USAID
FROM THE AMERICAN PEOPLE

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

Baseline Report

March 2017 (Revised September 2017)
This publication 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, Chantal Toledo, Steven Glazerman, Virginia Poggio, and Nancy Murray of Mathematica. Photo credit: Virginia Poggio.

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

Baseline Report

ABSTRACT

This report presents baseline results of 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. The impact evaluation uses a random assignment design of either children or communities depending on the size of the community eligible to receive the EpC. 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 baseline results confirm that the random assignment created groups with similar characteristics at baseline. Treatment and control groups are equivalent in their characteristics and in school enrollment rates. Based on programmatic intake data, children and communities in the evaluation treatment and control groups are equivalent prior to exposure to the interventions. Similarly, according to base year survey data for the second cohort of EpC rollout, household characteristics across the treatment and control groups are similar, as are child level outcomes such as current enrollment in school and reading behaviors in the household (such as a parent reading to the child). However, there are some differences in base year reading skills that may reflect early impacts of EpC or may be due to chance.

CONTENTS

ABSTRACT I

LIST OF ACRONYMS V

EXECUTIVE SUMMARY ES-1

A. Purpose and background ES-1

B. Baseline evaluation goals and design ES-1

C. Summary of baseline findings ES-2

D. Overall conclusions and limitations ES-2

I. BACKGROUND 1

A. THE INTERVENTION: Espacios para Crecer 1

B. The Evaluation 2

1. Evaluation questions 2

2. Evaluation design: Randomized controlled trial 3

C. THIS REPORT 6

II. DATA FOR THE BASELINE REPORT 7

A. Timing of data collection relative to random assignment 7

B. Baseline data sources 9

C. Sampling procedures 11

III. FINDINGS: CHARACTERISTICS OF EVALUATION COMMUNITIES, HOUSEHOLDS, AND CHILDREN 13

A. Baseline equivalence of educational communities in evaluation population 13

B. Children in evaluation population and in evaluation sample 15

IV. FINDINGS: COHORT 2 BASE YEAR SURVEY SAMPLE 17

A. Cohort 2 survey subsample respondents are representative of the Cohort 2 evaluation population 17

B. Cohort 2 background characteristics of households and children 20

V. FINDINGS: COHORT 2 BASE YEAR LITERACY ENVIRONMENT AND SKILLS 23

A. Child reading activities at home 23

B. Child literacy skills 24

VI. CONCLUSIONS 27

A. Conclusions 27

B. Limitations 28

REFERENCES 29

APPENDIX A DATA COLLECTION A.1

APPENDIX B BASE YEAR LITERACY ASSESSMENT B.1

APPENDIX C EPC IMPLEMENTATION ROLLOUT C.1

APPENDIX D EMPIRICAL APPROACH D.1

APPENDIX E SUBGROUP ANALYSES E.1

TABLES

III.1. Educational communities participating in Cohorts 1 and 2 of the EpC intervention 14

III.2. Characteristics of medium educational communities in Cohort 2 15

III.3. Characteristics of children in evaluation population 15

III.4. Characteristics of children in evaluation sample 16

IV.1. Comparison of intake data between evaluation population, evaluation sample, and respondent sample of Cohort 2 19

IV.2. Comparison of intake data between Cohort 2 treatment and control groups 20

IV.3. Cohort 2 household and primary caregiver characteristics from household survey 21

IV.4. Cohort 2 child background characteristics from household survey 22

V.1. Cohort 2 child reading activities at home 23

V.2. Literacy assessment description 24

V.3. Spanish literacy scores for Cohort 2 children 25

FIGURES

I.1. Location of the RACCS in Nicaragua and municipalities in the RACCS 2

I.2. Random assignment 5

II.1. Timing of EpC rollout, recruitment of evaluation sample, and data collection 8

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

LIST OF ACRONYMS

CARS	Community Action for Reading and Security
EGRA	Early-grade reading assessment
EpC	Espacios para Crecer
IDEL	Dynamic Indicators of Reading Success (in Spanish)
LAC	Latin America and the Caribbean
LAC/RSD	Latin America and the Caribbean/Regional Sustainable Development
NGO	Nongovernmental organization
RACCS	Southern Caribbean Atlantic Autonomous Region
USAID	United States Agency for International Development

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

EXECUTIVE SUMMARY

A. Purpose and background

This baseline report is part of the impact evaluation of Espacios para Crecer (EpC), a core component of the U.S. Agency for International Development (USAID)-funded Community Action for Reading and Security (CARS) intervention that is being implemented by DevTech Systems, Inc. and its partners in Nicaragua. The purpose of the evaluation is to generate rigorous evidence on the effectiveness of EpC at improving children outcomes. We are also conducting a performance evaluation of CARS to understand implementation of all CARS components.

The evaluation is part of a suite of five impact evaluations and two performance evaluations that are underway in Guatemala, Peru, Honduras, Nicaragua, and El Salvador under the LAC Reads project. As part of this project, the Latin America and the Caribbean/Regional Sustainable Development (LAC/RSD) team of the USAID contracted with Mathematica Policy Research to conduct independent impact evaluations and analyses of the effectiveness and costs of promising reading interventions and access to education interventions intended to prevent crime and violence in the LAC region.

The EpCs are designed to serve children who are having difficulties at school, have dropped out of school, or have never attended school. EpCs are led by teachers called "facilitators." For three hours before or after each school day, EpCs offer activities designed to support early reading (through reading time and help with homework), socialization (through playtime), and individual growth. The EpCs work with the Quantum Learning methodology, whose learning principles are based on education theories such as accelerated learning, neurolinguistic programming, experimental and cooperative learning, and effective instruction. In the EpCs, children work with reading materials that have been culturally and linguistically adapted to their mother tongue and local context.

B. Baseline evaluation goals and design

The main goal of the impact evaluation of the EpCs is to understand what impact EpCs have on reading and on other education, socioemotional, and security-related outcomes. The EpC evaluation is a randomized experiment conducted in five municipalities on Nicaragua's Caribbean coast to determine the effect of EpCs on student outcomes as well as their cost-effectiveness. For this baseline report, the goals are twofold: (1) to describe the population of children and communities eligible to participate in the EpC intervention and (2) to assess whether randomization produced equivalent evaluation groups. While random assignment ensures that an individual's or community's treatment status is not a result of specific characteristics, it could still result in chance differences between intervention groups on characteristics that might be correlated with the outcomes that the intervention is seeking to affect. Using baseline data allows us to check for those chance differences and control for them in the final analysis should they arise.

Because eligible educational communities and children were recruited over an extended period of time beginning in mid-2014 through the end of 2014 (Cohort 1) and continuing through 2015 (Cohort 2), the information we have available to describe the evaluation population in this baseline report varies, depending on when they were recruited. We have intake data on

educational communities and at-risk children that the CARS program collected to determine eligibility for evaluation participants which we use to both describe the evaluation populations and assess group equivalence, and which can be used to control for any chance differences across the evaluation groups. Furthermore, for the communities that were recruited in 2015, we have additional household survey data and child literacy assessment information which can be used to describe the evaluation populations in more detail, including children's early literacy outcomes 2-4 months after program rollout for Cohort 2.

C. Summary of baseline findings

The population of children and communities eligible to participate in EpC are struggling academically, are hard to reach, and are mostly Spanish-speaking. There are also several communities with Miskitu and Kriol speaking populations. The sample that will be used for estimating impacts on reading is representative of the larger eligible population from which it was drawn.

We find that treatment and control groups are equivalent in their characteristics and in school enrollment rates. The intake data demonstrate that children in cohorts 1 and 2 in the treatment and control groups are equivalent prior to exposure to the interventions. Similarly, according to base year survey data for Cohort 2, language, gender, and literacy are similar for households across the treatment and control groups, as are current enrollment in school and reading behaviors in the household (such as a parent reading to the child).

We do see some differences in reading outcomes for Cohort 2 children that were assessed in our late baseline. We conducted a survey of a sample of cohort 2 children after random assignment had taken place and after the EpC intervention had started in some communities. We find statistically significant differences in letter identification, familiar-word reading, and oral reading for these Cohort 2 children assessed during the base year. This is unsurprising, since we document these differences after some months of exposure and the differences may reflect early program effects. These differences also may be due to chance.

D. Overall conclusions and limitations

The evaluation groups are balanced for most variables, and we will incorporate the differences we found in literacy outcomes in the final analysis. Regardless of their cause, identifying the initial differences in reading skills between evaluation groups in the base year allows us to control for them in the estimation of EpC intervention impacts. In addition, we will conduct sensitivity analyses without controlling for these differences to verify whether controlling for them affects our impact findings at endline, and if so, by how much.

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

The response rate for the base year survey was lower than expected. Given the geographical isolation and mobility of many of the participants in our evaluation, response rates to the household survey were 79 percent, somewhat lower than the desired 90% response rates. If response rates are similar at follow-up then our ability to detect small impacts on key outcomes may be hindered.

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

I. BACKGROUND

Nicaragua is the poorest nation in Central America and the second poorest in Latin America, with 48 percent of its population living on less than U$1 a day and 76 percent on less than U$2 a day. The average number of years of schooling in Nicaragua is 5.8, the second lowest in Central America, and the national primary-level dropout rate is 9.5 percent (USAID/Nicaragua 2013). According to data reported by the World Bank, the nationwide homicide rate in 2012 was 12 per 100,000 inhabitants and this rate increased 72 percent from 1998 to 2009 (DevTech 2013).

The geographically remote Southern Caribbean Atlantic Autonomous Region (RACCS) has even higher poverty and lower education levels than the national average. The RACCS is the second-poorest of all Nicaragua's regions and departments (after the Northern Caribbean Atlantic Autonomous Region) and more than half of its population lives in extreme poverty. In the RACCS, the average number of years of schooling in the region is 2.9 years; 45 percent of school-age boys and 40 percent of girls are not in school, and 25.2 percent of boys and 25.6 percent of girls are illiterate (whereas in Managua, the richest of all the country's regions and departments, illiteracy rates are 7.8 percent for males and 8.5 percent for females). In addition, the primary-level dropout rate is 17 percent in the RACCS, compared to 9.5 percent nationally (USAID/Nicaragua 2013). Similarly, crime statistics in the RACCS are worse than the national average, with a rate of 39 homicides per 100,000 people. To address these issues, in 2011, the U.S. Agency for International Development (USAID) Nicaragua awarded the Community Action for Reading and Security (CARS) program to work in five municipalities in the RACCS. 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. These CARS activities include the Espacios para Crecer (EpC) after-school program that is the focus of this evaluation. Figure I.1 shows the location of the RACCS in Nicaragua and the RACCS municipalities working with CARS.

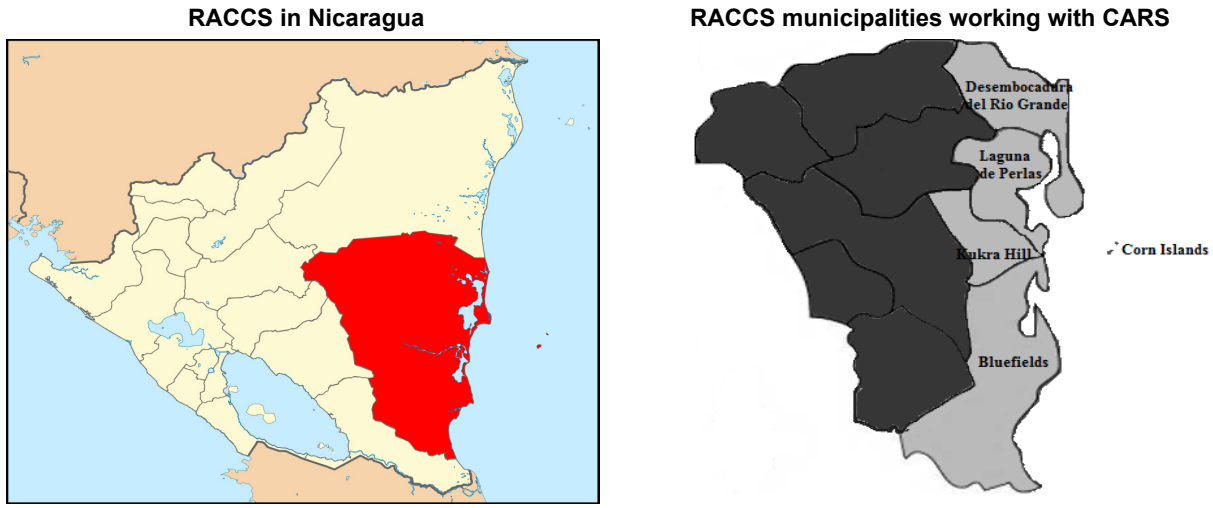
This baseline report is part of the impact evaluation of EpC, a core component of the USAID-funded CARS intervention that is being implemented by DevTech Systems, Inc. and its partners in Nicaragua. The evaluation is part of a suite of five impact evaluations and two performance evaluations that are underway in Guatemala, Peru, Honduras, Nicaragua, and El Salvador under the LAC Reads project. As part of this project, the Latin America and the Caribbean/Regional Sustainable Development (LAC/RSD) team of USAID contracted with Mathematica Policy Research to conduct independent impact evaluations and analyses of the effectiveness and costs of promising reading interventions and access to education interventions intended to prevent crime and violence in the LAC region. This project is known as the LAC Reads Evaluation. Seven evaluations are underway in Guatemala, Peru, Honduras, Nicaragua, and El Salvador: five of these are impact evaluations and two are performance evaluations.

