



Evaluation of the Guatemala Child Literacy Development Pilot BASELINE REPORT

SEPTEMBER 2018

Rebecca Stone | Kevin Kamto | Srinivasudevan | Adria Molotsky

MAKING RESEARCH RELEVANT



Evaluation of the Guatemala Child Literacy Development Pilot

Baseline Report

SEPTEMBER 2018

Rebecca Stone | Kevin Kamto | Srini Vasudevan | Adria Molotsky



1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000

www.air.org

Copyright © 2018 American Institutes for Research. All rights reserved.



Contents

	Page
Introduction	1
Overview/Objective	1
Context.....	1
Pilot Program Description.....	3
Theory of Change	6
Research Questions	9
Study Design.....	9
Analysis Plan.....	11
Baseline Data Collection	11
Instruments.....	12
Data Collection.....	14
Baseline Results	14
Sample Characteristics.....	14
Assessments.....	18
Baseline Equivalence	21
Discussion	25
Conclusion.....	27
Annex 1	30
Annex 2	32
References	34

Tables

	Page
Table 1. ELGI and EESNAJ Skills.....	12
Table 2. Household Roster	14
Table 3. Spanish Assessment	19
Table 4. K'iche' Assessment.....	20
Table 5. Spanish Language versus K'iche' Language Comparison	20
Table 6. Group 1A versus Group 1B Spanish Assessment	22

Table 7. Group 2A versus Groups 2B and 2C Spanish Assessment.....	23
Table 8. Group 1A versus Group 1B K'iche' Assessment	24
Table 9. Group 2A versus Group 2B and 2C K'iche' Assessment	24
Table 10: Ki'iche Correlations Language Assessments	30
Table 11: Spanish Correlations Language Assessments	32

Figures

	Page
Figure 1. Child Literacy Development Model Framework	3
Figure 2. Program Materials	5
Figure 3. Summary of Toolkit Skills by Language.....	6
Figure 4. Theory of Change	8
Figure 5. Study Design: New Schools.....	10
Figure 6. Study Design: Existing Schools.....	11
Figure 7. Baseline Characteristics	17

Introduction

Overview/Objective

This report presents the baseline results from a randomized control trial (RCT) of the Child Literacy Development (CLD) pilot program in Totonicapán Province in northern Guatemala. Funded by Catholic Relief Services (CRS) and developed in partnership with the American Institutes for Research (AIR), the AIR-CRS CLD pilot program was designed to complement the literacy component of Phase II of the U.S. Department of Agriculture Learning for Life project. The Learning for Life project uses methodologies including Kemom Ch'ab'al (KC) in primary grades to support reading comprehension in Spanish and K'iche' and Jardin de Letras (JdL) to teach bilingual literacy skills in Grade 1. The CRS-AIR pilot is an additional methodology using **classroom-based assessments of critical literacy skills to group students at the right level** followed by **simple, targeted, remedial teaching activities** to support children at their level. The assessments and the remedial teaching strategies are unique in three critical ways:

- ✓ They focus on pre-literacy skills.
- ✓ They are tailored to the language-specific factors of Spanish and K'iche'.
- ✓ They focus on bilingual/multilingual literacy acquisition.

AIR developed and piloted the toolkits in collaboration with CRS and PRODESSA staff. AIR designed and conducted the initial toolkit training for a small sample of teachers. For the full-scale pilot, PRODESSA provided coaches who were involved in training the teachers to use the toolkits, then provided support to teachers implementing the literacy toolkits during four visits throughout the school year. The coaches also collected fidelity-of-implementation data during the classroom visits. To evaluate the pilot, AIR designed the methodology used to assess the effectiveness of the literacy-related interventions in the Learning for Life project – KC, JdL and the CLD pilot toolkits. AIR analyzed the baseline data presented in this report in addition to a local evaluator who will analyze both baseline and endline results.

Context

Guatemala is a multilingual country where, in addition to Spanish, 22 Mayan languages are spoken as well as Xinca and Garífuna. In 2015, the population was around 16,176,133. At least 40% of the population is Mayan, but in areas like Totonicapán (where the CLD pilot and research is being conducted) the Mayan population is closer to 90%. According to the Third Regional Comparative and Explanatory Study (TERCE) results from 2013, over 30% of Guatemalan third graders scored at the lowest level of reading performance (Level 1 or below) in Spanish, and about 20% of students were still performing at Level 1 or below by sixth grade.

