percentage) (child report)	62.9	59.4	3.5	62.7	59.3	3.4	64.6	60.4	4.2
Child enrolled in school (percentage)	85.2	83.9	1.3	84.2	82.2	1.9	92.7	97.8	-5.2
Child ever repeated a grade (percentage)	44.6	48.9	-4.3	47.2	50.7	-3.5	24.2	38.3	-14.0
Attended preschool (percentage)	27.8	25.5	2.2	22.7	20.6	2.2	65.8	61.4	4.4
Attendance (number of days present in the last month)	73.1	72.2	0.9	73.1	70.2	1.1	85.6	88.4	-2.9
Child worked for pay in the last week (percentage)	2.5	4.1	-1.6	2.4	4.0	-1.6	3.1	5.1	-2.0
Child worked unpaid (percentage)	1.0	1.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0
Number of children	**542**	**514**		**476**	**456**		**66**	**58**	
Number of communities	**76**	**75**		**70**	**69**		**6**	**6**	

Source: Base-year household survey.

Notes: Columns A and B, C and D, and E and F present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Differences in sample size come from missing information for some variables. "Child worked unpaid" indicates whether the child performs any unpaid work such as collecting water, doing domestic chores, working in the field, or collecting wood.

Finally, we tested for balance between the EpC and control groups using literacy assessments administered to children in Cohort 2 a few months after implementation started. There were some base-year differences in literacy skills between EpC and control groups for children in the full sample and in small communities, but not for children in large communities (Table IV.7). In the full sample and in small communities, literacy skills were balanced across the evaluation groups for only 6 of the 10 measures of literacy skills. Children in the EpC group performed better than those in the control group on four measures: letter identification, familiar-word reading, oral reading fluency of a simple passage, and reading comprehension of a simple passage. Even though the effect sizes are not large (from 0.13 to 0.15 of a standard deviation, not shown in the table), the number of differences suggests the possibility that they may not be due to chance and may instead represent early effects of the EpC intervention. We discuss these differences in our discussion of interpretation of findings in Chapter VIII. That the differences in large communities between treatment and control groups are not statistically significant might be due to the fact that the number of large communities with base-year data is very small, and the number of children in large communities correspondingly small. However, the differences between treatment and control groups in large communities are smaller in magnitude than in small communities.[22]

There were some base-year differences in literacy skills between EpC and control groups for children in the full sample and in small communities. Even though the differences are not large, the number of differences suggests that they may represent early effects of the EpC intervention.

[22] We were able to collect base-year data only in Cohort 2 communities (which are disproportionately small communities); therefore, we have base-year data for 95 percent of small communities but only 12 percent of large communities.

Table IV.7. Base-year child literacy skills, by treatment group and community size (Cohort 2 only)

	All communities			Small communities			Large communities		
Literacy skills	**EpC (A)**	**Control (B)**	**Difference (A–B)**	**EpC (C)**	**Control (D)**	**Difference (C–D)**	**EpC (E)**	**Control (F)**	**Difference (E–F)**
Letter identification score (number of letters identified of 100 per minute)	33.4	28.8	4.6***	34.4	29.5	4.9***	25.1	24.3	0.8
Identification of initial sounds score (out of 12)	2.0	1.8	0.1	2.0	1.7	0.2	2.4	2.7	-0.3
Passage comprehension score (percentage correct out of 5)	63.4	62.7	0.7	63.2	62.5	0.7	65.2	64.5	0.7
Phonemic awareness score (percentage correct out of 10)	13.2	11.2	2.0	13.1	11.2	1.9	14.4	10.9	3.5
Familiar word reading score (number of words read per minute)	24.1	19.7	4.4***	24.9	20.5	4.4***	17.8	14.0	3.8
Vocabulary score (percentage correct out of 12)	35.4	35.1	0.3	35.3	34.4	0.8	35.8	41.6	-5.8
Oral reading fluency score, first reading: number of correct words read per minute	31.3	25.7	5.6**	32.7	26.9	5.8**	20.7	17.0	3.7
Reading comprehension, first reading: questions correct (percentage correct out of 5)	30.4	25.1	5.3**	31.7	26.1	5.5**	20.9	17.6	3.3
Oral reading fluency score, second reading: number of correct words read per minute	14.0	10.8	3.3	14.6	11.4	3.3	9.8	6.0	3.8
Reading comprehension, second reading: questions correct (percentage correct out of 5)	13.1	11.5	1.6	13.2	12.0	1.6	0.1	0.1	0.1
Number of children	**542**	513		**476**	**456**		**66**	58	
Number of communities	**76**	**75**		**70**	**69**		**6**	6	

Source: Base-year child literacy assessment.

Notes: Columns A and B, C and D, and E and F present ordinary least squares regression-adjusted group means and include covariates to account for the design (strata variables). Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to missing data. Literacy measures are described in Table F1.

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

***Difference in group means is statistically significant at the .01 level.

V. FINDINGS ON IMPLEMENTATION AND INTERMEDIATE OUTCOMES

In this chapter, we examine whether the EpC intervention was implemented as planned. We describe results separately for small and large communities because implementation differed in important ways by community size. We use data from educator surveys and observations of the school environment collected at the end of the evaluation period, as well as qualitative data collected from focus groups and project documents.

Implementation largely took place according to plan. CARS established all 166 planned EpC (of which 84 were in small communities), distributed more than the planned amount of materials, including more than 6,000 reading materials (including storybooks, modules, facilitator guides, and workbooks), and trained EpC facilitators.

However, there were implementation challenges that may have influenced whether EpC generated impacts that could be estimated in this evaluation. Specifically, there were delays in the distribution of reading materials in Cohort 1 and EpC facilitators received fewer support visits than planned.

A. Implementation findings in small communities

At end line, schools in small EpC communities were more likely than those in control communities to have most materials distributed by CARS for the EpC intervention, including musical instruments, musical equipment, sports equipment, and arts and crafts materials. Schools in small EpC communities were 69 percentage points more likely to have music instruments, 38 percentage points more likely to have a music player, 35 percentage points more likely to have sports equipment, and 24 percentage points more likely to have art supplies than schools in control communities (Table V.1).

Students have limited access to print material in these communities, and there was no difference between EpC and control communities in the availability of reading materials of any type for students' use. In both EpC and control communities, fewer than 50 percent of school directors reported having reading materials other than textbooks for students' use (Table V.1.). The difference of 12 percentage points between the two groups is not statistically significant. There are two potential explanations for these findings, including that CARS reading materials did not make it to all EpC, or that the materials were delivered to the EpC but not shared with the school[23]. There were also no statistically significant differences in the availability of textbooks for students' shared or exclusive use, or in students being allowed to take textbooks home. The distribution of textbooks was not a specific goal of EpC but is indicative of children's access to printed material (Table V.1.).

[23] EpC facilitators were not required to share learning materials with the schools, but sharing was feasible because the EpC often shared the physical space of the school.

Table V.1. Impact of EpC on school resources in small communities

Materials (percentages)	EpC (A)	Control (B)	Impact (A–B)	*p*-value
Availability of arts, sports, and music materials				
Music instruments	69.8	0.5	69.4***	0.000
Music player	39.8	2.1	37.7***	0.000
Sports equipment	43.2	8.3	34.8***	0.000
Art supplies	57.5	33.5	24.0***	0.003
Paper supplies	72.3	62.3	10.0	0.223
Availability of (non-textbook) reading materials	48.1	36.2	11.9	0.145
Availability of textbooks				
Of shared use	39.7	47.4	-7.7	0.352
Of exclusive use	8.3	12.2	-3.9	0.468
Students allowed to take textbooks home	20.9	25.4	-4.4	0.545
Number of communities	**74**	**73**		

Source: Follow-up school director survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for the urban/rural status of the community where the school is located.

***Difference in group means is statistically significant at the .01 level.

A larger percentage of facilitators in EpC communities than teachers in control communities were trained in teaching reading, writing, and the use of assessments. In small communities, the percentage of EpC facilitators who reported having participated in trainings on how to teach reading is 34 percentage points higher than teachers in the control group, and 56 percentage points higher for reporting having participated in trainings on the use of standardized assessments to monitor students' reading progress in the year after randomization (see Table V.2). That only about half of EpC facilitators reported having participated in trainings to teach reading the year after randomization could be due to the timing of follow-up, which made it difficult to locate some of the original facilitators for the cohorts included in this evaluation (Table A.1).

Trained educators received less support from CARS than expected. According to data collected during the performance evaluation in 2016, the average number of CARS visits to facilitators was 3.3, which is about half of the eight visits per facilitator that were planned. Twenty two of 23 interviewed EpC facilitators reported at least one CARS visit or classroom observation according to interviews conducted with a sample of active facilitators participating in the performance evaluation in 2016.

Table V.2. Educator training in small communities, by treatment status

Educator training	EpC (A)	Control (B)	Impact (A–B)	*p*-value
Received reading/writing training (percentage)				
Ever received training	84.4	73.5	10.9*	0.099

Educator training	EpC (A)	Control (B)	Impact (A–B)	*p*-value
Trained the year after randomization	50.6	16.3	34.3***	0.000
Trained the year of randomization	7.2	3.2	4.0	0.232
Received standardized child assessment training (percentage)				
Ever received training	76.9	19.0	57.9***	0.000
Trained the year after randomization	64.4	8.5	55.9***	0.000
Trained the year of randomization	2.0	1.4	0.6	0.792
Educator has participated in trainings on other topics since the year of randomization	100.0	84.3	15.7***	0.000
Number of other topics in which educator has been trained since the year of randomization	6.0	3.0	2.9***	0.000
Number of educators	**86**	**89**		
Number of communities	**70**	**68**		

Source: Follow-up educator survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level. Column A–B presents differences in the regression-adjusted group means. The analysis sample includes teacher-facilitators and facilitators in the treatment group and teachers in the control group. Robustness checks are shown in Appendix H. There are 138 communities in the analysis sample (as opposed to 147) because in 4 EpC communities we did not interview EpC facilitators (teacher-facilitators or facilitators) and in 5 control communities we did not interview any teachers that had never been a facilitator.

*Difference in group means is statistically significant at the .10 level.

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

***Difference in group means is statistically significant at the .01 level

B. Impacts on intermediate outcomes in small communities

Given the implementation findings, it is important to determine if educators changed their instructional practices as a result of the EpC intervention. According to the intervention's theory of change, training received by educators should change their instructional practices, which should in turn improve children's academic and social-emotional outcomes. Therefore, we asked EpC facilitators about their classrooms and instructional practices to determine if the knowledge gained by EpC training translated to changes in their own behaviors and whether non-facilitator teachers used similar practices. CARS recruited active teachers as well as non-teachers to serve as EpC facilitators. Therefore, our sample of EpC facilitators combines teacher-facilitators and facilitators only (details about the evaluation sample are discussed in Appendix B). The results reported in this chapter are robust to including teacher-facilitators only (more details are provided in Appendix H).

In small communities, EpC facilitators were more likely to use the instructional practices designed to increase student engagement and promote concept development. We compared self-reported behaviors of EpC facilitators in the treatment group to teachers in the control group (shown in Table V.3) and found large, statistically significant differences in the use of many techniques designed to engage students, including the use of music, traditions or routines, flip charts, didactic materials, visualization (asking children to imagine a place that promotes relaxation and concentration), and singing songs. However, we found no difference in other practices aimed to promote student engagement, such as creating a flexible space for children to transition easily from individual to small group or whole class work, using strategies to focus students' attention in classwork, or using breathing techniques.

Table V.3. Educators' self-reported use of instructional practices they were trained on in EpC in small communities, by treatment status

Instructional practices (percentage)	EpC (A)	Control (B)	Impact (A–B)	*p*-value
Practices to promote student engagement				
Used music in the classroom	51.5	12.1	39.4***	0.000
Used traditions or routines	99.6	66.7	32.9***	0.000
Used flip charts on the walls to teach	95.4	69.6	25.8***	0.000
Used didactic materials to teach	94.3	70.6	23.7***	0.000
Flexibility for children to move easily from individual work to small group or whole class work	92.9	92.3	0.6	0.869
Used strategies to focus children's attention on classwork	97.2	95.9	1.3	0.698
Used breathing techniques	60.6	58.3	2.3	0.781
Asked the children to imagine a place that promotes relaxation and concentration	91.8	57.3	34.5***	0.000
Sang songs with children	98.7	59.6	39.1***	0.000
Practices to promote concept development				
Students read or practiced reading	100.0	88.8	11.2***	0.001
Students read books other than textbooks	87.6	51.3	36.3***	0.000
Educator reads books or stories aloud in class	100.4	90.6	9.8***	0.005
Educator and students talked about books or stories that were read in class	95.6	85.2	10.4**	0.029
Educator talked about why characters in books or stories did what they did	96.0	85.9	10.1**	0.024
Students worked in pairs during class	99.5	84.8	14.8***	0.001
Students worked in small groups during class	96.2	98.1	-1.9	0.591
Students shared questions or ideas during class time	100.6	92.7	7.9***	0.009
Use of student reading assessments (e.g. mini-EGRA)	74.5	24.6	49.9***	0.000
Number of educators	**86**	**89**		
Number of communities	**70**	**68**		

Source: Follow-up educator survey

Notes: Columns A and B present ordinary least squares regression-adjusted means that account for the evaluation design (cohort and strata variables). The analysis sample includes teacher-facilitators and facilitators in the treatment group and teachers in the control group. Errors are clustered at the community level. Percentages greater than 100 result from regression adjustment, which was a linear probability model that can result in out-of-range predictions for binary outcomes.

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

***Difference in group means is statistically significant at the .01 level.

Compared to teachers in control communities, EpC facilitators were more likely to report using strategies to promote concept development, namely having students practice reading, read books other than textbooks, discuss the motivations underlying the behavior of characters in the stories read in class, work in pairs, and share questions or ideas in class (Table V.3). Differences between the two groups ranged in size from 9 to 38 percentage points. There was also a difference in whether or not educators discussed books or stories read in class, though the difference was not statistically significant at the 5 percent level. There were no statistically significant differences between treatment and control communities, however, in the number of educators who reported having students work in small groups.

Educators in EpC communities were also more likely than those in control communities to report administering reading assessments to their students (Table V.3). In these small communities, there was a statistically significant difference between the treatment and control group in the number of educators who reported administering reading assessments (such as Early Grade Reading Assessment [EGRA] or mini-EGRA) to their students. Three-fourths (77 percent) of EpC facilitators reported using reading assessments, compared to 22 percent of teachers in control communities (a difference of 55 percentage points). EpC facilitators were trained to use a mini-EGRA in their classrooms to help monitor children's learning progress and inform lesson planning. This would be expected to improve the instructor's ability to teach and the children's ability to learn.

Given the success with implementation and the promising results observed on intermediate outcomes, we might expect to see improvements on children's academic and non-academic learning outcomes, the primary goals of EpC.

C. Implementation findings and impacts on intermediate outcomes in large communities

As we described in earlier sections, the integrity of randomization in large communities was compromised by contamination and the contrast further affected by low take-up of EpC amongst those children assigned to receive the intervention. We present findings on implementation and intermediate outcomes for the subsample of large communities because implementation shortcomings or failure to influence educators' behaviors could help explain the lack of impacts observed on children, which we describe in Appendix J. However, we do not have two experimental groups that meaningfully differed in their EpC participation status in large communities, and this limits the level of confidence in our conclusions. Because randomization was done at the child level in these communities, we cannot compare the availability of school resources at the community level. Similarly, we do not have a true control group of educators to analyze differences in training and instructional practices. Instead, we compare EpC facilitators (who can be teacher-facilitators or facilitator-only) to teachers in the same schools who were not trained and whom we refer to as "teachers only" (see Appendix C for details on our analytic approach).[24] When interpreting these results, the reader should also bear in mind that endline data collection took place after most EpC activities had ended for Cohorts 1 and 2 (see Appendix

[24] Indeed, in our original evaluation design we intended to estimate the effects of spillover from EpC facilitators in large communities with child level random assignment. However, due to the non-adherence issues in such communities, we were unable to do so.

A). The time elapsed between the end of EpC and when the data were collected likely led to recall difficulties, attrition, and turnover by the educators who facilitated the EpC in the cohorts of interest.

Most EpC facilitators in large communities report having been trained in teaching reading, writing, and the use of assessments. In large communities, the percentage of EpC facilitators who reported having participated in trainings on how to teach reading is 16 percentage points higher than the percentage of regular teachers who were not responsible for EpC. Also, in the year after randomization, the percentage of EpC facilitators who reported having participated in trainings on the use of standardized assessments to monitor students' reading progress is 30 percentage points higher than for regular teachers (Table V.4).

Table V.4. Educator training in large communities, by educator type

Educator training	EpC facilitators	Teachers – only	Difference (A–B)	*p*-value
Received reading/writing training (percentage)				
Ever received training	83.9	68.0	15.9**	0.023
Trained the year after randomization	19.8	21.6	-1.8	0.766
Trained the year of randomization	4.5	6.4	-1.9	0.597
Received standardized child assessment training (percentage)				
Ever received training	73.3	43.5	29.8***	0.001
Trained the year after randomization	34.8	16.1	18.7**	0.011
Trained the year of randomization	1.6	0.9	0.7	0.724
Educator has participated in trainings on other topics since the year of randomization	100.0	94.5	5.5**	0.035
Number of other topics in which educator has been trained since the year of randomization	6.8	5.0	1.7***	0.000
Number of educators	67	110		
Number of communities	52	50		

Source: Follow-up educator survey.
Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). The group of EpC facilitators includes teacher-facilitators and facilitators. Errors are clustered at the community level.

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

***Difference in group means is statistically significant at the .01 level.

We found few differences in the instructional practices of educators (EpC facilitators and non-facilitator teachers) in large communities (Table V.5). We compared responses from EpC facilitators about their experiences in the EpC to the responses of educators who are teachers only (and therefore were not trained in EpC methodologies by CARS) about their experiences in the classroom. EpC facilitators were more likely than those who are only teachers to report using music in the classroom, using traditions and routines, and singing with children. These differences range from 16 to 39 percentage points and are statistically significant. There are no statistically significant differences between educators in large communities in the use of flip charts, didactic materials, and on asking students to imagine a place that promotes relaxation and concentration. Similarly, most educators in large communities reported implementing strategies to promote concept development, such as having students practice reading, reading books or stories aloud, and having students work in pairs and small groups. However, EpC facilitators were significantly more likely to report a couple of behaviors that reflect key components of the EpC intervention—and which are associated with improved reading/learning outcomes. First, EpC facilitators had students read books other than textbooks more than teachers-only. The difference of 21 percentage points (which corresponds to 0.70 standard deviations) is statistically significant. In addition, a larger percentage of EpC facilitators reported using reading assessments than educators who are teachers only. The difference of 19 percentage points is statistically significant.

The five statistically significant differences we observed between instructional practices of EpC facilitators and non-facilitator teachers in large communities decreases the likelihood that there would be impacts on child academic and non-academic outcomes. Because EpC facilitators and teachers in large communities engage in many similar instructional practices, we can conclude that the learning environment is not all that different between EpC and regular classrooms. However, it is still feasible that impacts on child outcomes could manifest in large communities, because EpC were opened, and facilitators did use assessments to identify gaps in children's reading skills, as well as music and other techniques to make learning fun, potentially increasing the amount of time that children spent in a high quality learning environment.

Table V.5. Educators' use of instructional practices in large communities, by educator type

Instructional practices (percentage)	EpC facilitators (A)	Teachers only (B)	Difference (A–B)	*p*-value
Practices to promote student engagement				
Used music in the classroom	48.7	23.5	25.2***	0.000
Used traditions or routines	100.3	61.7	38.6***	0.000
Used flip charts on the walls to teach	96.9	90.1	6.8*	0.056
Used didactic materials to teach	92.4	87.3	5.1	0.227
Arranged children in a flexible way in the classroom	95.5	93.6	1.9	0.639
Used strategies to focus children's attention on classwork	97.0	93.6	3.4	0.240
Used breathing techniques	56.3	64.4	-8.2	0.250
Asked the children to imagine a place that promotes relaxation and concentration	71.3	68.2	3.1	0.679
Sang songs with children	92.6	76.4	16.2***	0.001
Practices to promote concept development				
Students read or practiced reading	98.5	93.6	4.8	0.115
Students read books other than textbooks	81.8	60.9	20.9***	0.002
Educator read books or stories aloud in class	97.0	97.3	-0.2	0.927
Educator and students talk about books or stories that have been read in class	98.5	94.5	4.0	0.181
Educator talked about why characters in books or stories did what they did	97.0	93.6	3.3	0.308
Students worked in pairs during class	89.4	90.9	-1.4	0.776
Students worked in small groups during class	89.4	94.5	-5.1	0.205
Students shared questions or ideas during class time	96.9	96.4	0.6	0.842

Instructional practices (percentage)	EpC facilitators (A)	Teachers only (B)	Difference (A–B)	*p*-value
Use of student reading assessments	69.0	50.3	18.7**	0.022
Number of educators	**67**	**110**		
Number of communities	**52**	**50**		

Source: Follow-up educator survey

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). The group of EpC facilitators includes teacher-facilitators and facilitators. Errors are clustered at the community level. Sample sizes vary due to missing data.

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

***Difference in group means is statistically significant at the .01 level.

VI. IMPACTS ON CHILD OUTCOMES IN SMALL COMMUNITIES

The primary goal of the EpC program is to improve child academic and non-academic outcomes. In this chapter we answer the question about the impact of EpC on these outcomes, using the sample of children from small communities, where the experiment performed well and we were able to identify a treatment effect cleanly. The next chapter (Chapter VII) describes findings from the full sample of small and large communities combined, as well as a discussion of the generalizability of findings from small communities to large communities.

A. Impacts on children's attachment to school

One of EpC's primary goals is to improve children's attachment to school, because the program focuses on children who have never attended school, have dropped out, or are at risk of dropping out. We examined the impact of EpC on children's school attendance, a key indicator of school attachment identified as an important outcome in the EpC Theory of Change (Figure II.1) as well as in the literature (see Jimerson et al. 2003). We also explored its impacts on secondary measures of school attachment, school enrollment, and grade progression.

EpC had no impacts on school attendance or enrollment, but it did have a positive impact on grade progression. Students in both groups attended school about 70 percent of the days school was in session in the month before data collection (Table VI.1). The most common reasons for absenteeism in small communities were illness (32 and 36 percent in the treatment and control group, respectively) and agricultural or domestic responsibilities (22–23 percent as shown in Appendix K). These are barriers to attendance that EpC was not designed to address.

At follow-up, 79 percent of the children in the treatment group were enrolled in school, as were 81 percent of the children in the control group. This level of enrollment is slightly higher for both groups than it was at the time of intake (75 and 77 percent respectively, as shown in Table IV.5). However, the increase cannot be attributed to the EpC intervention, because there was an increase in both the treatment and control groups. Increasing enrollment may be partly explained by the fact that the out-of-school children in our sample were growing older. Forty-three percent of the children who were out of school at intake in small communities were between the ages of 5 and 7, and their families may have thought they were too young to be in school at the time of intake, even though children are eligible to enroll in pre-school starting at age three (See Table K.2 for more details). Alternatively, these children could have simply been late to enroll, as school enrollment in the RACCS continues for a few months after the official start of the school year. Conversations with CARS staff revealed that many older children refused the option of participating in EpC—and therefore were not included in the program intake data as eligible for participation in the program—because they had competing activities (such as work), and caregivers were unable to encourage or enforce participation as they did with younger children.

About the same percentage of children in the EpC and control group had progressed since intake (that is advanced at least one grade since recruitment[25]). However, children in EpC advanced slightly more in the number of grades progressed than children in the control group, a difference that is statistically significant at the 5 percent level. The difference is equivalent to 0.16 standard deviations in the sample of small communities, compared to 0.11 standard deviations in the full sample. This result is robust to using alternative measures of this outcome (see Appendix H).

Table VI.1. Impact of EpC on child attachment to school in small communities

Attachment to school	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Attendance (percentage of days present in the last month) (0 if not enrolled)	68.9	69.7	-0.8	0.800	-0.02	1,170
Enrollment (percentage of children enrolled)	78.8	81.0	-2.2	0.461	-0.08	1,172
Progression since recruitment (percentage of children who advanced at least one grade)	71.0	67.9	3.1	0.374	0.09	1,167
Progression (number of grades child progressed)	1.3	1.1	0.2**	0.017	0.16	1,167
Number of children	**593**	**579**				
Number of communities	**74**	**73**				

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item-level nonresponse. Progression since recruitment and the number of grades child progressed have a value of 1 for children who may have missing grade data but who were out of school at recruitment and enrolled at follow-up, and a value of 0 for children who were enrolled at recruitment and out of school at follow-up.

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

B. Impacts on children's literacy

A core objective of EpC was to improve children's literacy skills. We examined the intervention's impacts on children's ability to decode and read fluently—two foundational skills—as well as on their ability to comprehend what they read, a key milestone for achieving literacy.

[25] Progression since recruitment, the first progression variable in Table VI.1, equals 1 for children who advanced at least one grade since recruitment and 0 for children who did not advance. The number of grades child progressed, the second progression variable in the same table, was calculated as the child's grade at follow up minus their grade at recruitment. For children who were out of school at recruitment, we subtracted the grade they would have attended if they were to enroll. Children could have progressed 0, 1, or 2 grades from the time of recruitment to follow-up data collection, with expected progression equaling 2 grades.

EpC had a positive impact on children's decoding ability, equivalent to 0.12 standard deviations. Children in the treatment group were able to correctly read two more invented words (pseudo-words) in one minute than students in the control group were. Children in the treatment group correctly read an average of 18 invented words in one minute, whereas children in the control group read an average of 16 invented words (Table VI.2). This difference is equivalent to 0.12 standard deviations and is statistically significant at the 5 percent level.

EpC had a positive impact on children's reading fluency, equivalent to 0.11 standard deviations. The average child in the control group was able to read 37 words per minute, and in the treatment group, the average child could read 41 words per minute (Table VI.2). This difference of 4 words per minute is statistically significant at the 5 percent level, and reflects an effect size of about 0.1 standard deviations. The impact on fluency is close to the lower bound effect found for fluency outcomes in a recent review of early literacy interventions, which reported that the positive impacts programs have on fluency outcomes are as small as 0.14 standard deviations and as large as 0.73 standard deviations (Kim et al. 2017). Given the low level of reading fluency at which children in our sample started, a larger impact would be required for them to reach a level of fluency that would enable them to read with comprehension. The literature suggests that given the capacity of short-term memory, children should read a minimum of 45–60 words per minute to understand a simple passage (Abadzi 2011). In our full sample, children who read with comprehension[26] read an average of 64 words per minute and there is no difference in the number of words read by children in the EpC and control groups (children who read with comprehension read an average of 65 words per minute in EpC and 61 words per minute in the control group, a difference that is not statistically significant). Children who did not read with comprehension only read an average of 32 words per minute and there was no difference between the EpC and control groups (children read 31 words per minute in EpC and 33 words per minute in the control group, a difference that is not statistically significant).

EpC had positive impacts on reading comprehension. On average, children in the treatment group answered 42 percent of the questions (or 2.1 questions out of 5) about the third-grade level reading passage correctly, whereas children in the control group answered 37 percent of the questions (or 1.9 questions out of 5) correctly. Similarly, children in the treatment group answered 1.9 questions of the fifth-grade level reading passage correctly, whereas children in the control group answered 1.7 questions correctly on average (Table VI.2). These differences reflect an effect size of 0.1 standard deviations and are statistically significant at the 5 percent level. This magnitude is on par with existing research on many interventions that have yielded limited impact or mixed results in reading comprehension skills (Kim et al. 2017).

EpC had positive impacts on additional, secondary, literacy outcomes, including achieving a score above zero in the 3rd-grade level and 5th-grade level reading comprehension subtasks. In Appendix H we present findings for whether EpC had an impact on the percentage of children who achieved a score above zero in the decoding, fluency, and reading comprehension subtasks, as well as other additional measures of literacy skills. A higher percentage of children in EpC than in the control group obtained a score above zero in the two

[26] This group consisted of children who were able to answer four out of five reading comprehension questions on the first reading passage correctly.

reading comprehension tasks. For the third-grade level reading passage, the difference is significant at the 10 percent level, whereas for the fifth-grade level passage the difference is significant at the 5 percent level. There are no other statistically significant differences between the groups (see Appendix H, Table H.9).

There was no strong evidence that EpC had an impact on children's reading behavior at home. The EpC intervention was designed to foster children's enjoyment of reading and motivate them to read at home. There is a difference of 4 points between the two groups in the percentage of children who read at home alone (Table VI.2), but it is significant at the 10 percent level only. Caregivers in both groups reported that children read at home alone three days per week on average. In the base year sample, children in both EpC and control groups were reading at home 1.9 days per week on average as reported by caregivers (Bagby et al. 2017c).

Table VI.2. Impact of EpC on child literacy skills and behaviors in small communities

Literacy skills	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Pseudo-word decoding (# correct per minute)	18.1	16.0	2.1**	0.026	0.12	1,164
Reading fluency (# correct words per minute)	41.5	37.1	4.4**	0.035	0.11	1,162
Reading comprehension (third-grade level) percentage correct out of 5)	42.2	37.3	4.9**	0.017	0.13	1,162
Reading comprehension (fifth-grade level) (percentage correct out of 5)	37.7	33.3	4.3**	0.021	0.11	1,162
Child reads at home alone (percentage) (caregiver report)	81.2	77.0	4.2*	0.087	0.16	1,171
Number of days child read at home in last 7 days (caregiver report)	3.0	2.9	0.1	0.713	0.02	1,162
Number of children	**593**	**578**				
Number of communities	**74**	**73**				

Source: Follow-up household survey

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item-level nonresponse.

*Difference in group means is statistically significant at the .10 level.

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

C. Impacts on children's social-emotional skills and attitudes

EpC includes a personal growth component that emphasizes the internalization of principles, values, and rights. According to the EpC logic model, personal growth is expected to foster self-esteem and improve social-emotional skills and attitudes. In the long term, those enhancements, together with improved academic outcomes, are expected to reduce the risk of involvement in illicit activities and thereby contribute to building safer communities.

The evaluation did not detect any impacts of EpC on children's social-emotional skills or attitudes. There were no statistically significant differences between the treatment and control groups on caregiver's reports of children's social competence, our primary measure of social-emotional skills, or on children's self-reported self-esteem and intercultural competence (Table VI.3). The treatment-control differences were all fewer than 0.07 standard deviations.

There were also no statistically significant differences between the two groups of children on attitudes that are risk factors for engaging in illicit behavior (Table VI.3). There were no program impacts on moral disengagement, impulsive risk taking, attitudes towards delinquency, and bullying or peer victimization. Treatment-control differences were all fewer than 0.03 standard deviations. The levels for these risk factors are low—a good thing—and there was therefore little room for improvement.

Table VI.3. Impact of EpC on children's social-emotional skills and attitudes in small communities

Social-emotional skills and risk factors	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Skills (higher score is better)						
Social competence (range 0 to 3) (caregiver report)	2.3	2.2	0.0	0.290	0.06	1,151
Self-esteem (range 1 to 5)	4.1	4.1	0.0	0.797	-0.02	1,155
Intercultural competence (range 0 to 3)	2.1	2.1	0.1	0.138	0.07	1,164
Risk factors (lower score is better)						
Moral disengagement (range 0 to 3)	1.1	1.1	0.0	0.951	0.00	1,164
Impulsive risk taking (range 0 to 3)	1.0	1.1	0.0	0.546	-0.03	1,164
Attitudes towards delinquency (range 0 to 3)	0.8	0.8	0.0	0.622	-0.03	1,164
Bullying or peer victimization (range 0 to 3)	0.6	0.6	0.0	0.907	0.01	1,164
Number of children	**589**	**575**				
Number of communities	**74**	**73**				

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item-level nonresponse. Outcome definitions are included in Appendix G.

As is always the case, that the evaluation did not detect impacts on these measures could reflect that there is no impact of EpC on these outcomes, or could reflect limitations in measurement. In addition to social desirability and recall biases, which are common with self-report methods, the measures we used for social emotional skills and attitudes had not been used in Nicaragua before (though the self-esteem measure had been used with Mexican American children [Hess and Petersen 1996] and validated with Spanish children [Broc 2014]). In addition, internal reliability, defined as the extent to which the items on a scale measure the same construct or idea, was a challenge with most of the measures. (We discuss in more detail in Appendix G).

D. Impacts on additional outcomes of interest in small communities: safety perceptions and time use

EpC had an impact on children's perceptions of safety, but not on their caregivers' perceptions. Close to 9 in 10 (87.4 percent) of the children in EpC said they felt safe in their community (Table VI.4). This is 11 percentage points higher than the result for children in the control community (76.9 percent). This difference is statistically significant and equivalent to 0.44 standard deviations. Relatedly, about three in five (59.5 percent) of the children in EpC communities reported that they feel safe walking alone at night, which is 7 percentage points higher than the result for children in control communities and equivalent to 0.18 standard deviations. There are no statistically significant differences between the groups in children's sense of safety at school or going to school, but most children reported feeling safe in those contexts. Also, there were no differences between the groups in caregivers' perceptions of safety.

Table VI.4. Impact of EpC on perceptions of community safety in small communities

Perceptions of safety	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Community is very safe or safe (percentage) (child report)	87.4	76.9	10.5***	0.001	0.44	1,167
Child feels very safe or safe walking alone at night (percentage)	59.5	52.2	7.3***	0.045	0.18	1,165
Child feels safe at school (percentage)	95.1	95.3	-0.2	0.824	-0.03	1,155
Child feels safe going to school (percentage)	91.5	90.2	1.3	0.416	0.09	1,154
Community is very safe or safe (percentage) (caregiver report)	74.4	70.8	3.6	0.204	0.11	1,164
Caregiver feels very safe or safe walking alone at night (percentage)	40.0	38.2	1.7	0.603	0.04	1,164
Number of children	**589**	**578**				
Number of communities	74	73				

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item nonresponse.

***Difference in group means is statistically significant at the .01 level.

The evaluation did not reveal any impacts of EpC on child labor or other after-school activities. Children in both groups engaged in few other after- (or before-) school educational or recreational activities, and also engaged in some form of work (compensated or without compensation) for an average of 12 hours per week (Table VI.5).

Caregivers of EpC children were more likely to participate in EPM (parent schools) than caregivers of control group children were, on average. Caregivers of EpC group children were 19 percentage points more likely to report having participated in EPM, which were supported by CARS in EpC communities (control group caregivers do participate in EPM because they are available in all public schools) (Table VI.5). The level of parent participation in EPM could have influenced children's participation in EpC, or vice versa, as well as children's academic outcomes.