A. THE INTERVENTION: Espacios para Crecer

The EpC intervention is a half-day program, held each school day over an 18-month period in the RACCS, that provides social, development, and academic leveling for children on school days, either in the afternoon for children who go to school in the morning, or vice versa. EpC provide educational and recreational activities over an 18-month period, use what program

implementers refer to as "dynamic pedagogical techniques," and adapt materials to be culturally and linguistically appropriate. EpC are intended to strengthen children's attachment to school, improve performance, and promote socio-emotional development. To bring children to the same academic level, children do homework in groups by grade/skill level. EpC are offered to children who are at risk, defined as having difficulties at school, having dropped out of school, or having never attended school. EpC focus on improving reading skills but also target social and personal aspects of the child, providing a positive social environment for learning. The EpC intervention includes training for teachers to focus instruction on core reading skills and monitor students' mastery of these skills. In the EpC, children complete academic activities (such as working on their homework), and the EpC intervention promotes instruction in the mother tongue and adapts to the specific needs of the local community. Children at the EpC develop their socio-emotional skills through visual arts, dance and movement, and collaborative and teamwork activities. EpC were implemented by DevTech along with three different local nongovernmental organizations (NGOs).[1]

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



B. The Evaluation

The main purpose of the evaluation is to generate rigorous evidence on the effectiveness of EpC at improving children outcomes. The evaluation will determine the impact of EpC relative to prevailing practice on children's literacy skills, attachment to school, and socio-emotional outcomes. To help conceptualize the findings and policy recommendations from the impact evaluation of EpC, we will estimate the intervention's cost-effectiveness.

1. Evaluation questions

The evaluation of the impact of EpC on children's engagement in school and on their reading and socio-emotional skills will constitute the first rigorous evaluation of EpC in

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

Nicaragua and in the LAC region overall. The results of the evaluation will inform not only USAID's and its partners' continuing work to improve learning outcomes, but also the efforts of other government ministries and non-governmental organizations with similar goals that operate in the LAC region. The main evaluation questions are:

- What impact does the CARS EpC intervention have on reading and on other education, socio-emotional, and security-related outcomes?
- What are the costs of EpC? Are the effects of EpC large enough to justify the costs?
- What services were delivered at the child, school, teacher, and community levels?

We will answer the first two evaluation questions by using the randomized controlled trial described in detail in the evaluation plan (Bagby et al. 2016b) and summarized in the next section. We will conduct a descriptive analysis of the implementation of the EpC activity to answer the third evaluation question. This descriptive analysis will be complemented by the rich data that are being collected for the performance evaluation of CARS as a whole (Bagby et al. 2016a). Linking the descriptive analysis of EpC implementation with the impact evaluation findings will enhance our ability to understand how the program impacts were achieved.

2. Evaluation design: Randomized controlled trial

The design of the impact evaluation of the EpC consists of creating two experimental groups, one whose children are exposed to the EpC intervention (the "treatment" group) and another whose children are not (the "control" group). An unusual feature of this evaluation is that we randomly assigned different units—children or communities—depending on the size of the community. As explained in more detail below, in larger communities (with more eligible children), we randomly assigned *children* to a treatment or control group. In communities with fewer children, it was not possible to form two separate groups, so *communities* themselves were the unit of random assignment, with all children in the community assigned together to either the treatment or control group.

For the purposes of the EpC intervention and this evaluation, an "educational community" defined as a school and its immediate catchment area. Educational communities vary in size from just a few children to hundreds of children. This is important because EpCs themselves require a minimum number of children (about 20), to be cost-efficient, and become unwieldy if they have more than the maximum (about 34).[2] We 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 eligible children) and its ability to support a control group

[2] During the process of identifying and recruiting children eligible to participate in EpC and the study—which took approximately one and a half years total—recruiting a minimum number of children per community became more challenging, so the minimum thresholds had to be adjusted. Initially, an EpC intervention would not have fewer than 30 or more than 35 children. However, CARS' implementation plans included a minimum number of EpC to be open, and a minimum number of children to be served. To accommodate a larger proportion of children assigned to treatment, for later groups that participated in random assignment, Mathematica randomly assigned recruited children to smaller-sized EpC (with some having only 20 or even 17 children). Originally, there would be no more than 35 children per EpC to avoid overcrowding, but some exceptions were made later in the process to allow NGOs to hit their targets in terms of number of children being served (up to 37 children were allowed in one EpC group).

within the same community. Therefore the random assignment protocol was flexible enough to adapt to local conditions, with specific rules that characterized each educational community on the basis of the number of EpC it could support.

The team of staff implementing CARS, in consultation with local education authorities, identified educational communities in five municipalities in the RACCS to be eligible for the project, as well as children within those communities eligible to receive the project activities. To be eligible, 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) and have schools that offered first to third grades, served at-risk students (that is, students who have problems in reading performance, retention, dropout, and absenteeism, according to education authorities), and had at least 20 children who were eligible to participate in the EpC intervention. The CARS team defined the total number of children that an EpC could serve, balancing educational needs and the cost-effectiveness considerations of providing services to a minimum number of children. We describe below how the CARS team also defined the criteria for classifying a child as "at-risk".

Figure I.2 illustrates how we conducted random assignment in the two different ways described at the start of this sub-section, depending on the size of the community. For *large* communities, defined as those with 50 or more eligible children, it was possible for the community to support both an EpC treatment group and a control group. In those situations, we randomly assigned children within each community.[3] For *medium*-sized educational communities, defined as having 20 to 49 eligible children, it was necessary to assign all children to the same status—treatment or control.[4] In doing so in these communities with few eligible children, having all eligible children in the same evaluation group allowed us to maintain the integrity of the evaluation design. Therefore, in those cases, we assigned educational communities themselves. All children within these medium-sized communities were assigned together to either treatment or control. Within the medium communities, those with 35–49 children were subject to both levels of randomization, in order to form a waitlist from which program staff could fill slots as they became open. *Small* communities were defined as those with less than 20 eligible children and were not eligible to receive the EpC intervention, and therefore were not included in the evaluation.

We excluded waitlisted children from data collection, as they were not part of the evaluation. We classified children according to their originally assigned status, regardless of their ultimate participation status. To be eligible to participate in the EpC intervention, children had to meet the following requirements as defined by CARS:

- School age (6 to 16 years old, but 5-year-olds in first grade were eligible[5])
- Out-of-school with the equivalent of a 3rd grade education or less, or enrolled in first, second, or third grades
- If enrolled in school, classified as being "at risk," meaning they

[3] Siblings and children who traveled to school together were assigned to the same group.

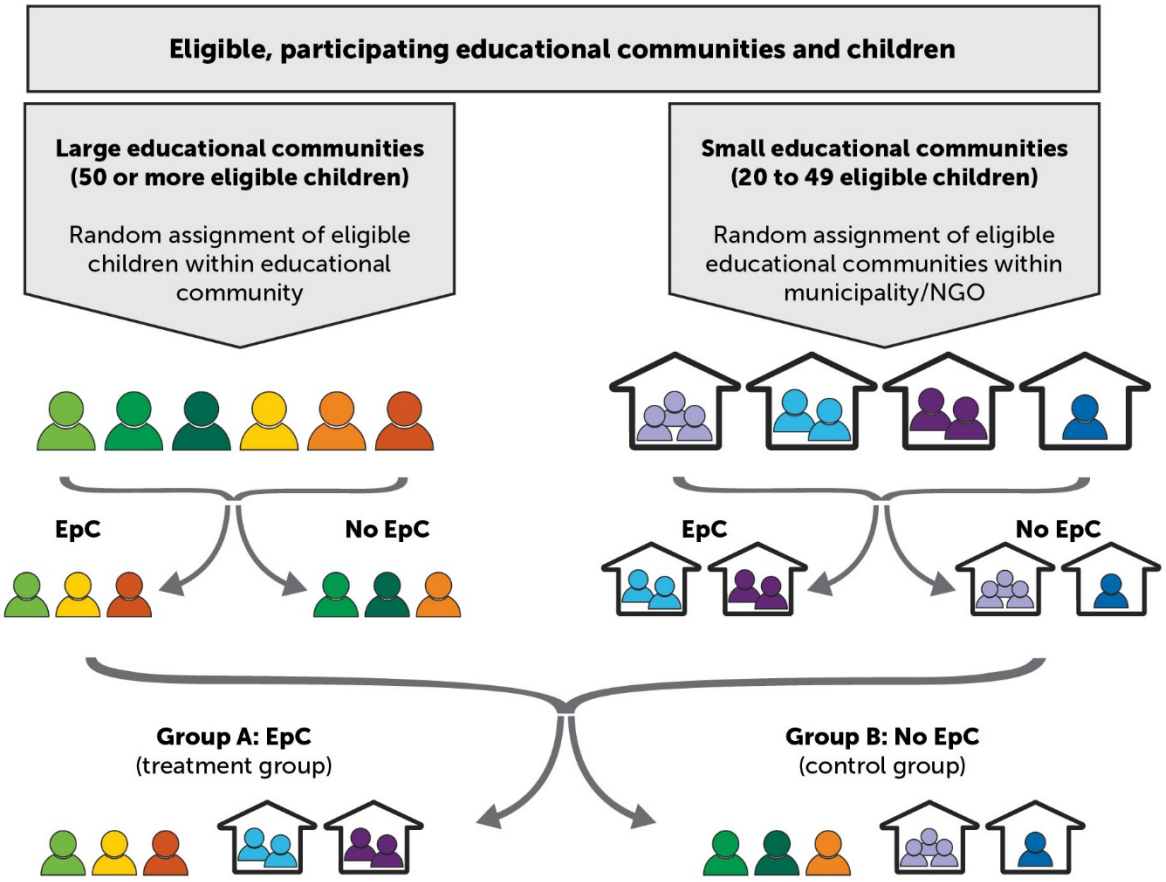
[4] Note that this language is updated in the final evaluation report.

[5] Five year olds enrolled in first grade accounted for only 1.3% of all children recruited to participate in the study.

- failed a grade or had poor academic performance (that is, low grades),
- showed high rates of absenteeism, or
- reported a mother tongue different from the language of instruction.

After random assignment, we have two groups of children: those in Group A (treatment) who are assigned to receive the EpC intervention and those in Group B (control), assigned not to receive the EpC intervention. We conducted random assignment on a rolling basis as children and communities were identified. At each round of random assignment, approximately half of children were assigned to the treatment group and half to the control group.[6] Of 9,223 recruited children across 219 educational communities, 4,542 are in the control group and 4,681 are in the treatment group. The impact of EpC will be estimated by comparing outcomes at follow-up for children in these two groups.

Figure I.2. Random assignment



[6] We used available data at each round of random assignment to stratify the sample. We stratified on municipality and other community characteristics in medium educational communities and on child characteristics such as age, gender, and risk status, in large educational communities.

C. THIS REPORT

The goals of this baseline report are twofold: (1) to describe the population of children and communities eligible to participate in the EpC intervention and (2) to assess whether randomization produced equivalent evaluation groups. While random assignment ensures that an individual's or community's treatment status is not a result of specific characteristics, it could still result in chance differences between intervention groups on characteristics that might be correlated with the outcomes that the intervention is seeking to affect. Using baseline data allows us to check for those chance differences and control for them in the final analysis should they arise.

The report first presents information on the intake and base year survey data, followed by background characteristics of the evaluation population communities, households and children. It then presents information about the cohort 2 survey sample, providing early information on the literacy environment and child literacy skills in the base year shortly after the intervention began. The report concludes with a summary of findings in the conclusions section.