Only 3% of third graders and 8% of sixth graders performed at the highest level of reading (Level 4).¹

Guatemala has a bilingual education system (EBI), established in Article 76 of the constitution, which states (in translation): “The administration of the education system should be decentralized. Schools in zones with a predominant indigenous population should preferably be taught bilingually.” In addition, Article 8 outlines the following language policy (in translation): “In the Guatemalan territory the Mayan, Garífuna and Xinka languages can be used in the linguistic communities where they are present, in all their forms, without restrictions in the public and private spheres, in educational, academic, social, economic, political or cultural activities.” Finally, Article 13 provides more specific detail about language use in the education system (in translation): “The national education system, in the public and private arenas, should apply all of the processes, modalities, and levels, respect, promotion, development and use of the Mayan, Garífuna, and Xinka languages, according to the particularities of each linguistic community.”²

Although Guatemala has a progressive language policy and bilingual education system, in practice the education system is missing many of the vital components that are necessary to make it fully functional. For example, the system lacks a teacher training system that teaches teachers *how* to teach early literacy in multiple languages; a teacher posting system, which matches teachers’ language skills to the languages of the communities where they will be teaching; a monitoring system to ensure that teachers are following the curriculum and teaching bilingually; a coaching system to support teachers in improving their instruction; sufficient materials in local languages and Spanish, and at the appropriate reading levels, to support reading practice; and a bilingual national assessment system (a system previously existed but has reverted to Spanish only). The CLD pilot and research aims to target teachers—a core component of the teaching and learning system—and focus on the tools and training they need to successfully develop and improve children’s early literacy skills in both K’iche’ and Spanish.

While there are a multitude of problems that contribute to the low quality of learning, large-scale meta analyses and other studies have highlighted one factor that is present in most high-impact learning programs: effective pedagogies that focus on the “teacher–learner interaction”.^{3,4} Although the precise details of the effective pedagogies may vary, teaching students at their level (skill-based teaching, or “teaching at the right level”) promises to have a significant impact on learning outcomes in varied contexts.⁴

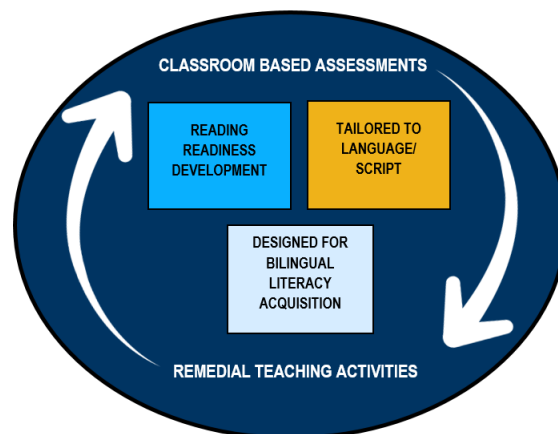
A second and often-overlooked issue is that millions of children are educated in languages they do not understand.^{4,5} Evidence on how to better support these children is extremely limited. In the areas where the Learning for Life project is being implemented, schools are considered bilingual in both K’iche’ and Spanish; thus, CRS and PRODESSA have promoted a bilingual literacy curriculum (KC and JdL). In addition, CRS and AIR developed a teacher training package for the pilot toolkits which supports teachers to use the formative classroom-based assessments, create student groupings and implement remedial teaching strategies tailored for the right level, and tailor instruction to bilingual learners in Guatemala.

Pilot Program Description

Based on the theoretical framework and linguistic descriptions presented above, CRS and AIR developed and implemented the CLD pilot program within the larger Learning for Life project.

AIR also developed the evaluation methodology to evaluate the program. The program is a simple yet comprehensive literacy teaching package, focused on providing Grade 1 teachers a set of classroom-based assessments and teaching tools that can be used to enhance student reading (and pre-reading) outcomes. The program is being implemented in 72 schools in Totonicapán during the 2018 school year in both urban and rural areas.

Figure 1. Child Literacy Development Model Framework



The model of the toolkit focuses on three main features that distinguish it from other reading packages: (1) development of “reading readiness” in Spanish and K’iche’, in addition to the basic Grade 1-appropriate early literacy skills; (2) *language-specific* assessments and teaching strategies; and (3) a focus on *bilingual/multilingual literacy acquisition*. Each of these targeted pedagogical practices is embedded within a framework of classroom-based data usage for teaching purposes.

1. **Reading readiness development:** We emphasize reading readiness development in our program due to children’s limited access to print materials (in any language) prior to the start of school, as well as the limited oral language skills in Spanish and/or K’iche’ that many

children bring to the classroom. Reading readiness is operationalized as the development of necessary oral language skills, as well as a “concept of print” (i.e., a very basic understanding of how print works).