Table VI.5. Impact of EpC on time use in small communities

Other child outcomes	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Non-EpC activity child attended before or after school (percentage)						
Academic reinforcement	2.2	1.8	0.5	0.569	0.15	1,172
Sports/recreation	1.2	0.3	0.9*	0.082	0.78	1,172
Art/music/theater	2.1	1.2	0.9	0.342	0.35	1,172
Religious activities	0.8	0.9	-0.1	0.891	-0.06	1,172
Other	1.1	0.7	0.5	0.362	0.33	1,172
Number of hours child spent in non-educational or non-recreational activities (e.g., labor) in the last week	12.7	12.2	0.6	0.590	0.03	1,172
Child's caregiver participated in EPM (percentage)	37.8	18.7	19.1***	0.000	0.59	1,094
Number of EPM meetings attended by child's caregiver in 2017	0.7	0.4	0.3***	0.000	0.23	1,094
Number of children	**589**	**578**				
Number of communities	**74**	**73**				

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item nonresponse.

***Difference in group means is statistically significant at the .01 level.

E. Impacts by subgroup

We present the subgroup findings for the primary subgroups of interest as defined in our evaluation design: gender and out-of-school status at intake.[27] We present findings for the set of primary outcomes, which we consider representative of the full set of findings. The findings confirm that the impacts for the full sample are significant for girls and for children who were out of school at intake. The differences in impacts by subgroup (between boys and girls, and

[27] We conducted analyses for other subgroups (specifically implementation cohort, implementing NGO, whether children's maternal language matches the school's language of instruction, child's age, and the household's education level and socio-economic status). We do not present those findings because they do not provide compelling evidence that changes our interpretation of the results.

between out-of-school and enrolled children) were not statistically significant, so we cannot conclude that there were differential impacts.

1. Gender

Subgroup analyses show effects of EpC on literacy skills for girls but not for boys, however they do not indicate that there are significant differences in effects by gender. Table VI.6 summarizes the impacts by gender for the set of primary outcomes. For most outcomes, the impact for girls is larger than the impact for boys, but the difference in impacts by gender is not statistically significant. As with the full sample, we find no detectable impacts on attendance or moral disengagement by gender. Girls in the treatment group were able to read correctly three more invented words (pseudo-words) in one minute than girls in the control group could. This difference is statistically significant. There are no corresponding impacts for boys. Girls in the treatment group were also able to read 5.4 more (non-invented) words in a minute than girls in the control group could, but this difference is statistically significant at the 10 percent level only. Girls in the treatment group correctly answered about half of a question more about the reading passages, and this difference is statistically significant at the 5 percent level. There are no statistically significant differences for boys. We also see a small difference in social competence for girls, but it is only significant at the 10 percent level, and there is no impact for boys.

Table VI.6. Impact of EpC on primary child outcomes in small communities, by gender

Child outcomes	Impact for girls (A)	*p*-value	Impact for boys (B)	*p*-value	Difference in impacts by gender (p-value)
Attendance (percentage of days present in the last month) (0 if not enrolled) (pp)	1.6	0.684	-2.9	0.432	0.310
Pseudo-word decoding (# correct words per minute)	2.9**	0.026	1.4	0.258	0.356
Reading fluency (# correct words per minute)	5.4*	0.068	3.6	0.188	0.643
Reading comprehension (combined) (percentage correct out of 10) (pp)	5.4**	0.031	3.9	0.133	0.684
Social competence (range 0 to 3)	0.1*	0.069	0.0	0.926	0.170
Moral disengagement (range 0 to 3)	0.1	0.122	-0.1	0.171	0.052
Number of children	**554**		**617**		
Number of communities	**147**		**147**		

Source: Follow-up household survey

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level. Impacts on attendance and reading comprehension are in percentage points (pp). All regressions include weights to account for oversampling of out-of-school children. Regressions include controls for children's characteristics. Sample sizes vary due to item nonresponse.

*Difference in group means is statistically significant at the .10 level.

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

2. School enrollment at intake

The findings confirm impacts of EpC on children who were not enrolled in school at the time of intake in our sample, but the results may only apply to children who had never been enrolled in school at intake. Among children who were reportedly not enrolled in school at intake, those in the treatment group correctly read 4.4 more invented words and 11 (non-invented) words than children in the control group could (Table VI.7). Also, they were able to answer correctly nearly one more question about the reading passages than children in the control group were, and these differences were statistically significant. There were no differences among children who were in school at the time of intake. Children who were out of school had more room to grow than those who were enrolled in school. There are no impacts for either group on attendance, social competence, or moral disengagement. As noted, 43 percent of the children who were out of school at intake were between 5 and 7 years old, an age when many children in the region are not yet enrolled in school despite being eligible to enroll. Only 13 percent of children who were out of school at intake were still out of school at follow-up, which supports the hypothesis that age was a key reason why these children may not have been enrolled at intake (see Table K.8.). Further, the majority of these children's caregivers (83 percent) reported that if their children were to enroll in school, they would attend first grade, suggesting that most of the children identified as out-of-school at intake had never attended school, instead of having attended and dropped out. Therefore, we should interpret results for the group of children who are "reportedly not enrolled at intake" carefully, and not conflate them with older school dropouts in this evaluation.

Table VI.7. Impact of EpC on children's primary outcomes in small communities, by school enrollment status at intake

Child outcomes	Impact for children out of school (A)	*p*-value	Impact for children in school (B)	*p*-value	Difference in impacts by enrollment status (p-value)
Attendance (percentage of days present in the last month) (0 if not enrolled) (pp)	3.2	0.610	-2.0	0.538	0.426
Pseudo-word decoding (# correct per minute)	4.4**	0.019	1.4	0.149	0.133
Reading fluency (# correct words per minute)	11.0***	0.010	2.5	0.264	0.055
Reading comprehension (combined) (percentage correct out of 10) (pp)	9.0**	0.035	3.3*	0.090	0.232
Social competence (range 0 to 3)	0.0	0.608	0.0	0.334	0.965
Moral disengagement (range 0 to 3)	-0.1	0.486	0.0	0.776	0.476
Number of children	**332**		**832**		
Number of communities	**147**		**147**		

Source Follow-up household survey

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level. Impacts on attendance and reading comprehension are in percentage points (pp). All regressions include weights to account for oversampling of out-of-school children. Regressions include controls for children's characteristics. Sample sizes vary due to item nonresponse.

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

***Difference in group means is statistically significant at the .01 level.

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

VII. IMPACTS ON CHILD OUTCOMES IN ALL COMMUNITIES

In section A of this chapter we answer the question about the overall impact of EpC on children's primary outcomes using the full evaluation sample, which includes children from small and large communities. These analyses follow the original analysis plan, but we anticipated impact estimates to be diluted compared to those from small communities, due to issues with non-compliance and low take-up in large communities, which we described in earlier chapters. In section B we explore the possibility that differential impacts by community size were not only due to features of the evaluation, but to differences in population characteristics or implementation processes, and discuss implications for generalizing results from small to large communities.

A. Impacts on child primary outcomes

EpC had a positive impact on children's decoding ability, but no impacts on other outcomes. Children in the EpC group were able to correctly read 1.4 more invented words (pseudo-words) in one minute than students in the control group were. Children in the treatment group correctly read an average of 19 invented words in one minute, whereas children in the control group read an average of 18 invented words (Table VII.1). This small difference is equivalent to 0.08 standard deviations, but it is statistically significant at the 5 percent level.

Students in both groups attended school about 77 percent of the days school was in session in the month before data collection (Table VII.1). This result is robust to using alternative measures of this outcome (see Appendix I). We also did not observe impacts of EpC on children's reading fluency or reading comprehension, two subtasks were we found significant differences in small communities. Similarly, there were no statistically significant differences between the treatment and control groups on caregiver's reports of children's social competence, our primary measure of social-emotional skills, or in children's self-reports of moral disengagement, our primary indicator of risk factors for engaging in illicit behavior (Table VII.1).

We examined and found no impacts on secondary outcomes such as the percentage of children who obtained zero scores in the reading subtasks, self-esteem, and impulsive risk taking, among others. These findings, as well as findings on alternative measures of some outcomes, are described in Appendix I.

Table VII.1. Impacts of EpC on child primary outcomes in all communities

	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Attendance (percentage of days present in the last month) (0 if not enrolled)	76.6	76.7	-0.1	0.942	0.00	2,356
Pseudo-word decoding (# correct per minute)	19.2	17.8	1.4**	0.039	0.08	2,351
Reading fluency (# correct words per minute)	41.4	39.2	2.2	0.119	0.06	2,349

	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Reading comprehension (combined) (percentage correct questions, out of 10)	40.5	38.3	2.2	0.109	0.06	2349
Social competence (range 0 to 3) (caregiver report)	2.1	2.1	0.0	0.415	0.03	2,334
Moral disengagement (range 0 to 3)	1.2	1.1	0.0	0.343	0.04	2,349
Number of children	**1,284**	**1,076**				
Number of communities	**126**	**125**				

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item-level nonresponse.

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

B. Differential impacts and generalizability of results from small to large communities

In this evaluation, the impact analysis for small communities is cleaner than it is for the full sample, for reasons of design and implementation that have been noted (community-level versus child-level randomization, and higher levels of non-compliance and low take-up in large communities). It is not surprising that findings from the two sets of analyses are different, because the treatment/control contrast was so much weaker in the large communities. There remains the question of whether EpC in large communities might have been just as effective as it was in small communities, but we simply could not detect its impacts, or whether small and large communities differ in fundamental ways that may have influenced take-up and effectiveness of the EpC intervention.

In this section, we compare small and large communities in a range of characteristics that could be related to intervention take-up as well as student outcomes. Our comparisons focus on the control group only, in cases in which the characteristics examined could have changed because of EpC. These comparisons do not aim to provide a definite answer as to why impacts differed between small and large communities, but aim to explore the likelihood that differences in findings could be attributed to differences in community characteristics or implementation processes. These analyses shed light on the generalizability of results from small to large communities and can inform implementers and other decision-makers’ future efforts in these and similar communities.

1. Child and school characteristics in large and small communities

Differences in the findings for small and large communities are probably due to features of the evaluation design; specifically assigning whole communities to EpC versus assigning individual children within communities. However, they could also have been due to differences in population characteristics. If characteristics of the two types of communities were remarkably dissimilar, EpC may simply have had different impacts for children in these different contexts (detailed results are included in Appendix K).

As a group, children in small communities differed from children in large communities in a range of socio-demographic characteristics that may have influenced intervention take-up and sensitivity to change. Children in small communities were more likely to be female (56 versus 41 percent), out of school at intake (27 versus 4 percent, although this difference is not statistically significant), and in lower grades (10 percent were in grade 3 versus 31 percent in large communities) than children in large communities. They were also much poorer than those in large communities (90 versus 55 percent), which placed them at higher risk of school failure and drop out.

Children in small communities were more likely to walk to school (83 versus 73 percent), less likely to walk to school with another child (20 versus 32 percent), and less likely to take public transportation than children in large communities were (16 percentage points less likely to take a bus or car, and 9 percentage points less likely to use a motorcycle). They spent almost twice as much time getting to school as children in large communities did (23 versus 15 minutes).

Children in small communities were significantly less likely to attend the afternoon shift at school (9 versus 27 percent), and were more likely to have a school-provided meal (75 versus 7 percent) than children in large communities. CARS staff observed that children in large communities were more likely to go home for lunch and rely on public transportation to get back to school. Therefore, participating in EpC in large communities was potentially not as straightforward as in small communities. Because children in large communities had to travel to school a second time in the day to participate, they probably incurred a higher cost (in transportation) to attend EpC.

Children in small communities participated in fewer extracurricular activities than children in large communities. In small communities, children in the control group were 6 percentage points less likely to attend academic reinforcement activities and 8 percentage points less likely to participate in sports and other recreational activities than their counterparts in large communities were. Greater access to academic and recreational activities outside of school could have reduced the perceived value of EpC and result in low take-up or attendance among children in large communities. This is in line with the CARS staff observation that compared to children in small communities, children in large communities had more access to resources and activities that competed with EpC, including access to internet and TV, as well as activities organized by the school.

Children in small communities spent more time in activities that were not recreational or educational. Compared to control children in large communities, control children in small communities spent about eight more hours per week, on average, in activities that were not recreational or educational. This suggests that they spent more time helping with household chores or working, with or without pay. This could be associated with the rurality of small communities and could also be related to the fact that children in small communities live in poorer households and have less access to extra-curricular and recreational activities than children in large communities. This finding may appear counterintuitive, as dedicating more time to domestic obligations could have posed barriers to EpC attendance. However, to the extent that those activities may have been informal, they would have allowed most children in small communities to participate in EpC. In the absence of detailed or complete EpC attendance records, we cannot ascertain whether household chores or other domestic obligations may have interfered with children's regular EpC attendance.

Children in small communities who were assigned to the control group had literacy outcomes similar but not equal to children in the control group in large communities (Table IV.7). Because most small communities were in rural areas and most large communities were in urban areas, we disaggregated the control group follow-up values by urban/rural status to better understand the contribution of community size and urbanicity (detailed results are included in Appendix K). We find that the differences in literacy are driven by children in *large urban* communities. Children in large urban communities have better literacy skills than children in all other areas, whereas children in large rural communities perform similarly to children in small communities (both urban and rural). The effect of the EpC intervention on children in small communities brought their skills closer to the level of children in large urban communities. These findings suggest that EpC's leveling activities were effective at improving the skills of underperforming children, but other children may require different kinds of support to achieve higher levels of performance.

Schools in small communities had fewer resources than schools in large communities. Therefore, students in small communities could have benefitted more from EpC than students in large communities. Small and large communities were similar in a range of characteristics, including the number of students per teacher, student attendance rates, and the materials classrooms were made of. However, there were some notable differences.

Schools in small communities were 34 percentage points more likely to have multi-grade classrooms than schools in large communities were (detailed results are included in Appendix K). Most schools in the small communities had only multi-grade classrooms (89 percent). Even though multi-grade classrooms can expand education access to marginalized populations, they pose pedagogical challenges for teachers who are likely trained to teach in single-grade classrooms and who may not have teaching materials designed to facilitate learning in multi-grade settings (Little 2005).

Schools in small communities were open 3.2 fewer days per month than schools in large communities were. Over the full academic year, this amounts to almost a month less of instruction for students in small communities, which likely contributes to their lesser academic achievement (Abadzi 2009).

All the schools in the evaluation lacked some basic infrastructure. However, schools in small communities were 25 percentage points less likely to have classrooms with enough furniture for students. Over half of schools in both large and small communities received support from external programs other than EpC. However, schools in small communities were 38 percentage points less likely than schools in large communities to have received textbooks and other learning materials. Teachers in large communities had eight more years of teaching experience on average, but the difference was not statistically significant and there are inconclusive studies on the association between teaching experience and student achievement in developing countries (Glewwe et al. 2011).

We cannot be certain about how or whether each of these differences between small and large communities influenced children's participation in EpC, or the likelihood that EpC would have an impact in large communities. However, the results help paint a picture of communities that differed in ways that could have reasonably influenced the take-up and effectiveness of EpC through several mechanisms. Implementers of interventions like EpC should carefully consider how to address the distinct barriers faced by children in rural versus urban communities, as well as build on the unique favorable conditions of the communities they target.

2. Comparing implementation in large and small communities

The difference in findings based on community size could also be due to differences in implementation, which may or may not be related to the characteristics of children and schools in these communities. We therefore compared implementation of the EpC intervention across community sizes (detailed results are included in Appendix K).

Overall, large communities had more reading materials than small communities did. There is a statistically significant 55 percentage point difference between large and small communities with EpC in the availability of reading materials other than textbooks (91 percent of all schools in large communities, compared to 35 percent of schools in small treatment communities). In small treatment communities, non-textbook reading materials were available at relatively the similar level as music players (38 percent), but lower level than music instruments (65 percent), for example. This suggests that not only are there difficulties in accessing small communities, but there also may be some interaction between access and the nature of the materials (i.e., reading materials are heavier).

There were delays in the distribution of reading materials to large communities. CARS reports indicate that the distribution of print materials appropriate to the RACCS did not happen until 10 months into the implementation of Cohort 1A and six months into the implementation for Cohort 1B, both of which are primarily made up of large communities. Therefore, children in Cohort 1 did not receive the full package of the EpC intervention, whereas children in Cohort 2 did. This may result in lower-than-expected impacts on children's literacy skills overall, and fewer impacts on literacy for children in Cohort 1 compared to Cohort 2. Given that Cohort 1 is mostly large communities, and Cohort 2 is mostly small communities, this could be a factor in explaining the difference in impacts we see by community size.

Teachers in large communities were more likely to have participated in trainings on topics covered by EpC. Overall, teachers in large communities were more likely to have participated in trainings to teach reading/writing and the use of standardized assessments than teachers in small communities were. The largest differences are for trainings on how to teach reading/writing that took place two years after randomization (23 percentage points) and trainings on how to use standardized assessments that took place one and two years after randomization (15 to 16 percentage points, respectively). The fact that teachers who were not facilitators participated in trainings on topics that were covered by the EpC intervention, especially in large communities, may be further evidence of contamination between treatment and control groups, resulting in a weaker contrast between experimental groups that would hamper the evaluation's ability to detect EpC impacts.

Caregivers in large communities reported more participation in EPM and more contact with teachers than caregivers in small communities. More caregivers in the control group in large communities reported participation in EPM (34 percent), as well as more communication with their child's teachers (52 percent), than caregivers in the control group in small communities did (18 percent and 25 percent respectively; Table K.2). In large communities, the control group's levels in both variables were on par with the treatment group's, suggesting that EpC did not have a significant impact on caregivers' participation in EPM or on caregivers' contact with teachers. Some local stakeholders anecdotally observed that the school and church are more often at the center of social life in rural than in urban communities and, therefore, may facilitate caregivers' participation in school-centered activities. Because large communities are predominantly urban and small communities are predominantly rural, one could expect more involvement from caregivers in small than in large communities. However, we find lower levels of caregiver participation in small communities than in large communities (Table K.2). There may be other factors that account for caregivers' relatively low levels of participation in small communities and that may encourage parents in large communities to participate more in school activities (such as distance to the school), counteracting the relative displacement of the school and church as focal points for community activities.

VIII. INTERPRETATION OF FINDINGS IN SMALL COMMUNITIES

In this chapter, we present several additional analyses to understand the robustness of the findings in small communities and discuss cost effectiveness. The analyses presented previously adjusted for a few characteristics that were recorded at intake and for the design, but did not adjust for a comprehensive set of background variables of the children. Specifically, base-year literacy assessment scores (that were different between treatment and control groups) were not included as controls in those analyses, because they were collected after the intervention had begun in some communities. In this chapter, we present findings that include some of these controls in small communities, to assess whether the intervention produced impacts over and above potential base-year differences in reading performance. To explore whether EpC was more effective for the youngest children in the sample, we examine impacts on the primary literacy outcomes by children's grade level at intake. We conclude by discussing program costs in light of the findings.

A. Robustness of findings in small communities after controlling for base-year differences

To test the robustness of our findings in small communities, we controlled for differences in the literacy skills of the treatment and control groups, which were observed at the time of the base-year measurement early in the intervention period. We would ideally have measured all of the evaluation's outcomes at baseline, before random assignment, but that was not feasible. Instead, we measured base-year literacy in the months after random assignment. Use of the base-year data as late baseline data could potentially generate bias in the post-test impact estimates, because pre-test scores could reflect the early effects of EpC, making the differences between pre- and post-tests for children in the treatment group look smaller than they are (Schochet 2010). However, pre-test scores can substantially increase the precision of the estimates of treatment impact. There is, therefore, a trade-off between higher precision and a downward bias in treatment impact estimates. For the analysis presented in Chapter VI, we chose the reduced precision, which yielded unbiased estimates that do not adjust for base-year literacy. Here, we present the other approach.

We found that when we adjusted for the differences in literacy outcomes between the groups during the base year, the impacts on literacy skills were attenuated and became non-significant, or significant at the 10 percent level, for reading comprehension (Table VIII.1). Thus, even though base-year differences in literacy outcomes were small (0.13 to 0.15 standard deviations), they may explain many of the follow-up differences between the treatment and control groups. This suggests that to the extent we found impacts of EpC, such impacts may have been generated rapidly (by the time base-year data were collected), and they did not increase over time. If stakeholders were interested in producing impacts similar to those found for the 18-month version of EpC, a shorter version of the program could be more cost-effective. Research on the effectiveness of different lengths of exposure to EpC is needed to determine the optimal duration of the program, as positive impacts may fade out with shorter exposure.

Table VIII.1. Impacts of EpC on primary child outcomes in small communities, adjusting for base-year literacy skills

Child outcomes	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Attendance (percentage of days present in the last month) (0 if not enrolled)	68.9	69.7	-0.9	0.777	-0.02	1,170
Pseudo-word decoding (# correct per minute)	17.5	16.4	1.1	0.185	0.06	1,164
Reading fluency (# correct words per minute)	40.1	38.0	2.1	0.238	0.05	1,162
Reading comprehension (combined) (percentage correct out of 10)	39.0	35.9	3.1*	0.060	0.09	1,162
Social competence (range 0 to 3)	2.3	2.2	0.0	0.266	0.06	1,151
Moral disengagement (range 0 to 3)	1.1	1.1	0.0	0.976	0.00	1,164
Number of children	**592**	**578**				
Number of communities	**74**	**73**				

Source: Base-year and follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, whether the child was overage for his/her grade at intake, and base-year literacy scores. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item nonresponse. Cohort 2 base-year mean scores were used to impute the scores for Cohort 1 children, for whom base-year data were not collected.

*Difference in group means is statistically significant at the .10 level.

B. Impact of EpC on children's primary literacy outcomes in small communities, by grade at intake

To explore whether EpC was more effective if implemented earlier, when children are learning the foundational skills that they will need as they move to higher grades, we examined the impacts of EpC separately for students at each grade level, including for children who were categorized as being out-of-school at the time of intake. The findings differ sharply across grades and indicate that EpC had positive impacts on out-of-school children and children who were in first grade at the time of intake, but not on children in second and third grades.

Among children who were out-of-school at intake, those assigned to EpC decoded 4.7 more invented words, read 11 more words per minute, and answered 8 percent more reading comprehension questions (or about half a question) correctly than children in the control group. These differences range from 0.21 to 0.28 standard deviations. The impacts on decoding and oral fluency were significant at the 5 percent level, while the impact on reading comprehension was significant only at the 10 percent level. We found smaller differences on the same outcomes for children who were in first grade at intake, ranging from 0.16 to 0.20 standard deviations. The differences for decoding and fluency were significant at the 10 percent level, whereas the difference in reading comprehension was significant at the 5 percent level. Children who were in second grade at intake and were assigned to EpC only differed from children in the control group in reading comprehension. This difference was equivalent to 0.18 standard deviations and was significant only at the 10 percent level. There were no discernable impacts for children who were in third grade at intake (Table VIII.2).

Table VIII.2. Impact of EpC on primary literacy outcomes in small communities, by grade at intake

	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Out-of-school						
Pseudo-word decoding (# correct per minute)	12.5	7.8	4.7**	0.019	0.28	331
Reading fluency (# correct words per minute)	29.0	17.9	11.1***	0.009	0.26	331
Reading comprehension (combined) (percentage correct out of 10)	27.6	19.5	8.0*	0.057	0.21	331
Number of children	**171**	**160**				
Number of communities	**58**	**58**				
First grade						
Pseudo-word decoding (# correct per minute)	12.6	10.0	2.5*	0.051	0.18	387
Reading fluency (# correct words per minute)	27.7	22.9	4.9*	0.086	0.16	385
Reading comprehension (combined) (percentage correct out of 10)	31.7	25.2	6.5**	0.029	0.20	385
Number of children	**179**	**208**				
Number of communities	**66**	**66**				
Second grade						
Pseudo-word decoding (# correct per minute)	25.1	22.8	2.3	0.215	0.17	257
Reading fluency (# correct words per minute)	57.1	51.8	5.3	0.162	0.18	257
Reading comprehension (combined) (percentage correct out of 10)	54.3	49.2	5.0*	0.088	0.18	257
Number of children	**134**	**123**				
Number of communities	**63**	**62**				
Third grade						
Pseudo-word decoding (# correct per minute)	30.6	31.5	-0.9	0.631	-0.06	188
Reading fluency (# correct words per minute)	71.3	74.9	-3.6	0.318	-0.12	188
Reading comprehension (combined) (percentage correct out of 10)	58.2	60.2	-2.0	0.631	-0.08	188
Number of children	**105**	**83**				
Number of communities	**54**	**49**				

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item nonresponse.

*Difference in group means is statistically significant at the .10 level.

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

***Difference in group means is statistically significant at the .01 level.

The impacts on out-of-school children were robust to the inclusion of base-year adjustments (for Cohort 2 only), but they were attenuated. Even though the magnitude of the effect was smaller, these results indicate that by end line children who were out-of-school at intake were able to perform at a level comparable to that of children who were in first-grade. This can be interpreted as EpC having the intended academic leveling effect among this subgroup of children. The impacts on first grade children became non-significant after including base-year adjustments, suggesting that the potential early boost for this group was not sustained over time (see Table VIII.3).

Table VIII.3. Impact of EpC on literacy outcomes in small communities by grade at intake, adjusting for base-year literacy skills

	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Out-of-school						
Pseudo-word decoding (# correct per minute)	12.2	8.2	3.9**	0.015	0.24	331
Reading fluency (# correct words per minute)	27.6	19.3	8.3**	0.015	0.20	331
Reading comprehension (combined) (percentage correct out of 10)	27.1	19.7	7.4**	0.041	0.20	331
Number of children	**171**	**160**				
Number of communities	**58**	**58**				
First grade						
Pseudo-word decoding (# correct per minute)	11.1	11.4	-0.3	0.763	-0.02	387
Reading fluency (# correct words per minute)	24.3	25.9	-1.6	0.498	-0.05	385
Reading comprehension (combined) (percentage correct out of 10)	29.0	27.7	1.3	0.632	0.04	385
Number of children	**179**	**208**				
Number of communities	**66**	**66**				
Second grade						
Pseudo-word decoding (# correct per minute)	24.7	22.8	1.9	0.180	0.14	257
Reading fluency (# correct words per minute)	56.0	51.9	4.1	0.156	0.14	257

	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Reading comprehension (combined) (percentage correct out of 10)	53.1	50.0	3.1	0.268	0.11	257
Number of children	**134**	**123**				
Number of communities	**63**	**62**				
Third grade						
Pseudo-word decoding (# correct per minute)	30.1	31.7	-1.6	30.1	0.264	188
Reading fluency (# correct words per minute)	70.5	75.1	-4.7*	0.05	0.052	188
Reading comprehension (combined) (percentage correct out of 10)	58.9	59.1	-0.2	58.9	0.962	188
Number of children	**105**	**83**				
Number of communities	**54**	**49**				

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item nonresponse.

*Difference in group means is statistically significant at the .10 level.

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

Early reading interventions can produce substantial improvements in children's decoding skills within a few months (Abadzi, 2011). Results from this evaluation suggest that EpC may have rapidly improved decoding skills for children with little prior exposure to school at intake (specifically out-of-school children and first graders), but not for children who were enrolled in school before EpC. At least two mechanisms may explain these results. First, it may be that EpC was best positioned to help children who had little to no decoding skills at intake, than children who had already acquired basic decoding skills. In other words, children's *initial ability* could have moderated the impact of EpC. We find some partial evidence in support of this hypothesis. At end line, there is a positive significant effect on invented word decoding for children in small communities who were unable to identify a single letter correctly in the base year (n=220, b=1.83, $p < 0.05$). However, we do not find a significant impact on children who were able to identify at least one letter correctly in the base-year assessment (n=700, b=1.32, $p < 0.25$). After that rapid leveling effect on children with low initial ability, EpC appears to have been unable to make a difference in children's performance (Table H.11). We do not find the same pattern for children who were able, or unable, to identify a single word correctly in the base year.

An alternative hypothesis is that EpC was more effective for younger children, who may have been more receptive to learning foundational literacy skills. In this case, *age* would moderate the impact of EpC. We do not find consistent evidence in support of differential impacts by age. We tested impacts on decoding separately for children who were 7 years old or younger and children who were 8 years old or older at intake (Table H.12). We found a positive effect on invented word decoding that was significant at the 10 percent level for the older cohort, but not the younger cohort; and a positive impact on reading fluency that was significant at the 10 percent level for the younger cohort, but not the other cohort. Together, these findings suggest that initial ability, and not age at intake, may be the main driver of the effects of EpC. However, more research is needed to generate evidence in support of this hypothesis. These results should be interpreted with caution, because the evaluation was not designed to test these hypotheses and breaking down the sample into smaller ability or age subgroups reduces our statistical power. In general, there is little empirical evidence suggesting that there are marked sensitive or critical periods for acquiring skills such as those related to literacy and numeracy (Thomas & Knowland, 2009).

C. Cost-effectiveness of the intervention

Policymakers need to know if the EpC intervention's impacts on children's learning were large enough to justify its costs. Using cost-effectiveness analysis, we can produce a statistic that directly compares the effects of an intervention to its costs. It can be stated as the ratio of the effects of an intervention to its costs—that is, the cost per unit of effect. We estimated that the cost-effectiveness of the EpC intervention was between $45 and $358 per 0.10 standard deviation increase in literacy scores[28]. In this section, we explain how we obtained these estimates, and discuss how they compare with those for other interventions.

1. Methods and data

We used two methods to estimate EpC costs. The costs incorporate materials delivery, facilitator salaries, transportation, and project overhead.

The first method is based on the original CARS budget and the planned rollout of EpC across three cohorts. Because Cohorts 1 and 2 are the focus of the impact evaluation, the goal was to determine the costs associated with EpC design and initiation, and with the rollout to Cohorts 1 and 2 (excluding Cohort 3). Cohorts 1 and 2 represent about 38 percent of targeted CARS child beneficiaries. Because the initial CARS contract from 2013-2017 was $9.96 million, we estimate that Cohorts 1 and 2 represent approximately $3.8 million (which is 38 percent of the costs). However, if the other CARS activities are relatively more expensive than EpC activities (for example, the formal schooling component), this approach may not be appropriate.

This process results in a cost estimate of $830 per beneficiary (or approximately $30,000 per EpC[29] assuming 35 beneficiaries per EpC). If we include Cohort 3, the cost per beneficiary for EpC is $466 (the total budget for EpC is $6.7 million, which makes sense because the upfront costs of EpC are included in the earlier cohorts). This is likely an upper bound of the estimate of costs, because many of these overall costs incorporate one-time activities, such as initial project setup, development of training materials, adaptation of training materials (such as the EpC modules) to the local context, training of and support to local NGOs to implement EpC, and recruitment to the project and evaluation, some of which may also benefit other project participants (including participants of the pre-primary activity and the formal schooling activity).

[28] These estimates are based on the total cost of implementing EpC for an average of 18 months. Estimated cost-effectiveness per beneficiary per year (12 months) are $30 and $239 USD, depending on the method used.

[29] The cost estimate of an EpC per year (12 months) is approximately $19,365.

The second method is based on estimates from the CARS project team and it uses project administrative data that aggregate actual costs from across the three NGOs. This method results in the estimate that each EpC costs approximately $2,070, which results in a cost per beneficiary of $60 (assuming 35 beneficiaries per EpC). These values do not incorporate initial project setup costs that are incorporated in the first method, but they do include transport costs to the remote communities. These costs likely represent those that would be incurred if CARS were to set up another EpC in the RACCS. Indeed, they are similar to the estimates for later cohorts of EpC that were opened through CARS.

2. Cost-effectiveness: discussion

If we assume that the intervention’s estimated impacts on small communities according to the analysis presented in Chapter VI are the true impact of the intervention, we can estimate cost-effectiveness of the EpC model. Cost-effectiveness is calculated by dividing incremental cost per beneficiary by the incremental improvement (or estimated impact) per beneficiary.

- Using the first method of calculating project costs, the cost-effectiveness of EpC is estimated to be $358 per 0.1 standard deviations in improvement in literacy skills, which we assume to be an upper bound on the estimate.
- Using the second method of calculating project costs, the cost-effectiveness of EpC is estimated to be $45 per 0.1 standard deviations in improvement in literacy skills, which we assume to be a lower bound on the estimate.

We can compare the cost-effectiveness estimates of the EpC intervention to those for other interventions focused on improving test scores. Although the range of $45 to $358 per 0.1 standard deviations in improvement in literacy skills is wide, it is on the middle to high end of the cost-effectiveness range that has been published for other education interventions designed to improve learning. We compared our estimates to those found in a systematic review of the impacts and cost-effectiveness of educational investments in developing countries by Evans and Ghosh (2008) and to other more recent evaluations of early reading programs. In making those comparisons, it is important to note the challenging context in which the intervention was rolled out, and the complexity of the intervention. The low-end estimate is on par with similar reading programs in low-income countries. For instance, LAC Reads Honduras found a cost-effectiveness of $52 and $57 for the two different interventions (Liuzzi et al. 2018), and the cost-effectiveness of Niger’s NECS early grade reading intervention was $24 (Bagby et al. 2017a).

Our approach follows common practices for the estimation of cost-effectiveness in education (Hummel-Rossi & Ashdown, 2002). We focus on the impact on children’s literacy outcomes after one year and a half of intervention and refrain from estimating potential long-term costs and benefits, impacts on other potential child outcomes, or costs and benefits for society at large. This approach facilitates the identification and measurement of inputs and outputs, and allows for comparisons with similar education programs. However, it does not consider other potential benefits, such as those accrued from building the capacity of local NGOs and educators.

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

IX. CONCLUSIONS, LIMITATIONS, AND RECOMMENDATIONS

The EpC theory of change hypothesizes that providing a space for children to participate in high quality academic, personal growth, and recreational activities that are adapted to the local context will improve children's literacy skills, readiness to learn, and their social-emotional skills and attitudes. By spending time in a safe learning environment after or before school, children have the opportunity to develop these desirable skills and attitudes rather than engaging in other (potentially less productive or even harmful) activities, such as child labor, or being exposed to violence and abuse. Over time, the increased attachment to school, accelerated learning in the early grades, and positive social-emotional skills and attitudes should result in improvements in educational achievements and employment, and reductions in risk factors for illicit or unsafe activities.

This evaluation is the first to test the causal impacts of EpC on children's outcomes. We implemented a mixed-methods evaluation, comprised of a rigorous impact evaluation and a qualitative performance evaluation, to answer three evaluation questions whose purpose was to (1) assess whether the EpC intervention was implemented as planned, (2) estimate the impact of EpC on children's academic and social-emotional outcomes, and (3) assess the cost-effectiveness of EpC. In this chapter, we summarize the findings corresponding to each evaluation question, discuss our conclusions and limitations of the evaluation, and offer a set of recommendations.

A. Summary of findings

1. Was the EPC intervention implemented as intended? What services were delivered at the child, school, educator, and community levels?