II. DATA FOR THE BASELINE REPORT

This baseline report draws on two data sources: program eligibility intake data and base year survey data for Cohort 2. The intake data come from administrative records that the CARS project used to identify whether children and communities were eligible for the EpC intervention. The survey data come from the household survey and child assessment conducted in conjunction with the household survey of Cohort 2 evaluation participants.

A. Timing of data collection relative to random assignment

Led by DevTech, the team implementing CARS followed a phased approach to implement the EpC intervention. We defined two cohorts of children in the evaluation on the basis of the year in which they began: Cohort 1 in 2014 and Cohort 2 in 2015.[7] Within each cohort, there was a further degree of phase-in, with one group starting early in the year and another group starting later in the year; we refer to these groups as Cohorts 1A, 1B, 2A, and 2B respectively. 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.[8,9] The timing of the recruitment and intake processes within cohorts varied slightly depending upon the NGO responsible for EpC rollout and the location (see Appendix C for more details).

The timing of data collection, shown in Figure II.1 along with the timing of rollout of the intervention for each group of communities, was planned to be the most efficient way to gather data. The intake data collected to identify the population of children and communities eligible to receive the EpC intervention comprise the "true" baseline data for the evaluation, as it was collected before evaluation participants were assigned to their treatment groups or participated in the program. This intake data has key socioeconomic info on all participants, which we use to look at the eligible evaluation population; and assess balance across evaluation groups. The timing of intake, rollout, and evaluation design allowed us to conduct a base-year household survey for the Cohort 2 evaluation sample only[10]. We plan to use the resulting household data to further describe evaluation population for Cohort 2, and will use it to improve the precision of our impact estimates at endline. We will discuss Cohort 1 and 2 survey and assessment results in an end-line report, which will combine follow-up data for both cohorts.

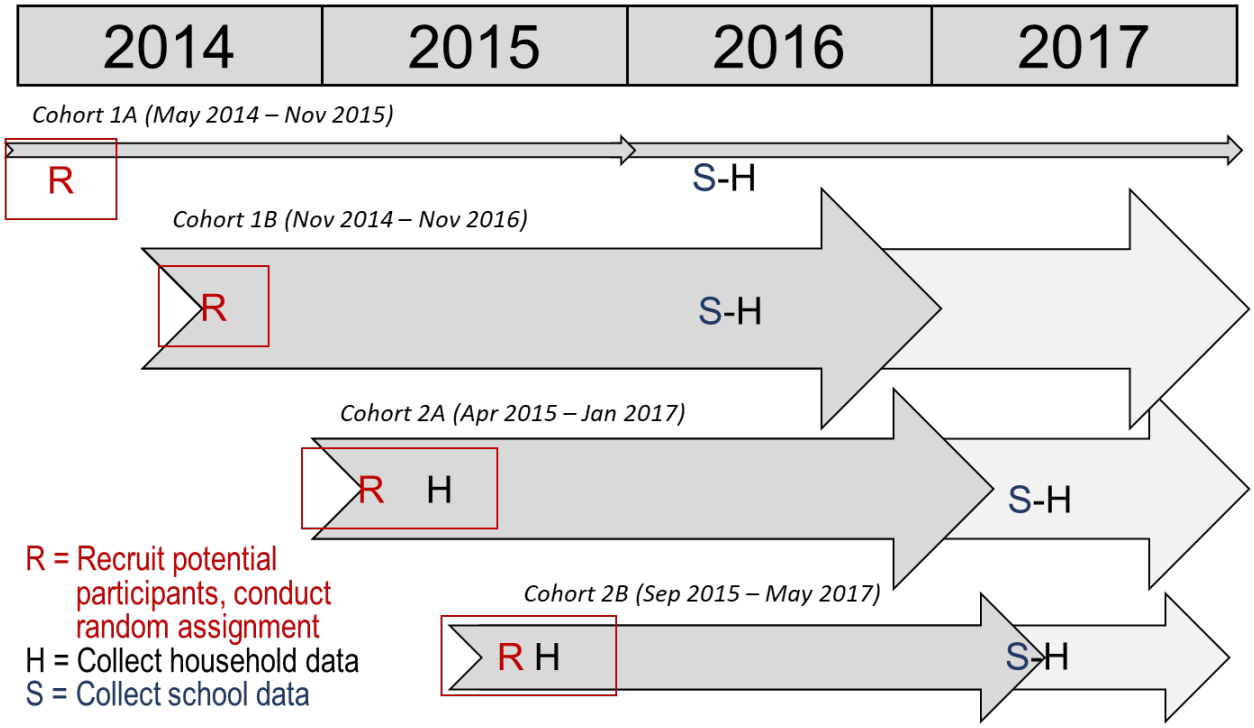
[7] The CARS team implemented EpC for a third cohort that began in 2016, but this was not part of the study.

[8] Implementation for Cohort 2 was set 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.

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

[10] Cohort 2 baseline data was collected within 0 to 5 months of program rollout, and follow-up data are being collected after approximately 18 months of exposure to the EpC intervention for the study sample of both Cohorts 1 and 2 (in 2016 and 2017, respectively).

Figure II.1. Timing of EpC rollout, recruitment of evaluation sample, and data collection



Note: The size of the arrows in the figure suggests the overall cohort size. Cohorts have different community sizes, different numbers of EpC, and different numbers of children by experimental group. Dates shown are approximate. Appendix C contains detailed information about each cohort.

The CARS team determined the overall rollout of the EpC on the basis of intervention requirements. We developed the evaluation design, including our data collection schedule, around this timeline to minimize any delays in program rollout related to evaluation activities. 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, and CARS or Mathematica notified communities 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 it determines the earliest point at which treatment effects could emerge.

This rolling process affected the timing of the start of EpC and of base year data collection in Cohort 2.[11] Base year survey data were collected for Cohort 2 after random assignment and as

[11] The intake process for Cohort 2 took many months because of the challenge of identifying enough communities for the study large enough to participate in the EpC intervention; twice the number of eligible children and communities needed to be identified than would be randomly assigned to participate in the EpC intervention. Because of this, the recruitment of Cohort 2 took longer than originally expected and therefore was broken up into two groups. To prevent implementation being delayed further, Mathematica visited the Cohort 2 communities for base-year data collection during two rounds of field work (forming Cohort 2A and Cohort 2B). Moreover, we

close as possible to the start of the EpC intervention in communities. To avoid delaying rollout of EpC, given the time required to conduct fieldwork for data collection, the timing of baseline data varied across communities. This means that base year (or pretest) scores were collected after children were exposed, even if for a relatively brief period of time, to the EpC intervention. Program participation began by the time of our data collection in approximately 57 percent of communities (40 percent had up to 2 months of exposure and 17 percent had approximately 5 months of exposure). Appendix C provides additional detail by cohort, on the number and size of educational communities, the number of EpC, the municipalities, the implementing NGOs, the lottery results notification date, and the start and expected end date of EpC activities.

Use of the base year data as late baseline data could potentially generate bias in the posttest impact estimates, because pretest scores could reflect the early effects of EpC, making the differences between pre- and posttests for children in the treatment group look smaller than they are (Schochet 2010). However, pretest scores can increase the precision of the estimates of treatment impact substantially. There is, therefore, a trade-off between higher precision and a downward bias in treatment impact estimates. We expect the potential bias in comparisons to be minimized by the relatively short lag between rollout and data collection. In only 29 communities (17 percent) the lag between rollout and data collection was about five months, and for the remaining communities it was between 0 and 2 months. When differences do exist, it is important to use base-year survey data to control for those differences in group means for the final analysis, as well as conduct sensitivity analyses to confirm results are not sensitive to the specification.[12]

B. Baseline data sources

As noted at the start of this chapter, data sources used in the baseline analysis in this report include intake data for the evaluation population and base year survey data for the Cohort 2 respondent sample. The implementation team collected intake data describing children and communities before random assignment. For Cohort 2 only, base-year household and child survey data were collected after the initial intake data and random assignment.

Intake data. The CARS implementation team developed preselection forms with which to identify children's and educational communities' eligibility to participate in the EpC intervention. These forms include data on children and the educational community—true baseline information that can be used to describe the evaluation's population as well as verify eligibility. Child data included children's sex and age in both cohorts and children's mother tongue in Cohort 2[13]. The CARS team collected information about child characteristics and key

couldn't procure the local data collection firm until the evaluation design was finalized (and the lengthy discussions around the design were further slowed down by changes in project leadership), in May 2015.

[12] If there are differences in outcomes between intervention groups at base year, it is important to control for such in the final analysis to ensure that the impacts measured reflect impacts of the intervention and not pre-existing differences. We will conduct additional analyses to verify the sensitivity of the findings to the specification.

[13] Mother tongue information for Cohort 1 was not explicitly collected by CARS when they identified children eligible to participate in the study. During the cohort 1 intake process, CARS included information about whether or not children's mother tongue is different from the language of instruction in the school they attend, since this was a risk factor CARS had identified to determine eligibility for EpCs.

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 school information such as quality measures of school infrastructure quality and number of teachers.

Base year survey data. We collected base year data from a household survey in Cohort 2 communities. Household visits consisted of a survey of primary caregivers, a short child survey, and a reading assessment. The goal of these visits was to collect data allowing us to characterize children's home environments and their reading skills during the base year of program participation.[14] Appendix A provides additional information about the data collection process.

- **Household 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, such as means of transportation and distance to school, child labor, and the home's literacy environment (for example, whether someone reads with the child and how often).
- **Child survey and reading assessments.** We used assessments of children's reading to measure their early literacy skills in the base year of program rollout. We can use these data at end line to reduce the amount of variation in children's reading scores unexplained by the intervention and thereby improve the evaluation's statistical power to detect program impacts.[15] The reading assessments were an adaptation of the Early Grade Reading Assessment (EGRA - Dubeck & Gove 2015) and the Dynamic Indicators of Reading Success (IDEL - Baker et al. 2006).[16] Instead of translating these assessments, Spanish and local language experts adapted them to the Nicaraguan linguistic and cultural contexts. All subtasks, instructions, skip patterns, and rules in the assessments follow those used in the EGRA and IDEL; we adapted the ordering of words and the content of the reading passages. The adaptation aimed to develop tasks with comparable levels of difficulty in Spanish and English, as indexed by word count, average word length, number of sentences, and average sentence length. Eight literacy skills were measured at baseline: (1) letter identification, (2) initial sound identification, (3) listening comprehension, (4) phonemic awareness (5) familiar-word reading, (6) vocabulary, (7) oral reading fluency, and (8) reading comprehension.

[14] Survey instruments are available upon request. Once the final report from the impact evaluation is released, the instruments and database will be posted on the Development Data Library.

[15] By including baseline information in the endline analyses for even just the cohort 2 study participants, the precision of our overall impact estimates will be improved, thereby allowing us to detect smaller differences between groups (and thus increasing the power of the study).

[16] We developed the survey difficulty using a combination of EGRA and IDEL tasks to avoid floor and ceiling effects, and properly capture the potential variation in reading skills across children in our sample. These kids cover a wide age range (5-15), and include children who are not enrolled in school.

Enumerators trained in applying the reading assessment administered literacy assessments to all children in the sample in their homes, regardless of their current enrollment status in school. We administered the reading assessments in Spanish during the base year for all children because Spanish was the official language of instruction in all communities participating at baseline. In addition, in some communities where many children's mother tongue was Kriol, we administered an additional oral reading fluency and comprehension task in English. The enumerators read the passage in English and children were given the option to answer the questions in Spanish, English, or Kriol. Appendix B provides information about the assessment development, scoring, and internal consistency.

C. Sampling procedures

The research team selected from the evaluation population 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 medium-sized community, we sampled 10 children. We over-sampled out-of-school children to ensure adequate representation of this group, which is smaller than the school-going population but of important interest to policymakers and stakeholders. We sampled up to four children per community who were not enrolled in school at the time of recruitment. Once four out-of-school children were sampled, we drew a proportional number of boys and girls (that is, 3 boys and 3 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 uneven (that is, 1 or 3 children), we randomly selected the next child without regard to sex, ensuring that the overall ratio was random.