The intervention was implemented largely as intended, but there were some important challenges. EpC were opened in targeted communities, and musical equipment and musical instruments, sports equipment, and arts and crafts materials were delivered. Enough EpC facilitators (who were, in some cases, also teachers) were trained to teach reading, apply active learning methodologies, and use reading assessments to monitor students' learning progress. Trained facilitators (both teacher-facilitators and facilitators-only) engaged in the instructional practices they were trained on for EpC (for example, using music, routines, didactic materials, flip charts, and having students read books other than textbooks and work in pairs), and were much more likely to do so than teachers who were not trained. CARS staff conducted follow-up visits with all EpC facilitators an average of 3.3 times during the 18-month intervention to support the facilitators in implementing the methodologies they learned during training.

In small communities we assigned communities as a whole to either the EpC or the control group, while in large communities we randomly assigned children to either group. Small educational communities had at least 20 and fewer than 50 eligible children, while large communities had 50 or more eligible children and could support both an EpC and a control group.

CARS confronted some general implementation challenges, as well as challenges that were more pronounced in large communities compared to small communities. First, even though the program had a strong desire to reinsert out-of-school children into school, most children in the sample who were categorized as out of school during the intake process can be best characterized as those that started school later than others (either later than recommended or were young at the time of identification during intake). The EpC intervention was initially designed to focus on decreasing child labor and enrolling out-of-school children into school (reinsertion), in addition to improving children's early literacy skills. However, most out-of-school children in the sample were too young to have enrolled and dropped out. Thus, it appears that the program was unable to engage a primary segment of the population it was intended to reach.

Second, while CARS successfully adapted the EpC training materials to the RACCS context, the project faced challenges in providing supplemental reading materials, such as story-books, that were adapted to the local context. Rather than creating reading materials that were adapted to the cultural and linguistic context of the RACCS, CARS purchased materials in languages spoken in the target communities in Cohorts 1 and 2.[30] Providing reading materials in students' languages is essential, but adapting materials to reflect students' contexts can further advance the goal of creating culturally responsive learning environments (Gay, 2002). In addition, there were delays in the distribution of those materials to communities. The delays reduced the time of exposure to reading materials for children in EpC, particularly in Cohort 1, the majority of which are large communities. Furthermore, surveys of school directors at end line showed that a lack of reading materials other than textbooks was a persistent issue in small communities, in both treatment and control schools. It is possible that CARS reading materials did not make it to all EpC in small communities, or that those materials were delivered to the EpC but not shared with the school. Therefore, in many EpC communities, a persistent lack of reading materials (non-textbook and textbook [only 35% of large communities have textbooks for shared use and 51% of small (treatment) communities have textbooks for shared use]) might be limiting the potential for children to improve their literacy skills.

In addition, follow-up visits to EpC facilitators happened less often than planned (an average of only 3.3 visits took place, in contrast with the planned 8 visits). For Cohort 1 in particular, follow-up visits only started taking place 9 months into the implementation of EpC. This is due in part to the geographical dispersion of target communities and the fact that CARS staff were tasked with recruiting communities while supporting newly established EpC. Evaluation requirements may have inadvertently contributed to this challenge, as CARS had to recruit enough communities to allow for the creation of a control group. Also, even though trained EpC facilitators reported using the techniques they were trained on, in large communities there were considerably fewer differences between EpC facilitators and teachers who were not EpC facilitators than there were in small communities. Specifically, non-EpC facilitator teachers in large communities reported they were engaging in many of the instructional practices used in EpC, which may have decreased the likelihood of detecting impacts on children's learning outcomes in large communities.

Lastly, children's participation levels in EpC were lower than expected and a non-trivial proportion of children in the control group participated in EpC. These problems were more acute in large communities, where treatment and control children lived in the same community. In

[30] CARS provided materials adapted to the local context to later cohorts, which were not included in this evaluation.

small communities, 73 percent of children assigned to EpC reported participating, and about 10 percent of children assigned to the control group did. However, in large communities, less than half of children assigned to EpC and 20 percent of those assigned to the control group reported participating. Non-adherence to the experimental protocol (reflected in lower than anticipated participation from those assigned to EpC, and higher than anticipated participation from those assigned to the control group) attenuates the contrast between the two groups in large communities and increases the probability of finding null impacts.

Several factors likely underlie the differences in adherence to the experimental protocol between small and large communities. Children in large communities were more likely than children in small communities to go home for lunch and to rely on public transportation to get back to school, which could decrease their likelihood of returning to the school for the EpC. Children in small communities were more likely to walk to school and more likely to have a school-provided meal than children in large communities were, which could make it more likely for them to stay at or return to school for the EpC. Also, in some small communities, children assigned to the control group needed to travel to a different community in order to attend the EpC, which helps to explain why fewer control group children participated in EpC than in large communities.

2. What impact does the CARS EpC intervention have on reading, attachment to school, social-emotional skills, and attitudes?

This report focuses primarily on findings from small communities because, as previously discussed, the actual contrast between the experimental groups in large communities was smaller than expected. We also describe findings from the full combined sample of small and large communities because the original analysis plan aimed to estimate the average impact of EpC across all communities. We anticipated findings from the full sample to be attenuated due to high levels of non-adherence with the experimental protocol in large communities.

In small communities, we find that EpC had positive impacts on child reading outcomes, but not on school attachment or social-emotional outcomes. Specifically, EpC had a positive impact of 0.12 standard deviations on decoding (2.1 invented words per minute) and of 0.11 standard deviations on reading fluency (4.4 words per minute), which are two of the first emerging literacy skills that EpC should affect. These impacts represent about a 10 percent difference from the control group’s mean for these outcomes. EpC also improved reading comprehension, with children in EpC scoring 4 to 5 percentage points better on this measure than control group children.

Subgroup analysis confirmed that the learning impacts were statistically significant for girls and for children who were out of school at intake, but not for boys and children who were already enrolled in school. Girls in the treatment group were able to read correctly 3 more invented words and 5.4 more words in one minute, and to answer about half of a question more about the reading passages than girls in the control group could. There were, however, no detectable impacts on boys. Among children who were not enrolled in school at intake, those in the treatment group correctly read 4.4 more invented words and 11 more words per minute than children in the control group could, and were able to answer correctly nearly one more question about the reading passages than children in the control group were. There were no significant impacts for children who were enrolled in school at intake, except for an impact in reading comprehension that was significant at the 10 percent level only. The differences in impacts by subgroup (between boys and girls, and between out-of-school and enrolled children) were not statistically significant, so we cannot conclude that there were differential impacts across subgroups.

After we controlled for base-year differences, however, we found that the intervention's impacts on literacy skills outcomes were attenuated and became non-significant, or significant at the 10 percent level in the case of reading comprehension. This suggests that to the extent that we see impacts of EpC, such impacts may have been generated rapidly (by the time base-year data were collected) and did not increase over time.

When we consider the full sample of small and large communities, the pattern of EpC-control differences suggests that any potential impacts on child outcomes were too small to detect, with only one outcome, decoding, showing a statistically significant difference. Children in EpC were able to decode 1.4 more invented words than children in the control group were. It is not surprising that findings from the two sets of analyses are different, because the treatment/control contrast was so much weaker in the large communities. These differences are probably due to features of the evaluation design; specifically assigning whole communities to EpC versus assigning individual children within communities. However, they could also have been due to differences in population characteristics. To explore this possibility, we compared small and large communities in a range of child and school characteristics.

We found that small and large communities differ in ways that could have influenced children's participation in EpC and children's likelihood of benefiting from the intervention.

- Children in small communities were more likely to walk to school instead of taking public transportation, and more likely to have a school-provided meal than children in large communities were. These factors could have made it easier for children in small communities to participate in the EpC.
- Children in small communities were from poorer families and participated in fewer extracurricular activities than children in large communities. Greater access to academic and recreational activities outside of school for children in large communities could have reduced the perceived value of EpC and result in low take-up or attendance.
- More children in large communities attended an afternoon shift at school. The low participation levels for EpC may be a consequence of this. For example, children or households may prefer afternoon shifts to avoid conflicts with work or family responsibilities, and this practice could have conflicted with the time of day in which the EpC was open in the community. Similarly, changing the time at which a child leaves for school might be more challenging than changing the time at which a child returns from school.
- Children in small communities who were assigned to the control group had literacy scores that were similar but not equal to children in the control group in large communities. Because most small communities were in rural areas and most large communities were in urban areas, we disaggregated the control group follow-up values by urban/rural status to

better understand the contribution of community size and urbanicity. We found that children in large urban communities had better literacy skills than children in all other areas, whereas children in large rural communities performed similarly to children in small communities (both urban and rural). The effect of the EpC intervention on children in small communities brought their skills closer to the level of children in urban large communities.

- Schools in small communities were more likely to have multi-grade classrooms than schools in large communities were. Even though multi-grade classrooms can expand education access to marginalized populations, they pose pedagogical challenges for teachers who are likely trained to teach in mono-grade classrooms and who may not have teaching materials designed to facilitate learning in multi-grade settings.
- Schools in small communities were open fewer days per month than schools in large communities were. Over the full academic year, the difference amounts to almost a month less of instruction for students in small communities, which likely contributes to their lesser academic achievement.
- In large communities, parents of treatment and control group children had the option to participate in classes for mothers and fathers (or EPM for its acronym in Spanish) that were supported by the CARS project. In small communities, only parents of children in the treatment group had that option[31]. The classes were aimed at encouraging parents to support their children's learning and wellbeing. Topics included strategies to support children's emerging literacy skills, bullying awareness and prevention, gender equality, and values education, all of which could have influenced the child outcomes that EpC aimed to improve. We find that parents of children assigned to the control group were more likely to participate in these classes if they lived in large communities than in small communities.
- In large communities, teachers who were not trained by CARS used many of the same instructional techniques used in EpC. In large communities, educators trained in EpC were more likely than those who were not trained to use music, traditions and routines, and songs. However, in contrast with small communities, there were no differences in the use of practices more likely to have a direct impact on academic outcomes, such as using visual aids and didactic materials, and having children work in pairs or small groups. Changes in the instructional practices of educators trained by CARS were therefore more likely to have an impact on children's outcomes in small communities, where the quality of instruction in the EpC differed significantly from what teachers who were not trained by CARS were doing in the classroom.

We cannot be certain whether null findings from the full sample are a consequence of the evaluation design or of pre-existing differences between small and large communities. However, that the characteristics of large communities are quite different from small communities limits our ability to conclude that findings in small communities are generalizable to large communities. Characteristics in large communities may have made it more difficult and less

[31] All parents in large and small communities could have participated in parent schools organized by the Ministry of Education (assuming they were functioning), but not all parents could have participated in parent schools supported by CARS. In large communities, CARS supported all parent schools organized by the Ministry, whereas in small communities it only supported parent schools in treatment communities.

relevant for children to participate in EpC than in small communities. Future implementation of EpC may require more fine-tuning to the distinct needs of different populations.

The lack of impacts on school attachment and child social-emotional outcomes can be due to multiple factors. We offer some hypotheses to be examined in future research. EpC likely did not have an impact on attendance (our main attachment indicator) because attendance was already high overall at baseline (over 70 percent in both treatment and control groups) and EpC was not designed to mitigate some common reasons for children's absenteeism in the RACCS, primarily illness and agricultural or domestic responsibilities (as per reports from caregivers in our sample). In addition, our performance evaluation found that long commutes and weather conditions may heavily affect children's attendance in the RACCS, neither of which was addressed in this intervention. Regarding social-emotional outcomes, it is possible that the lack of impacts is due to characteristics of the intervention. Our review of implementation materials indicates that although the EpC modules had a clear focus on the acquisition of specific literacy skills, the same does not appear to be the case for social-emotional skills. Instead, the modules included activities for a multiplicity of social-emotional and related domains, including self-esteem, emotion recognition, regulation, and expression, communication skills, as well as values such as honesty, solidarity, and friendship. While the EpC approach is built on evidence-based practices to improve the learning environment and level up student academic skills (ICF Macro 2011), it is unclear whether the combination of activities designed to improve social-emotional skills (including the time dedicated to each skill or value) is based on previous evidence and could reasonably generate an impact on the outcomes measured in this evaluation. The literature shows that sequenced interventions (those that lay out a sequence of coordinated activities designed to develop each social-emotion skill), as well as interventions that use evidence-based practices to develop each skill, are most likely to yield positive results (Durlak et al. 2010). Future implementation of EpC may benefit from focusing on a smaller number of discrete social-emotional skills for the full duration of the program and ensuring measurement of the targeted skills.

3. What are the costs of EpC? Are the effects of EpC large enough to justify the costs?

Cost-effectiveness estimates for the EpC intervention ranged from $45 to $358, depending on how EpC costs were estimated, per 0.1 standard deviations in improvement in literacy skills. These estimates are on the middle to high end of the range found for other education interventions, including early reading interventions in developing countries (see for example Liuzzi et al., 2018 and Bagby et al., 2017a).

Although the range of $45 to $358 per 0.1 standard deviations in improvement in literacy skills is wide, it is on the middle to high end of the cost-effectiveness range that has been published for other education interventions designed to improve learning. We compared our estimates to those found in a systematic review of the impacts and cost-effectiveness of educational investments in developing countries by Evans and Ghosh (2008) and to other more recent evaluations of early reading programs. In making those comparisons, it is important to note the challenging context in which the intervention was rolled out and the complexity of the intervention. The low-end estimate for EpC is on par with similar reading programs in low-income countries. For instance, LAC Reads Honduras found a cost-effectiveness of $52 and $57 for interventions on the use of summative and formative assessments, respectively (Liuzzi et al.

2018), and the cost-effectiveness of Niger's NECS early grade reading intervention was $24 (Bagby et al. 2017a).

Our approach follows common practices for the estimation of cost-effectiveness in education (Hummel-Rossi & Ashdown, 2002). Thus, we focus on the impact on children's literacy outcomes after one year and a half of intervention, instead of also estimating potential long-term costs and benefits, impacts on other potential child outcomes, or costs and benefits for society at large. This facilitates the identification and measurement of inputs and outputs and permits comparisons with similar education interventions. However, we do not consider other potential benefits, such as those accrued from building the capacity of local NGOs and educators.

B. Conclusions

EpC is a promising after-school approach to improve early grade reading outcomes in poor and hard-to-reach communities in Nicaragua. Although the evaluation did not to detect impacts of EpC on all hypothesized outcomes for all populations, there was evidence of improvement in word-level decoding, oral reading fluency, and comprehension skills in small, largely rural communities where there was a meaningful contrast in participation status. These findings were driven by statistically significant impacts for girls and for children who were out of school at intake. Even though the effects are relatively small, our findings indicate that EpC is a promising approach to improve the literacy skills of Nicaraguan children in remote, under-resourced communities. Policymakers should consider the cost, however, because our estimates include the possibility that EpC requires a heavier investment than comparable interventions, which have yet to be tested for cost-effectiveness in the RACCS. In fact, despite the challenges CARS faced, the lower-bound cost estimate of delivering EpC in the RACCS is less expensive than in the Dominican Republic ($30 compared to $120 per child for 12 months) (ICF Macro 2011). To optimize the cost-effectiveness of future investments, we recommend identifying the core components of EpC that underlie improvements in children's skills or reducing the duration of the program, as our results indicate that impacts may have taken place in the first months of the program and did not increase over time.

Offering EpC to those that are hardest to reach (often rural residents) can be expensive, but when done well can have impact. Children living in harder to reach areas (e.g., remote rural areas) often have lower educational outcomes and more room for improvement than children living in more accessible areas, and their families and communities are more likely to take up the intervention if there are few or no similar services available. In this evaluation, we found positive impacts of EpC on the reading outcomes of children living in hard to reach rural communities. In countries aiming to provide better access to high quality education, large investments may be warranted to improve the academic outcomes of the most disadvantaged populations, including those that are linguistically diverse, poor, and hard to reach.

EpC had challenges reaching older children who were not enrolled in school because they had either dropped out or had never enrolled. The intervention aimed to reach children who had dropped out of school or had never enrolled. However, most of the out-of-school children who were eligible to receive EpC and agreed to participate in the evaluation were 5 to 7 years old and most enrolled in school during the evaluation. This intervention was unable to recruit older children who had dropped out of school. The evaluation is unable to test whether the reason is that EpC did not mitigate the need for older children to work, but prior research in Nicaragua indicates that labor competes with children's participation in school (e.g., Del Carpio & Macours, 2009). Without the right incentives, after-school programs designed to make school attractive might have limited success in overcoming outside constraints to pursuing education, such as competing employment opportunities.

The findings from this evaluation can help us understand the implementation of EpC in the RACCS in its early stages. However because the program continued to evolve, it would be worthwhile to evaluate again while in steady state. Prior to this evaluation, the EpC intervention had been piloted in other countries and in other regions in Nicaragua. However, there were challenges specific to the RACCS that the implementing team discovered as implementation was underway. Because of these challenges CARS and USAID modified the EpC model in the Caribbean region for later cohorts, as well as for the extension of the project to the RACCN[32], largely to focus even more strongly on literacy. Therefore, this evaluation reflects impact findings for a version of EpC model that is not quite the same as the model that is currently being implemented in the RACCS. In addition, the implementing NGOs had no prior experience with EpC, so CARS trained the NGOs in the EpC process as the intervention was rolled out. Another evaluation of EpC in steady state in the RACCS would be informative and could show the program to be more cost-effective.

C. Evaluation limitations

Several limitations of the evaluation should be considered when interpreting the findings in this report.

A limitation of this evaluation is that we are unable to draw conclusions about the effectiveness of EpC in large, primarily urban communities in the RACCS. The evaluation used a hybrid randomization design, whereby within-community or between-community random assignment was employed. The experiment worked well in small communities, where all eligible children were assigned to the same experimental condition. However, in large communities, where we implemented within-community random assignment, a sizable proportion of children assigned to the control group reported participating in EpC. In addition, there was surprisingly low take-up of the EpC intervention in large communities among children assigned to participate. Low take-up of the intervention and contamination reduced the contrast between the experimental groups and increased the probability of finding null impacts in large communities, which represent 24 percent of the communities in the evaluation. We are therefore unable to conclude that the EpC approach was ineffective in large communities because we cannot rule out the possibility of positive impacts, had compliance with the experimental protocol been stronger.

Second, it was not possible to collect baseline data on all children's primary outcomes before the intervention was rolled out. Even though randomized control trials do not require pre-intervention data to draw unbiased conclusions about the impacts of an intervention (because baseline differences in randomized trials can be assumed to occur by chance), such data can

[32] RACCN stands for "Region Autónoma de la Costa Caribe Norte"

increase the statistical power to detect intervention impacts by reducing the amount of unexplained error in the estimation model.

Finally, the cost effectiveness analyses conducted in this evaluation are based on imprecise cost estimates due to the complexity of tracking costs across multiple implementing partners and because the program had other objectives aside from what we studied in this evaluation. The analyses provide a broad range for the overall cost effectiveness of EpC within the context of the RACCS in Nicaragua. They reflect costs ranging from those required to start a brand new program (based on the contract) to those required to open a new EpC once the program (according to the implementer) activities are at steady state. Opening EpC in another context, with different languages and a different social and geographic landscape, would likely have different costs.

D. Recommendations

A systematic exploration of different packages of materials delivered as a part of EpC and differing exposure periods to EpC would be important to inform future investments in EpC. The EpC model includes many activities that may contribute to the impacts observed, but that also increase program complexity and make it more difficult and costly to implement. Identifying the core package of materials and activities that are most cost-effective would be worthwhile, especially if the program continues to be implemented with hard to reach populations. In addition, findings from the evaluation suggest that the impacts of EpC on child literacy were generated within the first two months of implementation and did not increase over time. Thus, a shorter version of EpC might be more cost-effective than the 18-month version tested in this evaluation. Additional research to identify the optimal duration of the program is warranted, as positive impacts might not be sustained if the time of exposure is reduced. In addition, further research to compare different combinations of materials to support program implementation, such as books that are or are not adapted to the local context, musical instruments, and sports equipment could be useful in optimizing the program.

After-school interventions should be targeted to populations that are most in need of the intervention activities. Children in small communities were more disadvantaged than those in large communities. They participated in fewer academic and recreational activities outside of school and performed worse in the early reading assessments compared to children in large communities. Limited access to extracurricular activities might have increased the perceived value of EpC and contributed to higher take-up among children in small communities. Further, the fact that there was more room for improvement in the reading skills of children in small communities possibly made it easier to generate and detect impacts of EpC than in large communities.

Barriers to participation specific to the populations targeted by the intervention should be addressed or the targeted population should be narrowed to those most likely to benefit from EpC activities. The EpC intervention aimed to involve children who were at risk of dropping out of school, had dropped out, or had never enrolled. However, most of the out-of-school children identified as eligible to receive EpC were 5 to 7 years old and had never been to school. This indicates that the recruitment process was not successful with recruiting eligible older out-of-school children who had dropped out of or never enrolled in school. This evaluation is unable to test whether recruitment of older children was not successful because EpC did not mitigate the need for children to work, but prior research in Nicaragua indicates that labor competes with children's participation in school (e.g., Del Carpio & Macours, 2009). When the EpC model was first tested in Nicaragua, one of the recommendations from the qualitative study was to add income generation and technical training components to help families compensate for income lost if children left work to return to school (Macro International Inc. 2009). Alternatively, the EpC activities may have been less attractive to older children compared to younger children. In Colombia, the EpC model was implemented alongside "Spaces for Entrepreneurship" (Espacios para Emprender, in Spanish) to respond to the needs and interests of different age groups. Thus, while EpC focused on children between the ages of 6 and 14, Espacios para Emprender targeted adolescents between the ages of 15 and 18 (ICF Macro 2010). Future iterations of EpC in Nicaragua should consider focusing on a narrower age range or adjusting the intervention to respond to the needs and interests of older out-of-school children.

In randomized-controlled trials with high risk of low take-up in the treatment group and of contamination in the control group, group-level randomization should be preferred over individual-level randomization. Researchers always face tradeoffs in designing evaluations. Faced with trade-offs between the within-school and the between-school random assignment design (Glazerman 2012), it is tempting to use within-school random assignment because it enables the program to serve as many children as possible while also generating information about impact. Despite carrying a higher risk for contamination, within-school designs are cheaper and potentially more powerful than the between-school alternative. The contamination risk can potentially be managed by the implementer, though it can be challenging to do and is not always successful. We recommend conducting a pilot trial to estimate the level of contamination in the control group and the level of take-up in the treatment group. Pilot trials can also offer an opportunity to detect barriers to take-up and test mitigation strategies. If a pilot trial is not feasible, and the risk of contamination is deemed high, group-randomization is advisable despite the higher costs that come with using larger sample sizes. When there is uncertainty regarding the risks of contamination and low take-up, hybrid designs, like the one employed in this evaluation, are a good strategy to balance different risks.

REFERENCES

Abadzi, H. "Instructional Time Loss in Developing Countries: Concepts, Measurement, and Implications." *The World Bank Research Observer,* vol. 24, no. 2, 2009, pp. 267–290.

Abadzi, H. "Reading Fluency Measurements in EFA FTI Partner Countries: Outcomes and Improvement Prospects." Global Partnership for Education Working Paper Series on Learning, No. 1. Washington, DC: World Bank, 2011.

Bagby, E., K. Bos, A. Dumitrescu, N. Ingwersen, and M. Sloan. "Niger NECS Impact Evaluation Report." Final report submitted to the Millennium Challenge Corporation. Washington, DC: Mathematica Policy Research, July 2017a.

Bagby, E., S. Glazerman, and N. Murray. "Evaluation of Espacios para Crecer ("Room to Grow") After-School Program for At-Risk Youth." AEA RCT Registry. September 15, 2017b. Available at https://www.socialscienceregistry.org/trials/640/history/21448. Accessed July 16, 2018.

Bagby, E., S. Glazerman, N. Murray, V. Poggio, and J. Rodriguez. "Impact Evaluation of Espacios para Crecer: Final Evaluation Plan." Submitted to the U.S. Agency for International Development/Washington. Washington, DC: Mathematica Policy Research, September 2016.

Bagby, E., C. Toledo, S. Glazerman, V. Poggio, and N. Murray. "Evaluation of Espacios para Crecer (EpC) in Nicaragua: Baseline Report." Submitted to the U.S. Agency for International Development/Washington. Washington, DC: Mathematica Policy Research, March 2017c.

Baker, D. L., R. H. Good, N. Knutson, and J. M. Watson (eds.). *Indicadores Dinámicos del Éxito en la Lectura* (7a ed.). Eugene, OR: Dynamic Measurement Group, 2006.

Blair, R., V. Poggio, C. Torrente, E. Bagby, and N. Murray. "Performance Evaluation of the Community Action for Reading and Security (CARS) Intervention in Nicaragua: Final Report." Submitted to the U.S. Agency for International Development/Washington. Washington, DC: Mathematica Policy Research, 2018.

Bland, J., and D. Altman. "Statistics Notes: Cronbach's Alpha." *BMJ,* vol. 314, 1997, p. 572.

Bond, L., S. Wolfe, M. Tollit, H. Butler, and G. Patton. "A Comparison of the Gatehouse Bullying Scale and the Peer Relations Questionnaire for Students in Secondary School." *Journal of School Health,* vol. 77, no. 2, 2007, pp. 75–79.

Broc, M.A., "Harter's Self-perception Profile for Children: An Adaptation and Validation of the Spanish Version". *Psychological Reports*, vol. 115, no. 2, pp. 444-466, 2014.Castañeda, S. "Medición de la Pobreza a Través de los Indicadores de Necesidades Básicas Insatisfecha." Instituto Nacional de Información de Desarrollo (INIDE). 2017. Available at https://www.cepal.org/sites/default/files/presentations/2017-05-saurasani-castaneda-ni.pdf. Accessed July 16, 2019.

CARE. "La Réplica del Modelo Educativo: Espacios para Crecer en Escuelas de Jinotega y San Rafael del Norte de Nicaragua. Proyecto Primero Aprendo 2." 2010.

Collaborative for Academic, Social, and Emotional Learning (CASEL). "CASEL Guide: Effective Social and Emotional Learning Programs. Preschool and Elementary School Edition." Chicago, IL: CASEL, 2013.

Cortina J. "What is Coefficient Alpha: An Examination of Theory and Applications." *Journal of applied psychology,* vol. 78, 1993; pp. 98-104. doi:10.1037/0021-9010.78.1.98.

Creative Associates International and Mathematica Policy Research. "Findings from the Four Country School Dropout Prevention Pilot Program Impact Evaluation." Washington, DC: Creative Associates International, September 2015. Available at http://schooldropoutprevention.com/wp-content/uploads/2016/03/Summary_Findings_Report_English_FINAL.pdf. Accessed September 2015.

Del Carpio, X. V., and K. Macours. "Leveling the intra-household playing field: compensation and specialization in child labor allocation". Policy Research Working Paper. *The World Bank*, 2009.

Developmental Studies Center. Scales from Student Questionnaire, Child Development Project for Elementary School Students (Grades 3–6). 2005. Oakland, CA: Developmental Studies Center. Available at https://www.collaborativeclassroom.org. Accessed July 16, 2018.

DevTech. "Technical Proposal: Community Action for Reading and Security (CARS)." Request For Proposal No. SOL-524-13-000005, USA/Nicaragua. Arlington, VA: DevTech, 2013.

Dinarte, L. "Peer Effects in After-School Programs. Experimental Evidence in El Salvador." Doctoral dissertation. Santiago, Chile: Pontifical Catholic University of Chile, 2017.

Dubeck, M. M., and A. Gove. "The Early Grade Reading Assessment (EGRA): Its Theoretical Foundation, Purpose, and Limitations." *International Journal of Educational Development,* vol. 40, 2015, pp. 315–322.

Durlak, J. A., R. P. Weissberg, and M. Pachan. "A Meta-Analysis of After-School Programs that Seek to Promote Personal and Social Skills in Children and Adolescents." *American Journal of Community Psychology,* vol. 45, no. 3-4, 2010, pp. 294–309.

Esbensen, F. -A., and D. W. Osgood. "Gang Resistance Education and Training (GREAT): Results from the National Evaluation." *Journal of Research in Crime and Delinquency,* vol. 36, no. 2, 1999, pp. 194–225.

Esbensen, F. -A., D. Peterson, T. J. Taylor, and A. Freng. "Similarities and Differences in Risk Factors for Violent Offending and Gang Membership." *Australian and New Zealand Journal of Criminology,* vol. 42, no. 3, 2009, pp. 310–335.

Evans, David K., and Arkadipta Ghosh. "Prioritizing Educational Investments in the Developing World." RAND Working Paper 587. Santa Monica, CA: RAND, 2008. Available at https://www.rand.org/pubs/working_papers/WR587.html. Accessed July 16, 2018.

Feres, J. C., and X. Mancero. "El Método de las Necesidades Básicas Insatisfechas y sus Aplicaciones en América Latina." Serie Estudios Estadísticos y Prospectivos. División de Estadística y Proyecciones Económicas. Santiago, Chile: CEPAL-ECLAC, February 2001.

Gay, G. Preparing for culturally responsive teaching. *Journal of teacher education,* vol. 53, no. 2, 2002, pp. 106-116.

Glazerman, Steven. "Random Assignment Within Schools: Lessons Learned from the Teach For America Experiment." *Education Finance and Policy,* vol. 7, no. 2, 2012, pp. 124–142.

Glewwe, P. W., E. A. Hanushek, S. D. Humpage, and R. Ravina. "School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010." Cambridge, MA: National Bureau of Economic Research, 2011.

Gove, A. "Early Grade Reading Assessment Toolkit." Washington, DC: World Bank and RTI International, 2009.

Harter, S. "Supplementary Description of the Self-Perception Profile for Children: Revision of the Perceived Competence Scale for Children." Unpublished manuscript. Denver, CO: University of Denver, 1983.

Harter S. "Self-Perception Profile for Children: Manual and Questionnaires (Grades 3–8.") Denver, CO: University of Denver, 2012.

Hennigan, Karen M., Cheryl L. Maxson, David C. Sloane, Kathy A. Kolnick, and Flor Vindel. "Identifying High-Risk Youth For Secondary Gang Prevention." *Journal of Crime and Justice*, vol. 37, no. 1, 2014, pp. 104–128.

Henson, R. K. "Understanding Internal Consistency Reliability Estimates: A Conceptual Primer on Coefficient Alpha." *Measurement and Evaluation in Counseling and Development,* vol. 34, no. 3, 2001, pp. 177–189.

Hummel-Rossi, B., and J. Ashdown. "The state of cost-benefit and cost-effectiveness analyses in education." *Review of Educational Research* 72(1), 2002, pp. 1-30.

ICF Macro. "Independent Final Evaluation of the Combatting Exploitative Child Labor Through Education in Colombia Project: *Edúcame Primero Colombia.*" Fairfax, VA: ICF International, 2010.

ICF Macro. "Independent Final Evaluation of the Combatting Exploitative Child Labor Through Education Project in the Dominican Republic." Fairfax, VA: ICF International, 2011.

Kim, Y. -S. G., H. N. Boyle, S. Simmons Zuilkowski, and P. Nakamura. "Landscape Report on Early Grade Literacy." Washington, DC: U.S. Agency for International Development, 2017.

Kremer, K. P., B. R. Maynard, J. R. Polanin, M. G. Vaughn, and C. M. Sarteschi. “Effects of After-School Programs with At-Risk Youth on Attendance and Externalizing Behaviors: A Systematic Review and Meta-Analysis.” *Journal of Youth and Adolescence,* vol. 44, no. 3, 2015, pp. 616–636.

Little, A. W. “Learning and Teaching in Multigrade Settings.” Paper commissioned for the UNESCO 2005 EFA Global Monitoring Report. 2004. Available at http://multigrade.ioe.ac.uk/fulltext/fulltextLittle.pdf. Accessed July 16, 2019.

Liuzzi, S., S. Glazerman, N. Murray, and I. Cheban. “Data-Driven Instruction in Honduras: An Impact Evaluation of the EducAcción-PRI Promising Reading Intervention.” Submitted to the U.S. Agency for International Development/Washington. Washington, DC: Mathematica Policy Research, July 2018.

Loeber, R., D.P. Farrington, M. Stouthamer-Loeber, and W.B. Van Kammen. “Multiple Risk Factors for Multiproblem Boys: Co-Occurrence of Delinquency, Substance Use, Attention Deficit, Conduct Problems, Physical Aggression, Covert Behavior, Depressed Mood, and Shy/Withdrawn Behavior.” In Jessor, R. (Ed.), *New Perspectives on Adolescent Risk Behavior.* New York, NY: Cambridge University Press, 1998, pp. 90–149.

Macro International Inc. “Independent Final Evaluation of the Primero Aprendo Project: Combating Child Labor Through Education in Central America and the Dominican Republic.” CARE, Cooperative Agreement Number: E-9-K-0045. Calverton, MD: Macro International Inc., 2009.

McEwan, P. 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.

McHugh, M. L. “Interrater Reliability: The Kappa Statistic.” *Biochemia Medica,* vol. 22, no. 3, 2012, pp. 276–282.

Prinz, R.J., Foster, S., Kent, R.N., O'Leary, K.D., “Multivariate assessment of conflict in distressed and nondistressed mother-adolescent dyads.” *Journal of applied behavior analysis,* vol. 12, no. 4, 1979, pp. 691-700.

Rosenberg, M., “Society and the adolescent self-image”. Princeton University Press, Princeton, NJ., 1965.

RTI International. “Early Grade Reading Assessment Toolkit.” March 30, 2016. Available at https://pdf.usaid.gov/pdf_docs/pa00m4tn.pdf. Accessed July 17, 2018.

Schochet, Peter Z. “An Approach for Addressing the Multiple Testing Problem in Social Policy Impact Evaluations.” *Evaluation Review,* vol. 33, no. 6, 2009, pp. 540–567.

Schochet, Peter Z. “The Late Pretest Problem in Randomized Control Trials of Education Interventions.” *Journal of Educational and Behavioral Statistics,* vol. 35, no. 4, 2010, pp. 379–406.

Skinner, E.A., Kindermann, T.A., Furrer, C .J. “A Motivational Perspective on Engagement and Disaffection.” *Educational and Psychological Measurement,* vol. 69, no. 3, 2009, pp. 493-525.

Tavakol, M., and R. Dennick. “Making Sense of Cronbach’s Alpha.” *International Journal of Medical Education,* vol. 2, 2011, pp. 53–55.

UNICEF. “Child Friendly Schools Programming: Global Evaluation Report.” New York, NY: UNICEF, 2009. Available at https://www.unicef.org/spanish/evaldatabase/files/Global_CFS_Evaluation_Report_2009(1).pdf. Accessed July 16, 2018.

USAID. “Assessment And Design Recommendations For Improving Reading Performance Of Under-Performing Children In Primary Schools On Nicaragua’s Atlantic Coast.” Washington, DC: U.S. Agency for International Development, 2012. Available at https://pdf.usaid.gov/pdf_docs/PA00KGVN.pdf

USAID/Nicaragua. “Country Development Cooperation Strategy, FY 2013-FY 2017.” Washington, DC: U.S. Agency for International Development, 2013. Available at https://www.usaid.gov/sites/default/files/documents/1862/CDCS%203-3-14%20Public%20Version%20FINAL.pdf. Accessed July 18, 2018.

Viera, A. J., and J. M. Garrett. “Understanding Interobserver Agreement: The Kappa Statistic.” *Family Medicine,* vol. 37, no. 5, 2005, pp. 360–363.

What Works Clearinghouse. “Standards Handbook” (version 4.0.) Washington, DC: U.S. Department of Education, 2017. Available at https://ies.ed.gov/ncee/wwc/Handbooks. Accessed July 17, 2018.

World Bank. “An innovative approach: Transforming the preschool education quality model in Nicaragua.” Washington, DC: World Bank, August 18, 2017. Available at: https://www.worldbank.org/en/results/2017/08/18/an-innovative-approach-transforming-the-preschool-education-quality-model-in-nicaragua. Accessed November 4, 2020.

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

APPENDIX A:

EPC IMPLEMENTATION ROLLOUT

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

A. Random assignment and EpC rollout in Cohorts 1 and 2

Cohorts had different community sizes, different numbers of EpC, and different numbers of children by experimental group. In summary, during random assignment,

- Cohort 1A had 10 EpC and comprised nine large communities with a total of 586 children, 290 of whom were assigned to the treatment group and 296 to the control group.
- Cohort 1B had 65 EpC and comprised 37 large communities and 8 small communities with a total of 3,520 children, 1,850 of whom were assigned to the treatment group and 1,670 to the control group.
- Cohort 2A had 72 EpC and comprised 7 large communities and 121 small communities with a total of 4,154 children, 2,103 in the treatment group and 2,051 in the control group.
- Cohort 2B had 18 EpC and consisted of 37 small communities, with a total of 963 children, 438 in the treatment group and 525 in the control group.

Table A.1 provides additional information about each cohort. Table A.2 summarizes the time of exposure at the time of base-year data collection, by community size.

Table A.1. Implementation timeline of evaluation cohorts

Cohort	Educational community size	Number of EpC	Municipalities	Implementation organization	Date of notification of lottery results	Start date of EpC activities	End date of EpC	Base-year data collection	Months of exposure at 2015 base-year household survey	Follow-up data collection
1A	9 large	10	Bluefields, Kukra Hill, Corn Island	FHR, FZT, URACCAN	March 2014	May 2014	November 2015	n/a	n/a	May–June 2016
1B	23 large, 7 small	42	Bluefields, Kukra Hill, Desembocadura, Laguna de Perlas	FHR, FZT, URACCAN	October/ November 2014	November 2014	June 2016	n/a	n/a	May–June 2016
1B	14 large, 1 small	23	Bluefields, Corn Island	FZT	October/ November 2014	November, 2014	November 2015	n/a	n/a	May–June 2016
2A	2 large, 27 small	16	Kukra Hill	FHR	March 2015	April 2015	November 2016	July–September 2015	5	Mar–June 2017
2A	3 large, 37 small	25	Bluefields	FZT	June/July 2015	June 2015	January 2017	July–September 2015	2	Mar–June 2017
2A	2 large, 57 small	31	Bluefields, Laguna de Perlas	URACCAN	June/July 2015	August 2015	January 2017	July–September 2015	0	Mar–June 2017
2B	26 small	13	Bluefields	FZT	August/ September 2015	September 2015	April 2017	November, 2015	2	Mar–June 2017
2B	11 small	5	Bluefields, Laguna de Perlas	URACCAN	November 2015	February 2016	May 2017	November 2015	0	Mar–June 2017

Notes: FHR is the Fundación Hermanamiento Rama, URACCAN is Universidad de las Regiones Autónomas de la Costa Caribe Nicaragüense, and FZT is the Fundación Zamora Terán.

Table A.2. Number of communities, educators and children by months of exposure at 2015 base year (cohort 2)

	Small communities				Large communities			
	Treatment		Control		Treatment		Control	
	Number	Percentage of total	Number	Percentage of total	Number	Percentage of total	Number	Percentage of total
Communities								
	Months of exposure at 2015 base-year household survey							
Zero months	34	49%	34	49%	2	33%		
Two months	22	31%	22	32%	2	33%		
Five months	14	20%	13	19%	2	33%		
Total number of communities	**70**		**69**		**6**			
Educators								
	Months of exposure at 2015 base-year household survey							
Zero months	72	49%	46	53%	4	22%		
Two months	50	34%	25	29%	8	44%		
Five months	25	17%	15	17%	6	33%		
Total number of educators	**147**		**86**		**18**			
Children								
	Months of exposure at 2015 base-year household survey							
Zero months	271	47%	276	50%	18	25%	15	22%
Two months	185	32%	175	32%	39	53%	38	55%
Five months	116	20%	98	18%	16	22%	16	23%
Total number of children	**572**		**549**		**73**		**69**	

Sources: Follow-up educator and household surveys. Administrative information: implementation timeline

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

APPENDIX B:

EVALUATION SAMPLE

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

A. Sampling procedures

In this section, we discuss our approach to drawing the sample for the evaluation. All the educational communities that CARS had identified as eligible to receive the program early in the implementation process (Cohorts 1 and 2) were included in the evaluation sample; therefore, we do not describe sampling procedures at the community level.

Sampling of children. The evaluation team selected from the evaluation population of treatment and control groups a random sample of children to visit for the evaluation's primary data collection. For each of the large communities, we drew a random sample of 20 children per EpC intervention, 10 from the treatment group and 10 from the control. From each small community, we sampled 10 children. We oversampled out-of-school children to ensure adequate representation of this group, which is smaller than the school-going population but of particular interest to policymakers and stakeholders. In each community, we sampled four children who were not enrolled in school at the time of recruitment where possible. Once we had sampled four out-of-school children, we drew the same number of boys and girls (that is, three boys and three girls) among those currently enrolled in school to reach a sample size of 10. In communities with fewer than four out-of-school children and where the number of out-of-school children was not even (that is, one or three children), we randomly selected the next child without regard to gender, ensuring that the overall ratio was random. A total sample of 3,050 children from 219 communities were selected to participate in the evaluation (1,282 in Cohort 1, and 1,768 in Cohort 2).

Replacing certain children. Children were eligible to participate in EpC if they were 6 to 16 years old at intake, or were 5 years old but were in first grade or out of school at intake. The evaluation team replaced children who were selected for the main evaluation sample when the children could not be found (that is, they were unknown to key informants in the community), were found to be ineligible to participate in EpC, or could not be assessed due to physical, behavioral, or learning disabilities reported by the primary caregiver at the time of data collection. A list of potential replacements from the same learning community and gender group was used to replace the ineligible or missing children. Potential replacements were selected at the same time as the main evaluation sample was drawn. The replacement rate for the follow-up sample was 5 percent (n =110).

Sampling educators. To select teachers and EpC facilitators to participate in the evaluation, the evaluation team used information from the school register about the number and characteristics of educators in the school and educational community. In schools with three teachers or fewer, all teachers were invited to participate. In schools with more than three teachers, the team selected three teachers if the community had one EpC or two teachers if the community had two EpC.

In schools with more than three educators, the team prioritized interviewing teachers who taught students in the target grades (that is, second through fourth grades) and educators who were EpC facilitators the year after randomization (2015 for Cohort 1 and 2016 for Cohort 2). If more educators met the criteria than the number required to complete the sample, the team selected a random sample of teachers and facilitators. If, on the contrary, not enough educators met the criteria, the team selected a purposeful sample following criteria developed by

Mathematica. For example, if a teacher who taught the target grades in the year after randomization was not available, the team tried to interview a teacher who taught other grades in the same year.

Table B.1. Educator sample

	Teachers only		Teacher-facilitators		Facilitators only		Teacher-facilitators and facilitators only		Total
	Number	Percentage (%)	Number	Percentage (%)	Number	Percentage (%)	Number	Percentage (%)	Number
Number of educators									
All communities	270	62.9	66	15.4	93	21.7	159	37.1	429
Small communities	160	63.5	44	17.5	48	19.1	92	36.5	252
Small communities (treatment only)	71	45.2	38	24.2	48	30.6	86	54.8	157
Large communities	110	62.2	22	12.4	45	25.4	67	37.9	177
Number of communities									
All communities	173	86.9	63	31.7	78	39.2	128	64.3	199
Small communities	123	83.7	42	28.6	41	27.9	76	51.7	147
Small communities (treatment only)	55	74.3	36	48.7	41	55.4	70	94.6	74
Large communities	50	96.2	21	40.4	37	71.2	52	100.0	52

Source: Follow-up school director surveys.

B. Attrition

Children were grouped within educational communities in this evaluation's design. Therefore, bias in the impact estimates could be generated not only if communities dropped from the evaluation, but also if children were dropped from the evaluation.

Educational community attrition. At follow-up, the evaluation team was unable to collect data from 19 Cohort 2 communities due to security threats in the Punta Gorda area, which has been beset by social unrest related to plans to build a canal. In addition, school or educator data were not collected from an additional community in Cohort 2 in the same area.[33] Therefore, overall attrition (or loss) at the community level is 9 percent. Attrition rates are similar in the

[33] Data were collected from two out of 10 children sampled in that community.

treatment and control groups in small communities (12 percent in the treatment group and 11 percent in the control group; see Table B.2). Together, the overall rate of attrition (9 percent) and the difference in attrition between the groups (1 percent) are considered low, which means they are unlikely to lead to bias in the impact estimates that is greater than .05 standard deviations on the outcome (What Works Clearinghouse 2018).

Table B.2. Community sample size and attrition rates

	Total number of communities	EpC (A)	Control (B)	Difference (A-B)
All communities				
Recruited to participate	219	NA	NA	NA
Completed end year data collection	199	NA	NA	NA
Attrition rate (percent)	9	NA	NA	NA
Small communities				
Recruited to participate	166	84	82	2
Completed end year data collection	147	74	73	1
Attrition rate (percent)	11	12	11	1
Large communities				
Recruited to participate	53	NA	NA	NA
Completed end year data collection	52	NA	NA	NA
Attrition rate (percent)	2	NA	NA	NA

Source: Intake and Follow-up 1 and 2 data.

Notes: Large communities are those that have more than 50 eligible children (approximately); communities with 20 to 49 eligible children are considered small in size. The attrition rate was estimated as the difference between the number of communities that were recruited and those that were assessed at follow-up, divided by the number communities recruited to participate.

NA = not applicable.

Attrition of evaluation children. Table B.3 shows the number of children sampled to participate in the evaluation[34] (n = 2,820) and the number of children assessed at follow-up (n = 2,371). The attrition rate for the follow-up child survey and assessment was 16 percent (14 percent for Cohort 1 and 18 percent for Cohort 2) and did not differ between the treatment and control groups (16 percent in both groups). Together, overall attrition (16 percent) and differential attrition (0) in this evaluation represent a tolerable threat of bias under both optimistic and cautious assumptions (What Works Clearinghouse 2018).

[34] Excludes 230 children from the 20 communities that could not be visited at follow-up.

Table B.3. Sample size and attrition rates: children

Number of children	Selected to be sampled	EpC	Control	Cohort 1	Cohort 2
All communities					
Evaluation sample	2820	1539	1281	1282	1538
Completed interviews	2371	1290	1081	1108	1263
Attrition rate (percent)	15.9	16.2	15.6	13.6	17.9
Small communities					
Evaluation sample	1455	738	717	77	1378
Completed interviews	1182	598	584	61	1121
Attrition rate (percent)	18.8	19.0	18.5	20.8	18.7
Large communities					
Evaluation sample	1365	801	564	1205	160
Completed interviews	1189	692	497	1047	142
Attrition rate (percent)	12.9	13.6	11.9	13.1	11.3

Source: Follow-up household survey.

Notes: Children in the respondent sample are those who completed the interview at follow-up. The attrition rate is estimated as the difference between the number of children who were sampled and those who were assessed at follow-up, divided by the number children sampled. Following guidelines by What Works Clearinghouse 2018, attrition at the child level is based on the communities remaining in the sample. That is, the denominator only includes sampled children in the communities where data were collected at follow-up.

APPENDIX C:

ANALYTICAL APPROACH

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

A. Empirical specification

We estimate impacts using regression analysis.

1. Main impacts

A regression model controls for chance differences between treatment groups in children's base-year characteristics. These characteristics include enrollment status and age in the base year, as well as other variables. Including covariates in our regression model allows us to increase the statistical power of the evaluation. We also include dichotomous variables for the strata used in random assignment as well child-level time invariant covariates. For each outcome we analyze, the model can be expressed as follows:

$$Y_{ict} = a + \beta x_{ic0} + \lambda T_{ic} + \delta C_c + \eta_c + \varepsilon_{ict} \qquad [1]$$

where Y_{ict} is the outcome measure (such as reading comprehension) for child *i* in educational community *c* at time t. The vector x_{ic0} represents the base-year (t=0) child-level characteristics, including age, gender, and enrollment status at intake. The variable T is the treatment indicator for child *i* in educational community *c*; it equals "1" if the child was assigned to the treatment group and "0" if he or she was assigned to the control group. The parameter λ indicates the impact of the EpC intervention. The variable C equals "1" if the community is in Cohort 2 and "0" if it is in Cohort 1. The term η_c is an educational community-level-specific error term (a group or cluster effect), and ε_{ict} is a random error term for child *i* in educational community *c* at time *t*. We use sampling weights to adjust for the oversampling of out-of-school children. Our main analysis estimates impacts for small communities but we also run robustness checks with the full sample or small and large communities combined, and separately with large communities.

The intervention effect estimates represent what is known as an "intent-to-treat" effect. This effect estimates the average difference in outcomes between all eligible children in intervention communities, regardless of whether a child participated in or completed all program activities. Likewise, the effect quantifies the average difference in outcomes regardless of whether the community implemented the intervention as intended or whether it was assigned to the intervention group but did not follow through in implementing the intervention. Such an estimate can be interpreted as the average impact on the population of children in communities that had the option of implementing and participating in the intervention. The intent-to-treat approach preserves the main advantage of random assignment, which is to guarantee that there are no systematic differences between treatment groups in terms of what we can measure and the myriad factors that we cannot measure. If we were to focus only on children who remained in EpC during the evaluation period, the intervention effects would be confounded with any factors that determine a child's persistence in EpC. For example, suppose that a child's assignment to EpC predicts participation in EpC, but children who participate in EpC also happen to be among those underperforming in reading. If we compared only children in EpC at follow-up, the average reading scores for children in the control group would appear artificially lower than the average scores for children in EpC. By measuring outcomes for all children, including those who did not participate or did not advance in grade level with their cohort, and by following an intent-to-treat approach, we are able to obtain an unbiased estimate of the impact of the offer of the intervention, regardless of whether a child actually received the intervention during the full evaluation period.

We also conduct sensitivity analyses to include pre-test reading scores for some children in the sample.[35] Because pre-test individual scores are often highly correlated with post-test individual scores, including these pre-test scores in the design allows the evaluation team to obtain more precise post-test impact estimates than would have been the case had we implemented a post-test random assignment design only. When pre-test scores are available, a pre-test score will be added as a covariate in equation 1 (where these scores can be considered part of vector x_{ic0} in the model). Despite the gains in precision, this specification can potentially generate bias in the post-test impact estimates because the pre-test measures were obtained after children were aware of the treatment group to which they have been assigned, or had actually received the services offered by EpC. In other words, the potential for bias arises because pre-test scores could contain the early effects of EpC, so the differences between pre- and post-test scores in treatment group children would be smaller than they would be, had early exposure not occurred. The higher precision comes at the expense of potential bias in the estimates of the intervention effect (Schochet 2010). We administered pre-tests as early as possible to diminish the potential for late pre-test biases for Cohort 2, but the base-year data collection for Cohort 2 was conducted after random assignment, and reading assessment scores are thus "late pre-tests" instead of "pre-tests."

2. Subgroups

Exploring the variation in impacts by subgroups is of interest to the evaluation. Key subgroups include those defined by gender and by school enrollment status. We evaluate the impacts for a particular subgroup by restricting the sample used to estimate equation (1) or by including appropriate interaction terms in the equation. We also explore variation in impacts by other subgroups of interest, such as age or grade at intake.

3. Robustness checks

We estimate the robustness of our main analyses using different empirical specifications, including:

- No pre-test or covariate adjustments
- No weights accounting for oversampling of out-of-school children
- Alternative outcome measures (for example, we asked additional questions to measure self-esteem and intercultural competence in the 2017 follow-up. Therefore, we run the analysis with items that are common across both follow-ups (2016 and 2017), but also with those that were only collected at one point in time).

4. Educational community and educator analyses in small communities

We use similar models for analyses conducted at the educational community and educator levels. However, these regressions apply to small communities only (large communities are not

[35] We were unable to collect pre-test scores for Cohorts 1A or 1B, for example, as shown in Figure II.2.

included) because random assignment at the community level was undertaken only for small communities. Because random assignment was done at the individual level for large communities, large communities are in both the treatment and control groups.

For each outcome we analyze at the community level, the model can be expressed as follows:

$$W_{ct} = \alpha + \gamma U_c + \lambda T_c + \gamma C_c + u_{c0} \qquad [2]$$

where W_{ct} is the outcome of interest for educational community c at time t, the variable U indicates whether the community is urban or rural, and T_c is an indicator equal to "1" for communities assigned to EpC and "0" for those assigned to the control group. The parameter λ indicates the impact of the EpC intervention. The variable C_c is an indicator equal to "0" for communities in Cohort 1 and "1" for communities in Cohort 2. u_{c0} is an educational community-level-specific error term. Regressions are unweighted and standard errors are not clustered. We conduct sensitivity analysis including children's pre-test scores aggregated at the educational community level.

For outcomes at the educator-level, the regression model we use to test for differences between EpC facilitators and teachers in the control group can be expressed as follows:

$$Y_{ict} = \alpha + \delta C_c + \lambda T_{ic} + \eta_c + \varepsilon_{ict} \qquad [3]$$

The parameters are the same as in equation [1] but regressions are unweighted. To estimate impacts on educators trained by CARS, our main analysis compares EpC facilitators (both teachers who are facilitators and facilitators only) in the treatment group to teachers in the control group. We also conduct sensitivity analyses comparing teacher-facilitators in the treatment group to teachers in the control group, excluding facilitators who are not teachers. Teachers in the two groups did not statistically significantly differ in a set of background characteristics that could impact their instructional practices, specifically years of teaching experience and qualifications, whereas facilitators who are not teachers significantly differed from teachers in those characteristics.

In small communities, educators who are not trained in EpC techniques are in the same community as educators who are trained. Therefore, untrained educators are likely to be exposed to the effects of EpC without participating in EpC training, through their interactions with other educators. We test for the potential spillover among educators within communities, by examining the effect on untrained educators in EpC and control communities. We use the same model as in equation [3], but exclude trained educators (teacher-facilitators and facilitators) from the analytic sample.

5. Educator analyses in large communities

In large communities, we compare the same groups of educators as in small communities, but do not make comparisons across treatment groups. Thus, for our main analysis, we compare EpC facilitators (both teachers who are facilitators and facilitators only) to "teachers only". We use the same model as in equation [3], but exclude the λ treatment parameter and replace it with a variable that equals "1" if the educator is an EpC facilitator and "0" if the educator is a teacher only.

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

APPENDIX D:

BALANCE ACROSS EVALUATION GROUPS

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

Randomized controlled trials (RCTs) ensure that, on average, the treatment and control groups members' characteristics before an intervention are not correlated with their receipt of the intervention. Typically, after random assignment, treatment groups are similar in the distribution of observed and unobserved characteristics, but there may be some differences between the two groups that are due to chance. We assessed differences between EpC and control groups in characteristics and outcomes that we measured at intake and in the base year, as well as characteristics that are unlikely to change as a result of EpC at follow-up, to verify that the randomization resulted in balanced groups.

In this appendix, we present the assessment of balance between treatment and control groups for a set of child characteristics, in addition to those presented in chapter IV. We also present our assessment of balance on school and educator characteristics. We present this assessment for small communities only where there is a true control group to compare to the treatment group. In large communities, where random assignment occurred at the child level, there is no analogous control group of schools or educators.

A. Balance on child characteristics in all communities

As discussed in Chapter IV, most observed children's characteristics were balanced across EpC and control groups in the full sample and in small and large communities. Using survey data for the respondent sample (base-year data for Cohort 2 and follow-up data for Cohort 1), the groups were balanced for 23 out of 27 characteristics that are unlikely to have changed as a result of the EpC program (Table D.1).

In the full sample, we found that children in the EpC group were slightly less likely to speak Kriol and attend school in the evening than children in the control group (1.9 and 0.2 percentage points difference, respectively), but these differences were significant only at the 10 percent level. Children in EpC were also more likely to use a boat to get to school (a difference of 1.2 percentage points), while children in the control group were more likely to use a horse or mule (a difference of 2.5 percentage points). Results in small and large communities largely reflect those in the full sample, with a few exceptions. In small communities, we found no differences between the two groups in the percentage of children who speak Kriol or attend the evening shift, but found that children in the EpC group were 5.4 percentage points more likely to walk to school, a difference significant only at the 10 percent level. Children in the control group were also slightly more likely to live in households with a car or truck (a difference of 1.2 percentage points). In large communities, the groups were balanced on 26 of 27 characteristics. We found that children in the EpC group were less likely to speak Kriol (4.3 percentage points). Finding one difference is what is expected to arise due to chance. There was also a difference of 3.6 percentage points in use of a motorcycle to get to school, which was significant at a 10 percent level.

Table D.1. Other child characteristics, by treatment group and community size

	All communities			Small communities			Large communities		
Child characteristics (percentages)	EpC (A)	Control (B)	Difference (A–B)	EpC (C)	Control (D)	Difference (C–D)	EpC (E)	Control (F)	Difference (E–F)
Maternal language									
Spanish	80.3	80.2	0.1	99.0	99.4	-0.4	61.8	61.0	0.8
English	2.8	2.5	0.3	0.0	0.4	-0.4	5.4	4.5	1.1
Miskitu	6.6	5.8	0.9	0.0	0.0	0.0	13.2	11.4	2.1
Ulwa	1.2	1.0	0.2	0.0	0.0	0.0	2.3	2.0	0.4
Kriol	8.5	10.4	-1.9*	0.6	0.3	0.3	16.4	20.8	-4.3*
Rama-Kriol	0.6	0.2	0.4	0.6	0.1	0.5	0.6	0.4	0.2
Maternal language is language of instruction	87.4	87.0	0.4	99.0	99.4	-0.4	75.9	74.4	1.5
Parents live in the same home as child	59.3	60.4	-1.1	73.2	77.1	-3.9	47.6	45.9	1.7
Transportation to school									
Walks to school	79.0	76.3	2.7	78.7	73.3	5.4*	79.5	79.1	0.4
Rides bus/car to school	5.3	5.7	-0.4	0.0	0.3	-0.3	9.8	10.5	-0.7
Takes motorcycle to school	1.7	1.9	-0.2	0.0	0.2	-0.2	3.1	3.4	-0.3
Takes boat to school	1.5	0.3	1.2**	3.2	0.8	2.4**	0.0	0.0	0.0
Rides animal to school	2.0	4.5	-2.5**	3.5	8.9	-5.4**	0.6	0.4	0.1
Rides a bicycle	1.3	0.8	0.5	0.7	-0.1	0.8	1.7	1.5	0.2
Time required to get to school (minutes) (children enrolled in school only)	17.9	18.9	-1.0	25.7	26.2	-0.5	12.4	13.7	-1.3
Shift attended (children enrolled in school only)									
Morning	81.3	80.4	0.9	96.5	96.3	0.2	68.4	67.2	1.2

	All communities			Small communities			Large communities		
Child characteristics (percentages)	EpC (A)	Control (B)	Difference (A–B)	EpC (C)	Control (D)	Difference (C–D)	EpC (E)	Control (F)	Difference (E–F)
Afternoon	18.4	19.0	-0.6	3.2	3.0	0.2	31.4	32.2	-0.8
Evening	0.1	0.3	-0.2*	0.0	0.3	-0.3	0.2	0.4	-0.3
Saturday (distance education)	0.1	0.3	-0.2	0.3	0.4	-0.2	0.0	0.2	-0.1
Child reports that school provides a meal	67.6	67.7	0.0	70.8	69.8	1.1	43.6	51.6	-8.0
Family poverty	70.1	72.3	-2.2	94.7	94.8	-0.1	50.5	54.3	-3.8
Household asset ownership									
Bicycle	17.3	16.9	0.5	3.3	4.3	-1.0	29.5	27.2	2.3
Motorcycle	5.6	4.5	1.1	1.5	2.7	-1.2	9.2	5.6	3.6*
Car or truck	3.3	3.2	0.1	0.1	1.3	-1.2**	5.8	4.8	1.1
Boat	12.4	12.7	-0.3	6.4	3.5	3.1	17.6	21.1	-3.5
Household migration (Cohort 2)									
Adult	23.3	24.2	-0.9	21.5	20.4	1.1	35.5	54.9	-19.4
Child	2.5	3.0	-0.5	2.5	3.0	-0.5	2.5	3.0	-0.6
Number of children	**1,290**	**1,081**		**598**	**584**		**685**	**486**	
Number of communities	**126**	**125**		**74**	**73**		**52**	**52**	

Source: Base-year and follow-up household survey. Intake data was used when base-year and follow-up language were missing.

Notes: Columns A and B, C and D, and E and F present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Differences in sample size come from missing information for some variables. The transportation to school variables are 0 for children who are not enrolled in school, and do not sum to 100. Household migration was collected from the Cohort 2 sample only. We define the family poverty indicator as equal to one if two or more of the following are met: overcrowding, unsound housing, and economic dependence. Time required to get to school and the shift attended (the only variable using follow-up data) include children enrolled in school only. Other samples sizes vary due to missing data.

*Difference in group means is statistically significant at the .10 level.

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

B. Balance on school characteristics in small communities

Of the 26 comparisons of school characteristics made between schools in EpC and control groups, only one difference was significant at the 5 percent level, which is within the range that would be expected to arise due to chance. EpC and control communities were similar in most basic school characteristics (see Table D.1). Schools in the sample are public, have multi-grade classrooms, and have approximately 28 to 30 students per teacher. Schools were open 14 to 16 days per month, and school days were cancelled 2 to 3 days per month in March and April. In addition, few schools offer bilingual instruction. Schools in both groups lacked some basic infrastructure. About one-fifth of classrooms could not be used in extreme weather conditions (such as heavy rain) and more than 16 percent lacked furniture sufficient to accommodate all students. In addition, most schools had toilets that were inadequate in number or function. Schools assigned to the EpC group were 16 percentage points more likely to have classrooms made of rudimentary materials than schools in the control group.

More than 40 percent of schools in both groups received support from programs other than CARS. Of those schools, about one-third received textbooks or other teaching materials and 36 to 47 percent of schools received other kinds of support, such as tutoring, teacher training, scholarships, and assistance for improving infrastructure. Most schools had a feeding program in the year of data collection but most offered snacks rather than breakfast or lunch. The majority of teachers in both groups had more than three years of teaching experience, however control group schools have teachers with somewhat less experience overall than treatment group schools (the difference is not significant at a 5 percent level but is significant at a 10 percent level).

Table D.1. School characteristics in small communities

School characteristics	EpC (A)	Control (B)	Difference (A–B)	*p*-value
School is public (percentage)	93.4	88.9	4.5	0.302
School has multi-grade classrooms only (percentage)	98.7	97.2	1.4	0.340
Student-teacher ratio	27.8	29.8	-2.0	0.218
Average number of days per month the school was open two years after randomization (year of data collection)[a]	14.3	13.6	0.7	0.208
Average number of days per month the school was open one year after randomization (the year before data collection)	16.4	16.5	-0.1	0.875
Number of cancelled school days in April, two years after randomization (year of data collection)	2.2	2.6	-0.4	0.538
Number of cancelled school days in March two years after randomization (year of data collection) (Cohort 2 only)	3.1	3.3	-0.2	0.668
School is bilingual (percentage)	1.4	0.0	1.4	0.217
School infrastructure				
Classrooms made of rudimentary materials (percentage)	76.0	60.1	15.9**	0.017
Classrooms without blackboard or flipchart (percentage)	4.9	1.7	3.2	0.242
Classrooms that cannot be used under extreme weather conditions (percentage)	26.0	24.8	1.2	0.869

School characteristics	EpC (A)	Control (B)	Difference (A–B)	*p*-value
Classrooms without enough desks/tables/chairs for students (percentage)	16.7	21.1	-4.3	0.465
There are no functioning toilets (percentage)	10.2	14.3	-4.0	0.440
Toilets work but do not have locks/doors (percentage)	36.5	27.3	9.2	0.241
Toilets work and have doors/locks but are not segregated by gender (percentage)	27.8	34.8	-7.0	0.375
Has an active external program two years after randomization (in the year of data collection) (percentages)	49.4	43.1	6.2	0.431
External program offers textbooks/materials	33.3	33.4	-0.1	0.987
External program offers other, non-EpC, support or resources	47.3	35.6	11.7	0.138
Has a meal program two years after randomization (in year of data collection) (percentages)	97.6	94.2	3.3	0.337
Meal program offers breakfast	5.6	1.2	4.4	0.181
Meal program offers a snack	93.8	89.8	4.0	0.389
Meal program offers lunch	2.4	3.1	-0.7	0.803
Average teacher experience (years)				
Fewer than 3 years of experience	23.9	29.2	-5.2	0.447
3–5 years of experience	24.1	36.9	-12.8*	0.073
6–10 years of experience	26.9	20.4	6.6	0.327
More than 10 years of experience	25.1	13.6	11.5*	0.052
Number of communities	**74**	**73**		

Source: Follow-up school director survey; CARS M&E data.

Note: Average number of days per month the school was open in the year of data collection is calculated using data from February to April for Cohort 1 and February to June for Cohort 2. Average number of days per month the school was open in the year before data collection is calculated using data from February to November. Sample sizes vary due to missing administrative data in schools (for example, a number of school directors did not have records of number of days the school was open).

[a] Not a full year, but number of months at the time of data collection.

B. Balance on educator characteristics in small communities

Making treatment-control comparisons is not straightforward when looking at educators because educators were not randomly selected to become EpC facilitators and there could be differences, not attributed to CARS, between educators who became EpC facilitators and those who did not. Also, this group includes regular classroom teachers in control schools, but in treatment schools it could include regular classroom teachers as well as EpC facilitators who were not regular classroom teachers. In many cases, the classroom teacher was also the EpC facilitator.

We sampled three educators per school, which included teachers and EpC facilitators for schools in treatment communities, and included only teachers in control schools (see Appendix B for details). Across treatment and control groups, we interviewed 92 EpC facilitators, of whom 86 were in EpC communities, and 6 were in control communities at follow-up (Table D.2). Fifty-two percent of facilitators were only facilitators (and not teacher-facilitators). Note that some teachers could have received training in one community and moved to another. The group of facilitators includes educators who were not facilitators at the time of data collection, but who had been facilitators at some point over the course of the evaluation. We interviewed 160 educators who were only regular classroom teachers, of whom 71 were in the EpC group and 89 in the control group.

Our main educator analyses compare teacher-facilitators and facilitators in the EpC group (n = 86) to teachers only in the control group (n = 89). Therefore, we focus on comparing these two groups of educators, but make additional comparisons between teachers across the EpC and control group, excluding educators who were only facilitators, to help with interpretation of the results.

Table D.2. Educator sample for small communities, by treatment group

	EpC		Control		
Educator type	Number	Percent (%)	Number	Percent (%)	Sample size
EpC facilitators (teacher-facilitators and facilitators only)	86	93.5	6	6.5	92
Teachers only	71	44.4	89	55.6	160
Number of communities	**74**		**73**		
Number of communities with EpC facilitators	**70**		**6**		
Number of communities with teachers-only	**55**		**68**		

Source: Follow-up educator survey.

Note: Estimates presented for small communities only. We were unable to interview EpC facilitators in four EpC communities and teachers who had never been a facilitator in five control communities at follow-up. In small communities, 52 percent of facilitators are not teachers.

EpC facilitators in EpC communities were similar in many characteristics to educators who were only teachers in the control group; however, they had several important differences. The EpC and control groups were balanced in terms of educators' gender, maternal language, time living in the community, and the proportion of highly educated educators (three- to five-year university degrees) (Table D.3). Teachers in the control group had more years of teaching experience and higher teaching qualifications (they were more likely to have a university degree in an area related to education; unlike empirical teachers or teachers certified to teach elementary school only) than EpC facilitators (who may or may not have also been teachers). This likely due to the recruitment process of EpC facilitators, where some municipalities trained non-teachers to be facilitators (it was not always feasible to train teachers as facilitators). On the other hand, we find that teacher-facilitators and teachers-only have similar teaching experience and level of education on average. In other words, those teachers that were chosen to be trained as facilitators have similar levels of experience and education as teachers that did not receive the training, but EpC facilitators who were not teachers differ from teachers in the control group (see Table D.4).

Table D.3. Characteristics of educators in small communities, by treatment group, including all EpC facilitators

Educator characteristics	Teacher-facilitators and facilitators in EpC (A)	Teachers-only in control (B)	Difference (A-B)	*p*-value
Female (percentage)	54.7	57.3	-2.6	0.739
Maternal language is Spanish (percentage)	97.6	98.9	-1.3	0.217
Time living in the community (years)	9.0	7.0	2.0	0.117
Teacher experience in any school (in years)	3.7	5.8	-2.1***	0.005
Teacher experience in current school (in years)	2.5	3.6	-1.1*	0.065
Teaching qualifications (percentages)				
Empirical teacher (high school degree or less, or post-secondary education in a field unrelated to education)	55.6	31.6	24.0***	0.001
Qualified to teach elementary school or preschool, as certified by a “Normal School”	37.7	56.8	-19.1**	0.018
Qualified to teach secondary education (3-year university degree in education or related field)	1.2	2.2	-1.0	0.578
Education graduate (5-year university degree in education or related field)	5.5	9.3	-3.9	0.397
Number of educators	**86**	**89**		
Number of communities	**70**	**68**		

Source: Follow-up educator survey.

Note: Estimates presented for small communities only. Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level. Educators encompasses teachers, facilitators, and educators who are both teachers and facilitators. Teachers only refers to educators who are not facilitators. Sample sizes vary because there are 4 EpC communities with no EpC facilitators (teacher-facilitators or facilitators) and 5 control communities with no teachers-only. Normal Schools (or Escuelas Normales, in Spanish) are institutes in charge of training primary education teachers in Nicaragua (Reglamento de las Escuelas Normales de Nicaragua. Decreto No. 8 de 1968. Retrieved from http://legislacion.asamblea.gob.ni on June 11, 2018.)