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

III. FINDINGS: CHARACTERISTICS OF EVALUATION COMMUNITIES, HOUSEHOLDS, AND CHILDREN

We present intake data in this report to describe the evaluation population and to test for differences between intervention groups. The intake data are available for all 219 communities; 7,650 households; and 9,223 children for both Cohorts 1 and 2 and contain information on community-level characteristics (for example, language of instruction, access to services, and number of teachers in the schools) and individual level characteristics (for example, child's gender, age, grade, and mother tongue). These data were collected in advance of the intervention and allow us to determine baseline equivalence of our evaluation groups.

A. Baseline equivalence of educational communities in evaluation population

Table III.1 describes the 219 educational communities included in the evaluation and their distribution across treatment (EpC) and control groups. (As explained in section I.B, educational communities incorporate the community around a school, including children who live in the area but are not enrolled). Because they have children who were randomly assigned to treatment (Group A) as well as children who were randomly assigned to control (Group B), each large community is counted both in the number of Group A communities and the number of Group B communities. Medium communities that were randomly assigned to either treatment or control would be counted in only one group. The evaluation population includes a mix of communities by geography, cohort, and size.

Table III.1. Educational communities participating in Cohorts 1 and 2 of the EpC intervention

Educational community characteristic	EpC (A)	Control (B)	Total number of communities
Municipality			
Bluefields	77	76	128
Kukra Hill	25	24	41
Laguna de Perlas	27	27	42
Desembocadura de Rio Grande	4	4	4
Corn Island	4	4	4
Size of community			
Large (50 or more eligible children)	53	53	53
Medium (20 to 49 eligible children)	84	82	166
Cohort			
Cohort 1	50	50	54
Cohort 2	87	87	165
Primary Language of Instruction			
Spanish	126	124	208
English	6	6	6
Miskitu	4	4	4
Ulwa	1	1	1
Cohort 1			
Spanish	39	39	43
English	6	6	6
Miskitu	4	4	4
Ulwa	1	1	1
Cohort 2			
Spanish	87	85	165
English	0	0	0
Miskitu	0	0	0
Ulwa	0	0	0
Number of communities			**219**

Source: Impact evaluation of Espacios para Crecer—Recruitment 2014 & 2015

Note: Large communities are included in both the treatment and control groups because random assignment was implemented at the child level in those communities. Medium communities are either in the treatment group or the control group because random assignment was implemented at the community level for medium communities. Two NGOs recruited an odd number of medium communities, so it was decided that they would have an extra EpC each (which would help them reach their target number of EpCs).

Baseline equivalence. The data suggest that treatment and control communities of medium size—within which either all children were assigned to the treatment group or all were assigned to the control group—were equivalent at the time of intake into the evaluation, as shown in Table III.2. We found no statistically significant differences in either the size of the school (in terms of teachers, classrooms, or multigrade status) or the infrastructure available at the school, suggesting that random assignment produced equivalent groups of medium communities.

Table III.2. Characteristics of medium educational communities in Cohort 2

Medium educational community characteristic	EpC (A)	Control (B)	A–B	P-value
Number of classrooms in school	1.2	1.2	0.0	(0.685)
Multigrade school (percentage)	100.0	100.0	0.0	n/a
Number of teachers in school	1.2	1.2	0.0	(0.789)
Poor infrastructure[a] (percentage)	10.1	9.0	1.2	(0.627)
Access to services – water (percentage)	13.4	11.9	1.5	(0.725)
Access to services – power (percentage)	4.2	3.4	0.7	(0.787)
Number of Communities				**158**

Source: Impact evaluation of Espacios para Crecer—Recruitment 2015

Note: Columns A and B present the regression-adjusted group means. Community-level regressions include medium communities only because treatment is not uniquely defined at the community level for large communities. Intake data in this table are available for communities in Cohort 2 only.

[a]Poor infrastructure is an indicator equal to one if the school has no water, no electricity, and low-quality building.

n/a = Not applicable

B. Children in evaluation population and in evaluation sample

Table III.3 describes the children in the evaluation population. Looking across all 9,223 children in the evaluation (medium and large communities, Cohorts 1 and 2), we found that the treatment and control groups were balanced in terms of sex, age, enrollment status, and grade.

Table III.3. Characteristics of children in evaluation population

Child's characteristic	EpC (A)	Control (B)	A–B	P-value
Sex (percentage of girls)	46.9	46.6	0.3	(0.792)
Age (years)	8.7	8.7	0.0	(0.845)
Current grade (percentages)				
Out-of-school	16.7	15.8	0.9	(0.370)
Grade 1	38.2	37.9	0.3	(0.840)
Grade 2	24.0	25.8	−1.7	(0.063)
Grade 3	20.5	19.4	1.1	(0.282)
Grade 4	0.1	0.1	0.0	(0.322)
Number of Children	**4,681**	**4,542**		

Source: Impact evaluation of Espacios para Crecer—Recruitment 2014 & 2015

Note: Columns A and B present the regression-adjusted group means to account for the evaluation design No statistically significant differences at the .05 level.

As described in Chapter II, the evaluation will focus on a sample of children from the population, and we find balance in characteristics of the evaluation sample as with the evaluation population. Table III.4 describes the children in the evaluation sample. Looking across all sampled children in the evaluation (medium and large communities, Cohorts 1 and 2), we found that the treatment and control groups were balanced in terms of sex, age, enrollment status, and grade.

Table III.4. Characteristics of children in evaluation sample

Child's characteristic	EpC (A)	Control (B)	A–B	P-value
Sex (percentage of girls)	49.4	49.3	0.2	(0.855)
Age (years)	8.7	8.7	0.0	(0.961)
Current grade (percentages)				
Out-of-school	20.8	19.7	1.1	(0.336)
Grade 1	36.1	36.5	-0.5	(0.798)
Grade 2	22.4	24.5	-2.1	(0.110)
Grade 3	20.7	19.1	1.6	(0.273)
Grade 4	0.0	0.0	0.0	
Number of Children	**1,660**	**1,640**		

Source: Impact evaluation of Espacios para Crecer—Recruitment 2014 & 2015

Note: Columns A and B present the regression-adjusted group means to account for the evaluation design No statistically significant differences at the .05 level.

IV. FINDINGS: COHORT 2 BASE YEAR SURVEY SAMPLE

We complement the description of evaluation participants using the base year household and child assessment survey data described in Chapter II, which contains a larger set of information than the intake data, including results from reading assessments that the evaluation team administered together with the household survey. As explained in section II.B, the survey data describe Cohort 2 only. The survey data contain detailed information such as employment, education, and literacy information on the household head and primary caregiver, on child-level characteristics (such as sex or age), and on child reading practices and assessments. In the endline impact evaluation report we will present a complete description of the survey sample for both cohorts 1 and 2 since we will have data for both cohorts at that point.

Most of the remainder of this report focuses on Cohort 2, which makes up 76 percent of the communities and 55 percent of the children in the sample. The Cohort 2 sample of 165 communities includes 158 medium-sized communities and 7 large communities. Cohort 2 communities have a smaller percentage of children in the sample than communities in the sample because only 7 of the 53 large communities participating in the evaluation are in Cohort 2.

A. Cohort 2 survey subsample respondents are representative of the Cohort 2 evaluation population

To generalize from the base year survey analysis sample to the full evaluation population, one needs to compare respondents to the *full sample* (respondents and nonrespondents), as well as to the *full population* from which the sample was (randomly) drawn.

The response rate for the Cohort 2 base year household survey was 79 percent, which did not differ between the treatment and control groups (and did not differ for urban and rural settings). The response rate was lower than the target of 90 percent planned in the evaluation design, a difference we attribute to the greater difficulty of finding respondents at home than the evaluation team had anticipated, despite repeated visits. The large majority of nonrespondents were not found, either because they were not at home during any of the visits, or the household was simply not found (but we could not determine that they were not eligible for the evaluation so they were not replaced) and this phenomenon was evenly distributed across treatment and control groups. The evaluation population was widely distributed geographically, in hard-to-reach, mostly rural areas with limited communication; therefore arranging visits and return visits was challenging. That the response rates were similar between treatment and control groups suggests that nonresponse was random with respect to treatment status, and nonresponse should not bias the estimates of differences between treatment and control groups. A significant proportion of households visited for baseline were located in the Punta Gorda area, which has been the location of some social unrest related to the plan to build a canal through the area. This may explain in part why the response rate was lower than expected, and it may also affect our follow-up response rate.

To assess whether the survey sample is representative of the recruited, or evaluation, population, we compare intake data from the evaluation population, the evaluation sample, and the respondent sample for Cohort 2 (see Table IV.1). The evaluation population consists of the universe of children recruited to participate in the evaluation for Cohort 2, the evaluation sample

consists of the subset of the evaluation population from Cohort 2 chosen to be in the main interview list (excluding replacement list), and the respondent sample is the subset of the evaluation population from Cohort 2 that was successfully interviewed.

We find that the Cohort 2 respondent sample is representative of the recruited population for Cohort 2 (Table IV.1). We compare means across the different samples for child characteristics, whether or not the child walks to school with another child, whether the child failed a class, and whether the child had a low performance. When comparing the means between the recruited population and the respondent sample we have to keep in mind that 63% of the respondent sample is in Bluefields municipality. However, we find that the sample and recruited population means are balanced in all but three variables. Children in the respondent sample are slightly younger than children in the overall population (0.2 years younger). Children are also more likely to be out-of-school (2.6 percentage points), which is the expected results of our intentional oversampling of out-of-school children. We incorporate weights adjusting for this oversampling when comparing means between treatment and control groups. Parents of respondent children are slightly more likely (2.9 percentage points) than those in the overall population to have attended a community assembly about the evaluation and EpC intervention before random assignment; however the difference is substantively small (less than 5 percent of the population mean). When we compare the respondent sample to the evaluation sample, there are no differences in means.

Table IV.1. Comparison of intake data between evaluation population, evaluation sample, and respondent sample of Cohort 2

Child's characteristic	Evaluation population (A)	Evaluation sample (B)	Respondent sample (C)	A-C	B-C	Sample size (A)	Sample size (B)	Sample size (C)
Sex (percentage of girls)	47.2	48.4	48.6	−1.4	−0.2	5,117	1,780	1,414
Age (years)	9.2	9.1	9.0	0.2*	0.1	5,116	1,779	1,413
Parent attended community assembly (percentage)	67.8	67.8	70.7	−2.9*	−2.8	4,777	1,645	1,305
Current grade level (percentages)								
Out-of-school	25.4	30.3	28.1	−2.6*	2.3	5,117	1,780	1,414
Grade 1	36.5	33.2	33.9	2.6	−0.7	5,117	1,780	1,414
Grade 2	21.2	19.6	21.5	−0.3	−1.9	5,117	1,780	1,414
Grade 3	16.8	16.9	16.5	0.3	0.4	5,117	1,780	1,414
Walks to school with another child (percentage)	46.0	44.6	45.7	0.4	−1.1	5,117	1,780	1,414
Failed a class, current school year (percentage)	28.5	29.0	30.4	−1.9	−1.4	3,815	1,240	1,017
Poor performance, current school year[a] (percentage)	74.5	74.8	74.9	-0.4	-0.1	3,815	1,240	1,017
Number of communities						**165**	**165**	**165**
Number of children						**5,117**	**1,780**	**1,414**

Source: Impact evaluation of Espacios para Crecer—Recruitment 2015 – Cohort 2

Note: Column A shows means for the universe of children recruited to participate in the evaluation. Column B shows means for the subset of the recruited population chosen to be the evaluation sample. Column C shows means for the subset of the recruited population that is in the respondent sample (that is, the sample of successfully interviewed respondents in Cohort 2).

[a] Poor performance was defined by the CARS team during the intake process by talking with teachers about their student's performance (low grades) and was not collected for all students.

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

Baseline equivalence. We also find that these same child characteristics are balanced across treatment and control groups for children in Cohort 2 (Table IV.2). The groups were balanced in terms of sex, age, parent participation in the evaluation community assembly, academic risk factors like not being enrolled or being enrolled but having poor performance, grade, and whether or not the child travels to school with another eligible child.

Table IV.2. Comparison of intake data between Cohort 2 treatment and control groups

Child's characteristic	EpC (A)	Control (B)	A–B	P-value	Sample size
Sex (percentage of girls)	47.1	47.3	−0.2	(0.907)	5,117
Age (years)	9.2	9.2	0.0	(0.992)	5,116
Parent attended community assembly (percentage)	67.1	68.4	−1.3	(0.619)	4,777
Current grade level (percentages)					
Out-of-school	25.9	25.0	1.0	(0.601)	5,117
Grade 1	36.1	37.0	−0.9	(0.645)	5,117
Grade 2	20.7	21.8	−1.1	(0.412)	5,117
Grade 3	17.3	16.3	1.0	(0.472)	5,117
Walks to school with another child (percentage)	45.8	46.3	-0.5	(0.891)	5,117
Failed a class (percentage)	27.7	29.3	−1.7	(0.564)	3,815
Poor performance (percentage)[a]	76.9	72.2	4.7	(0.103)	3,815

Source: Impact evaluation of Espacios para Crecer—Recruitment 2015-Cohort 2.