*Difference in group means is statistically significant at the .10 level.

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

***Difference in group means is statistically significant at the .01 level.

Table D.4. Characteristics of educators by treatment group, excluding facilitators who were not teachers: small communities

Educator characteristics	Teacher-facilitators in EpC (A)	Teachers-only in control (B)	Difference (A-B)	*p*-value
Female (percentage)	48.1	59.3	-11.2	0.364
Maternal language is Spanish (percentage)	98.4	98.4	0.0	NA
Time living in the community (years)	9.7	7.5	2.2	0.190
Teacher experience in any school (in years)	6.8	5.8	1.0	0.400
Teacher experience in current school (in years)	4.9	3.7	1.2	0.237
Teaching qualifications (percentages)				
Empirical teacher (high school degree or less, or post-secondary education in a field unrelated to education)	37.0	31.4	5.6	0.597
Qualified to teach elementary school or preschool, as certified by a “Normal School”	49.2	57.7	-8.5	0.442
Qualified to teach secondary education (3-year university degree in education or related field)	1.8	2.6	-0.8	0.819
Education graduate (5-year university degree in education or related field)	12.1	8.3	3.8	0.466
Number of educators	**38**	**89**		
Number of communities	**36**	**68**		

Source: Follow-up educator survey.

Note: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level. Educators encompasses regular classroom teachers and educators who are both teachers and facilitators. Normal Schools (or Escuelas Normales, in Spanish) are institutes in charge of training primary education teachers in Nicaragua (Reglamento de las Escuelas Normales de Nicaragua. Decreto No. 8 de 1968. Retrieved from http://legislacion.asamblea.gob.ni on June 11, 2018.)

NA = not applicable.

APPENDIX E:

DATA COLLECTION

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

A. Data sources

As described in Chapter IV, we used data from several sources in our analyses, including CARS administrative data collected during the intake process. The base year and follow-up survey instruments will be available online upon request at USAID's Development Data Library website: https://data.usaid.gov/. The data are restricted use housed by USAID.

Intake data. The CARS implementation team developed preselection forms that were used to identify children's and educational communities' eligibility to participate in the EpC intervention. These forms included data on children and the educational community—true baseline information that could be used to describe the evaluation's population as well as to verify eligibility. Data on children included their gender and age for both cohorts and the children's maternal language for Cohort 2. The CARS team and its local partners collected information about children's characteristics and key educational outcomes such as enrollment status; community characteristics such as number of eligible children and language of instruction in the community; and, in Cohort 2, some additional information on schools, such as quality measures of school infrastructure and number of teachers. In addition, community assemblies were conducted where feasible in order to communicate about the EpC intervention and obtain community members' informed consent to participate in random assignment and have the chance to receive the EpC intervention. When possible, we and CARS tracked the participation of parents of eligible children in these assemblies. When parents were not able to participate, school directors provided consent for the community and children to participate in random assignment.

Base year survey data. We collected base-year data through a household survey in Cohort 2 communities. Household visits consisted of a survey of primary caregivers, a short child survey, and a child literacy assessment. The goal of these visits was to collect data allowing us to characterize children's home environments and their literacy skills during the base year of program participation.

Follow-up survey data. We collected follow-up data from household and school visits in Cohorts 1 (in 2016) and 2 (in 2017). Similar to what was done in the base year, household visits consisted of a survey of primary caregivers, a child survey, and a child literacy assessment. School visits consisted of a survey of school directors, collecting information from school registers, observations of the school infrastructure and the physical environment of the classrooms, and surveys of educators (that is, teachers and EpC facilitators). We describe each of these sources of information below.

- **Household (primary caregiver) survey.** In the household survey, researchers asked primary caregivers about the family's composition, adults' education level and occupation, household assets (for example, number of bedrooms, income, livestock ownership), and recent migration patterns (in 2014 and 2015). Caregivers also reported on their children's educational background (for example, school enrollment, grade at the time of data collection, grade repetition, and attendance), as well as other factors that may influence school attendance and academic success (for example, means of transportation and distance to school, child labor, and whether someone reads with the child at home and how often). We also asked primary caregivers about the quality of their relationship with children as a way to measure children's relationship skills, their sense of safety in the community, and parenting practices such as monitoring, supervision, and discipline.

- **Child survey.** Researchers asked children about their maternal language, their reading habits at home, and the literacy environment in the household (for example, by asking children whether someone reads with them at home). The survey also included questions about children's socio-emotional skills and attitudes reflecting their self-esteem, intercultural competence, and academic engagement, and about security risk factors such as victimization by school peers (or bullying), impulsive risk taking, and moral disengagement. A description of each of these measures is included in Appendix G.

- **Child reading assessments.** We measured children's early literacy skills using an adaptation of the Early Grade Reading Assessment (Dubeck and Gove 2015) and the Dynamic Indicators of Reading Success (Baker et al. 2006). Spanish and local language experts adapted subtasks from these assessments to the Nicaraguan linguistic and cultural contexts. The adaptation aimed to develop tasks with comparable levels of difficulty in Spanish, English, and Miskitu, as indexed by word count, average word length, number of sentences, and average sentence length. We measured four skills: (1) pseudo-word (or invented word) decoding, (2) oral comprehension (for children who could not decode), (3) oral reading fluency, and (4) reading comprehension. We used two passages to assess reading comprehension. After the passages, children were asked factual and inferential questions to measure more advanced skills.

  We administered literacy assessments to all children in the sample in their homes, regardless of their current enrollment status in school. We administered the assessments in the official language of instruction of the school in each community. For Cohort 1, follow-up assessments were administered in Spanish (87 percent), English (7 percent), and Miskitu (6 percent). For Cohort 2, all assessments were administered in Spanish. Appendix F provides information about the literacy assessment development, scoring, and internal consistency.

- **School director survey.** Researchers asked school directors about general school characteristics (for example, whether the school is public, private, or subsidized; languages of instruction; number of school shifts; grades offered; number of sections per grade; number of teachers), school and community activities (for example, whether the school has an active parent association or a parent school (Escuela para Padres y Madres, in Spanish), and what community action plans are in place), school and community safety (for example, instances of bullying or children being physically or sexually abused), number of effective school days and reasons why the school was closed if it was, financial and instructional resources (for example, whether the school offers free meals or receives support from outside programs, whether students have textbooks for exclusive or shared use and in what languages, availability of materials for arts and sports), characteristics of the students and teachers (for example, students' means of transportation to school, teachers' years of experience, teacher absenteeism, participation in EpC), and teachers' participation in professional development activities (for example, training in reading methodologies or other topics).

- **School register.** School directors used administrative records to report on the number of administrators and teachers at the school, the grades they taught, and whether teachers and administrators were EpC facilitators. Researchers asked questions about the enrollment status of children in the evaluation sample, attendance on the day of data collection and in the prior week and month, and enrollment and participation in EpC. When school directors did not have information about children's EpC enrollment and participation, the evaluation team attempted to obtain the information from EpC facilitators. For Cohort 1, researchers also asked school directors about the number of students enrolled per grade and by gender in the year of data collection and the year before data collection. For Cohort 2, enrollment information was not collected from administrative records and was not disaggregated by grade and gender in order to reduce participant burden and the costs of data collection. Instead, school directors reported the overall number of students enrolled in the school as part of the school director survey.

- **Educator (teacher and EpC facilitator) survey.** Educators (who could be teachers-only, EpC facilitators-only, or both), answered questions about their background (for example, languages spoken, time living in the community, teaching qualifications, years of teaching experience, workload), participation in professional development activities to teach reading and other topics, perceptions about obstacles to teaching in the classroom and teaching in the EpC (for example, not enough time for instruction, lack of training, lack of support from parents), general instructional practices and practices specific to EpC (for example, reading out loud, having children work in pairs or small groups, using assessments for formative purposes, dividing children by age groups [peces y robles, in Spanish]), using the tutor-child component, singing the welcoming song, and the children's use of reading materials.
- **School observations.** After completing the school director survey, researchers asked the school director for authorization to conduct a brief observation of the school building. If the observation was not feasible, the researcher gathered the information from the school director or another informant familiar with the school. The observation included items about the number of classrooms in use, number of classrooms made of finished or rudimentary construction materials, number of classrooms that cannot be used under extreme climate conditions, availability of blackboards or flip charts, availability of enough desks and chairs for students, and the availability and condition of toilets.

  In addition, after completing the educator surveys, researchers conducted brief observations of educators' regular classrooms and EpC classrooms, when appropriate. The observation included items about the physical condition of the classroom, including infrastructure (for example, whether the classroom had a roof, floor, and walls; and what the room's temperature, lighting, and ventilation were), the adequacy of the classroom's space and furniture, and other factors that could affect children's learning or are necessary to follow the EpC model (for example, external noise, availability of a blackboard visible to all students, walls decorated with instructional materials, exemplars of children's work posted on the walls).

B. Data collector selection and training

This section focuses on training for follow-up (end-year) data collection. Details about personnel selection and training for base-year data collection are included in the baseline report (Bagby et al. 2017).

Mathematica collaborated with Fundación Internacional para el Desafío Económico Global (FIDEG) to conduct data collection activities. FIDEG, in turn, collaborated with the Centro de

Investigación y Acción Educativa Social (CIASES), a Nicaraguan education and reading firm. We conducted two rounds of end-line data collection: one between May and July of 2016 for Cohort 1, and another from March to June of 2017 for Cohort 2. Each round involved training and field practice (from April 25 to May 28 of 2016 for Cohort 1, and from February 20 to March 15 of 2017 for Cohort 2).

In both rounds, FIDEG's technical team and Mathematica staff trained enumerators and supervisors on the data collection instruments and procedures. Training consisted of reading the training manual aloud; practicing administration of the instruments in pairs, small groups, and on the board; practicing with the chronometer; and administering the reading assessments to children in communities near Managua that were not part of the evaluation sample. All enumerators completed a written examination and participated in field testing. During the field testing, FIDEG's technical team used a rubric to detect administration errors and assess rapport and pace of administration. The rubric provided a standard metric to score enumerators and inform personnel selection. Inter-rater reliability was assessed during classroom practice or as part of field testing. Written examinations, field tests, and inter-rater reliability scores were used to select the most qualified personnel and to provide actionable feedback before actual data collection.

There were a few differences between the first and second rounds of follow-up data collection. In the first round, enumerators practiced administering all instruments (that is, household, child, educator, and school director surveys) in schools and households near Managua. In the second round, to reduce costs (and because 51 percent of enumerators had participated in the first round), enumerators practiced administering the educator and school director surveys in mock-up sessions in which FIDEG's technical team reenacted scenarios from the first round. Also, the first round of training included training for bilingual personnel because the official language of instruction in some Cohort 1 communities was English/Kriol or Miskitu. A reinforcement session was held for bilingual enumerators in round one to develop their familiarity with the bilingual instruments because the general training was conducted in Spanish. Field testing for English/Kriol and Miskitu speakers took place in Laguna de Perlas and Bilwi, respectively. In the second round, reinforcement was unnecessary because the official language of instruction in all Cohort 2 communities was Spanish. Field testing for Spanish speakers took place in Managua or Leon.

During the field work, FIDEG organized personnel into five to seven teams. Each team consisted of one supervisor, one reviewer, one expert in the reading assessments, and three or four enumerators. Supervisors were staff who demonstrated leadership, organizational, and management skills; had supervisory experience; and were fluent in English/Kriol or Miskitu, in the case of bilingual teams.

C. Quality assurance visits during data collection

The field coordinator and team supervisors were responsible for ensuring compliance with data collection protocols and planned logistics. They monitored adherence to the schedule and verified and delivered completed hard-copy instruments to FIDEG's office staff. Reviewers detected omissions or inconsistent information and, in coordination with the supervisor, instructed enumerators to go back to the school or home to recover or verify the information.

Supervisors conducted observations of every enumerator on their team at least once a week for quality assurance. If an enumerator did not follow the protocols, the supervisor provided feedback and prioritized observing that enumerator until errors were corrected.

D. Data quality assurance

Before data entry, FIDEG's office editor checked and organized the questionnaires and performed an additional manual check to detect inaccuracies (such as incorrect unique identifiers or inconsistent information). Data were then entered using CSPro 6.1. In the base year and in the first round of follow-up, double data entry was performed on 100 percent of the survey responses. To reduce costs, double data entry was performed in the second round of follow-up for a subset of sections from each instrument (that is, three sections from the household and child surveys, including the full reading assessment; four sections in the school director survey; four sections in the educator survey, and all sections in the school register). Inconsistencies were corrected by going back to the paper-and-pencil records. Once entry errors had been corrected, a secondary editing program (editing.bch) available in CSPro and customized for the evaluation, was used to perform additional consistency checks.

Once we received the raw data from FIDEG, Mathematica conducted data consistency checks and reviewed frequencies by variable. To verify the quality of the data sets, we checked that all the fields in the instrument were represented as variables in the data set and that the values in the hard-copy instruments matched the values in the data files. We consulted with FIDEG to clarify any inconsistencies.

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

APPENDIX F:

LITERACY OUTCOMES MEASUREMENT AND ASSESSMENT CHARACTERISTICS

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

A. Assessment development and administration

To measure children's literacy skills, we created an assessment based on the Early Grade Reading Assessment (EGRA) and the Dynamic Indicators of Reading Success (IDEL in Spanish). To mitigate floor and ceiling effects that can arise from sample heterogeneity in age, grade, and exposure to schooling, we used a range of subtasks that cover skills from different stages of literacy development. The four skills measured were (1) decoding pseudo-words or unfamiliar words; (2) oral comprehension (administered only to children unable to decode pseudo-words); (3) oral reading fluency; and (4) reading comprehension (assessed twice, by using passages with different levels of difficulty). Local stakeholders reviewed all assessment materials to ensure that the questions were appropriate for the cultural and linguistic context of the RACCS. Mathematica and FIDEG thoroughly vetted and piloted the questionnaires and assessments to improve face validity. Table F.1 includes a description of each subtask.

Enumerators administered the assessments orally in the school's language of instruction in each educational community: Spanish, English, or Miskitu. The assessments took 5 to 15 minutes to administer, depending upon the child's ability. Within each subtask, the enumerators recorded the correct number of responses in each line or section, the time remaining (in seconds) for timed tasks, and the total number of correct responses. Enumerators marked an "autostop" check box at on the test form if the child was unable to correctly answer an item in the first row or section of the subtask. This "early stop rule" is consistent with EGRA procedures. Having a time limit for some subtasks also is standard for EGRA; it makes the assessment shorter and helps assess how automatic the children's responses are (RTI International 2016).

Table F.1. Description of literacy assessment subtasks

Subtask	Description	Number of items	Timed task
Decoding pseudo-words (also known as invented words)	This task measured children's ability to decode pseudo-words with speed and accuracy. Pseudo-words are vowel-consonant combinations that follow the target language's phonological and spelling rules but are not actual, meaningful words. Children read aloud as many pseudo-words as they could in 60 seconds. Enumerators discontinued the task if the child failed to read the first five pseudo-words. We used the task as a universal screener to determine if the fluency and reading comprehension tasks were to be administered.	Fifty pseudo-words	Yes
Oral comprehension	This task used recall and inferential questions to measure the receptive language of an orally read passage. The recall questions were based on information taken directly from the text, whereas the inferential questions asked children to draw conclusions based on the passage. The score shows the percentage of correct responses out of five questions. Enumerators administered this task only to children who were unable to decode a single pseudo-word correctly.	Five questions	No
Oral reading fluency	This task measured children's ability to read a 3rd grade level passage with speed and accuracy. The oral reading fluency score shows the number of words read correctly per minute. When children stopped on a word for three seconds, they were asked to read the next word. Enumerators administered this task to children who decoded at least one pseudo-word correctly.	Spanish: 196 words English: 190 words Miskitu: 200 words	Yes
Reading comprehension	This task measured children's ability to answer recall and inferential questions about a passage. The recall questions were based on information taken directly from the text, whereas the inferential questions asked children to draw logical conclusions based on the passage. Children were presented with two passages. The first passage was appropriate for 3rd grade students and the second was appropriate for 5th graders (because it was longer and the words were more difficult than those in the first passage). The reading comprehension scores show the percentage of correct responses to each passage out of five questions. Enumerators administered this task to children who decoded at least one pseudo-word correctly.	Passage 1: 176 words in Spanish, 172 in English, and 179 in Miskitu Passage 2: 192 words in Spanish, 188 in English, and 184 in Miskitu	No, but four minutes are set as the limit to prevent children from reading the passage twice and to limit participant burden

Table F.2 presents summary statistics for each subtask for the full sample, as well as by community size. For timed subtasks (decoding pseudo-words and oral reading fluency), we calculated the number of correct pseudo-words or words read per minute. For all other subtasks we calculated the percentage of correct responses in the subtask. To minimize the children's burden, we did not administer the reading fluency and reading comprehension subtasks to children who were unable to decode a single pseudo-word correctly. In light of the high correlation between decoding pseudo-words and reading fluency ($r = 0.92$ in our sample), we assigned a score of zero in the reading fluency and reading comprehension subtasks to those children. In addition, we only administered the oral comprehension subtask to children who were unable to decode a single pseudo-word correctly.

Table F.2. Summary statistics for literacy skills at endline

Literacy skills	Minimum value	Maximum value	Mean	Standard deviation	Number of children	Number of communities
Panel A: All communities						
Decoding pseudo-words (number correct per minute)	0	93.8	18.6	15.7	2,356	199
Oral comprehension (percentage correct out of five)	0	100.0	47.6	26.6	602	199
Reading fluency (number of correct words per minute)	0	163.8	40.3	33.8	2,354	199
Reading comprehension, 3rd grade level (percentage correct out of five)	0	100.0	43.2	33.6	2,354	199
Reading comprehension, 5th grade level (percentage correct out of five)	0	100.0	35.6	33.4	2,354	199
Panel B: Small communities						
Decoding pseudo-words (number correct per minute)	0	79.6	17.1	15.4	1,169	147
Oral comprehension (percentage correct out of five)	0	100.0	46.5	26.9	343	147
Reading fluency (number of correct words per minute)	0	153.2	39.2	34.3	1,167	147
Reading comprehension, 3rd grade level (percentage correct out of five)	0	100.0	39.7	33.3	1,167	147
Reading comprehension, 5th grade level (percentage correct out of five)	0	100.0	34.9	33.2	1,167	147
Panel C: Large communities						
Decoding pseudo-words (number correct per minute)	0	93.8	20.1	15.9	1,187	52
Oral comprehension (percentage correct out of five)	0	100.0	49.0	26.2	259	52
Reading fluency (number of correct words per minute)	0	163.8	41.5	33.2	1,187	52
Reading comprehension, 3rd grade level (percentage correct out of five)	0	100.0	46.7	33.5	1,187	52
Reading comprehension, 5th grade level (percentage correct out of five)	0	100.0	36.3	33.6	1,187	52

Source: Follow-up household survey.

Note: This table shows unadjusted summary statistics for literacy skills for children who were interviewed during the year-end data collection. The oral comprehension task was applied to nonreaders, or children who could not decode sufficient numbers of pseudo-words, which is why the sample size was small. Other sample sizes varied due to item nonresponse.

B. Language adaptation

The EGRA is not intended to be comparable across languages (Dubeck and Gove 2015). However, in order to assess the emergent skills of a linguistically diverse population as part of this evaluation, we adapted the subtasks for administration in Spanish, English, and Miskitu.

Local consultants referenced textbooks, grade-level word lists, and prior assessments to determine the appropriate content, text length, and vocabulary for each subtask. Unlike Spanish, there is limited formal knowledge of Miskitu. Consultants emphasized the lack of vocabulary complexity and shortness of words (number of letters per word) in Miskitu, relative to Spanish or English. Spanish and English assessments were developed first; Miskitu passages were then adapted to match the difficulty levels of the other languages. All subtasks underwent revisions for content and length based on field tests.

We analyzed the level of difficulty of the reading passages for the oral comprehension, reading fluency, and reading comprehension subtasks by using number of words, average word length, number of sentences, and average sentence length. Table F.3 shows that the passages have comparable levels of difficulty across the three languages.[36] Given the similar difficulty levels, we conducted an impact analysis by using the sample as a whole, as opposed to separately by language. This analytical choice preserved the evaluation's statistical power to detect the impact of EpC. It is important to note, however, that we did not formally test and cannot assume measurement invariance across the three languages.

Table F.3. Statistics for reading passages in Spanish, English, and Miskitu

	Oral comprehension			Reading fluency			Reading comprehension—3rd grade level			Reading comprehension—5th grade level		
Test feature	Spa	Eng	Mis	Spa	Eng	Mis	Spa	Eng	Mis	Spa	Eng	Mis
Word count	99	101	100	196	190	201	176	172	179	192	188	184
Average word length	4.4	4.4	4.1	4.1	4.2	4.4	4.4	4.3	4.6	4.1	4.3	4.7
Number of sentences	7	7	6	13	12	14	12	13	12	12	13	12
Average sentence length	14.1	14.4	16.7	15.1	15.8	14.4	14.7	13.2	14.9	16	14.5	15.3

Eng = English; Mis = Miskitu; Spa = Spanish.

C. Internal consistency and reliability

Cronbach's alpha is one of the most widely used measures of internal reliability for multiple-item tests. It calculates the intercorrelation between test items: the higher the coefficient, the more the items measure a given concept in the same way (Tavakol and Dennick 2011). Scores range from 0 (items within the test are completely uncorrelated) to 1 (items are perfectly correlated). The literature on Cronbach's alpha cites 0.7 to 0.95 as an acceptable range

[36] There are several indexes designed to analyze text complexity, including the Flesch-Kincaid Reading Ease, Gunning-Fog Score, and Coleman-Liau Index. However, these indexes are based on the English language and do not translate well across the other two languages in this study. For example, syllables are counted differently in English and Spanish, so direct comparisons are inappropriate. The statistics that we used to evaluate passage difficulty (for example, word count, word length, and so on) were preferred because they were not specific to a given language.

for establishing internal consistency within test items (Tavakol and Dennick 2011). Bland and Altman (1997) specify that a range of 0.7 to 0.8 is sufficient when comparing groups in social science, whereas a Cronbach's alpha above 0.9 is critical in clinical settings. For this reason and in accordance with previous early grade reading studies, we consider 0.7 or higher an acceptable Cronbach's alpha—that is, it reflects a high degree of internal consistency across the test items.

At follow-up, Cronbach's alpha for the overall literacy assessment was 0.88,[37] indicating a high level of internal consistency reliability. Table F.4 shows that internal consistency is high for children whose mother tongue is Spanish, English or Kriol, and Miskitu.[38]

Table F.4. Internal consistency reliability (Cronbach's alpha)

	Alpha			
	All	Spanish	English or Kriol	Miskitu
Overall literacy assessment	0.88	0.88	0.89	0.90
Number of children	**2,356**	**1,889**	**315**	**148**

In addition to calculating Cronbach's alpha for the overall assessment, we analyzed the correlation between subtasks. The subtasks were arranged in increasing order of difficulty. Therefore, we expected adjacent subtasks to be closely correlated, meaning that students scoring high on one subtask would also likely score high on the preceding and succeeding subtasks. Our findings confirmed that for the most part adjacent subtasks were highly correlated with each other. All of the correlations were statistically significant (see Table F.5).

Table F.5. Correlations of literacy assessment subtasks

	Subtask 1	Subtask 3	Subtask 4a	Subtask 4b
Subtask 1: Decoding pseudo-words (number correct per minute)	1.000			
Subtask 3: Reading fluency (number of correct words per minute)	0.918	1.000		
Subtask 4a: Reading comprehension, 3rd grade level (percentage correct out of five)	0.655	0.679	1.000	
Subtask 4b: Reading comprehension, 5th grade level (percentage correct out of five)	0.650	0.707	0.708	1.000

Source: Follow-up household survey.

Note: Correlations are presented for 2,358 children who were assessed at follow-up. Subtask 2, oral comprehension, was omitted because it was only administered to 603 children, who were unable to decode a single pseudo-word correctly. Those children were not administered Subtasks 3, 4a, or 4b.

[37] According to the EGRA toolkit, it is improper to calculate Cronbach's alpha for individual subtasks because timed or time-limited measures inflate the reliability score and the degree to which the score is inflated is unknown. The toolkit recommends using summary scores of subtasks to calculate an overall Cronbach's alpha across the subtasks (RTI International 2016).

[38] Cronbach's alpha was similar across the three languages of administration of the assessment, which was not based on children's mother tongue but on the language of instruction in the community. Alphas ranged from 0.88 to 0.90. Results are available on request.

D. Inter-rater reliability

We estimated the magnitude of agreement between enumerators for approximately five percent of the assessments administered at follow-up. We used two measures of agreement: (1) an index of observed agreement and (2) Cohen's kappa statistic. We calculated Cohen's kappa by using the Spanish assessments only because the samples sizes were too small to calculate it for the English and Miskitu assessments.

Observed agreement is the percentage of items for which the enumerators assigned the same score, whereas Cohen's kappa quantifies the difference between observed agreement and the degree of agreement expected by chance alone (expected agreement). Kappa ranges from -1 to 1, where 1 is perfect agreement, 0 is exactly what would be expected by chance, and a negative value indicates less agreement than would be expected by chance (Viera and Garrett 2005). Kappa values above 0.90 are considered almost perfect agreement, values above 0.80 but below 0.90 are considered strong agreement, values of 0.60 to 0.79 are considered moderate agreement, values of 0.40 to 0.59 are considered weak agreement, and values of 0.30 or less represent minimal to no agreement (McHugh 2012).

Table F.6 shows perfect or nearly perfect agreement (at least 95 percent) between raters in 14 of 20 subtask-by-language comparisons. For Cohort 1, oral comprehension in Miskitu showed moderate levels of agreement between raters during the first follow-up, as did the reading comprehension subtasks in Spanish and English. In both the oral and reading comprehension subtasks, enumerators used their judgment to score children's answers to open-ended questions, which likely explains the lower levels of agreement.

Table F.6. Percentage agreement between raters in the literacy assessment

	Decoding pseudo-words	Oral comprehension	Reading fluency	Reading comprehension, 3rd grade level	Reading comprehension, 5th grade level
Follow-up, Cohort 1					
Spanish					
Percentage agreement	98	100	99	71	80
Number of children	51	9	42	37	37
English					
Percentage agreement	97	100	95	73	73
Number of children	11	5	6	6	6
Miskitu					
Percentage agreement	96	67	99	97	87
Number of children	9	3	6	6	6
Follow-up, Cohort 2					
Spanish					
Percentage agreement	98	98	99	99	98
Number of children	64	19	45	45	45

Source: Follow-up household survey.

Note: Languages refer to the language of administration of the assessments. Three raters, instead of two, administered Miskitu assessments. The table shows the average agreement between pairs of raters.

Kappa statistics were consistent with those from the percentage agreement analysis (see Table F.7). There was moderate to perfect agreement in all subtasks during the second follow-up for Cohort 2. In addition, there was strong to perfect agreement in decoding pseudo-words, oral comprehension, and reading fluency in the first follow-up for Cohort 1; however, agreement was weak for reading comprehension.

Table F.7. Inter-rater reliability in the literacy assessment

	Decoding pseudo-words	Oral comprehension	Reading fluency	Reading comprehension, 3rd grade level	Reading comprehension, 5th grade level
Follow-up, Cohort 1					
Kappa statistic	0.892[a]	1.000	0.835[b]	0.434	0.468
Number of children	**51**	**9**	**42**	**37**	**37**
Follow-up, Cohort 2					
Kappa statistic	0.838[c]	0.963	0.706[d]	0.963	0.948
Number of children	**64**	**19**	**45**	**45**	**45**

[a] Only 46 items out of 50 were included in the pooled kappa estimate.
[b] Only 70 items out of 196 were included in the pooled kappa estimate.
[c] Only 43 items out of 50 were included in the pooled kappa estimate.
[d] Only 68 items out of 196 were included in the pooled kappa estimate.

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

APPENDIX G:

ADDITIONAL CONSTRUCTS

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

A. Measurement of social-emotional skills and attitudes

We focused on social-emotional skills and attitudes explicitly targeted by EpC (for example, self-esteem), skills and attitudes considered predictors of later academic and life success (for example, social competence), and risk factors for engaging in illicit behavior (for example, attitudes toward delinquency). Even though EpC implementation materials do not directly address delinquent behavior, the logic model proposes that, with time, program activities will mitigate the impact of factors that increase a youth's likelihood of joining street gangs and engaging in other delinquent activities. We assessed risk factors for delinquency, as opposed to delinquent behaviors, because our sample included young children for whom the questions about delinquency would had been inappropriate (children were ages 7 to 15 in the base year), and detecting long-term impacts on delinquency would require collecting data beyond the time frame of this evaluation.

To measure these constructs, we used a combination of validated surveys, including the American Institutes for Research's Conditions for Learning survey (UNICEF 2009) and the Conflict Behavior Questionnaire (Prinz et al., 1979), along with a few items developed by the evaluation team (see Table G.1). Measures that were originally in English were adapted to Kriol and to Spanish, and then adapted from Spanish to Miskitu. In adapting the items to align with Nicaraguan idioms and the education level of the evaluation population, we used input from FIDEG staff, and we piloted the measures with Nicaraguan children and adults in the RACCS.

Table G.1. Description of measures of social-emotional skills and attitudes and risks for engaging in illicit behavior

Socio-emotional skills and attitudes	Description	Data source	Number of items
1. Relationship skills: social competence	We used an adaptation of the Conflict Behavior Questionnaire (CBQ, Parent Version, Prinz et al., 1979) to assess children's relationship skills, a dimension of social competence that encompasses listening, self-regulation, and communication skills that are fundamental to establishing and maintaining positive relationships. Caregivers were presented with 14 items that described their interactions with the child (for example, "Your child listens to what you tell him/her; gets angry with you; cares about what you think"), and rated the frequency of such interactions using a four-point scale (ranging from never = 0 to always/almost always = 3). Responses were coded so that higher scores denoted more positive relationships. All items were averaged to create a score for each child. Analysis of Follow-up 1 data showed that the number of items could be reduced while maintaining high internal reliability. To reduce participants' burden and costs, three items were removed for Follow-up 2. Our main analysis uses the 11-item measure. Results using the 14-item measure are presented in Appendix I, where we show additional analysis for large communities, where most Follow-up 1 data were collected.	Caregiver	14 in Follow-up 1; 11 in Follow-up 2

Socio-emotional skills and attitudes	Description	Data source	Number of items
2. Self-esteem	We assessed children's self-esteem using five items adapted from the Global Self-Worth subscale from Harter's Self Perception Profile for children (Harter 2012). The scale had been previously validated in Spanish (Broc, 2014), but had never been used in Nicaragua. The items assess how much children like themselves. On each item, the child is asked to decide which kind of kids he or she is most like: kids who feel good/happy with themselves, or kids who do not feel good/happy with themselves. Children get a score of 1 if they identify with kids who feel good with themselves, and 0 if they identify with kids who don't feel good about themselves. We coded all items so that higher scores denote higher self-esteem. We then summed all the items to create a total score. Follow-up 1 analysis indicated that the measure was not internally consistent for our evaluation sample. In Follow-up 2, we also used five items adapted from Rosenberg's self-esteem scale (Rosenberg, 1965) as an alternative measure of self-esteem. Items include: "you have a positive attitude about yourself," "you are able to do things as well as most people," "you feel that you have a lot to be proud of," "you feel that you have a number of good qualities," "in general, you are satisfied with yourself." Children were asked to rate how true each statement was for themselves (0 = not at all true, 3 = very true). We averaged the five items to obtain a total score for each child. Our main analysis uses Harter's measure because it was collected at both Follow-up 1 and 2. Results using Rosenberg's scale are reported in Appendix H, where we report additional findings for small communities, most of which have data from Follow-up 2. Results using Rosenberg's scale are not reported for large communities in Appendix I because most Follow-up 2 data were collected in small communities.	Child	5 items in Follow-up 1; 10 items in Follow-up 2
3. Intercultural competence	We used two items adapted from the Safe, Inclusive, and Respectful Climate subscale from the Conditions for Learning Survey (that is, children in this community: "help each other even if they are not friends; treat each other with respect") (UNICEF 2009). The two items had been previously used in Nicaragua. We developed two additional items to gauge children's openness to peers from different backgrounds (that is, your community is a welcoming place for "children who speak different languages; "come from different regions"). Preliminary analysis with Follow-up 1 data indicated that internal reliability was low. We added two items ("Children in your community like talking to people from other places," "You'd like to be friends with children who speak differently") to improve internal reliability. Those items were administered in Follow-up 2 only. The main analysis uses the four items used in both follow-ups.	Child	4 items in Follow-up 1; 6 in Follow-up 2

Socio-emotional skills and attitudes	Description	Data source	Number of items
4. Academic engagement and community support for academic goals	We measured academic engagement and children's perceptions of support for academic goals by adults in their community using five items adapted from the Conditions for Learning Survey (UNICEF 2009). Children were asked to use a scale from 0 ("not at all true") to 3 ("very true") to rate whether they were interested in going to school, completing secondary school, and whether adults in the community encourage them to take school seriously and believe that all children can learn and do well in life. All items were averaged to obtain an overall score for each child that ranged from 0 to 3.	Child	5
Risk factors			
5. Moral disengageme nt	We adapted six items from the Neutralization subscale in the Youth Services Eligibility Tool (Hennigan et al. 2014). Items included: "It is okay to lie to keep your friends from getting in trouble," "It is okay to hit people if they hit you first," and "It is okay to take something from someone who can easily replace it." Children were read the statements and were asked to indicate their level of agreement by pointing to a laminated scale ranging from 0 (disagree a lot) to 3 (agree a lot). Responses were coded so that higher scores represented higher levels of moral disengagement. We averaged all items to compute an overall score for each child that ranged from 0 to 3.	Child	6
6. Impulsive risk taking	We used four items adapted from the Impulsive Risk Taking subscale in the Youth Services Eligibility Tool (Hennigan et al. 2014). Children were read the statements ("Sometimes you like to do something dangerous just for the fun of it," "Sometimes you find it exciting to do things that might get you in trouble, "Sometimes you do things without stopping to think if you will get in trouble for it," and "You like to have fun when you can, even if you will get in trouble for it later") and were asked to indicate if the statement was 0 (not at all true) to 3 (very true) for him/herself. An average of all the items was computed as the overall score for each child ranging from 0 to 3.	Child	4
7. Bullying or peer victimization	We assessed children's overt and covert victimization by peers using an adaptation of the Gatehouse Bullying Scale (Bond et al. 2007). Children were asked about four types of situations: "Anyone made fun of you or called you nicknames," "Someone hurt you, or told you they were going to hurt you," "Your friends did not invite you to play," and "Anyone talked bad about you recently." For each situation, they were asked to rate how frequently it happened (a lot, a little) and how much it upset them (a lot, a little, not at all). The severity of victimization was computed for each of the four situations using a scale from 0 to 3 (0 = not bullied, 1 = bullied but not frequently and not upset, 2: bullied, either frequently or upset, but not both, 3 = bullied frequently and upset. We computed an overall score (ranging from 0 to 3) by taking the mean across the four situations.	Child	4

Socio-emotional skills and attitudes	Description	Data source	Number of items
8. Attitudes toward delinquency	We used four items adapted from the Pittsburgh Youth Study (Loeber et al. 1998), asking children about attitudes that legitimize engaging in delinquent behavior (that is, it is OK to: "disobey or talk back to adults such as parents, teachers, or others," "purposely damage or destroy property that belongs to someone else," "use force to get money or things from people," and "miss school without an excuse.”). Children answered using a 4-point scale (0 = agree a lot–3 = disagree a lot). We coded answers so that higher scores reflect attitudes that favor delinquency. An average of all the items was computed as the overall score for each child ranging from 0 to 3.	Child	4
9. Sense of safety in the school and on the way to school	We measured children's sense of safety in and around the school using two items adapted from the Conditions for Learning Survey Safe, Inclusive, and Respectful Climate subscale (UNICEF 2009). Children were asked whether they felt safe at school and going to or coming back from school. Our main analysis uses each individual question.	Child	2
10. Caregiver/ child perception of safety in the community	We asked caregivers and children to rate the safety of their community and how safe they felt walking alone at night using a four-point scale (0 = very safe, 1 = a little safe, 2 = a little unsafe, and 3 = very unsafe). We coded answers as 0/1 dummy variables, where higher scores reflect feeling more unsafe (0 = very unsafe or a little unsafe; 1 = a little safe or very safe). Our main analysis uses each individual item.	Caregiver/Child	2
11. Community safety and abuse	We used seven items from the CARS Socio-educational Diagnostic Assessment (Instrumento para Diagnóstico Socio-educativo, in Spanish) to assess whether the school has instances of bullying, children being threatened, children who show extreme signs of anxiety or fear in the school, etc. School directors answered yes or no to each item.	School director	7

Note: Survey items are worded in Kriol instead of standard English.