Note: Columns A and B present regression-adjusted group means.

[a] Poor performance was defined by the CARS team during the intake process by talking with teachers about their student's performance and was not collected for all students.

No statistically significant differences at the .05 level.

B. Cohort 2 background characteristics of households and children

Using survey data from the cohort 2 respondent sample, we find that the average household has six to seven members and the average monthly income is under 4,000 Nicaraguan Cordoba (C$) or around US$140 (Table IV.3). Approximately 50 percent of the heads of household are literate and 92 percent are employed or self-employed. Approximately 96% of our sample is extremely poor according to Nicaragua standards as described in Appendix E. The primary caregiver is the person who cares for the sampled focal child primarily and may also be the household head. In some households with more than one child in the evaluation, there may be more than one primary caregiver. Around 9 percent of primary care givers are literate and around 90 percent are employed or self-employed. Around 25 percent of adults in sampled households migrated (or intended to migrate) in 2014 or 2015 and 2 to 3 percent of children migrated (or intended to migrate) in 2014 or 2015. Across all comparisons, one variable—whether the household head completed primary school—has statistically different means across evaluation groups. Having some statistically significant differences between groups is expected. In this case, one difference is actually less than expected given the number of comparisons made. The characteristics of the households in which evaluation children are living are balanced across evaluation groups.

Table IV.3. Cohort 2 household and primary caregiver characteristics from household survey

Household characteristic	EpC (A)	Control (B)	A–B	P-value
Household				
Household size	6.5	6.4	0.0	(0.882)
Household asset index (0–4)[a]	1.8	1.8	0.0	(0.916)
Household monthly income (Nicaraguan Córdoba)	3908.9	3887.8	21.1	(0.931)
Inadequate housing composite (percentage)[b]	94.7	95.4	−0.7	(0.165)
House is owned (percentage)	87.5	86.0	1.5	(0.469)
Household head characteristics				
Age (years)	41.0	41.6	−0.5	(0.452)
Speaks Spanish (percentage)	95.5	96.2	−0.6	(0.479)
Speaks Kriol (percentage)	3.3	3.7	−0.4	(0.715)
Speaks Miskitu (percentage)	0.0	0.2	−0.1	(0.320)
Is literate (self-reported) (percentage)	55.6	54.4	1.2	(0.705)
Completed primary school[c] (percentage)	15.2	11.4	3.7*	(0.037)
Currently employed[d] (percentage)	92.1	92.3	−0.2	(0.904)
Child's primary care giver				
Age	36.9	37.4	−0.5	(0.443)
Speaks Spanish (percentage)	95.5	96.2	−0.8	(0.395)
Speaks Kriol (percentage)	3.5	3.8	−0.3	(0.781)
Is literate (self-reported) (percentage)	9.6	8.9	0.7	(0.677)
Completed Primary School[c] (percentage)	0.2	0.0	0.2	(0.169)
Currently Employed[d] (percentage)	93.3	92.3	1.0	(0.567)
Migration				
Any adult in the child's household has migrated in 2014 or 2015 (percentage)	23.2	24.3	−1.1	(0.723)
Child has migrated in 2014 or 2015 (percentage)	2.1	2.7	−0.7	(0.407)
Child attends school in destination (percentage)	0.5	0.9	−0.4	(0.388)
Number of households	**628**	**605**		
Number of Children	**716**	**698**		

Source: Impact evaluation of Espacios para Crecer—Cohort 2 Base Year Household Survey 2015

Note: Columns A and B present regression-adjusted group means. All regressions include weights to account for oversampling of out-of-school children. There are 1,233 households in the survey, with 1,414 children in the sample. Differences in sample size are due to missing information for some variables.

[a] Following the literature (Montgomery 2003), we use asset indices as a proxy for wealth. The household asset index is defined as follows: It is a count of how many of 4 assets (cell phone, radio, television, and refrigerator) the household has.

[b] The housing composite is defined as follows: It is equal to 1 (indicating "inadequate") if it has only one acceptable material (or none) in walls, roofs, or floors; it is considered adequate when it has 2 adequate materials in either walls, roofs, or floors.

[c] Completed primary school" is defined as having completed sixth grade.

[d] Employment is defined by a self-report of being either self-employed or doing work for pay.

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

As far as the characteristics of children in the respondent sample, forty-eight to fifty percent of the sample are girls, and children are 9 years of age on average (Table IV.4). Around 97 percent of children have Spanish as a mother tongue, and 70 to 75 percent of children's parents live in the same home as the child. Approximately 60 percent of children say that they know how to read, and their school attendance in the last week is at around 66 percent of days school was open. Seventy-three to 77 percent of children walk to school; other modes of transport to school

include riding a bus or animal or taking a motorcycle or boat. Children spend approximately 25 minutes traveling to school. In the past week, three percent of children worked for pay, while 100 percent of children did some kind of unpaid work, such as collecting water, doing domestic chores, working in the field, or collecting wood. Child characteristics are balanced across evaluation groups: of the 25 comparisons made, two are statistically significant at the 5 percent level, which is less than one would expect to arise by chance.

Table IV.4. Cohort 2 child background characteristics from household survey

Child's characteristic	EpC (A)	Control (B)	A–B	P-value
Sex (percentage of girls)	48.3	49.6	−1.3	(0.497)
Age (years)	9.3	9.1	0.2	(0.094)
Child's maternal language (percentages)				
Spanish	96.5	97.3	−0.8	(0.198)
Kriol	2.6	2.7	−0.1	(0.894)
English	0.5	0.0	0.5	(0.252)
Rama Kriol	0.2	0.0	0.3	(0.169)
Parents live in the same home as child (percentage)	70.4	75.2	−4.8	(0.076)
Child can read according to child (percentage)	62.0	58.8	3.2	(0.225)
Child enrolled in school (percentage)	84.5	83.1	1.4	(0.582)
Child ever repeated a grade (percentage)	45.1	47.5	−2.4	(0.365)
Attended preschool (percentage)	24.9	25.1	-0.2	(0.942)
Attendance in last week (percentage)	68.3	65.2	3.0	(0.316)
Attendance in last month (percentage)	73.6	71.8	1.8	(0.478)
Receives school meal (percentage)	80.5	78.8	1.7	(0.703)
Walks to school	77.8	73.3	4.4	(0.152)
Rides bus to school	2.8	2.8	0.1	(0.830)
Takes motorcycle to school	0.6	0.7	−0.1	(0.513)
Takes boat to school	3.4	0.5	2.9*	(0.020)
Rides animal to school	4.2	7.8	−3.7*	(0.036)
Time required to get to school (minutes)	25.3	26.0	−0.7	(0.643)
Child worked for pay in the past week (percentage)	2.1	3.3	−1.2	(0.209)
Child worked unpaid[a] (percentage)	100.0	100.0	0.0	n/a
Number of Children	**716**	**698**		

Source: Impact evaluation of Espacios para Crecer—Cohort 2 Base Year Household Survey 2015

Note: Columns A and B present regression-adjusted group means. All regressions include weights to account for oversampling of out-of-school children. Differences in sample size come from missing information for some variables.

[a] "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.

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

V. FINDINGS: COHORT 2 BASE YEAR LITERACY ENVIRONMENT AND SKILLS

This section focuses on base year reading behaviors and skills for children and their families in Cohort 2. Here we define reading behavior as time being spent reading in the household by parent, another household member, or child. It is useful to remember that we measured base year literacy skills after children were exposed to the EpC intervention for 0 to 5 months, depending on how early the community entered the evaluation relative to when data collection in that community could be completed. In addition, the vast majority of children in both treatment and control groups (approximately 97 percent of children) in Cohort 2 are Spanish speakers. The remaining children are Kriol or Rama-Kriol speakers and live in communities where Spanish is the language of instruction.

A. Child reading activities at home

Approximately 57 percent of children in the base year survey regularly read books and were read to, and approximately 55 percent of children report being able to read. This relatively high level of literacy and reading activity may be due to the fact that 63% of the sample for Cohort 2 is drawn from Bluefields. Reading activities in the household are balanced across treatment and control groups for all measured variables, as shown in Table V.1.

Table V.1. Cohort 2 child reading activities at home

Child reading activities	EpC (A)	Control (B)	A–B	P-value
Parent reports reading to the child (percent)	62.6	61.1	−1.5	(0.650)
Stories	11.6	13.5	−1.9	(0.307)
Newspapers	0.2	0.8	−0.5	(0.114)
The Bible	23.0	25.1	−2.1	(0.414)
Textbooks	69.2	65.8	3.4	(0.256)
Child reports that someone in the household reads to him or her (percentage)	58.5	58.2	0.2	(0.933)
Number of days in the last week that someone read to child at home	1.5	1.5	0.0	(0.788)
Parent reports that child reads to him or herself at home (percentage)	61.0	58.5	2.5	(0.326)
Child reports being able to read (percentage)	55.8	53.4	2.4	(0.402)
Number of days in the last week that the child read by him or herself at home	1.9	1.9	0.0	(0.760)
Number of children	**716**	**698**		

Source: Impact evaluation of Espacios para Crecer—Cohort 2 Base Year Household Survey 2015

Note: Columns A and B present regression-adjusted group means. All regressions include weights to account for oversampling of out-of-school children. Some variations in sample size arise because of "don't know" responses coded as missing.

No statistically significant differences at the .05 level.

B. Child literacy skills

We have 10 measures of child literacy in the base year because eight literacy skills were measured at baseline, two of which have two measures each.[17] Figure V.1 provides details about the measurement of literacy.

Table V.2. Literacy assessment description

Reading skill section	Description
1. **Letter identification**	Measures knowledge of letter names. The letter identification score shows the correct number of letters identified per minute by the child, out of 100 letters. The task is discontinued if none of the first 10 letters are read correctly.
2. **Initial sound identification**	Measures the ability to discriminate beginning sounds. The identification of initial sounds score identifies the number of correct initial sounds in a word identified by the child, out of 12 words. No time limit.
3. **Listening comprehension**	Measures receptive language of an orally read passage, with both explicit and inferential questions. The listening comprehension score shows the percentage of correct responses to questions about a short text read to the child, out of 5 questions. No time limit.
4. **Phonemic awareness**	Measures knowledge of letter–sound correspondences. The phonemic awareness score counts the percentage of correctly identified words when the different sounds of a word are read to the child, out of 10 words. No time limit.
5. **Familiar-word reading**	Measures the ability to identify individual words from grade-level text. The familiar-word reading score shows the number of correct words read by the child per minute, out of 50 words.
6. **Vocabulary**	Measures the ability to use a given word in a sentence. The vocabulary score shows the percentage of correctly formed sentences with a given word, out of 12 words. No time limit.
7. **Oral reading fluency**	Measures the ability to read grade-level passages with speed and accuracy. The oral reading fluency score shows the number of words read correctly per minute. The child is shown two texts, the first of which has 66 words and is appropriate for grades 1 and 2. The second text is longer at 76 words and harder, being appropriate for grades 3 and 4. Children who could read and understand the first text (read a minimum of 34 words correctly and answered at least three out of five comprehension questions) were presented with a second reading passage, which was accompanied by four comprehension questions. Children who did not complete the second reading received a zero score.
8. **Reading comprehension**	Measures the ability to answer explicit and inferential questions about grade-level passages. The reading comprehension score, for the first and second readings, show the percentage of correct responses to questions about the text that the child read, out of 5 questions for the first reading and out of 4 questions for the second reading. Children who did not complete the second reading received a zero score.

Note: The vocabulary section was developed from the Dynamic Indicators of Reading Success literacy assessment. The other reading sections were based on the Early Grade Reading Assessment.

[17] Kriol speakers were also assessed in English in oral reading fluency and reading comprehension to obtain a baseline measure of comprehension. Because only 6 children were assessed in the sample, we do not report mean values in the findings.

On average, evaluation children correctly identified around 30 letters per minute, around 2 initial sounds (of 12), were able to blend 11 percent of sounds "read" to them to identify a common word, and were able to read around 20 familiar words per minute. Children's oral reading fluency was 25 to 30 words per minute (and about 30 percent were not able to read any word from a short text, as shown in Appendix B). Literacy skills are balanced across evaluation groups for 6 of 10 measures of reading skill assessed, as shown in Table V.2.

Children in the treatment group performed better than those in the control group on four measures: letter identification, familiar-word reading, oral reading fluency of the simpler passage, and reading comprehension of the simpler passage. These differences are largely explained by differences in the percentages of children who are able to complete the tasks at all (as shown in Appendix B, approximately 20 percent of children could not identify a single letter, and larger percentages could not complete more difficult tasks). These differences are documented in more detail in Appendix B. That there are this many differences, and that they were 15 to 18 percent of the control group means, suggests the possibility of early effects of the EpC intervention. However it is also possible that these differences are due to chance. In the final impact analyses we will control for these differences in group means. In addition, we will conduct sensitivity analyses to confirm that our findings are not sensitive to the specification.

Table V.3. Spanish literacy scores for Cohort 2 children

Score on reading assessment task	EpC (A)	Control (B)	A–B	P-value
Letter identification score (number of letters identified per minute out of 100 letters)	33.1	27.8	5.4**	(0.001)
Identification of initial sounds score (number of initial sounds identified out of 12)	1.9	1.8	0.1	(0.606)
Listening comprehension score (% out of 5)	64.4	62.0	2.5	(0.173)
Phonemic awareness score (blending) (% out of 10)	12.5	10.9	1.5	(0.198)
Familiar-word reading score (number of words read per minute out of 50)	23.2	19.1	4.1**	(0.003)
Vocabulary score (% out of 12)	35.8	34.2	1.5	(0.522)
Oral reading fluency score, first reading: Number of correct words read per minute (out of 66)	30.0	25.0	4.9*	(0.017)
Reading comprehension, first reading: Questions correct (% out of 5)	29.3	25.1	4.2*	(0.029)
Oral reading fluency score, second reading: Correct words per minute (out of 76)	13.8	10.8	3.0	(0.062)
Reading comprehension, second reading: Questions correct (% out of 4)	13.2	11.2	2.0	(0.227)
Number of children assessed	**716**	**698**		

Source: Impact evaluation of Espacios para Crecer—Cohort 2 Base Year Household Survey 2015

Note: Columns A and B present regression-adjusted group means. All regressions include out-of-school children weights. Reading measures are described in Figure VI.1. Please refer to Table B.3 in Appendix B for conditional results for the second reading passage.

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

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

We found the evidence to be inconclusive on whether the differences are due to chance or early impacts. We estimated impacts separately for communities that had had less opportunity

for exposure to the intervention before data collection (less than five months) and those that had had more opportunity for exposure (approximately five months). The differences arose in the communities with less exposure to EpCs, which is inconsistent with the early impacts hypothesis. However, we did find weak evidence, because the sample sizes become small, that within the communities with shorter exposure, the treatment–control differences were largest among those with about two months exposure compared to those with almost no exposure. Duration of exposure, however, is confounded with quality of implementation of the implementing NGO and other influences in the sub region, so it is difficult to conclude that duration of exposure explains differences rather than either of those factors. Therefore, at endline we will conduct analyses with and without controlling for such differences to explore the robustness of the results.

VI. CONCLUSIONS

In this report, we document baseline characteristics, verify balance in evaluation group characteristics, and identify a few initial differences in outcomes between evaluation groups of the EpC intervention evaluation sample. We use data collected for the evaluation population during the intake process and for Cohort 2 children during the first year of implementation.

A. Conclusions

The evaluation population and the subsample we will use to estimate impacts reflect the intended population of the evaluation. Evaluation children are almost 9 years old on average, and about 47 percent of the sample are girls. Approximately 45 percent of the evaluation children enrolled are in first grade. The official language of instruction is Spanish in 95 percent of communities in the evaluation, and English, Miskitu, and Ulwa are the instructional languages in the remaining 5 percent of communities. In the Cohort 2 sample, children have Spanish, Kriol, or Rama Kriol as their mother tongue.

We find that children in the evaluation are poor, hard to reach, and have poor educational outcomes, which is unsurprising since CARS targeted disadvantaged children. By Nicaraguan standards, 96 percent of the households in the sample are considered to be "extremely poor." Children in the evaluation also have poor educational outcomes, with approximately 17 percent found not to be enrolled in school during the intake process. In addition, only 60 percent report being read to at home, and evaluation children overall have poor reading skills. Evaluation children assessed identified on average only 29 correct letters per minute, and thirty percent were unable to read any word from a short reading passage.

We find that the evaluation groups are balanced on measures of child characteristics and outcomes using the intake data, and we conclude that the random assignment process produced comparable groups. We find fewer statistical differences between characteristics of treatment and control groups than we would expect to see given the number of comparisons made.

In addition, when we take a deep dive and look at base year information for a sample of cohort 2 children after the EpC activities had begun in communities we find that the evaluation groups are balanced on measures of child and household characteristics, and on most measures of child outcomes. Communities, households and children are similar in treatment and control groups for the cohort 2 survey sample in the base year. The evaluation groups were balanced in terms of most measures of early literacy for cohort 2, but there were some exceptions (4 out of 10 measures). Data were inconclusive in determining whether these differences reflect early impacts or chance differences. In particular, letter identification, familiar-word reading, oral reading fluency and reading comprehension for the first reading passage show statistically significant differences in means across groups. In all cases, the treatment group shows higher scores. These early reading skills are precursors to reading comprehension and a primary outcome of interest for the EpC impact evaluation. The early reading skills may be influenced by EpC and may reflect early effects of the EpC intervention because some (approximately 17 percent of the sample) treatment communities had participated in 5 months of EpC before data collection was completed. However, there are no differences

between evaluation groups among children in those communities, and there are differences between evaluation groups among children who were exposed to the EpC intervention for 0 to 2 months. Given the inconsistent relationship between exposure duration and differences in early literacy outcomes, and the confounding of exposure duration with the types of communities and implementing organizations, it is not possible to determine with certainty the reason for the early differences that we found.

While the evaluation groups are balanced for most variables, we will incorporate the differences we found in literacy outcomes in the final analysis. Regardless of their cause, identifying the initial differences in reading skills between evaluation groups in the base year allows us to control for them in the estimation of EpC intervention impacts. This will ensure that the estimated impacts reflect the effect of the intervention since base year data collection and not the effect of any initial differences between evaluation groups. In addition, we will conduct sensitivity analyses without controlling for these differences to verify whether controlling for them affects our impact findings at endline, and if so, by how much.

B. Limitations

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

The response rate for the base year survey was lower than expected. The response rate for the Cohort 2 baseline household survey was 79 percent, which is lower than the 90 percent planned for in the evaluation design. We attribute this low response rate to the difficulty of finding respondents at home, despite repeated visits. If response rates are similar at follow-up then our ability to detect small impacts on key outcomes may be hindered.

REFERENCES

Bagby, E., R. Blair, V. Poggio, and N. Murray. "Community Action for Reading and Security Performance Evaluation Plan." Washington, DC: Mathematica Policy Research, August 2016a.

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

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.

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

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

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.

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 de Chile: CEPAL-ECLAC. February 2001.

Good, R. H., and R. A. Kaminski (eds.). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement, 2007. Available at https://dibels.uoregon.edu/assessment/dibels.

Montgomery, M. R. "Measuring living standards: household consumption and wealth indices." Quantitative Techniques for Health Equity Analysis–Technical Note 4, 2003.

RTI International 2009. "Early Grade Reading Assessment Toolkit." March 30, 2009. Available at https://www.edu-links.org/sites/default/files/media/file/EGRA%20Toolkit%20Second%20Edition.pdf.

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. Available at https://ies.ed.gov/ncee/pdf/20094033.pdf.

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

USAID/Nicaragua. "Country and Development Cooperation Strategy, FY 2013-FY 2017." Available at https://www.usaid.gov/sites/default/files/documents/1862/CDCS%203-3-14%20Public%20Version%20FINAL.pdf.

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

APPENDIX A

DATA COLLECTION

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

A. Data collector selection and training

Mathematica partnered with Fundación Internacional para el Desafío Económico Global (FIDEG) to conduct data collection activities. FIDEG partnered with the Centro de Investigación y Acción Educativa Social (CIASES), a local education and reading firm. FIDEG's technical team interviewed and selected the candidates to participate in the training, giving preference to those who had prior experience in data collection and had worked with children or educators.

We conducted two data collection efforts for baseline in 2015 for Cohort 2—one for 2A and another for 2B. For that effort, we conducted two rounds of training and field practice, one from June 22 to July 6 and another from October 26 to November 6. Although the second round of fieldwork used the same instrument, and most of the enumerators and supervisors at the second round had conducted the first round of fieldwork training, a full training was conducted the second time to ensure that all team members understood all instruments and protocols.

FIDEG leaders and Mathematica staff trained enumerators and supervisors on the data collection instruments and protocols. We used a training manual and undertook group training. We practiced administering the instruments in pairs or small groups and on the board and practiced using the chronometer. Enumerators also practiced administering the reading assessments with children in communities near Managua, including children enrolled in first to third grade and children not enrolled in school (during the first training only). Finally, enumerators completed a written examination and field tests. On the last day of training, bilingual personnel (Spanish and English/Kriol speakers) were trained in administering the oral comprehension subtask in English/Kriol (first training only).

Field tests were conducted in communities that were not part of the evaluation sample. Two field tests were conducted at each training: one for the whole group and another for bilingual personnel (in the first training) and new personnel (in the second training). Information from the field tests enabled us to adapt the instruments to the Nicaraguan and Caribbean cultural and linguistic contexts and was used in personnel selection. For the latter, FIDEG's technical team observed interviewers, using a rubric designed to detect administration errors and assess rapport with the respondent and fluidity of administration. The rubric provided structure to the observations and facilitated the provision of concrete feedback to the interviewers.

FIDEG used results from the written exam, reading assessment practice, and field tests to select field staff. For fieldwork, FIDEG organized selected staff into five teams, each consisting of one supervisor, one reviewer, one expert in the reading assessments, and four interviewers. Staff who demonstrated leadership, organization and management skills, and fluency in English/Kriol (for the bilingual teams) and who had prior supervision experience were selected to supervise field teams. A representative from Mathematica attended and supported all staff selection and training activities.

B. Quality assurance visits during data collection

A field coordinator and the team supervisors were responsible for ensuring the field staff's compliance with the data collection protocols and planned logistics. They monitored the team's adherence to the work schedule and verified and delivered completed hard-copy instruments to FIDEG's office staff. Reviewers were in charge of detecting omissions or inconsistent information and, in coordination with the supervisor, instructed the interviewers to go back to the school or home to recover or verify the information. Supervisors conducted quality assurance

observations at least once a week with every interviewer on their team. If an interviewer did not follow the protocols, the supervisor provided feedback and prioritized observing that interviewer until errors were corrected. In addition, six percent of the households were visited twice to assess data accuracy. The second visit was conducted by a supervisor or by FIDEG leadership. Discrepancies were within an acceptable range.

C. Data quality assurance

Before data entry, a FIDEG office editor checked and organized the questionnaires and performed an additional manual check to detect inaccuracies (such as an erroneous unique identifier) or incoherent information. Data were then entered using CSPro 6.1. Double data entry was conducted for 100 percent of the survey responses. 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.

Upon receiving the raw data from FIDEG, a data quality assurance team from Mathematica conducted data consistency checks and a review of 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. When inconsistencies were identified, we interacted with FIDEG to answer all questions.

APPENDIX B

BASE YEAR LITERACY ASSESSMENT

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

A. Assessment development

The impact evaluation of EpC will provide an estimate of their effect on reading comprehension. Although an evaluation can be useful for looking into several outcomes related to reading comprehension, several outcomes are highly predictive of reading comprehension and are the most useful for measuring at baseline. Because reading skills typically improve as children grow older and progress in school, we measured their skills relative to their ages and schooling levels. For example, for young children who are unable to master reading, oral comprehension can be a good predictor of future reading comprehension and therefore can be an appropriate skill to measure, whereas for older children it may be better to measure reading comprehension directly.

We created literacy assessments that focus on skills predictive of reading comprehension, such as oral comprehension, vocabulary knowledge, letter identification, phonemic awareness, listening comprehension, and oral reading fluency. These assessments were developed on the basis of the Early Grade Reading Assessment (EGRA) and the Dynamic Indicators of Reading Success (IDEL in Spanish). To account for the context of the sample and education in the RACCS (Southern Caribbean Atlantic Autonomous Region) and to mitigate against floor and ceiling effects, we included a range of subtasks that cover different stages of reading ability.