We used instruments that rely on self- and caregiver-reports because they are more efficient, affordable, and easier to administer than direct observation or performance measures. However, this mode of administration has known limitations, including social desirability bias. We employed some strategies to minimize its influence. For example, we selected a measure of self-esteem designed to minimize social desirability bias (Harter 2012). Children were asked to choose which of two hypothetical groups they identify with the most. Framing the question as a choice between two existing groups of children makes it more acceptable for children in the evaluation to endorse socially undesirable views. For the other measures, the instructions underscored that there are no right or wrong answers. Even though social desirability likely reduced score variability, a nontrivial number of children endorsed socially undesirable items (for example, 19 percent of children said that sometimes they like doing something dangerous because it is fun; 32 percent agreed that it is OK to lie to prevent your friends from getting into trouble).

In addition to known limitations of self-report measures, there are challenges related to the cultural and linguistic heterogeneity in our sample. Mathematica researchers who are native Spanish speakers and fluent in English reviewed the English-to-Spanish translations to ensure translation accuracy. The English versions were adapted to Kriol by a group of data collectors who are native Kriol speakers. A native Miskitu speaker translated the measures from Spanish to Miskitu, and the translation was then reviewed by a group of data collectors who were native Miskitu speakers. All versions were refined to match the linguistic and cultural context of the RACCS based on input from the local teams and pilot activities. Even though these processes improved the instruments' ecological validity, they could have led to losses of fidelity or precision that the researchers are not aware of (especially for the Miskitu instruments), which could affect construct validity. These limitations may underlie the low internal reliability in Table G.3. It is also possible that the constructs measured are not identical across cultures, or that there are relevant behaviors or beliefs associated with the constructs that are unique to a given group and were not represented in the instruments.

Table G.2. Summary statistics for social-emotional skills and attitudes and risk factors

Social-emotional skills and attitudes	Count	Minimum observed value	Maximum observed value	Mean	Standard deviation	Skewness
Relationship skills: social competence	2,341	0.4	3.0	2.1	0.5	-0.4
Intercultural competence (4 items)	2,356	0	3.0	2.1	0.7	-0.4
Intercultural competence (6 items, C2 only)	1,252	0	3.0	2.1	0.6	-0.4
Academic engagement and community support for academic goals	2,354	0	3.0	2.5	0.5	-1.1
Self-esteem (5 items)	2,338	0	5.0	4.0	1.1	-1.0
Self-esteem (5 items, C2 only)	1,252	0	3.0	2.0	0.7	-0.5
Risk factors						
Moral disengagement	2,356	0	3.0	1.1	0.6	0.2
Impulsive risk taking	2,356	0	3.0	1.1	0.7	0.4
Bullying/peer victimization (construct without "not invited to play")	2,356	0	3.0	0.7	0.9	1.0
Attitudes toward delinquency	2,356	0	3.0	0.8	0.7	0.6
Sense of safety in school and going to school (range: 0–1), child report	2,342	0	1.0	10.4	24.7	2.4
You feel safe at school (child report)	2,338	0	1.0	92.6	26.1	-3.3
You feel safe going and coming back to school (child report)	2,340	0	1.0	86.7	34.0	-2.2
Your community is very safe or safe (child report)	2,356	0	1.0	62.6	48.4	-0.5
You feel very safe or safe walking alone at night (child report)	2,356	0	1.0	35.4	47.8	0.6

Social-emotional skills and attitudes	Count	Minimum observed value	Maximum observed value	Mean	Standard deviation	Skewness
Your community is very safe or safe (caregiver report)	2,361	0	1.0	67.5	46.8	-0.7
You feel very safe or safe walking alone at night (caregiver report)	2,359	0	1.0	49.8	50.0	0.0

Source: Follow-up household survey. C2 stands for Cohort 2.

B. Internal consistency and reliability for social-emotional skills, attitudes, and risk factors

Cronbach's alpha (see Appendix F for a definition of Cronbach's alpha) scores in Table G.3 show that the measures of social-emotional skills, attitudes, and risk factors had varying levels of internal reliability in our sample. The measures of social competence and bullying/peer victimization were internally consistent ($\alpha = .71$ in both cases).[39] Intercultural competence (as measured in Cohort 2, including six items), moral disengagement, attitudes toward delinquency, and academic engagement had alphas between .60 and .64, which are below but close to the lower threshold of .70. Low alpha values can result from having few items in the scale, poor inter-relatedness between the items, or items that measure multiple constructs as opposed to a single common construct (Cortina 1993). The scales with alphas slightly below the threshold had only four to six items so it is possible that the number of items, and not poor inter-relatedness, is driving the results. We conducted impact analysis using summary scores from these scales, but caution should be exercised when interpreting these results, because low reliability may affect the statistical power to detect intervention impacts (Henson 2001). Self-esteem (Harter's and Rosenberg's scales), intercultural competence (using four items), and impulsive risk taking have alphas below .60 and are therefore not internally consistent for this sample. Also, alphas are generally lower for the English/Kriol and Miskitu samples (for which we do not present separate analyses). However, the measure of social competence, our primary measure of social-emotional skills, is internally consistent for the three groups. In light of these findings, we conducted exploratory impact analysis on the individual items that constitute these measures, in addition to examining impacts on their summary scores. The conclusions from the two specifications did not differ; therefore, we present results using summary scores for the sake of brevity.

[39] Note that "not being invited to play" was excluded from the peer victimization scale.

Table G.3. Internal consistency reliability (Cronbach's alpha)

	Alpha				Sample size			
	Overall (A)	Spanish (B)	English/ Kriol (C)	Miskitu (D)	(A)	(B)	(C)	(D)
Relationship skills: social competence (11 items)	0.71	0.70	0.72	0.67	2339	1871	316	148
Relationship skills: social competence (14 items))	0.74	0.74	0.76	0.71	1105	685	309	148
Self-esteem (Harter) (5 items)	0.40	0.42	0.38	0.08	2336	1874	312	145
Self-esteem (Rosenberg) (5 items)	0.59	0.59	0.54	n/a	1250	1202	44	n/a
Intercultural competence (4 items)	0.49	0.51	0.41	0.39	2354	1888	314	147
Intercultural competence (6 items)	0.62	0.63	0.50	n/a	1250	1202	258	n/a
Academic engagement and community support for academic goals (5 items)	0.64	0.66	0.52	0.67	2352	1886	314	147
Moral disengagement (6 items)	0.63	0.64	0.53	0.36	2354	1888	314	147
Impulsive risk taking (4 items)	0.55	0.56	0.38	0.48	2354	1888	314	147
Bullying or peer victimization[1] (4 items_	0.71	0.72	0.63	0.55	2354	1888	314	147
Attitudes toward delinquency (4 items)	0.60	0.60	0.56	0.58	2354	1888	314	147

Source: Follow-up household survey.

Note: There were no Miskitu speakers in Cohort 2, so internal consistency cannot be computed using the intercultural competence items that were used in that cohort only.

C. Measures of instructional practices and classroom characteristics

To examine whether EpC strategies were implemented or if perceptions or behaviors may have changed as a result of EpC, we used a combination of previously validated items and a few items developed by the evaluation team (see Table G.4). We asked educators about their instructional practices, including those that promote student engagement and concept development, their classroom management practices, the use of literacy assessments to monitor students' progress, as well as other EpC-specific strategies. Finally, we asked educators about their perceptions of obstacles to teaching, as their perceptions may have changed as a result of CARS or EpC activities.

Table G.4. Description of measures of instructional practices and classroom characteristics

Instructional practices	Description	Data source	Number of items
1. Practices to promote student engagement	We developed eight items asking educators if they used strategies that can improve student engagement in academic activities. Items include "use of music," "use of traditions and routines," "arranging children in a flexible way in the room," "use of didactic materials," etc. Educators answered "yes" or "no," and when the answer was yes, they were asked how often they implemented the strategy (once a month, once every two weeks, once a week, once a day, more than once a day.) The report uses the yes/no answers.	Educator	8
2. Practices to promote concept development	We adapted eight items developed by the Developmental Studies Center (2005) for students in grades 3–6, to be administered to educators. The items focused on academic activities expected to promote students' reading and comprehension skills. Items include having students "read or practice reading during class hours," "read books other than textbooks," "talk about why characters in books or stories did what they did," etc. Educators answered "yes" or "no," and when the answer was yes, they were asked how often they implemented the strategy (once a month, once every two weeks, once a week, once a day, more than once a day.) The report uses the yes/no answers.	Educator	8
3. Use of reading assessments	We asked educators if they had ever applied standardized reading assessments such as the EGRA and mini-EGRA to their students, and which months they had applied them in in the current and previous school years.	Educator	3
4. Classroom behavior management	We asked educators about the time they spent managing misbehavior (0 = most of the time–4 = none of the time) and how often misbehavior interrupts classroom activities (0 = most of the time–3 = almost never).	Educator	2
5. Classroom learning environment	We adapted 12 items from the “Safe and Welcoming Classroom Environment” Classroom Observation in the Conditions for Learning Survey (UNICEF 2009). Items included whether the classroom was “protected from the elements,” “has adequate ventilation,” “is a comfortable temperature,” “has adequate lightning,” etc. Two of the items are considered part of the EpC model, namely the presence of “posters, artwork, or maps on the walls,” and “examples of student work or projects visible in the classroom.” We developed an additional item to ask if “rules of behavior were visible in the classroom,” which is also part of the EpC model.	Classroom observation	13

Instructional practices	Description	Data source	Number of items
6. Implementation of EpC activities	We asked EpC facilitators whether they implemented 11 EpC-specific activities including "splitting the children into *peces* and *robles*," "promoting the *child tutor* component," "promote activities of *tradition,"* "guarantee that homework assigned by the teacher was completed," etc. Educators answered "yes" or "no" and when the answer was yes, they were asked how often they implemented the strategy, using the same response scale as for the items in boxes 1 and 2 in this table.	Educator	11
7. Obstacles to teaching	We developed eight items to assess educators' perceptions of factors that make it difficult to teach. Items include "the small amount of teaching or classroom time," "absence of support from parents," "frequent student absences," and "the absence of student motivation."	Educator	8

D. Measure of household socioeconomic status

We constructed a household poverty indicator using an adaptation of the "Unsatisfied Basic Needs" methodology (*Necesidades Basicas Insatisfechas*, in Spanish), previously used by the Nicaraguan National Institute of Statistics and Census (Castañeda 2017). The methodology uses five categories: overcrowding, unsound housing, insufficient services, low education, and economic dependence. We used three of these categories to construct the poverty indicator included in our end-line analysis:[40]

- Overcrowding: A household is considered overcrowded in urban areas if there are four or more people per bedroom; in rural areas this standard is five or more people per bedroom. If there are no bedrooms, the house is considered one-bedroom.
- Unsound housing: A house is considered "sound" if it has two acceptable combinations of materials in its walls, roof, or floor. If it has only one acceptable combination of materials or none in its walls, roof, or floor, the house is considered "unsound."
- Economic dependence: This is defined using a binary variable equal to "1" if the household head has an education level of incomplete primary school or less and two or more people in urban areas (or three or more people in rural areas) depend on her or his income. If the household does not meet these conditions, the variable equals "0".

The family poverty indicator equals "1" if two or more of these conditions are met. If only one of these conditions is met, the household is considered poor but not extremely poor, and if none of these conditions are met, the household is not considered poor. By this definition, which is consistent with Nicaraguan standards, most households in the sample are considered extremely poor.

[40] We did not collect information on insufficient services at follow-up, because questions about services raised concerns in some communities. Also, low education (a dummy for whether there is at least one child age 7–14 who does not currently attend school) was not included in the indicator because base year information was not available for Cohort 1 households, and follow-up child education level was endogenous to the intervention.

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

APPENDIX H:

ADDITIONAL FINDINGS ON EDUCATOR AND CHILD OUTCOMES FOR SMALL COMMUNITIES

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

A. Implementation findings

In Chapter V, we used educator self-reports to compare the percentage of educators trained in reading, writing, and the use of reading assessments across EpC and control communities. In this section, we used school director reports to compare the percentage of teachers trained in other EpC methodologies[41] and used educators' self-reports to explore in more depth the timing of training (Table H.1).

Table H.1. Teacher training in EpC methodologies in small communities

	EpC (A)	Control (B)	Impact (A–B)	*p*-value
Received training in gender equality				
Trained the year after randomization	21.5	11.9	9.6*	0.092
Trained the year of randomization	25.5	4.3	21.2	0.193
Received training about the transition to kindergarten				
Trained the year after randomization	9.9	7.5	2.4	0.583
Trained the year of randomization	3.7	2.9	0.8	0.760
Received training about active learning methodology				
Trained the year after randomization	46.8	26.1	20.8***	0.004
Trained the year of randomization	9.7	5.3	4.4	0.269
Received training in APA				
Trained the year after randomization	36.7	18.2	18.5***	0.009
Trained the year of randomization	17.1	11.2	6.0	0.298
Received a certificate for teaching 1st to 3rd grade				
Received in the year after randomization	15.0	5.8	9.2*	0.062
Received in the year of randomization	7.7	5.6	2.0	0.605
Received training about monitoring reading progress using EGRA a year before data collection				
Trained the year after randomization	26.5	6.9	19.6***	0.001
Trained the year of randomization	2.7	4.1	-1.4	0.609
Number of communities	**74**	**73**		

Source: Follow-up school director survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for the urban/rural status of the community where the school is located.

*Difference in group means is statistically significant at the .10 level.

***Difference in group means is statistically significant at the .01 level.

[41] In this section we refer to "teachers" instead of "educators" more broadly because school director reports refer to educators who were teachers, and may have also been EpC facilitators (so not educators who were EpC facilitators only).

Similar to results in Chapter V, we found that a larger percentage of teachers in EpC communities than in control communities were trained in EpC methodologies the year after randomization. The differences for training in active learning methodologies, the Learn, Practice, and Apply method (Aprendo, Practico, Aplico [APA], in Spanish), and the use of assessments to monitor reading progress are statistically significant. However, as with results in Chapter V, there were no statistically significant differences in the percentage of teachers trained in the year of randomization, which could indicate that teachers received training later than expected.

Educators' self-reports indicate that a larger percentage of facilitators in EpC communities than teachers in control communities were trained in teaching reading, writing, and the use of assessments in the year after randomization (Table V.2). Here we examine educators' self-reports about training received two years after randomization and before randomization. We find that the percentage of teachers in the control group who reported having participated in trainings to teach reading and writing is 29 percentage points higher than facilitators. Differences in participation in trainings on the use of standardized assessments two years after randomization, as well as differences in participation in all trainings before the year of randomization were not statistically significant (Table H.2).

Table H.2. Educator training in small communities, by treatment status

Educator training	EpC (A)	Control (B)	Impact (A–B)	*p*-value
Received reading/writing training (percentage)				
Trained two years after randomization	25.4	54.1	-28.7***	0.000
Trained before the year of randomization	1.2	-0.1	1.3	0.259
Received standardized child assessment training (percentage)				
Trained two years after randomization	9.2	9.1	0.0	1.000
Trained before the year of randomization	1.2	-0.1	1.3	0.259
Number of educators	86	89		
Number of communities	70	68		

Source: Follow-up educator survey.
Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level. Column A–B presents differences in the regression-adjusted group means. The analysis sample includes teacher-facilitators and facilitators in the treatment group and teachers in the control group. There are 138 communities in the analysis sample (as opposed to 147) because in 4 EpC communities we did not interview EpC facilitators (teacher-facilitators or facilitators) and in 5 control communities we did not interview any teachers that had never been a facilitator.

*Difference in group means is statistically significant at the .10 level.

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

***Difference in group means is statistically significant at the .01 level

To understand the counter-intuitive result that two years after randomization more teachers in control communities than in EpC communities were trained, we excluded facilitators who were not teachers from the analysis to determine if the results were driven by those facilitators or by facilitators who were also teaches. The difference between the groups was equal to 10 percentage points and no longer statistically significant (Table H.3). Nearly 50 percent of teachers in both groups received training to two years after randomization, but few facilitators who were not teachers did. Therefore, including facilitators who were not teachers resulted in a significant difference between the groups.

Table H.3. Educator training in small communities, excluding facilitators who were not teachers

Educator training	EpC (A)	Control (B)	Impact (A–B)	*p*-value
Received reading/writing training (percentage)				
Ever received training	89.7	75.2	14.5*	0.086
Trained two years after randomization	45.6	55.8	-10.2	0.335
Trained the year after randomization	39.4	16.9	22.5**	0.020
Trained the year of randomization	1.8	2.6	-0.8	0.808
Trained before the year of randomization	2.9	-.01	3.0	0.347
Received standardized child assessment training (percentage)				
Ever received training	84.0	20.3	63.8***	0.000
Trained two years after randomization	9.4	10.6	-1.2	0.834
Trained the year after randomization	72.0	8.6	63.4***	0.000
Trained the year of randomization	-0.2	1.2	-1.4	0.418
Trained before the year of randomization	2.9	-0.1	3.0	0.347
Educator has participated in trainings on other topics since the year of randomization	97.6	85.3	12.3***	0.004
Number of other topics in which educator has been trained since the year of randomization	6.2	3.1	3.1***	0.000
Number of educators	**38**	**89**		
Number of communities	**36**	**68**		

Source: Follow-up educator survey.
Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level. Column A–B presents differences in the regression-adjusted group means. The analysis sample includes teacher-facilitators in the treatment group and teachers in the control group.

*Difference in group means is statistically significant at the .10 level.

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

***Difference in group means is statistically significant at the .01 level

B. Impacts on intermediate outcomes

The EpC model called for preparing the learning environment in a particular way, but there were no significant differences between EpC and control communities in the classroom characteristics we measured. In small communities, we did not find statistically significant differences between the two groups regarding the percentage of classrooms that have posters, artwork, maps on the walls, or visible examples of children's work or projects. There was a small difference in the presence of rules of behavior (which could include acting on one's own values, using mistakes as opportunities to learn, taking advantage of the present, being responsible for one's actions, and so on, as well as any other rules), but it was significant only at the 10 percent level (Table H.4). It is important to note that in several communities, the EpC had already closed (was no longer receiving CARS support or funding for the EpC facilitator's time) at the time of follow-up data collection.

Table H.4. Learning environment in schools prepared according to the EpC model in small communities

Characteristics of the learning environment (percentages)	EpC (A)	Control (B)	A–B	*p*-value
There are posters, artwork, or maps (commercially produced or handmade) on the walls of the classroom	89.5	86.9	2.6	0.645
There are examples of student work or projects visible in the classroom	79.0	72.0	6.9	0.367
The rules of behavior are visible in the classroom	37.4	26.0	11.4*	0.100
Number of communities	**71**	**73**		

Source: Follow-up school and classroom observation.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for the urban/rural status of the community where the school is located. Classrooms were observed by enumerators. When direct observation was not possible, enumerators relied on reports from educators. If more than one classroom was observed, an average at the school level was computed. The number of communities is 144 (as opposed to 147) because EpC classroom observations were not performed in three treatment communities.

*Difference in group means is statistically significant at the .10 level.

We also examined and found no statistically significant differences in classroom attributes that were not expected to be affected by EpC but that could have been impacted by other CARS activities. Those include whether the classroom had adequate ventilation and lightning, comfortable temperature, was clean and orderly, and had adequate space and furniture for children to work. In both groups, most classrooms met the basic physical conditions for learning (Table H.5).

Table H.5. Learning environment in small communities

	EpC (A)	Control (B)	Difference (A–B)	*p*-value
The classroom is protected from the elements	83.7	76.1	7.6	0.19
The classroom has adequate ventilation	87.8	87.7	0.1	0.99
The classroom has a comfortable temperature	83.2	85.9	-2.7	0.65
The classroom lighting is adequate for students to work	87.9	80.3	7.5	0.20
The classroom is clean and orderly	75.1	70.0	5.2	0.47
Outside noise affects communication within the classroom (reversed)	91.7	89.7	2.0	0.68
Each student has sufficient space to work	69.1	78.7	-9.7	0.17
Each student has a chair or bench to sit on while working	74.8	76.8	-2.0	0.75
Furniture is the right size for students to work comfortably	84.2	86.5	-2.2	0.71
There is a blackboard/whiteboard in the classroom that all students can see clearly from their seats	91.8	95.5	-3.6	0.34
Number of communities	**74**	**73**		

Source: Follow-up school and classroom observation.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for the urban/rural status of the community where the school is located. Classrooms were observed by enumerators. When direct observation was not possible, enumerators relied on reports from educators. If more than one classroom was observed, we computed an average at the school level.

Even though EpC did not directly target educators' classroom management strategies, well managed classrooms are known to facilitate learning. We found that EpC facilitators (teacher-facilitators and facilitators-only) reported spending significantly more time managing student behavior than teachers in the control group. The difference was equivalent to 0.4 standard deviations and may be due to increased awareness by EpC facilitators about the importance of creating a classroom environment conducive to learning. However, we found no statistically significant differences on educator reports of the frequency with which student misbehavior interrupted classroom activities (Table H.6).

Table H.6. Classroom behavior management in small communities

Behavior management	EpC (A)	Control (B)	Difference (A–B)	*p*-value
Time spent managing misbehavior (0=most of the time; 4=none of the time)	1.2	1.5	-0.3**	0.036
Educator reports that misbehavior interrupts classroom activities (0=most of the time; 3=almost never)	2.0	1.8	0.2	0.127
Number of educators	**86**	**89**		
Number of communities	**70**	**68**		

Source: Follow-up educator survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level.

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

We also asked educators about their perceptions of barriers to teaching, including parental support and student motivation, as EpC (or other CARS activities) could have led to improvements to such barriers. There were few statistically significant differences between educators in EpC and control communities (Table H.7). In small communities, fewer EpC facilitators (teacher-facilitators and facilitators-only) than teachers in control communities reported that little time for instruction was a barrier to teaching. The difference of 16 percentage points was significant at the 5 percent level. There were no other statistically significant differences between the groups. The lack of significant differences and overall low endorsement of items regarding teachers' knowledge of local language and students' knowledge of the language of instruction are likely due to the fact that small communities in the sample were relatively linguistically homogenous (see Appendix J for findings in large communities, which were more linguistically diverse).

Table H.7. Educators' perceptions of barriers to teach in small communities

	EpC (A)	Control (B)	Difference (A–B)	*p*-value
Number of perceived obstacles to teaching	4.7	4.6	0.1	0.805
Percentage of educators reporting the following perceived barriers				
Small amount of teaching or classroom time	59.5	75.1	-15.7**	0.038
Absence of training in intercultural bilingual education	60.7	66.1	-5.4	0.530
Teachers have little knowledge of local language	35.3	36.7	-1.4	0.855
Absence of support from parents	72.6	64.7	8.0	0.316
Frequent student absences	79.9	68.9	11.0	0.146
Absence of student motivation	69.7	63.0	6.7	0.414
Students don't understand the language of instruction	32.0	27.5	4.5	0.558
Parents don't help with school activities	61.1	60.1	1.0	0.910
Number of educators	**86**	**89**		
Number of communities	**70**	**68**		

Source: Follow-up educator survey.

Note: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level.

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

We compared the instructional practices of teacher-facilitators in the treatment group to those of teachers in the control group, to test whether the results in Table V.3 were robust to excluding facilitators who were not regular classroom teachers. The interpretation of results did not differ for practices to promote student engagement nor for practices to promote concept development, except for talking about books and stories read in class and for talking about the motivations of characters in books and stories. We found significant differences between the groups when facilitators only were included (see Table V.3) but not when they were excluded (Table H.8).

Table H.8. Educators' self-reported use of instructional practices they were trained on in EpC in small communities

Instructional practices (percentage)	EpC (A)	Control (B)	Impact (A–B)	*p*-value
Practices to promote student engagement				
Used music in the classroom	70.3	10.4	59.9***	0.000
Used traditions or routines	99.4	67.7	31.8***	0.000
Used flip charts on the walls to teach	95.0	71.8	23.1***	0.002
Used didactic materials to teach	96.0	72.5	23.5***	0.000
Flexibility for children to move easily from individual work to small group or whole class work	87.7	91.8	-4.0	0.570
Used strategies to focus children's attention on classwork	96.5	95.9	0.6	0.877
Used breathing techniques	65.3	58.6	6.7	0.571
Asked the children to imagine a place that promotes relaxation and concentration	79.3	60.5	18.7*	0.076
Sang songs with children	99.7	59.7	40.1***	0.000
Practices to promote concept development				
Students read or practiced reading	99.9	88.8	11.0***	0.007
Students read books other than textbooks	79.8	52.5	27.3**	0.011
Educator reads books or stories aloud in class	98.7	91.6	7.1*	0.056
Educator and students talked about books or stories that were read in class	87.3	87.4	-0.2	0.981
Educator talked about why characters in books or stories did what they did	89.0	87.9	1.1	0.861
Students worked in pairs during class	98.3	85.0	13.4***	0.007
Students worked in small groups during class	94.3	97.9	-3.6	0.478
Students shared questions or ideas during class time	100.4	93.1	7.2**	0.034
Use of student reading assessments	87.2	24.6	62.6***	0.000
Number of educators	**38**	**89**		
Number of communities	**36**	**68**		

Source: Follow-up educator survey.

Notes: Columns A and B present ordinary least squares regression-adjusted means that account for the evaluation design (cohort and strata variables). The analysis sample includes teacher-facilitators in the treatment group and teachers in the control group. Errors are clustered at the community level. Percentages greater than 100 result from regression adjustment, which was a linear probability model that can result in out-of-range predictions for binary outcomes.

*Difference in group means is statistically significant at the .10 level.

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

***Difference in group means is statistically significant at the .01 level.

C. Impacts on children's attachment to school

We report here findings on children's reports of academic engagement and the grade that children attended at follow up, which are additional indicators of school attachment. We also present additional findings for attendance and grade progression using a method to handle missing data that differs from what we use in the body of the report. In the version presented in Chapter VI, we used available information to replace missing data. In the version presented here, we kept missing data as missing, therefore excluding children with missing data from the analysis. The results presented here are overall consistent with results in Chapter VI; we found no significant impacts on attendance and a significant impact on the number of grades children progressed since recruitment. The only exception is the percentage of children who advanced a grade, where results vary depending on the version of the variable used.

We find that EpC had a positive impact on the proportion of children who advanced a grade since recruitment, but only when we restrict the analysis to children with complete data, as we do here. In Chapter VI, where we do not find a significant impact on the proportion of children who advanced a grade, we assigned a value of '1' to children with missing grade information who were out-of-school at intake but enrolled at follow-up, and a value of '0' to children who were enrolled at intake and out-of-school at follow-up. This method leads to more children classified as *not* having progressed a grade, compared to the variable used in Table H.8. Further, this difference is larger in the control group than in the treatment group, making the two groups more similar and likely explaining the lack of significant findings in Chapter VI. We also found a significant positive impact on the grade children attended at follow-up, but not when we assigned a grade equal to zero to children who were out of school at follow-up (the two versions of this variable are included in Table H.8). Finally, there was a small negative impact on children's self-report of academic engagement, but it is significant only at the 10 percent level.

Table H.8. Children's attachment to school in small communities

	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Attendance (percentage of days present in the last month) (missing if child was not enrolled)	87.6	86.1	1.5	0.444	0.06	932
Progression since recruitment (percentage of children) (missing if not enrolled at intake or follow-up)	89.8	83.7	6.1**	0.032	0.33	928
Progression (number of grades child progressed) (missing if not enrolled at intake or follow-up)	1.6	1.3	0.3***	0.000	0.26	928
Grade at follow-up (missing if child was not enrolled)	3.1	2.9	0.3***	0.001	0.16	934
Grade at follow-up (zero if child was not enrolled)	2.5	2.3	0.1	0.250	0.07	1,172
Academic engagement and community support for academic goals	2.5	2.5	-0.1*	0.050	-0.10	1,164
Number of children	**593**	**579**				
Number of communities	**74**	**73**				

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was over age for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. The progression and number of grades progressed variables are missing for children who were missing grade information at intake or follow-up. Academic engagement refers to children's connection or involvement with academic activities, goals, and values. It includes children's perception of support for academic goals from community members (Skinner, Kindermann, & Furrer, 2009).

*Difference in group means is statistically significant at the .10 level.

***Difference in group means is statistically significant at the .01 level.

D. Impacts on children's literacy

In addition to examining whether EpC led to improved children's decoding, reading fluency and reading comprehension, we explored whether it had an impact on additional measures of reading skills, including oral comprehension, a combined measure of the 3rd-grade-level and 5th-grade-level reading comprehension subtasks, the percentage of children who were unable to correctly answer at least one question in each of the literacy assessment subtasks, and children's reading habits.

EpC did not have a positive impact on children's oral comprehension. Children who were unable to decode were administered the oral comprehension subtask. In both groups, these children were able to answer only about half of the questions correctly about the passage they heard.

EpC had an impact on reading comprehension for children in small communities when using alternate measures of the constructs. These results are consistent with those presented in chapter VI, when using the 3rd- and 5th-grade level measures separately. Children in the treatment group answered 40 percent of the questions (or 4 questions out of 10) correctly, whereas children in the control group answered 35 percent of the questions (or 3.5 questions out of 10) correctly.

More EpC children than control children achieved a score above zero in the 3rd-grade level and 5th-grade level reading comprehension subtasks. Even though many children in both groups were still unable to answer a single question correctly in the reading comprehension subtasks (29 to 40 percent), we found a difference of 6 percentage points between the two groups in the percentage of children who had a score above zero in the 3rd-grade-level reading comprehension subtask. This difference was statistically significant only at the 10 percent level. On 5th-grade-level reading comprehension, the percentage of children with a score above zero was 6 percentage points higher than children in the control group. This difference was statistically significant at the 5 percent level.

Table H.9. Impacts on additional constructs of children's literacy in small communities

	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Oral comprehension (percentage correct questions, out of 5)	46.6	47.3	-0.7	0.827	-0.02	342
Reading comprehension (combined) (percentage correct questions, out of 10)	39.9	35.3	4.6**	0.010	0.13	1,162
Pseudo-word decoding: percentage achieving a score above zero	72.9	68.9	4.0	0.135	0.12	1,164
Oral comprehension: percentage achieving a score above zero	92.2	89.7	2.5	0.514	0.18	342
Reading fluency, words per minute: percentage achieving a score above zero	72.8	68.6	4.3	0.110	0.13	1,162
Reading comprehension, 3rd-grade-level: percentage achieving a score above zero	70.9	65.5	5.5*	0.061	0.15	1,162
Reading comprehension, 5th-grade-level: percentage achieving a score above zero	66.3	60.2	6.1**	0.031	0.16	1,162
Reading comprehension, 3rd-grade-level and 5th-grade-level combined: percentage achieving a score above zero	72.0	67.0	5.0*	0.078	0.14	1,162
Number of children	**589**	**575**				
Number of communities	**74**	**73**				

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was over age for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. The oral comprehension task was administered only to children who were unable to decode. Sample sizes vary due to item-level nonresponse.

*Difference in group means is statistically significant at the .10 level.

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

E. Impacts on children's social-emotional skills and attitudes

Preliminary analysis of Follow-up 1 data indicated that the measures of self-esteem and intercultural competence were not internally reliable for our sample (see Appendix G). To improve internal reliability, we included a new measure of self-esteem and wrote two new items to assess intercultural competence in Follow-up 2. The findings using these measures were similar to those using the measures from Follow-up 1.

EpC had no impacts on children's self-esteem and intercultural competence, as measured in Follow-up 2. The treatment-control differences were less than 0.04 standard deviations for both outcomes.

Table H.10. Impact of EpC on alternative measures of children's social-emotional skills and attitudes in small communities (in Cohort 2)

Social-emotional skills and risk factors (higher score is better)	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Self-esteem (range of 0-3) (alternative measure)	2.0	2.0	0.0	0.365	0.04	1,106
Intercultural competence (range of 0- 3) (alternative measure)	2.1	2.1	0.0	0.495	0.03	1,106
Number of children	**563**	**543**				
Number of communities	**70**	**69**				

Source: Follow-up household survey.

Note: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item-level nonresponse. Outcome definitions are included in Appendix G.

F. Impacts on children's primary literacy outcomes in small communities, by base-year literacy skills

Results from this evaluation suggest that EpC may have rapidly improved decoding skills for children with or little prior exposure to school at intake (specifically out-of-school children and first graders), but not for children who were enrolled in school before EpC (see Table VIII.3). To explore whether children's initial ability moderated the impact of EpC, we tested whether the program was more effective for children who had no letter and word decoding skills at baseline compared to children who were able to decode at least one letter or word. We only find positive impacts of EpC for children who were unable to identify a single letter at baseline and no impacts for those who were able to identify at least a letter correctly. There were no significant impacts on children who were unable (or able) to read a single word correctly.