All materials were reviewed by relevant stakeholders to verify the appropriateness of the questions within the context of the RACCS. Mathematica and FIDEG thoroughly vetted and piloted the questionnaires and assessments, providing a high level of confidence in their face validity and reliability, as discussed below. The assessments are sufficiently short to limit respondent burden, are tightly linked to the EpC intervention, and allow for sufficient variation in overall test scores. In Table 1 we show the reading outcomes measured as part of the baseline data collection of the impact evaluation of EpC.

Within each subtask, the enumerators mark the correct number of responses in each line or section, as well as the time remaining (in seconds) and the total number of correct responses. Enumerators are directed to mark an “autostop” if the child is unable to correctly answer an item in the first row or section of the subtask. This is consistent with EGRA procedure and is sometimes referred to as an “early stop rule.” Making some subtasks time-limited is standard for EGRA, as it makes the assessment shorter and helps with assessing how automatic the responses are (RTI International 2009).

Table B.1 presents summary statistics for each of the subtasks described in Figure VI.1 of the memo. For timed score variables (subtasks 1, 5, 7, and 8), we calculated the number of correct letters or words per minute. For all other subtasks we calculated the percentage of correct responses in the subtask. The scores for items in subtasks include 0 (incorrect) and 1 (correct), and we present findings in terms of percent correct.

Table B.1. Summary statistics for reading assessment subtasks

Reading skill	Count	Minimum value	Maximum value	Mean	Standard deviation
Subtask 1. Letter identification: Correct letters per minute (CLPM)	1,414	0	122.6	30.1	27.2
Subtask 2. Initial sound identification: Percentage of correct answers (out of 12)	1,414	0	100.0	15.0	20.0
Subtask 3. Listening comprehension: Percentage of correct answers (out of 5)	1,414	0	100.0	63.0	30.0
Subtask 4. Phonemic awareness: Percentage of correct answers (out of 10)	1,414	0	100.0	12.0	19.0
Subtask 5. Familiar word reading: Correct words per minute	1,414	0	127.8	21.1	25.1
Subtask 6. Vocabulary: Percentage of correct answers (out of 12)	1,414	0	100.0	35.0	40.0
Subtask 7a. Oral reading fluency: Correct words per minute	1,414	0	188.6	27.4	34.2
Subtask 7b. Reading comprehension: Percentage of correct answers	1,414	0	100.0	27.0	35.0
Subtask 8a. Oral reading fluency: Correct words per minute	1,413	0	147.1	12.5	28.1
Subtask 8b. Reading comprehension: Percentage of correct answers	1,414	0	100.0	12.0	29.0
Sample size	1,414				

Source: Impact evaluation of Espacios para Crecer— Cohort 2 Base Year Household Survey 2015

Note: This table shows unadjusted summary statistics for the main reading outcome measures for 1,414 children who were interviewed during the baseline data collection for the impact evaluation of Espacios para Crecer.

There are many children in the sample who are unable to read, and there is a high level of floor effects (i.e., zero correct responses) in the data at baseline. Table B.2 presents the number of children who were unable to identify or read a single letter, familiar word, or word from a written text.

Table B.2. Percentage of Cohort 2 children with zero correct responses on baseline reading subtasks

	EpC (A)	Control (B)	A–B	P-value
Letter identification: Percentage of children identifying 0 letters	20.0	26.4	−6.5**	(0.006)
Familiar word reading: Percentage of children reading 0 words	36.2	44.1	−7.9**	(0.004)
Oral reading fluency score, first reading: Percentage of children reading 0 words	27.9	31.9	−4.1	(0.126)
Oral reading fluency score, second reading: Percentage of children reading 0 words	80.0	84.3	−4.3*	(0.049)
Number of children assessed				1,414

Source: Impact evaluation of Espacios para Crecer—Cohort 2 Base Year Household Survey 2015

Note: Columns A and B present regression-adjusted group means. All regressions include out-of-school children weights. Reading skills are described in Figure VI.1.

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

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

During the test, children were asked to read the second passage only if they correctly read at least 34 words in the first reading and answered correctly 3 out of 5 questions in the first reading, making sample size for the second reading smaller than the overall sample. It was assumed that children who were not asked to read the second reading would not be able to read any words or answer any questions. With these assumptions, more than 80 percent of children in the sample could not read any words (Table B.2). Moreover, children correctly answered around 10 percent of questions (Table B.3). However, if we only consider children who were asked to read the second passage, they could correctly answer around 70 percent of the questions. See Table B.3 for the comparison of the number of questions correctly answered with the two different samples.

Table B.3. Reading comprehension of Cohort 2 children, conditional and unconditional values

	EpC (A)	Control (B)	A–B	P-value	Sample Size
Reading comprehension, second reading: questions correct (% out of 4)	13.21	11.2	2.017	(.227)	1414
Reading comprehension, second reading: questions correct (% out of 4), conditional	65.66	71.13	-5.467	(.253)	256

Source: Impact evaluation of Espacios para Crecer—Cohort 2 Base Year Household Survey 2015

Note: Columns A and B present regression-adjusted group means. All regressions include out-of-school children weights. Reading skills are described in Figure VI.1.

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

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

B. Internal consistency and reliability

Cronbach's alpha is one of the most widely used measures of internal consistency reliability for multi-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 for establishing internal consistency within the test items (Tavakol and Dennick 2011). Bland and Altman (1997) specify that an alpha of 0.7 to 0.8 is sufficient when comparing groups, whereas an 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 alpha; that is, it reflects a high degree of internal consistency across the test items.

Nonetheless, the use of Cronbach's alpha as a measure of internal consistency reliability involves some drawbacks. The value of alpha is affected by the length of the test, and alpha may underestimate the reliability of the test if different test items measure different underlying concepts (Tavakol and Dennick 2011). In addition, according to the EGRA toolkit (RTI International 2009), the fact that some language assessment subtasks have a time limit is likely to inflate the alpha score.[18] However, the extent of the associated bias is not known, and Cronbach's alpha continues to find widespread use for calculating the internal consistency of early-grade reading tests.

We calculated Cronbach's alpha for each subtask and for the assessment as a whole. This measure was calculated by using individual item responses for all subtasks. An item refers to one question. For instance, in the letter identification subtask, whether or not a child responded correctly for each letter was recorded. Similarly, in the listening comprehension subtask, for each of the five questions posed, children received a 1 if they correctly answered the question and a 0 if they incorrectly answered the question. In Table B.4, we display the alpha scores for each subtask and for the overall test in each language.

The alpha scores presented in Table B.4 indicate that, for the most part, the assessments developed have a high degree of internal consistency. Cronbach's alpha is around or above 0.7 for most subtasks except the reading fluency/comprehension ones. This probably reflects the fact that there is a wide range of skills being measured within the text. Across the entire test, the Cronbach's alpha is high, at 0.773.

[18] This assumes that students would score higher without a time limit, producing a greater distribution of scores that would lower the alpha. In the case of this study, particularly in word reading and oral reading fluency, so few students were able to identify words that it is unlikely scores would have been much higher without a time limit. The tasks are indeed measuring the students' ability rather than the speed at which they can complete the task.

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

Reading skill	Alpha
Subtask 1: Letter identification	0.952
Subtask 2: Initial sound identification	0.777
Subtask 3: Listening comprehension	0.597
Subtask 4: Phonemic awareness (blending)	0.761
Subtask 5: Familiar word reading	0.881
Subtask 6: Vocabulary	0.901
Subtask 7a: Oral reading fluency 1	0.965
Subtask 7b: Reading comprehension 1	0.428
Subtask 8a: Oral reading fluency 2	0.580
Subtask 8b: Reading comprehension 2	0.522
Overall test	0.773

Source: Impact evaluation of Espacios para Crecer— Cohort 2 Base Year Household Survey 2015

Note: This table shows the internal consistency reliability (Cronbach's alpha) for the items applied to 1,414 children who were interviewed during the baseline data collection for the impact evaluation of Espacios para Crecer.

In addition to calculating Cronbach's alpha for each subtask and the overall assessments, we analyzed the correlation between subtasks. We would expect adjacent subtasks to be the most closely correlated, meaning that students scoring high on one subtask would also likely score high on the preceding and succeeding subtasks, given that the subtasks are arranged in increasing order of difficulty. Our findings confirm that, for the most part, adjacent subtasks are highly correlated with each other (Table B.5) and that the correlations are statistically significant.

Table B.5. Correlations of reading assessment subtasks

.	Subtask 1	Subtask 2	Subtask 3	Subtask 4	Subtask 5	Subtask 6	Subtask 7a	Subtask 7b	Subtask 8a	Subtask 8b
Subtask 1: Letter identification	1									
Subtask 2: Initial sound identification	0.364***	1								
Subtask 3: Listening comprehension	0.443***	0.293***	1							
Subtask 4: Phonemic awareness (blending)	0.482***	0.348***	0.304***	1						
Subtask 5: Familiar word reading	0.870***	0.328***	0.378***	0.477***	1					
Subtask 6: Vocabulary	0.606***	0.429***	0.448***	0.421***	0.564***	1				
Subtask 7a: Reading fluency 1	0.822***	0.311***	0.352***	0.451***	0.943***	0.538***	1			
Subtask 7b: Reading comprehension 1	0.802***	0.334***	0.395***	0.460***	0.851***	0.572***	0.834***	1		
Subtask 8a: Reading fluency 2	0.550***	0.227***	0.259***	0.335***	0.659***	0.398***	0.671***	0.753***	1	
Subtask 8b: Reading comprehension 2	0.503***	0.246***	0.259***	0.311***	0.576***	0.413***	0.591***	0.708***	0.887***	1

Source: Impact evaluation of Espacios para Crecer—Base Year Household Survey with Child Assessment 2015 of Cohort 2

Note: This table shows the correlations between the summary statistics for 1,414 children who were interviewed during the baseline data collection for the impact evaluation of Espacios para Crecer.

APPENDIX C

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 have different community sizes, different numbers of EpC, and different numbers of children by experimental group. In summary,

- Cohort 1A has 10 EpC and comprises 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 has 65 EpC and comprises 37 large communities and 8 medium-sized 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 has 72 EpC and comprises 7 large communities and 121 medium-sized communities with a total of 4,154 children, 2,103 in the treatment group and 2,051 in the control group.
- Cohort 2B has 18 EpC and consists of 37 medium size communities, with a total of 963 children, 438 in the treatment group and 525 in the control group.

Table C.1 provides additional information about each cohort.

Table C.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	Expected end date of EpC	Base Year data collection	Months of exposure at 2015 base year household survey
1A	9 large	10	Bluefields, Kukra Hill, Corn Island	FHR, FZT, URACCAN	March, 2014	May, 2014	November, 2015	n/a	n/a
1B	37 large, 8 medium	65	Bluefields, Kukra Hill, Desembocadura, Corn Island, Laguna de Perlas	FHR, FZT, URACCAN	October/November, 2014	November, 2014	November, 2015	n/a	n/a
2A	2 large, 27 medium	16	Kukra Hill	FHR	March, 2015	April, 2015	November, 2016	July–September, 2015	5
2A	3 large, 37 medium	25	Bluefields	FZT	June/July, 2015	June, 2015	January, 2017	July–September, 2017	2
2A	2 large, 57 medium	31	Bluefields, Laguna de Perlas	URACCAN	June/July, 2015	August, 2015	March, 2017	July–September, 2017	0
2B	26 medium	13	Bluefields	FZT	August/September, 2015	September, 2015	April, 2017	November, 2015	2
2B	11 medium	5	Bluefields, Laguna de Perlas	URACCAN	November, 2015	February, 2016	May, 2017	November, 2016	0

Note: FHR is the Fundación Hermanamiento Rama, URACCAN is the Universidad de las Regiones Autónomas de la Costa Caribe Nicaragüense, and FZT is the Fundación Zamora Terán. n/a = not applicable

APPENDIX D

EMPIRICAL APPROACH

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

To assess differences (contrasts) between evaluation groups in measurable characteristics of educational communities, households, and children, we conduct regression analyses that adjust for design features and community and child characteristics. We estimate community and individual (child) regressions.