Table H.11. Impact of EpC on children's primary literacy outcomes in small communities, by base-year literacy skills

	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Children unable to identify a letter in the base-year assessment						
Pseudo-word decoding (# correct per minute)	3.8	1.9	1.8**	0.046	0.24	220
Reading fluency (# correct words per minute)	8.3	4.6	3.7*	0.076	0.21	219
Reading comprehension (combined) (percentage correct out of 10)	10.1	5.1	5.0*	0.052	0.26	219
Number of children	**101**	**119**				
Number of communities	**59**	**59**				

	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Children unable to identify a word in the base-year assessment						
Pseudo-word decoding (# correct per minute)	6.9	5.7	1.2	0.315	0.10	375
Reading fluency (# correct words per minute)	14.6	13.0	1.6	0.524	0.06	374
Reading comprehension (combined) (percentage correct out of 10)	16.6	14.6	2.0	0.485	0.08	374
Number of children	**173**	**202**				
Number of communities	**64**	**66**				
Children able to identify at least one letter in the base-year assessment						
Pseudo-word decoding (# correct per minute)	23.9	22.6	1.3	0.251	0.09	700
Reading fluency (# correct words per minute)	54.4	52.1	2.3	0.368	0.07	700
Reading comprehension (combined) (percentage correct out of 10)	50.6	46.7	3.9	0.088	0.13	700
Number of children	**367**	**333**				
Number of communities	**70**	**69**				
Children able to identify at least one word in the base-year assessment						
Pseudo-word decoding (# correct per minute)	27.2	25.4	1.8	0.166	0.12	545
Reading fluency (# correct words per minute)	62.3	59.1	3.1	0.304	0.10	545
Reading comprehension (combined) (percentage correct out of 10)	56.5	51.4	5.2	0.041	0.18	545
Number of children	**295**	**250**				
Number of communities	**69**	**69**				

Source: Base-year and Follow-up household surveys.

Note: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item-level nonresponse.

*Difference in group means is statistically significant at the .10 level.

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

We also examined the alternative hypothesis that EpC was more effective for younger children. However, we do not find consistent evidence in support of differential impacts by age (Table H.12). We found a positive effect for the older cohort but not the younger cohort on invented word decoding; and a positive effect for the younger cohort, but not the other cohort, on reading fluency. In both cases, the effects are significant at the 10 percent level.

Table H.12. Impact of EpC on children's primary literacy outcomes in small communities, by age at intake

Child primary outcomes	Impact for children between 5-7 years at intake (A)	*p*-value	Impact for children between 8-15 years at intake (B)	*p*-value	Difference in impacts by age (*p*-value)
Attendance (percentage of days present in the last month) (0 if not enrolled) (pp)	-5.1	0.228	1.1	0.774	0.222
Pseudo-word decoding (# correct per minute)	2.2	0.124	2.0*	0.070	0.928
Reading fluency (# correct words per minute)	5.6*	0.064	3.9	0.125	0.643
Reading comprehension (combined) (percentage correct out of 10) (pp)	6.3**	0.042	3.9*	0.077	0.505
Social competence (range 0 to 3)	0.1	0.240	0.0	0.593	0.476
Moral disengagement (range 0 to 3)	0.0	0.585	0.0	0.768	0.552
Number of children	**357**		**813**		
Number of communities	**147**		**147**		

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level. Impacts on attendance and reading comprehension are in percentage points (pp). All regressions include weights to account for oversampling of out-of-school children. Regressions include controls for children's characteristics. Sample sizes for different outcomes vary due item-level nonresponse.

*Difference in group means is statistically significant at the .10 level.

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

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

APPENDIX I:

ADDITIONAL FINDINGS ON CHILD OUTCOMES FOR ALL COMMUNITIES

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

We report here additional impacts on children's outcomes in the full sample, to complement the results presented in Chapter VII. We do not present additional implementation findings or impacts on intermediate outcomes because it is not possible to distinguish EpC and control communities when all communities are included in the estimation model.

A. Impacts on children's attachment to school

This section describes findings from additional indicators of school attachment. Specifically, we examine impacts on enrolment, the proportion of children who advanced a grade since recruitment, the number of grades children advanced, children's grade at endline, and children's self-report of academic engagement. As we did for small communities in Appendix H, we also present findings on attendance and grade progression using different approaches for handling missing data.

Consistent with the findings for small communities, we find no impact of EpC on children's school attendance when data are missing for children who were not enrolled in school. We also find a statistically significant difference on the proportion of children who advanced a grade since recruitment when we focus on children with complete data, but not when we replace missing data. Children in EpC advanced slightly more in the number of grades progressed than children in the control group, regardless of the variable used. Finally, we find no impacts of EpC on children's school enrolment or academic engagement. Eighty-six percent of the children in the treatment group were enrolled in school, as were 87 percent of the children in the control group. This level of enrollment is about the same as it was at the time of intake (84 and 85 percent respectively, as shown in Table IV.5).

Table I.1. Children's attachment to school in all communities

	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Attendance (percentage of days present in the last month) (missing if child was not enrolled)	89.5	88.6	0.9	0.394	0.05	2026
Enrollment (percentage of children enrolled)	85.7	86.6	-0.9	0.593	-0.05	2,360
Progression since recruitment (percentage of children) (missing if no data are available at intake or follow-up)	91.9	87.9	4.0**	0.030	0.27	2014
Progression since recruitment (percentage of children) (1 if grade information is missing but child was out of school at intake and enrolled at follow-up)	78.9	76.2	2.7	0.230	0.09	2,355
Progression (number of grades child progressed) (missing if no data are available at intake or follow-up)	1.6	1.5	0.1***	0.002	0.17	2014
Progression (number of grades child progressed) (1 if grade information is missing but child was out of school at intake and enrolled at follow-up)	1.4	1.3	0.1**	0.022	0.11	2,355
Grade at follow-up (0 if child was not enrolled)	2.8	2.7	0.1	0.191	0.05	2360
Grade at follow-up (missing if child was not enrolled)	3.3	3.1	0.1***	0.004	0.10	2030

	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Academic engagement and community support for academic goals	2.5	2.5	0.0	0.494	-0.03	2347
Number of children	**1,276**	**1,071**				
Number of communities	**126**	**125**				

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was over age for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Academic engagement refers to children's connection or involvement with academic activities, goals, and values. It includes children's perception of support for academic goals from community members (Skinner, Kindermann, & Furrer, 2009).

*Difference in group means is statistically significant at the .10 level.

***Difference in group means is statistically significant at the .01 level.

B. Impacts on children's literacy

In addition to examining whether EpC led to improved children's decoding, reading fluency and reading comprehension, we explored whether it had an impact on additional measures of reading skills, including oral comprehension, a measures of the 3rd-grade-level and 5th-grade-level reading comprehension subtasks, the percentage of children who were unable to correctly answer at least one question in each of the literacy assessment subtasks, and children's reading habits. We did not observe impacts on any of these outcomes, except for reading comprehension.

EpC had a positive impact on 5th-grade level reading comprehension, but it was significant only at the 10 percent level. On average, children in the treatment group answered 37 percent of the questions (or about 1.9 questions) about the fifth-grade level reading passage correctly, whereas children in the control group answered 34 percent of the questions (or 1.7 questions) correctly. This difference reflects a small effect size of 0.07 standard deviations and is significant only at the 10 percent level.

Table I.2. Impacts on additional constructs of children's literacy in all communities

	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Oral comprehension (percentage correct questions, out of 5)	48.2	47.6	0.7	0.808	0.02	601
Reading comprehension (third-grade level) percentage correct out of 5)	44.1	42.2	1.9	0.223	0.05	2,349
Reading comprehension (fifth-grade level) (percentage correct out of 5)	37.0	34.5	2.5*	0.079	0.07	2,349
Pseudo-word decoding: percentage achieving a score above zero	75.2	73.6	1.6	0.419	0.05	2351
Oral comprehension: percentage achieving a score above zero	93.6	88.7	4.8	0.202	0.37	601

	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Reading fluency, words per minute: percentage achieving a score above zero	75.1	73.3	1.8	0.374	0.06	2349
Reading comprehension, 3rd-grade-level: percentage achieving a score above zero	72.8	70.6	2.3	0.311	0.07	2349
Reading comprehension, 5th-grade-level: percentage achieving a score above zero	65.8	63.4	2.3	0.270	0.06	2349
Reading comprehension, 3rd-grade-level and 5th-grade-level combined: percentage achieving a score above zero	73.9	72.0	1.9	0.373	0.06	2349
Child reads at home alone (percentage) (caregiver report)	80.9	80.0	1.0	0.582	0.04	2,359
Number of days child read at home in last 7 days (caregiver report)	2.9	2.9	0.0	0.829	-0.01	2,346
Number of children	**1,280**	**1,071**				
Number of communities	**126**	**125**				

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was over age for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. The oral comprehension task was administered only to children who were unable to decode. Sample sizes vary due to item-level nonresponse.

*Difference in group means is statistically significant at the .10 level.

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

C. Impacts on children's social-emotional skills and attitudes

In Chapter VII we examine impacts of EpC on children's social competence and moral disengagement, the primary measures of social-emotional skills and risk factors for engaging in illicit behavior, respectively. Here we examine the impacts of EpC on secondary measures of children's social-emotional skills and risk factors. We also explore whether EpC had impacts on alternative measures of self-esteem and intercultural competence, because preliminary analysis of Follow-up 1 data indicated that self-esteem and intercultural competence were not measured reliably in our sample (see Appendix G). To improve internal reliability, we included a new measure of self-esteem and wrote two new items to assess intercultural competence in Follow-up 2.

We find that EpC had no impacts on the secondary measures of social-emotional skills and risk factors. There are no statistically significant differences between the two groups in any of the secondary measures of social-emotional skills and attitudes or children's risk factors. In addition, we find no significant differences between the groups when using alternative measures of self-esteem and intercultural competence. In all instances, the treatment-control differences were all close to zero standard deviations (Table I.3).

Table I.3. Impact of EpC on additional measures of children's social-emotional skills and attitudes in all communities

Social-emotional skills and risk factors	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Skills (higher score is better)						
Self-esteem (range 1 to 5)	4.0	4.0	0.0	0.921	0.00	2,331
Intercultural competence (range 0 to 3)	2.1	2.0	0.0	0.209	0.05	2,349
Self-esteem (range of 0-3) (Cohort 2 only)	2.1	2.0	0.0	0.408	0.04	1248
Intercultural competence (range of 0- 3) (Cohort 2 only)	2.1	2.1	0.0	0.426	0.04	1248
Risk factors (lower score is better)						
Impulsive risk taking (range 0 to 3)	1.1	1.1	0.0	0.723	0.02	2,349
Attitudes towards delinquency (range 0 to 3)	0.8	0.8	0.0	0.812	0.01	2,349
Bullying or peer victimization (range 0 to 3)	0.8	0.7	0.0	0.834	0.01	2,349
Number of children	**636**	**612**				
Number of communities	**76**	**75**				

Source: Follow-up household survey.

Note: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item-level nonresponse. Outcome definitions are included in Appendix G.

D. Impacts on additional outcomes of interest: safety perceptions and time use

EpC had an impact on children's perceptions of safety, but not on their caregivers' perceptions. Over two thirds (69.3 percent) of the children in EpC said they felt safe in their community, which is 4 percentage points higher than the result for children in the control group (65.2 percent). This difference is equivalent to 0.11 standard deviations and is significant only at the 10 percent level. About half of the children in EpC communities also reported that they feel safe walking alone at night, which is 6 percentage points higher than the result for children in control communities (and equivalent to 0.15 standard deviations). There are no other statistically significant differences between the groups in children's sense of safety at school or going to school, but most children reported feeling safe in those contexts. Also, there were no differences between the groups in caregivers' perceptions of safety (Table I.4).

Table I.4. Impact of EpC on perceptions of community safety in all communities

Perceptions of safety	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Community is very safe or safe (percentage) (child report)	69.3	65.2	4.1*	0.097	0.11	2,354
Child feels very safe or safe walking alone at night (percentage)	52.8	46.7	6.1**	0.012	0.15	2,352
Child feels safe at school (percentage)	92.9	92.6	0.3	0.815	0.03	2,331
Child feels safe going to school (percentage)	86.2	87.1	-0.8	0.637	-0.04	2,333
Community is very safe or safe (percentage) (caregiver report)	62.1	62.6	-0.5	0.830	-0.01	2,349
Caregiver feels very safe or safe walking alone at night (percentage)	36.3	34.8	1.5	0.525	0.04	2,349
Number of children	**1,279**	**1,075**				
Number of communities	**126**	**125**				

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item nonresponse.

***Difference in group means is statistically significant at the .01 level.

The evaluation revealed an impact of EpC on children's participation in academic reinforcement but no impacts on child labor or other after-school activities. Children in the EpC group were 3 percentage points more likely to participate in academic reinforcement than children in the control group. Children in both groups engaged in few other after- (or before-) school educational or recreational activities and also engaged in some form of work for an average of 8 hours per week (Table I.5).

Caregivers of EpC children were more likely to participate in EPM (parent schools) than caregivers of control group children were, on average. Caregivers of EpC group children were 11 percentage points more likely to report having participated in EPM, which were supported by CARS in EpC communities (control group caregivers do participate in EPM because they are available in all public schools) (Table I.5). The level of parent participation could have influenced children's participation in EpC, school, and reading.

Table I.5. Impact of EpC on time use in all communities

Other child outcomes	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Non-EpC activity child attended before or after school (percentage)						
Academic reinforcement	7.8	4.7	3.1**	0.029	0.33	2,360
Sports/recreation	4.6	4.2	0.4	0.657	0.06	2,360
Art/music/theater	2.2	1.8	0.4	0.536	0.14	2,360
Religious activities	0.9	1.1	-0.2	0.664	-0.13	2,360
Other	1.4	0.8	0.6	0.177	0.35	2,360
Number of hours child spent in non-educational or non-recreational activities (e.g., labor) in the last week	8.2	8.3	-0.1	0.892	-0.01	2,356
Child's caregiver participated in EPM (percentage)	36.9	25.6	11.3***	0.000	0.32	2,193
Number of EPM meetings attended by child's caregiver in 2017	0.8	0.6	0.2***	0.002	0.15	2,190
Number of children	1,284	1,076				
Number of communities	**126**	**125**				

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item nonresponse.

***Difference in group means is statistically significant at the .01 level.

APPENDIX J:

FINDINGS ON EDUCATOR AND CHILD OUTCOMES FOR LARGE COMMUNITIES

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

We report here findings on implementation outcomes, intermediate outcomes and child outcomes in large communities. Results in this section should be interpreted with caution because our ability to detect the true impacts of EpC in large communities was compromised by lower than expected levels of compliance with the evaluation design (see Chapter V for a detail description of differences in compliance between small and large communities).

A. Implementation findings

In Chapter V we compare EpC facilitators and teachers who were not facilitators on whether they participated in training in teaching, reading, writing and the use of assessments in the year of randomization and one year later (Table V.4). Here we examine differences in training between these groups two years after randomization and before the year of randomization.

We found that two years after randomization, the percentage of EpC facilitators who reported having participated in trainings to teach reading and writing is 26 percentage points higher than teachers only. Yet, not all facilitators reported receiving training two years after randomization and nearly a third of teachers who were not EpC facilitators also reported having participated in trainings to teach reading and writing two years after randomization (Table J.1). These somewhat counter-intuitive findings may be due to the timing of data collection, which made it difficult to locate some of the facilitators for the cohorts included in the evaluation[42]. In some cases, we surveyed teachers in control communities who may have been selected and trained to be facilitators for later EpC cohorts.

Table J.1. Educator training in large communities, by educator type

Educator training	EpC facilitators	Teachers – only	Difference (A–B)	*p*-value
Received reading/writing training (percentage)				
Trained two years after randomization	58.2	31.8	26.4***	0.004
Trained before the year of randomization	1.4	8.2	-6.8***	0.010
Received standardized child assessment training (percentage)				
Trained two years after randomization	35.6	17.4	18.2**	0.007
Trained before the year of randomization	1.4	9.1	-7.7**	0.011
Number of educators	**67**	**110**		
Number of communities	**52**	**50**		

Source: Follow-up educator survey.
Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). The group of EpC facilitators includes teacher-facilitators and facilitators. Errors are clustered at the community level.

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

***Difference in group means is statistically significant at the .01 level.

[42] In 44 percent of Cohort 1 communities (n=33), the EpC closed in late 2015, while the remaining 56 percent closed in June of 2016. However, we were unable to interview educators who had been facilitators in 2015. Seventy eight percent of those interviewed (twenty five percent from Cohort 1 and fifty three percent from Cohort 2) had been facilitators in 2016 and the remaining had only started their role in 2017.

B. Impacts on intermediate outcomes

We found that EpC facilitators (teacher-facilitators and facilitators-only) reported spending about as much time managing student behavior as teachers who were not facilitators did. This differs from findings in small communities, where EpC facilitators spent more time managing student behavior than teachers in the control group did. As in small communities, we found no statistically significant differences in large communities between EpC facilitators and teachers' reports on how often student misbehavior interrupted classroom activities (see Table J.2).

Table J.2. Classroom behavior management in large communities, by type of educator

Behavior management	EpC facilitators (A)	Teachers-only (B)	Difference (A–B)	*p*-value
Time spent managing misbehavior (0 = most of the time; 4 = none of the time)	1.5	1.4	0.1	0.443
Educator reports that misbehavior interrupts classroom activities (0 = most of the time; 3 = almost never)	1.7	1.6	0.1	0.452
Number of educators	**67**	**110**		
Number of communities	**52**	**50**		

Source: Follow-up educator survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level.

Overall, most educators in both groups (EpC facilitators and teachers) reported that lack of support from parents and children's absenteeism and lack of motivation were barriers to effective teaching. There were few statistically significant differences between EpC facilitators and teachers regarding their perceptions of barriers to teaching. EpC facilitators reported fewer overall barriers than teachers did, and this difference is statistically significant at the 5 percent level. EpC facilitators were also less likely than teachers to report as barriers insufficient training in intercultural bilingual education and students' lack of understanding the language of instruction. These differences were statistically significant at the 5 and 1 percent level, respectively. In addition, there was a difference of 15 percentage points in teachers regarding their limited knowledge of the local language as a barrier to instruction. This difference was significant at the 10 percent level only. These results align with the EpC model, which was designed to adapt learning and teaching materials to the language of the RACCS.

Table J.3. Educators' perceptions of barriers to teach in large communities, by type of educator

	EpC facilitators (A)	Teachers-only (B)	Difference (A–B)	*p*-value
Number of perceived obstacles to teaching	4.7	5.5	-0.8**	0.017
Percentage of educators reporting the following perceived barriers				
Small amount of teaching or classroom time	48.8	51.2	-2.4	0.410
Absence of training in intercultural bilingual education	51.8	69.8	-18.0**	0.032
Teachers have little knowledge of local language	39.0	56.4	-17.4*	0.056
Absence of support from parents	75.0	83.6	-8.6	0.286
Frequent student absences	70.2	75.5	-5.3	0.210
Absence of student motivation	67.0	70.1	-3.0	0.905
Students don't understand the language of instruction	35.9	56.9	-20.9***	0.010
Parents don't help with school activities	79.6	84.6	-5.0	0.638
Number of educators	**64**	**110**		
Number of communities	**51**	**50**		

Source: Follow-up educator survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level.

*Difference in group means is statistically significant at the .10 level.

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

***Difference in group means is statistically significant at the .01 level.

B. Impacts on children's attachment to school

In large communities, the evaluation did not find evidence of EpC impacts on children's school attendance, enrollment, grade progression or academic engagement. As shown in Table J.4, children in both groups had similar levels of attendance, 92 percent were enrolled in both groups, a similar percentage of students progressed to a higher grade since recruitment, and students reported similar levels of academic engagement.

Table J.4. Impact of EpC on child attachment to school in large communities

Child primary outcomes	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Attendance (percentage of days present in the last month) (missing if child was not enrolled)	91.2	90.7	0.5	0.621	0.03	1,094
Attendance (percentage of days present in the last month) (0 if not enrolled)	84.1	83.7	0.4	0.807	0.01	1,186
Enrollment (percentage of children enrolled)	92.3	92.2	0.1	0.962	0.01	1,188
Progression since recruitment (percentage of children) (missing if no data are available at intake or follow-up)	93.6	92.2	1.4	0.410	0.13	1,086
Progression since recruitment (percentage of children) (1 if grade information is missing but child was out of school at intake and enrolled at follow-up)	86.5	85.1	0.4	0.580	0.07	1,188
Progression (number of grades child progressed) (missing if no data are available at intake or follow-up)	1.6	1.6	0.0	0.726	0.03	1,086
Progression (number of grades child progressed) (1 if grade information is missing but child was out of school at intake and enrolled at follow-up)	1.5	1.5	0.0	0.726	0.02	1,188
Grade at follow-up (missing if child was not enrolled)	3.4	3.4	0.0	0.783	0.01	1,096
Grade at follow-up (0 if child was not enrolled)	3.2	3.1	0.0	0.783	0.02	1,188
Academic engagement and community support for academic goals	2.5	2.5	0.0	0.341	0.06	1,183
Number of children	**691**	**497**				
Number of communities	**52**	**52**				

Source: Follow-up household survey

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item nonresponse; two children in the EpC group are missing attendance data.

C. Impacts on children's literacy

In large communities, the evaluation did not find evidence of EpC impacts on children's ability to decode, read fluently, or comprehend what they read. Students in both groups correctly read an average 20 pseudo-words in one minute (Table J.5). They had similar fluency results, reading about 40 words correctly on average. Also, students in both groups answered 46 percent of questions correctly about the third grade level reading passage and 36 percent of questions correctly of the fifth grade level reading passage.

Table J.5. Impact of EpC on children's literacy skills and behaviors in large communities

Literacy skills	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Primary outcomes						
Pseudo-word decoding (# correct per minute)	20.2	19.8	0.4	0.654	0.02	1,187
Reading fluency (# correct words per minute)	41.2	41.8	-0.6	0.733	-0.02	1,187
Reading comprehension (combined) (percentage correct questions, out of 10)	41.1	42.0	-0.9	0.635	-0.03	1,187
Secondary outcomes						
Oral comprehension (percentage correct questions, out of 5)	51.2	45.8	5.4	0.193	0.20	259
Reading comprehension (third grade level) (percentage correct out of 5)	45.8	47.7	-1.9	0.398	-0.05	1,187
Reading comprehension (fifth grade level) (percentage correct out of 5)	36.3	36.2	0.1	0.963	-0.00	1,187
Pseudo-word decoding: percentage achieving a score above zero	77.2	79.2	-2.0	0.471	-0.07	1,187
Oral comprehension: percentage achieving a score above zero	95.2	85.9	9.4	0.254	0.72	259
Reading fluency, words per minute: percentage achieving a score above zero	77.0	79.0	-2.0	0.477	-0.07	1,187
Reading comprehension, 3rd-grade level: percentage achieving a score above zero	74.5	76.6	-2.1	0.539	-0.07	1,187
Reading comprehension, 5th-grade level: percentage achieving a score above zero	65.0	67.9	-2.9	0.328	-0.08	1,187
Reading comprehension (combined) (percentage achieving a score above zero)	75.7	78.1	-2.5	0.421	-0.08	1,187
Reading comprehension (combined) (percentage correct questions, out of 10)	41.1	42.0	-0.9	0.635	-0.03	1,187
Child reads at home alone (percentage) (caregiver report)	80.7	83.3	-2.5	0.267	-0.10	1,188
Number of days child read at home in last 7 days (caregiver report)	2.8	3.0	-0.1	0.347	-0.05	1,184
Number of children	**691**	**496**				
Number of communities	**52**	**52**				

Source: Follow-up household survey

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item-level nonresponse. The oral comprehension task was administered to only children who were unable to decode.

As in small communities, we explored whether EpC had an impact on alternative measures of reading skills, including oral comprehension, a measure combining the 3rd-grade-level and 5th-grade-level reading comprehension subtasks, and the percentage of children who were unable to correctly answer at least one question in each of the literacy assessment subtasks.

EpC did not have a positive impact on children's oral comprehension. In both groups, children who were unable to decode answered correctly only about half of the questions about the passage they listened to.

There were no impacts of EpC on the alternate measure of reading comprehension. Results for reading comprehension were consistent with those that use the combined measure of 3rd- and 5th- grade level reading passages. Children in both groups answered correctly fewer than half of the reading comprehension questions in each of the passages. As expected, children in both groups performed more poorly on the 5th-grade passage compared to the 3rd-grade passage.

We found no differences in the percentage of children achieving scores above zero across all literacy subtasks. Also, 21 to 36 percent of children across both groups were still unable to answer a single item correctly in all subtasks except for oral comprehension, where more than 86 percent children obtained a score above zero.

D. Impacts on children's social-emotional skills and attitudes

The findings for social-emotional skills and attitudes in large communities were similar to those in small communities.

In large communities, the evaluation did not find evidence of EpC impacts on children's measured social-emotional skills and attitudes. There were no statistically significant differences between the treatment and control groups on children's social competence, our primary measure of social-emotional skills (Table J.6). Results are consistent when we use a longer version of the measure[43]. We also explored and found no impacts of EpC on children's self-esteem and intercultural competence. The treatment-control differences were less than 0.05 standard deviations.

Similarly, the evaluation found no evidence of an impact of EpC on children's skills and attitudes that can protect them from engaging in illicit activities, including moral disengagement, impulsive risk taking, attitudes towards delinquency, and bullying or peer

[43] Preliminary analysis of Follow-up 1 data indicated that the number of items used to measure social competence could be reduced and still be internally consistent for our sample. To reduce the time of administration in Follow-up 2, we eliminated 3 of the 14 items we used in Follow-up 1. The alternative measure uses the 14 items from Follow-up 1.

victimization. Treatment-control differences were each less than 0.06 standard deviations. In both groups, there might be little room for improvement because the levels for these risk factors are low.

Table J.6. Impact of EpC on child social-emotional skills and attitudes in large communities

Social-emotional skills and attitudes	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Skills (higher score is better)						
Social competence (range 0 to 3) (11 items)	2.0	2.0	0.0	0.704	0.02	1,183
Social competence (range of 0-3) (14 items)	2.0	2.0	0.0	0.397	0.06	1,044
Self-esteem (range 1 to 5)	3.8	3.8	0.0	0.955	0.00	1,176
Intercultural competence (range 0 to 3)	2.1	2.0	0.0	0.892	0.01	1,185
Risk factors (lower score is better)						
Moral disengagement (range 0 to 3)	1.2	1.1	0.1	0.158	0.10	1,185
Impulsive risk taking (range 0 to 3)	1.2	1.1	0.0	0.292	0.08	1,185
Attitudes towards delinquency (range 0 to 3)	0.9	0.8	0.0	0.321	0.06	1,185
Bullying or peer victimization (range 0 to 3)	0.9	0.9	0.0	0.810	0.02	1,185
Number of children	**689**	**496**				
Number of communities	**52**	**52**				

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item-level nonresponse. Outcome definitions are included in Appendix G.

Impacts on additional outcomes of interest: safety perceptions and time use

In large communities, we did not find evidence that EpC had an impact on child or caregivers' perceptions of community safety. As shown in Table J.7, the perception of community safety was comparable in both groups. A similar percentage of children in both groups reported feeling safe walking alone at night in their community. We also found no statistically significant differences between the groups in children's sense of safety at school or going to school, but most children reported feeling safe in those contexts. Also, there were no differences between the groups in caregivers' perceptions of safety.

Table J.7. Impact of EpC on perceptions of community safety in large communities

Perceptions of safety	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Community is very safe or safe (percentage) (child report)	52.0	54.2	-2.2	0.545	-0.05	1,187
Child feels very safe or safe walking alone at night percentage)	46.4	41.7	4.7	0.133	0.12	1,187
Child feels safe at school (percentage)	90.9	89.9	1.0	0.700	-0.07	1,176
Child feels safe going to school (percentage)	81.1	84.5	-3.4	0.324	-0.15	1,179
Community is very safe or safe (percentage) (caregiver report)	50.1	55.5	-5.4	0.207	-0.13	1,185
Caregiver feels very safe or safe walking alone at night (percentage)	32.8	31.8	1.0	0.761	0.03	1,185
Number of children	**691**	**496**				
Number of communities	**52**	**52**				

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item nonresponse.

EpC did appear to affect how children spent their time. Children in the EpC group were 6 percentage points more likely to engage in academic reinforcement activities (before or after school) than children in the control group were. This might reflect EpC activities and not other academic reinforcements (Table J.8). There were no differences in participation in labor activities.

Caregivers of children in both groups participated in EPM at similar levels. About one-third of children’s caregivers in both groups reported participating in EPM (Table J.8), a level similar to the one found in small communities.

Table J.8. Impact of EpC on time use in large communities

Other child outcomes	EpC (A)	Control (B)	Impact (A–B)	*p*-value	Effect size	Sample size
Non-EpC activity child attended before or after school (percentage)						
Academic reinforcement	13.4	7.2	6.2**	0.031	0.42	1,188
Sports/recreation	8.1	8.1	0.0	0.998	0.00	1,188
Art/music/theater	2.5	2.4	0.1	0.919	0.03	1,188
Religious activities	1.0	1.4	-0.4	0.670	-0.18	1,188
Other	1.0	1.4	-0.4	0.670	-0.18	1,188
Number of hours child spent in non-educational or non-recreational activities (e.g., labor) in the last week	3.8	4.6	-0.8	0.289	-0.08	1,184
Child's caregiver participated in EPM (percentage)	36.5	33.6	2.9	0.441	0.08	1,099
Number of EPM meetings attended by child's caregiver in 2017	0.9	0.8	0.2	0.245	0.09	1,096
Number of children	**691**	**497**				
Number of communities	**52**	**52**				

Source: Follow-up household survey.

Notes: Estimates presented for large communities. Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for children's gender, maternal language, age, grade at recruitment, and whether the child was overage for his/her grade at intake. Errors are clustered at the community level. All regressions include weights to account for oversampling of out-of-school children. Sample sizes vary due to item nonresponse.

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

Impacts by subgroup

We present here the findings for the primary subgroups of interest as defined in the design: gender and out-of-school status at intake.

1. Gender

In large communities, there were no differential impacts by gender, but the intervention may have had a small impact on boys' moral disengagement, because boys in the treatment group are slightly more likely than boys in the control group to endorse beliefs that justify wrongdoing (Table J.9). However, this difference was significant at the 10 percent level only, and the difference between genders was not statistically significant. This finding for large communities is different from what we found with small communities, where impacts on decoding, reading fluency, reading comprehension, and social competence were confirmed for girls but not boys.

Table J.9. Impact of EpC on primary child outcomes in large communities, by gender

Child primary outcomes	Impact for girls (A)	*p*-value	Impact for boys (B)	*p*-value	Difference in impacts by gender (*p*-value)
Attendance (percentage of days present in the last month) (0 if not enrolled) (pp)	0.8	0.661	0.1	0.978	0.836
Pseudo-word decoding (# correct per minute)	1.4	0.300	-0.6	0.610	0.250
Reading fluency (# correct words per minute)	1.6	0.566	-2.6	0.239	0.249
Reading comprehension (combined) (percentage correct out of 10) (pp)	1.2	0.682	-2.9	0.213	0.251
Social competence (range 0 to 3)	0.0	0.273	0.0	0.590	0.243
Moral disengagement (range 0 to 3)	0.0	0.660	0.1*	0.092	0.378
Number of children	**591**		**596**		
Number of communities	**52**		**52**		

Source: Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level. Impacts on attendance and reading comprehension are in percentage points (pp). All regressions include weights to account for oversampling of out-of-school children. Regressions include controls for children's characteristics. Sample sizes for different outcomes vary due item-level nonresponse.

*Difference in group means is statistically significant at the .10 level.

2. School enrollment at intake

In large communities, EpC had a positive impact on moral disengagement for children who were out of school at intake, but no impact for children who were enrolled at intake. Among children who were out of school at intake, moral disengagement was 0.3 percentage points higher for children in the treatment group than in the control group. This difference is statistically significant (Table J.10). Further, the impact for out-of-school children is statistically significantly different from the impact for enrolled children. There were no other differences by enrollment status in large communities, which is different from what we found in small communities, where there were differences on the three primary reading outcomes (that is, decoding, reading fluency, and reading comprehension).

Table J.10. Impact of EpC on primary child outcomes in large communities, by school enrollment status at intake

Child primary outcomes	Impact for children out of school (A)	*p*-value	Impact for children in school (B)	*p*-value	Difference in impacts by enrollment status (*p*-value)
Attendance (percentage of days present in the last month) (0 if not enrolled) (pp)	9.3	0.209	0.0	0.983	0.242
Pseudo-word decoding (# correct per minute)	-0.1	0.963	0.4	0.642	0.843
Reading fluency (# correct words per minute)	-4.3	0.410	-0.4	0.826	0.465
Reading comprehension (combined) (percentage correct out of 10) (pp)	1.9	0.734	-1.1	0.592	0.616
Social competence (range 0 to 3)	0.2	0.444	0.0	0.951	0.451
Moral disengagement (range 0 to 3)	0.3**	0.030	0.0	0.287	0.062
Number of children	**80**		**1107**		
Number of communities	**52**		**52**		

Source Follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level. Impacts on attendance and reading comprehension are in percentage points (pp). All regressions include weights to account for oversampling of out-of-school children. Regressions include controls for children's characteristics. Sample sizes for different outcomes vary due to item-level nonresponse.

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

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

APPENDIX K:

ADDITIONAL DESCRIPTIVE FINDINGS

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

A. Comparison of characteristics of large and small communities

The difference in findings based on community size could be due to the differences in assigning EpC status at the community level versus assigning children within communities. However, the difference in findings could also have been due to differences in community and population characteristics. If characteristics of the two types of communities were remarkably dissimilar, EpC may simply have had different impacts on children in these different contexts. Consequently, we carefully examined the characteristics of our samples in small and large communities.

1. Comparing children and schools in large and small communities

As a group, children in small communities were more disadvantaged than children in large communities. Children in small communities were more likely to be female, out of school at intake (this difference is not statistically significant), and in lower grades; they were less likely to walk to school with another child than children in large communities were. Spanish was more likely to be their maternal language, and they were more likely to receive instruction in their maternal language. The spent almost twice as much time en route to school as children in large communities did, were more likely to walk to school, and less likely to take a bus or car or use a motorcycle. They were also significantly less likely than their counterparts in large communities to attend the afternoon shift at school, and were more likely to have a school-provided meal, both of which may have affected take-up of the EpC intervention. Children in small communities were also much poorer than those in large communities, which placed them at higher risk of school failure and drop out.