A. Community-level regressions

For each contrast we examine, the regression model we use to test for differences between evaluation groups in *educational community characteristics* in the base year can be expressed as follows:

$$(1)\quad W_{c0} = \alpha + \beta_1 T_c + \beta_2 S1_c + \beta_3 P_c + u_{c0}$$

where W_{c0} is the characteristic of interest for educational community c at baseline (that is, time 0) and T_c is an indicator for treatment, equal to zero for educational communities assigned to the control group and one for educational communities assigned to participate in the EpC intervention. Accordingly, the parameter β_1 is our coefficient of interest, which indicates the difference in the community-level characteristic between the EpC intervention and control group. The variable $S1_c$ includes a vector of indicator variables for the strata within which random assignment of communities was conducted. The variable P_c is an indicator equal to zero for communities in Cohort 1 and one for communities in Cohort 2. We use Cohort 2 indicators when we are estimating pooled regressions that use both Cohort 1 and Cohort 2 data. The term u_{c0} is an educational community-level error term. Community-level regressions do not include variables that define the strata themselves (for example, municipality). Regressions are unweighted and standard errors are not clustered. Community-level regressions include medium-size communities only (large communities are not included) because random assignment at the community level was undertaken only for medium communities. Because random assignment was done at the individual level for large communities, large communities appear in both the treatment and control groups.

B. Individual (child)-level regressions

For each contrast we examine, the regression model we use to test for differences between evaluation groups in *children's characteristics and outcomes* during the base year can be expressed as follows:

$$(2)\quad Y_{ic0} = \alpha + \beta_1 T_{ic} + \beta_2 S1_c + \beta_3 S2_{ic} + \beta_4 P_c + u_{c0} + e_{ic0}$$

where Y_{ic0} is the characteristic or outcome of interest for child *i* in educational community c at baseline, α is a constant term, and T_{ic} is an indicator for treatment, equal to zero for children assigned to the control group and one for children assigned to EpC. Accordingly, the parameter β_1 is our coefficient of interest, which indicates the difference in the outcome between the EpC intervention and the control group. The variable $S1_c$ includes a vector of indicator variables for the strata within which random assignment of communities was conducted, and $S2_{ic}$ is a vector of indicator variables for the strata within which random assignment of children was conducted. Because random assignment was undertaken at the community level for medium communities, all children in the same medium community will

be in the same stratum. The terms $S1_c$, P_c, and u_{c0} are defined the same way as in Equation (1). The term e_{ic0} is a child-level residual. Standard errors are clustered at the EpC level and weights to adjust for the oversampling of out-of-school children are included. Household-level regressions are treated in the same manner as individual (child)–level regressions. Subgroups are analyzed using the main regression on the separate subgroup samples.

Strata. In both regression models we include community stratum dummies (which are collinear with community size) instead of community dummies. Because the evaluation randomized educational communities in some cases and children in other cases (depending on community size), educational community and treatment are orthogonal in some cases but not in others. Because we conducted random assignment within groups (or strata) of similar communities and groups of similar children, we pay the small penalty associated with the need to account for the block effect associated with each stratum of similar communities and children and include the terms $S1_c$ and $S2_{ic}$ in the regression. Across the rounds of random assignment, there are 40 community-level strata and 223 child-level strata.[19] For individual-level regressions that use recruitment data, we include child-level strata from Cohort 1. For individual-level regressions that only use Cohort 2 (survey) data, we do not include these strata. We adjust standard errors to reflect the clustering of children in the 219 educational communities.

Sampling weights. In individual-level regressions, we include weights to correct for over-representation of out-of-school children, relative to the population of eligible children in the educational community. We include these weights because we purposefully sampled equal numbers of boys and girls and over-sampled out-of-school children. By using such weights, the results from our sample approximate what would have occurred in the population of at-risk children in these communities. We do not include weights for sex, given that we did not oversample females.

Robustness. In addition to the main analyses, we conduct several robustness checks using alternative specifications. We estimate separate regressions by evaluation design (random assignment of children and random assignment of communities), which is equivalent to estimating regressions by community size (large and medium, respectively). We estimate findings with and without weights, as well as with and without strata. Overall findings are consistent, regardless of the specification used.

[19] There are 45 child-level strata in Cohort 1A, 157 in Cohort 1B, 21 in Cohort 2A, and 0 in Cohort 2B (given that there were no large communities in this cohort). Similarly, there are 0 community-level strata in Cohort 1A (all communities in Cohort 1A are large, so there is no community-level random assignment), 2 in Cohort 1B, 26 in Cohort 2A, and 12 in Cohort 2B.

APPENDIX E

SUBGROUP ANALYSES

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

Children with high-risk status are of particular interest to this evaluation, not only because they often have poor educational outcomes, but also because they may be at higher risk of experiencing violence. This evaluation will therefore look at impacts across three primary subgroups: sex, out-of-school children, and households' socio economic status (as defined by a poverty indicator). Thus, we examine reading outcomes across evaluation groups within each of these three key subgroups of interest for balance.

Sex. We find a similar pattern with the sample of girls as was found with the overall sample, as shown in Table E.1. We do not find balance on 3 of the 10 literacy skills measured for girls—girls in the treatment group performed better on letter identification, familiar-word reading, and oral reading fluency than girls in the control group. We do, however, find balance among boys.

Table E.1. Spanish literacy scores for Cohort 2 children by sex

	Female			Male		
Score on reading assessment task	EpC (A)	Control (B)	Difference	EpC (A)	Control (B)	Difference
Letter identification score (number of letters identified per minute out of 100 letters)	35.8	29.1	6.7**	30.7	26.5	4.2
Identification of initial sounds score (number of initial sounds identified out of 12)	1.9	2.0	−0.1	1.9	1.6	0.2
Listening comprehension score (% out of 5)	65.0	61.7	3.3	63.9	62.3	1.6
Phonemic awareness score (% out of 10)	13.4	12.4	1.0	11.6	9.5	2.1
Familiar-word reading score (number of words read per minute out of 50)	25.7	21.0	4.7*	20.9	17.2	3.7
Vocabulary score (% out of 12)	39.3	37.3	2.1	32.5	31.2	1.2
Oral reading fluency score, first reading: Number of correct words read per minute (out of 66)	34.1	27.7	6.3*	26.1	22.3	3.8
Reading comprehension, first reading: Questions correct (% out of 5)	31.1	26.0	5.0	27.7	24.3	3.5
Oral reading fluency score, second reading: Correct words per minute (out of 76)	14.6	11.8	2.8	13.1	9.8	3.3
Reading comprehension, second reading: Questions correct (% out of 4)	15.5	11.7	3.9	11.1	10.7	0.3
Number of children assessed	**343**	**346**		**373**	**352**	

Source: Impact evaluation of Espacios para Crecer—Cohort 2 Base Year Household Survey 2015

Note: Columns A and B present regression-adjusted group means. All regressions include out-of-school children weights. Reading measures are described in Figure VI.1.

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

Out-of-school children. We find a similar pattern among out-of-school and in-school children as with the overall sample as shown in Table E.2. We do not find balance on 5 of the 10 literacy skills measured for out-of-school children, and on 2 of 10 measures for in-school children. Out-of-school children in the treatment group performed better on letter identification, familiar-word reading, and oral reading fluency of both reading passages, compared to the control group. Out-of-school children in the treatment group also performed better on reading comprehension of the second passage. In-school children in the treatment group performed better on letter identification and on familiar-word reading compared to the control group. We do not test for differences between the out-of-school and in-school children. At endline, we will test for differences in impacts between those two groups.

Table E.2. Spanish reading scores for Cohort 2 children by out-of-school status (at intake)

Score on reading assessment task	Out-of-school			In-school		
	EpC (A)	Control (B)	Difference	EpC (A)	Control (B)	Difference
Letter identification score (number of letters identified per minute out of 100 letters)	18.6	13.0	5.6*	37.8	33.3	4.5**
Identification of initial sounds score (number of initial sounds identified out of 12)	1.0	1.3	−0.2	2.2	2.0	0.2
Listening comprehension score (% out of 5)	57.3	51.2	6.0	67.7	66.2	1.6
Phonemic awareness score (% out of 10)	5.2	5.6	−0.3	15.1	12.7	2.4
Familiar-word reading score (number of words read per minute out of 50)	11.5	6.7	4.8*	27.6	23.9	3.7*
Vocabulary score (% out of 12)	21.1	17.3	3.9	41.8	39.8	2.0
Oral-reading fluency score, first reading: Number of correct words read per minute (out of 66)	15.5	8.9	6.6*	35.3	31.3	4.0
Reading comprehension, first reading: Questions correct (% out of 5)	13.7	9.6	4.1	35.4	31.4	4.0
Oral reading fluency score, second reading: Correct words per minute (out of 76)	8.0	2.3	5.7*	16.6	14.0	2.5
Reading comprehension, second reading: Questions correct (% out of 4)	7.3	2.6	4.7*	15.7	14.6	1.1
Number of children assessed	**209**	**188**		**507**	**510**	

Source: Impact evaluation of Espacios para Crecer—Cohort 2 Base Year Household Survey 2015

Note: Columns A and B present regression-adjusted group means. All regressions include out-of-school children weights. Reading measures are described in Figure VI.1.

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

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

Socioeconomic status of household. Approximately 96% of our sample is classified as extremely poor. To assess whether the full sample results are the same when we focus on the extremely poor, we constructed a poverty indicator, intended to capture whether the child's household is a poor household, taking into account the Nicaraguan context and following standards from the National Statistics Institute in Nicaragua. When a specific indicator was not available in our survey, we replaced it with the closest substitute. We define family poverty by using the NBI (Spanish acronym for Unsatisfied Basic Needs) definition in Nicaragua (Feres and Mancero 2001) poverty indicator, as follows:

- Overcrowding: A household is considered overcrowded in urban (rural) areas if there are four or more people (five or more) per 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 "less sound."
- Insufficient services refers to households that do not have an appropriate source of water and sewage system.
- Low education is defined using a dummy for whether there is at least one child aged 7–14 years old who does not currently attend school.

- Economic dependence is defined using a dummy if the household head has an education level of primary school or less and 3 or more people depend on her or his income.
- We define the family poverty indicator as equal to one if two or more of these conditions are met.[20] By Nicaraguan standards, these households are considered extremely poor.

Again we find a similar pattern of differences in literacy skills for children from extremely poor households as we did with the full sample. We find balance on 7 of the 10 literacy skills measured for children from extremely poor households and on all measures for children not from extremely poor households, as shown in Table E.3. As with the overall sample, poor children in the treatment group performed better on letter identification, familiar-word reading, and oral reading fluency than those in the control group.

Table E.3. Spanish literacy scores for Cohort 2 children by family socioeconomic status

	Poor			Nonpoor		
Score on reading assessment task	EpC (A)	Control (B)	Difference	EpC (A)	Control (B)	Difference
Letter identification score (number of letters identified per minute out of 100 letters)	32.7	27.5	5.3**	38.2	31.9	6.3
Identification of initial sounds score (number of initial sounds identified out of 12)	1.8	1.7	0.1	3.1	3.1	−0.1
Listening comprehension score (% out of 5)	63.8	61.2	2.6	72.1	72.8	−0.8
Phonemic awareness score (% out of 10)	11.9	10.8	1.1	20.3	13.0	7.3
Familiar-word reading score (number of words read per minute out of 50)	22.8	18.9	3.9**	29.1	22.2	7.0
Vocabulary score (% out of 12)	34.7	32.9	1.9	48.9	53.0	−4.1
Oral-reading fluency score, first reading: Number of correct words read per minute (out of 66)	29.6	24.9	4.8*	34.3	27.3	7.0
Reading comprehension, first reading: Questions correct (% out of 5)	28.8	24.6	4.2*	36.5	32.7	3.8
Oral reading fluency score, second reading: Correct words per minute (out of 76)	13.3	10.6	2.7	20.1	12.8	7.3
Reading comprehension, second reading: Questions correct (% out of 4)	12.5	10.9	1.6	22.6	14.6	7.9
Number of children assessed	**660**	**653**		**56**	**45**	

Source: Impact evaluation of Espacios para Crecer—Cohort 2 Base Year Household Survey 2015

Note: Columns A and B present regression-adjusted group means. All regressions include out-of-school children weights. Reading measures are described in Figure VI.1.

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

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

[20] The National Statistics Institute in Nicaragua uses the following definition: Households that have all needs met (all of these indicators are equal to 0) are considered nonpoor. Households that have one need that is not satisfied (1 of these 5 indicators is equal to 1) are considered poor but not extremely poor. Households that have two or more needs that are not satisfied (2 or more of these 5 indicators are equal to 1) are considered extremely poor.

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

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