Table K.1. Children's characteristics, by community size

Child characteristics	Small communities (A)	Large communities (B)	Difference (A-B)	*p*-value	Sample size
Female (percentage)	56.4	41.3	15.1***	0.008	2,371
Age at intake (years)	8.8	8.3	0.5	0.517	2,371
Parent attended community assembly (percent)	65.7	67.8	-2.1	0.673	1,766
Grade level at intake (percentage)					
Out of school	26.8	4.1	22.7	0.217	2,371
Grade 1	40.7	35.3	5.4	0.488	2,368
Grade 2	22.1	29.9	-7.8	0.646	2,368
Grade 3	9.9	31.1	-21.2***	0.000	2,368
Walks to school with another child (percentage)	20.1	31.5	-11.4***	0.000	2,371
Maternal language (percentage)					
Spanish	84.2	76.3	7.9***	0.004	2,371
English	2.6	2.6	0.0	0.000	2,371
Miskitu	6.2	6.2	0.0	0.000	2,371

Child characteristics	Small communities (A)	Large communities (B)	Difference (A-B)	*p*-value	Sample size
Ulwa	-2.9	5.0	-7.9***	0.004	2,371
Kriol	9.3	9.3	0.0	0.000	2,371
Rama-Kriol	0.4	0.4	0.0	0.000	2,371
Maternal language is language of instruction in community (percentage)	91.1	83.2	7.9***	0.004	2,371
Parents live in the same home as child (percentage)	65.8	54.7	11.0	0.532	2,166
Transportation to school (percentage)					
Walks to school	83.6	72.9	10.7**	0.033	2,165
Rides bus/car to school	-3.2	13.0	-16.2***	0.001	2,165
Takes motorcycle to school	-2.8	5.7	-8.5***	0.000	2,165
Takes boat to school	2.1	-0.1	2.1	0.332	2,165
Rides animal to school	0.5	5.3	-4.8	0.467	2,165
Rides a bicycle	1.9	0.3	1.7	0.340	2,165
Time required to get to school (minutes) (children enrolled in school only)	22.5	15.4	7.2**	0.046	1,870
Shift attended (percentage) (children enrolled in school only)					
Morning	91.1	72.2	18.9*	0.051	2,030
Afternoon	8.5	27.4	-18.9**	0.048	2,030
Evening	0.2	0.3	-0.1	0.591	2,030
Saturday (distance education)	0.3	0.1	0.2	0.387	2,030
Child reports that school provides a meal (percentage)	75.0	6.7	68.3***	0.001	1,014
Family poverty	89.7	55.2	34.4***	0.004	2,178
Household asset ownership (percentage)					
Bicycle	30.6	5.5	25.1**	0.014	2,176
Motorcycle	13.5	-2.2	15.7***	0.000	2,176
Car or truck	-3.1	8.7	-11.8**	0.036	2,176
Boat	13.3	11.9	1.4	0.333	2,176
Household migration (Cohort 2 only)					
Adult	24.8	15.6	9.2***	0.010	1,070
Child	2.7	2.7	0.0	0.000	1,019
Number of children	**1,182**	**1,189**			
Number of communities	**147**	**52**			

Source: CARS intake data; Base-year and follow-up household survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Errors are clustered at the community level. Sample sizes vary due to missing data. Child reports that school provides a meal was only collected for Cohort 2, and therefore the data are available for only a few large communities.

***Difference in group means is statistically significant at the .01 level.

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

*Difference in group means is statistically significant at the .10 level.

Anecdotally, CARS and USAID representatives have two compelling potential explanations that help to explain why children in large communities participated less in EpC than in small communities, both of which are supported by these data. First, children in large communities had more alternative options than children in small communities had, including access to internet and TV, as well as activities organized by the school. Second, because children in large communities were more likely to go home for lunch and to rely on public transportation to get back to school, participating in EpC was not as straightforward as in small communities. Children had to travel to school a second time in the day to participate, and therefore probably incurred a higher cost (in transportation) to attend EpC.

We also see differences between large and small communities in outcomes that could have been affected by EpC, such as caregivers' participation in parent schools and children's participation in extra-curricular activities. More caregivers in the control group in large communities reported participation in EPM and caregivers reported more communication with their child's teachers than caregivers in small communities did. The control group's levels in both variables were on par with the treatment group's in large communities (Table K.2). Also, children in large communities participate in more extracurricular non-EpC activities than their counterparts in small communities, and they are more likely to report attending the afternoon shift at school. Greater access to extracurricular activities in large communities could reduce the perceived value of EpC and result in low take-up or attendance. In addition, children spent less time on non-academic activities (for example, working) in large communities than in small communities. One member of the CARS team reported that working was one of the main reasons why older children refused to enroll in EpC in large and small communities.

Table K.2. Other child and caregiver characteristics, by community size

	Small communities			Large communities			Statistical significance of difference between control group means
Other child outcomes	EpC (A)	Control (B)	Sample Size	EpC (A)	Control (B)	Sample Size	
Child's caregiver participated in EPM (percentage)	38.4	18.1	1,094	35.8	34.0	1,099	***
Number of EPM meetings attended by child's caregiver in 2017	0.7	0.3	1,094	0.9	0.8	1,096	***
Caregiver has talked with child's teacher (percentage)	22.6	24.8	1,169	54.9	51.7	1,186	***
Number of times caregiver has talked with child's teacher	0.5	0.6	1,169	2.0	1.8	1,183	***
Parent reports knowing where the child is (percentage)							
Never	0.0	0.3	1,154	0.4	0.5	1,185	
Rarely	0.0	0.1	1,154	2.2	1.1	1,185	**

Other child outcomes	Small communities EpC (A)	Small communities Control (B)	Small communities Sample Size	Large communities EpC (A)	Large communities Control (B)	Large communities Sample Size	Statistical significance of difference between control group means
Half the time	1.1	1.9	1,154	3.4	4.8	1,185	***
Most of the time	14.9	15.4	1,154	19.4	18.0	1,185	*
Always	84.0	82.4	1,154	74.4	75.6	1,185	***
Non-EpC activity child attended before or after school (percentage)							
Academic reinforcement	2.2	1.8	1,172	13.2	7.6	1,188	***
Sports/recreation	1.2	0.3	1,172	8.0	8.3	1,188	***
Art/music/theater	2.0	1.4	1,172	2.5	2.5	1,188	**
Religious activities	0.9	0.8	1,172	1.0	1.1	1,188	
Other	1.1	0.7	1,172	1.6	0.6	1,188	
Number of hours child spent in non-educational or non-recreational activities (e.g., labor) in the last week	12.4	12.4	1,172	3.8	4.7	1,187	***
Number of communities	**74**	**73**		**52**	**52**		

Source: Follow-up household survey.

Notes: Unadjusted mans are reported for all variables. Sample sizes vary due to item nonresponse. T-tests of the difference between the means in the control group in small communities compared to the means in the control group in large communities were conducted.

***Difference in group means is statistically significant at the .01 level.

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

*Difference in group means is statistically significant at the .10 level.

The control group literacy outcomes for small and large communities were similar, but not the same (see Tables IV.7). Given this difference and the fact that the majority of small communities were in rural areas, while large communities were in urban areas, we disaggregated the control group follow-up values by urban/rural status (Table K.3). First, it is clear that control group children in urban areas had higher literacy levels. Second, child literacy in small communities (both urban and rural) was similar to child literacy in large rural communities. Third, children in large urban communities had better literacy skills than children in other areas. The effect of the EpC intervention on children in small communities brought their skills closer to the level of children in urban large communities.

Table K.3. Unadjusted literacy outcomes for children in the control group, by community size

	All communities		Medium communities		Large communities	
Literacy Outcome	Urban	Rural	Urban	Rural	Urban	Rural
Pseudo-word decoding (# correct/minute, out of 50)	20.8	16.1	13.6	16.0	21.5	16.3
Reading fluency (# correct words per minute, all languages)	46.1	35.3	29.6	37.0	47.7	30.5
Reading comprehension, 3rd grade level (percentage correct out of five)	50.7	38.0	32.7	37.7	52.5	38.7
Reading comprehension, 5th grade level (percentage correct out of five)	38.7	32.5	24.7	33.5	40.1	29.6
Reading comprehension (combined) (percentage correct out of 10)	44.7	35.2	28.7	35.6	46.3	34.1
Number of children	**332**	**743**	**30**	**549**	**302**	**194**
Number of communities	**34**	**91**	**4**	**69**	**30**	**22**

Source: Follow-up household survey.

Notes: Unadjusted means are presented for each variable for each sample. Sample sizes vary due to missing values.

Schools in large and small communities were similar on most characteristics, but differed in some important ways. Schools in small communities were 54.8 percentage points more likely to be located in a rural area (Table K.4). This is reflected in the amount of time children report taking to get to school and the more infrequent use of public transportation.

Schools in small communities were 34 percentage points more likely to have multi-grade classrooms than schools in large communities were. Most schools in the small communities had only multi-grade classrooms. Schools in small communities also had fewer students, and were less likely to offer bilingual instruction. Small and large communities were similar in a range of characteristics, including the number of students per teacher, student attendance rates, and the materials classrooms were made of. Schools in small communities were, however, open 3.2 fewer days per month than schools in large communities were.

All the schools in the evaluation lacked some basic infrastructure. However, schools in small communities were 25 percentage points less likely to have classrooms with enough furniture for students, but 31 percentage points more likely to have functioning toilets segregated by gender—though these differences were only significant at the 10 percent level. Over half of schools in both large and small communities received support from external programs other than EpC. However, schools in small communities were significantly less likely than their counterparts in large communities to have received textbooks and other learning materials. Also, the majority of schools in the sample had a feeding program in the year of data collection, but schools in small communities were less likely to offer lunch than were schools in large communities, though this difference was significant at the 10 percent level only. Teachers in large communities had more years of teaching experience on average, but the difference was not statistically significant.

Table K.4. School characteristics at end line, by community size

Characteristic	Small communities (B)	Large communities (A)	Difference (A-B)	*p*-value
School is public (percentage)	89.5	81.6	7.8	0.527
School is in a rural area (percentage)	94.7	39.9	54.8***	0.000
School has multi-grade classrooms only (percentage)	88.9	54.5	34.4***	0.000
Student-teacher ratio	29.9	33.6	-3.7	0.402
School enrollment two years after randomization (in year of data collection) (number of students)	63.3	183.8	-120.5***	0.001
School attendance (percentage of students present on the day of data collection out of the total number of enrolled students)	79	80	-1.0	0.846
Average number of days per month the school was open two years after randomization	13.9	17.1	-3.2***	0.007
School is bilingual (percentage)	-3.0	35.5	-38.6***	0.000
Classrooms made of rudimentary materials	54.2	47.0	7.2	0.609
Classrooms without blackboard or flip chart	97.7	95.9	1.8	0.739
Classrooms that can't be used under extreme climate	20.5	39.3	-18.9	0.263
Classrooms without enough desks/tables/chairs for students	89.5	64.6	25.0*	0.056
There are no functioning toilets	8.4	16.6	-8.2	0.475
Toilets work but do not have locks/doors	26.1	51.2	-25.1	0.174
Toilets work and have doors/locks but are not segregated by gender	28.2	26.0	2.2	0.901
Toilets work, have doors/locks, and are segregated by gender	37.3	6.2	31.1*	0.076
School has an active external program two years after randomization (in the year of data collection) (percentage)	52.3	69.4	-17.1	0.325
External program offers textbooks/materials (percentage)	29.9	67.5	-37.7**	0.049
External program offers other, non-EpC support or resources (percentage)	45.4	73.6	-28.2	0.111
Has a meal program two years after randomization (in year of data collection) (percentage)	97.1	94.8	2.3	0.759
Breakfast	6.5	14.8	-8.4	0.477
Snack	89.7	80.8	8.9	0.513
Lunch	2.6	21.9	-19.3*	0.078

Characteristic	Small communities (B)	Large communities (A)	Difference (A-B)	*p*-value
Educator's teaching experience (years)	25.0	16.6	8.4	0.556
Fewer than 3	33.1	10.1	23.1	0.118
3-5	21.3	33.0	-11.6	0.426
6-10	20.5	40.3	-19.8	0.143
More than 10	25.0	16.6	8.4	0.556
Number of communities	**147**	**52**		

Source: Follow-up school director survey; CARS administrative data.

Notes: School attendance is based on counts of students for Cohort 1 and school director reports for Cohort 2. Average number of days per month the school was open in the year of data collection is calculated using data from February to April for Cohort 1 and February to June for Cohort 2. Average number of days per month the school was open in the year before data collection is calculated using data from February to November. Sample sizes vary due to missing data and nonresponse (for example, a number of school directors did not have records of the number of days the school was open).

***Difference in group means is statistically significant at the .01 level.

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

*Difference in group means is statistically significant at the .10 level.

2. Comparing the intervention in large and small communities

In addition to the relative disadvantage of children and schools in small communities compared to large communities, the difference in findings based on community size could also be due to differences in implementation. We therefore compared implementation of the EpC intervention across community sizes.

Overall, large communities had more EpC-related materials than small communities did. Even though only one of nine comparisons was statistically significant at the 5 percent level (availability of art supplies), there were more materials in large communities (Table K.5). This may be because many small communities are difficult to reach (due to the geography in the region) and therefore have less access to materials than large communities did. The lack of statistical significance might be due to small sample sizes. However, there is a statistically significant 55 percentage point difference between large and small communities with EpC in the availability of books that are not textbooks. We can see that in small communities, non-textbook reading materials were distributed at relatively the similar level as music players, but lower level than music instruments, for example. This suggests that not only are there difficulties in accessing small communities, but there also may be some interaction between access and the nature of the materials (i.e., reading materials are heavier).

Table K.5. School resources in schools with EpC, by community size

Materials (percentage)	Large communities (A)	Small treatment communities (B)	Difference (A-B)	*p*-value
Availability of arts, sports, and music materials				
Music instruments	62.0	64.5	-2.5	0.925
Music player	34.3	38.1	-3.8	0.888
Sports equipment	34.2	40.9	-6.7	0.774
Art supplies	72.2	43.9	28.4	0.294
Paper supplies	74.0	54.8	19.2	0.486
Availability of (non-textbook) reading materials	90.6	35.0	55.7**	0.021
Availability of textbooks				
Of shared use	35.3	50.9	-15.5	0.536
Of exclusive use	19.8	3.6	16.2	0.332
Students allowed to take books home	32.7	23.0	9.7	0.686
Number of communities	**52**	**74**		

Source: Follow-up school director survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Sample sizes vary due to nonresponse.

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

The distribution of print materials appropriate to the RACCS did not happen until 10 months into the implementation of Cohort 1A and six months into the implementation for Cohort 1B, both of which are primarily made up of large communities. Therefore, children in Cohort 1 did not receive the full package of the EpC intervention, whereas children in Cohort 2 did. This may result in lower-than-expected impacts on children’s literacy skills overall, and fewer impacts on literacy for children in Cohort 1 compared to Cohort 2. Given that Cohort 1 is mostly large communities, and Cohort 2 is mostly small communities, this could be a factor in explaining the difference in impacts we see by community size.

EpC facilitators received similar training in both large and small communities. We found no differences between large and small communities in EpC facilitators’ reports of having participated in trainings to teach reading/writing and the use of standardized child assessments, nor in the timing of such trainings (Table K.6). Although on average facilitators in large communities had participated in 2.4 more trainings to teach other topics compared to facilitators in small communities, this difference is not significant at the 5 percent level (it is significant at the 10 percent level). Small communities had fewer EpC facilitators who did not also teach, with 52 percent of EpC facilitators being facilitators-only. In large communities, this figure was 67 percent.

However, the amount of training received by teachers who were not EpC facilitators differed depending on the size of their community. Overall, teachers in large communities were more likely to have participated in trainings to teach reading/writing and the use of standardized assessments than teachers in small communities were (Table K.6). The largest differences are for trainings on how to teach reading/writing that took place two years after randomization (23 percentage points) and trainings on how to use standardized assessments that took place one and two years after randomization (15 to 16 percentage points, respectively). The fact that teachers who were not facilitators participated in trainings on topics related to the EpC intervention, especially in large communities, may be further evidence of contamination between treatment and control groups, resulting in a weaker contrast between experimental groups that would hamper the evaluation's ability to detect EpC impacts.

Table K.6. Educator training, by community size

Educator training	Large communities (A)	Small communities (B)	Difference (A-B)	*p*-value
EpC facilitators				
EpC facilitator received reading/writing training (percentage)	87.5	80.9	6.6	0.804
EpC facilitator received standardized child assessment training (percentage)	75.3	73.4	2.0	0.940
EpC facilitator had participated in trainings on other topics since the year of randomization	100.0	100.0	0.0	N/A
Number of other topics in which EpC facilitator had been trained since the year of randomization	7.6	5.2	2.4*	0.056
Teachers-only				
Teacher received reading/writing training (percentage)	88.4	58.6	29.9**	0.033
Teacher received standardized child assessment training (percentage)	54.2	14.0	40.2***	0.000
Teacher had participated in trainings on other topics since the year of randomization	95.4	86.3	9.0	0.479
Number of other topics in which teacher had been trained since the year of randomization	4.6	3.7	-0.9	0.408
Number of EpC facilitators	**67**	**92**		
Number of teachers-only	**110**	**160**		
Number of communities	**52**	**123**		

Source: Follow-up educator survey.

Notes: Columns A and B present ordinary least squares regression-adjusted group means and include covariates to account for the design (cohort and strata variables). Regressions also adjust for educators' gender, years of experience, qualifications, number of years at the current school, number of years living in the community, Spanish as maternal language, and whether the educator teaches in his/her maternal language. Errors are clustered at the community level. Column A–B presents differences in the regression-adjusted group means. Sample sizes vary due to nonresponse.

*Difference in group means is statistically significant at the .10 level.

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

***Difference in group means is statistically significant at the .01 level.

Most trained EpC facilitators in both large and small communities reported implementing all measured EpC-specific methodologies. Except for organizing children by grade level and promoting activities that value local traditions, which were significantly more common in small than large communities, there were no significant differences by community size on facilitators' use of EpC instructional practices (Table K.7.).

Table K.7. EpC facilitator implementation of EpC methodologies, by community size

EPC methodologies	Small communities		Large communities		Statistical significance of difference between group means
	Mean	Standard deviation	Mean	Standard deviation	
Educator used methodology in the year after randomization					
Used didactic materials	97.6	15.4	96.9	17.5	
Split children into separate levels by grade ("*peces*" and "*robles*")	100.0	0.0	95.3	21.3	**
Promoted the child-tutor component	90.4	29.7	89.1	31.5	
Received the boys and girls that arrived at EpC with the typical greeting	100.0	0.0	98.4	12.5	
Promoted the activities of valuing local traditions ("*tradición*")	73.5	44.4	93.8	24.4	***
Promoted motivation of the boys and girls	98.8	11.0	100.0	0.0	
Carried out daily planning	96.4	18.8	98.4	12.5	
Guaranteed that homework assigned by the teacher was completed	98.8	11.0	96.9	17.5	
Provided feedback on the homework completed by the boys and girls	100.0	0.0	100.0	0.0	
Assigned planned activities to boys and girls that don't have homework	97.6	15.4	100.0	0.0	
Motivated the boys and girls using the welcome song	100.0	0.0	100.0	0.0	
Number of EpC facilitators	**83**		**64**		
Number of communities	**73**		**50**		

Source: Follow-up educator survey.

Notes: Unadjusted means are reported for all variables. T-tests of the difference between the means in small communities and the means in large communities were conducted. Sample sizes vary due to item nonresponse. All variables are from one or two years after randomization, and questions about the year after randomization or two years after randomization were only asked of EpC facilitators who were facilitators in each of those years.

***Difference in group means is statistically significant at the .01 level.

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

B. Children's school enrollment status, by age at intake

EpC was designed to serve children who had dropped out of school or had never attended. However, the program had difficulties recruiting children who had dropped out of school. About 40 percent of the children recruited for the evaluation who were classified as "out of school" at intake, were under age 8 (Table K.8). These children may have been out-of-school because their caregivers thought they were too young, or for other reasons that may differ from those of children who were enrolled but dropped out. In fact, only 13 percent of out of school children identified at intake in small communities and 16 percent in large communities were still out of school at follow-up. Moreover, in small communities 83 percent of caregivers of children who were out-of-school at intake indicated that if their child were to enroll in school, they would attend first grade in small communities. In large communities, 65 percent of caregivers reported their child would attend first grade if they were to enroll. Also, CARS staff reported that during intake, older children often refused the option of participating in EpC, leading to their underrepresentation in the evaluation population or sample.

Table K.8. Children's enrollment status at intake and follow up, by age at intake

	Out of school at intake (A)	Out of school at follow up (B)	Percentage out of school at intake (C)	Percentage who remained out of school at follow up (D)
Small communities				
Age 5 to 7	143	19	42.6	13.3
Age 8 to 10	108	38	32.1	35.2
Age 11 or older	85	52	25.3	61.2
Number of children	**336**	**109**		
Number of communities	**116**	**116**		
Large communities				
Age 5 to 7	31	5	38.8	16.1
Age 8 to 10	27	5	33.8	18.5
Age 11 or older	22	13	27.5	59.1
Number of children	**80**	**23**		
Number of communities	**22**	**22**		

Source: Intake data and follow-up household survey. Missing intake age was calculated using base-year or follow-up data.

Notes: Unadjusted means are reported for all variables. Sample sizes vary due to item nonresponse and because there were communities with no children in a given age-group at intake. For example, there were 15 small communities that had no children between the ages of 5 and 7 in the sample. Column B (and D) shows the number (and percentage) of children who were out of school at follow-up, of those who were out of school at intake. Nineteen additional children in small communities and 13 in large communities were out of school at follow-up but were enrolled at intake. Those children are not included in the table. Column C shows the percentage of children who were out of school for a given age group, out of all children who were out of school at intake.

C. Reasons for children's absences from school

Illness was by far the most common reason for school absenteeism, according to children's caregivers. In both small (34 percent) and large (49 percent) communities, the plurality of caregivers reported that children were absent for that reason. In small communities, but not in large ones, agricultural and domestic or household responsibilities were the next most common reason for absenteeism: 22 percent of the children who had been absent were reportedly absent for those reasons. Bad weather was the second most common reason in large communities, at 14 percent. Children in large communities were more likely to ride a bus or car and to take a motorcycle to go to school than children in small communities were. These means of transportation may be difficult to use under bad weather conditions.

Table K.9. Reasons for children's absences, by community size

	Small communities		Large communities	
	Number of children	Percentage of children	Number of children	Percentage of children
Illness/sickness	167	33.5	255	48.9
Illness of a household member (including pregnancy, temporary or permanent disability)	31	6.2	25	4.8
Paid work	11	2.2	1	0.2
Agricultural responsibilities	62	12.5	5	1.0
Domestic tasks or home responsibilities	48	9.6	15	2.9
Transportation issues	4	0.8	5	1.0
Child does not want to attend school (for example, s/he is bullied, s/he is made fun of, s/he doesn't like it)	21	4.2	28	5.4
Temporary migration to another region (for example, for work in another part of the country)	23	4.6	5	1.0
Parents do not want to send child to school because they are afraid he/she will become infected with a disease	0	0.0	0	0.0
Parents do not want to send child to school because they are afraid of gangs or for safety reasons	3	0.6	0	0.0
Parents do not want to send him/her to school because of bad weather	24	4.8	73	14.0
School is closed or is under construction	8	1.6	0	0.0
Problems with a teacher or principal (for example, frequent absence of teachers)	44	8.8	19	3.6
Distance/child doesn't have anyone to travel with	9	1.8	3	0.6
Financial reasons	6	1.2	34	6.5
Has repeated a grade/has learning issues	2	0.4	1	0.2
Death of a family member/friend	5	3.6	2	0.4
Vacation/family visits	12	1.0	19	3.6
Misinformation: did not know classes were being held	0	0.0	2	0.4
Other reason	18	2.4	30	5.8
Number of children	**498**		**522**	
Number of communities	**137**		**50**	

Source: Follow-up household survey.

Notes: Denominator is the number of children who were absent at least one day, according to caregiver reports, in the month before data collection. There were 10 small communities and 2 large communities in which no children were reported by their caregivers as being absent in the month before data collection.

APPENDIX L:

STAKEHOLDER COORDINATION AND CONSULTATION

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

In this appendix, we summarize the lengthy stakeholder coordination and consultation process followed for this evaluation, and key decisions made during the substantial amount of time invested in the evaluation design. Successful evaluations take into account a multiplicity of contextual factors as well as the intervention's program logic. Evaluation design and implementation needs to be inclusive, attentive to cultural nuances, and responsive to the constraints of the local implementation team. It was important for the Mathematica team to have a sound understanding of the program logic, activities, roll-out, intake processes, timing, and other nuances of the program design, implementation as well as its expected impacts, in order to propose a feasible and useful evaluation. In addition, facilitating stakeholder understanding of the key features of the rigorous evaluation protocol enabled us to have substantive conversations with USAID and CARS in order to finalize the evaluation's questions, the evaluation design, and evaluation implementation activities.

A. Consultation process and timeline

Consultations occurred through inter-institutional meetings by phone, video teleconferencing, in-country visits, and the exchange of documents and other sources of information between the United States and Nicaragua. The process had three initial phases: (1) meetings and information gathering in the United States while program staff were planning the roll-out, (2) travel to Nicaragua, and (3) the drafting and finalization of the evaluation plan. Once the evaluation plan was finalized, the process continued as part of the evaluation implementation in the fourth phase, and as part of the dissemination of findings in the fifth phase.

Phase 1: Information gathering by Mathematica via phone, email, and meetings in Nicaragua (2013 and 2014)

- Met with the USAID LAC education team to understand their expectations and the processes for communicating with CARS, USAID/Nicaragua, and other stakeholders.
- Reviewed programmatic reports on CARS activities to learn about what the program in general and the EpC activity in particular include, the implementation status to date, and plans for future implementation.
- Summarized early design options in a memo for discussion with stakeholders.
- Drafted the preliminary evaluation design to inform in-country stakeholder meetings and discussions.
- February 2014: conducted the first local consultation discussions with key stakeholders, including USAID/Nicaragua staff, CARS Chief of Party and other staff, and CARS partners. Met with potential local evaluation and data collection partners.

Phase 2: Creation and refinement of the evaluation design (2014 and 2015)

- Prepared presentation of the draft evaluation design.

- Solicited comments and feedback from stakeholders (USAID/Washington, USAID/Nicaragua, CARS, and Mathematica's local evaluation partners) on the draft evaluation design.
- May 2014: held conference call with the second CARS Chief of Party to address any concerns and to refine preliminary evaluation design.
- July 2014: met with the third CARS Chief of Party to address any concerns and to refine preliminary evaluation design. Met with regional educational authorities (SEAR) to brief them about the evaluation activities and get their approval.
- September 2014: met with CARS staff to support the process of recruiting communities for the evaluation.
- November 2014: met with the CARS team to support the finalizing of the activities for the rollout and impact evaluation of Cohort 1B and planning for Cohort 2. Led an impact evaluation workshop for CARS, USAID/Nicaragua, and local NGOs to explain the evaluation design and the resulting recruitment needs.
- Finalized evaluation design based on initial in-person meetings and follow-up with stakeholders.
- Engaged with NGOs working on the implementation of CARS activities in Nicaragua.
- February 2015: led a workshop for CARS and NGO staff that focused on impact evaluation principles, the evaluation of EpC, and the recruitment process (including informed consent). Supported CARS and local NGO staff in recruitment efforts and the consent process for recruitment of Cohort 2 participants for the evaluation.
- Throughout 2014 and 2015, Mathematica coordinated extensively with the CARS team and their local NGO counterparts in the recruitment of the necessary number of participants/communities. Mathematica and CARS held weekly calls to learn about implementation details and to coordinate with the evaluation needs. This approach helped to ensure that although the CARS team and the local NGOs met their implementation goals, the evaluation design would be preserved.
- Because of the regional characteristics in terms of geography, climate, and political sensitivities, recruitment was extended from mid-2014 until the end of 2015. To accommodate, from an implementation point of view, the need to start EpC activities, Mathematica agreed to stagger the random assignment of evaluation participants to be invited to the EpC by forming four groups: 1A, 1B, 2A, and 2B.

Phase 3: Finalize Evaluation Plan (2016)

- Revisited evaluation design and implementation scope, given anticipated funding shortfalls in USAID funding for the LAC Reads Evaluation contract.
- Finalized the evaluation plan, including staffing and the budget.
- Shared the plan with evaluation stakeholders (USAID/Washington, USAID/Nicaragua, CARS, and Mathematica's local evaluation partners).
- Incorporated suggested changes into the final evaluation design.

Phase 4: Continuous Consultation throughout Evaluation Implementation (2016-2018)

- Coordinated random assignment and supported communication of random assignment results.
- Coordinated data collection activities in the field for the qualitative implementation evaluation and impact evaluation with implementation activities and CARS studies.
- Kept abreast of implementation fidelity.

Phase 5: Dissemination (2019)

- Shared and discussed draft findings with USAID and CARS.
- Incorporated revisions to the report based on comments from USAID and CARS.

B. Key decisions made

The consultation process facilitated a common understanding of (1) the evaluation questions and design, (2) EpC eligibility requirements and random assignment decisions, and (3) data collection planning for the evaluation. A summary of key decisions that were made jointly between USAID, CARS and Mathematica is below.

1: Evaluation questions and design

- Conduct an implementation evaluation of the full set of CARS activities, as well as an impact evaluation of one or more CARS components to contribute to the USAID LAC Reads learning agenda. Impact evaluation of EpC decided.
- Use of random assignment for the impact evaluation. CARS expected excess demand for the EpC program, and expressed a preference for child level random assignment. After Cohort 1a, while the CARS team was able to identify sufficient numbers of children for the Phase 1 EpC to conduct child random assignment, it did not seem feasible to use only child level random assignment due to the small size of many communities in the RACCS. Given implementation requirements that an EpC serve 20 to 33 children we discussed using both child random assignment and community random assignment depending upon size of the eligible population. Consensus was achieved on the 'hybrid' design option, with random assignment at both the child and community level.
- Mathematica understood that CARS could recruit sufficient communities and could replicate targeting so we could have control communities, and that CARS could recruit sufficient children so we could have control children in the larger schools. We discussed that with child random assignment, it would be important to assign children that live or travel to school together as a unit, for safety reasons. Mathematica indicated that we understood the approach would be burdensome and would provide support by developing protocols and scripts for obtaining consent, and by contracting a local researcher that could work with the CARS team to ensure the protocols were understood and followed.
- The intake and random assignment process was tested for 10 EpC in communities with child random assignment, and feedback about the process was positive. Mathematica

understood that this process was acceptable for the CARS team moving forward, and that the next step was to clarify details of sorting out which communities would have child random assignment and which communities would have community random assignment.

- February 2016: updated sample sizes across cohorts 1A, 1B, 2A, and 2B.

2: Eligibility and random assignment

- The process of recruitment for the evaluation was long, as it was not easy to identify communities with a sufficient number of children eligible to participate in the EpC program. CARS continued with implementation (opening EpC), during the recruitment process, so random assignment was done in stages: Cohort 1A, Cohort 1B, Cohort 2A, and Cohort 2B.
- In each round of random assignment, recruiting a minimum number of children per community became more and more of a challenge, so the thresholds were adjusted.
  - Initially, an EpC would not have fewer than 30 or more than 35 children. However, the implementation plans of CARS covered a minimum number of EpCs to open, and a minimum number of children to be served. To accommodate a larger proportion of children assigned to treatment, for Cohort 1B onwards, we randomly assigned recruited children to smaller-sized EpCs (with some only having 20 or even 17 children).
  - Originally, there would be no more than 35 children per EpC to avoid overcrowding, but some exceptions were made in Cohort 2 to allow NGOs to hit their targets in terms of number of children being served (up to 37 children were allowed in one EpC).
- In Cohort 2, rather than identify eligibility at the child level in smaller communities that would likely participate in community random assignment, a community risk level was determined.
- At each stage of random assignment, we agreed with CARS as to choices that were made regarding the number of communities that would be selected to receive an EpC in each region, including in cases where odd numbers were identified.

3: Primary outcomes to measure and their measurement

- Given the wide range of potential outcomes that could be affected by EpC, we needed to limit the set of constructs to be measured at endline as a part of the evaluation to minimize respondent burden. Mathematica requested USAID and CARS to weigh in on our recommendations to measure constructs that were: (1) well-aligned with the EpC intervention and where we expected to observe an impact of EpC activities based on the ToC; (2) outcomes of particular interest to USAID (in particular, reading and security); and (3) process outcomes that were critical for understanding what was delivered as a part of EpC.
- Where possible, we used existing assessments/instruments that had been used and validated previously. We prioritized those that had been used with a wide range of ages, were available in Spanish, and had been used in low- and middle-income countries. For

constructs where such measurements were not available, we wrote and field-tested new items to capture constructs of interest to USAID or that were central to the EpC ToC.

- At endline, we used an adaptive test to measure literacy outcomes, with pseudo-word decoding as a screener to sort children into decoders vs. non-decoders. For decoders, we measured oral reading fluency and reading comprehension, two key outcomes of interest for USAID and CARS. For nondecoders, we measured oral comprehension to examine if those children had other foundational reading skills that are predictive of future reading comprehension. We considered but decided not to measure familiar word reading, vocabulary, and letter identification to minimize participant burden and because more advanced skills were deemed more informative for USAID's purposes.
- We measured a number of social-emotional skills and attitudes that were targeted by EpC (such as self-esteem) and that are thought to contribute to later academic and life success (such as social competence). We considered measuring other characteristics of children that have been associated with academic performance and involvement in criminal activity, but chose to focus on skills and attitudes that were malleable, likely to be positively affected by EpC, and that could be assessed using self- or third-party reports (due to cost considerations).
- USAID was interested in understanding the effect of EpC on security. At the child level, engagement in criminal behavior in the long term was the primary security outcome of interest. However, most children in EpC were young at the time of follow-up data collection and they were likely not yet engaging in criminal behaviors (such as gang activity and drug trafficking). In addition, some of the existing measures would not have been appropriate for children in that age range. Therefore, we measured a limited set of factors associated with engagement in criminal behaviors later in life, such as moral disengagement and impulsive risk taking.
- Household surveys were administered in the language most comfortable for the respondents, which was typically their mother tongue.
- We assessed children's literacy skills in the official language of instruction in their educational community (Miskitu, English, or Spanish) to answer the policy-relevant question of whether children were meeting national reading standards. In addition, we anticipated that if EpC were to achieve one of its intended goals and increased enrollment, the proportion of out-of-school children in EpC communities would have become significantly lower than in control communities. In that scenario, assessing out-of-school children in their mother tongue would have resulted in a confound between language of administration and treatment, inasmuch as significantly fewer children would have been assessed in their mother tongue in EpC schools relative to control schools.
- We decided not to administer reading assessments in Ulwa because of the considerable effort needed to translate our data collection instruments and the fact that Ulwa was the language of instruction in only one of the communities in the sample. We gathered information on other key outcomes such as enrolment and attendance in that community.

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