- 2. Language-specific factors:** The CLD toolkits are tailored to the Spanish and K’iche’ languages and writing system. Both are alphabetic languages with a transparent orthography and a relatively “restricted” symbol set size. This led to several assumptions, including the fact that decoding (learning the sound–symbol correspondences) will be less of a cognitive demand, compared to the acquisition of oral language comprehension subskills. Both languages use alphabetic writing systems, so we prioritized comprehension in the new language, given the relative ease with which the so-called “code-processing” skills can transfer between two similar writing systems.
- 3. Bilingual or multilingual literacy acquisition:** The project is specifically tailored to children who may be learning to read a language they do not speak or understand well at the start of literacy instruction. For this reason, it dedicates more time to the development of oral language skills both through the reading readiness toolkits and in the formative assessment toolkits and remediation activities. We do not assume that children come to school only speaking K’iche’ or only speaking Spanish and in fact data shows that many of these children come to school being either monolingual Spanish speakers, monolingual K’iche’ speakers, or bilingual to some extent in both languages. Therefore, the assessments seek to determine in which language oral language skills need to be further developed and that becomes a focus of the remediation activities.

Initial Pilot of Toolkits

Once the initial toolkits and supplementary materials were developed, AIR and CRS conducted a small pilot of the materials with teachers in 2017 to get their feedback and improve upon the materials and their usability. Dr. Rebecca Stone from AIR, along with colleagues Melchor Aguate from CRS and Marleny Tzicap, a consultant, facilitated a 3-day teacher training from the 11th to the 13th of July, 2017. Five Grade 1 teachers, several school directors and two CTAs (Coordinador Técnico Administrativo) attended the training. The training was designed to introduce the teachers to the reading readiness and formative assessment toolkits and supplementary materials, and to give them practice administering the assessments, tracking scores, and conducting remediation activities.

After the training, the consultant visited teachers twice per month for the remaining 2 months of the school year to support them in implementing the assessments and conducting the remedial activities. The consultant also sent regular updates on the teachers' progress, including suggestions for improving the toolkits and/or coaching support and training, based on what she had learned while working with the teachers. All these recommendations—including recommendations from teachers after the initial training—were used to improve the toolkits and training design to prepare for full implementation in 2018.

Program Materials

Each teacher that participated in the pilot toolkit intervention received the following materials:

- ✓ Bilingual Reading Readiness Toolkit
- ✓ Bilingual Formative Assessment Toolkit
- ✓ Set of alphabet dominos (K'iche')
- ✓ Syllable wheel (Spanish)
- ✓ Alphabet wheel (Spanish)
- ✓ Laminated poster scene with question prompts (Spanish)
- ✓ Laminated poster scene with question prompts (K'iche')

Teachers implement the reading readiness toolkits at the start of the year to help determine if children are ready to begin print instruction in K'iche' and/or Spanish. They then conduct remedial activities with children who do not demonstrate readiness before moving on. The formative assessment toolkit measures competence in various subskills: oral vocabulary knowledge, phonological awareness, letter names, decoding, and reading comprehension. Throughout the school year, teachers are expected to implement the formative assessment toolkit three to four times, each time following up with remedial activities for students who are struggling with a particular skill. Figure 3 provides a summary of the skills covered in each of the two toolkits.

Figure 2. Program Materials

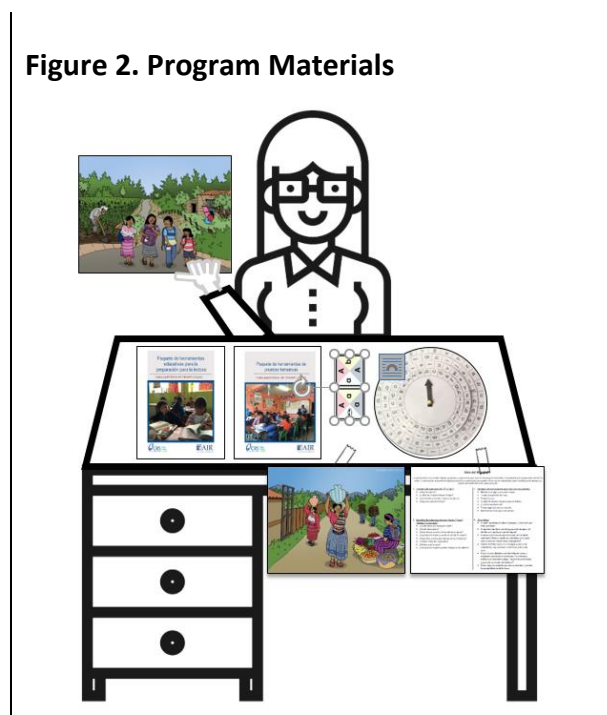




Figure 3. Summary of Toolkit Skills by Language

Toolkit		Literacy Skill	Toolkit		Literacy Skill
Reading Readiness Toolkit		Oral Language	Formative Assessment Toolkit		Oral Vocabulary Knowledge
		Concept of Print			Phonological Awareness
		Letter Names			
		Decoding			
		Reading Comprehension			

All skills were assessed in both languages, except letter names, which was only assessed in Spanish and concept of print which is not language specific.

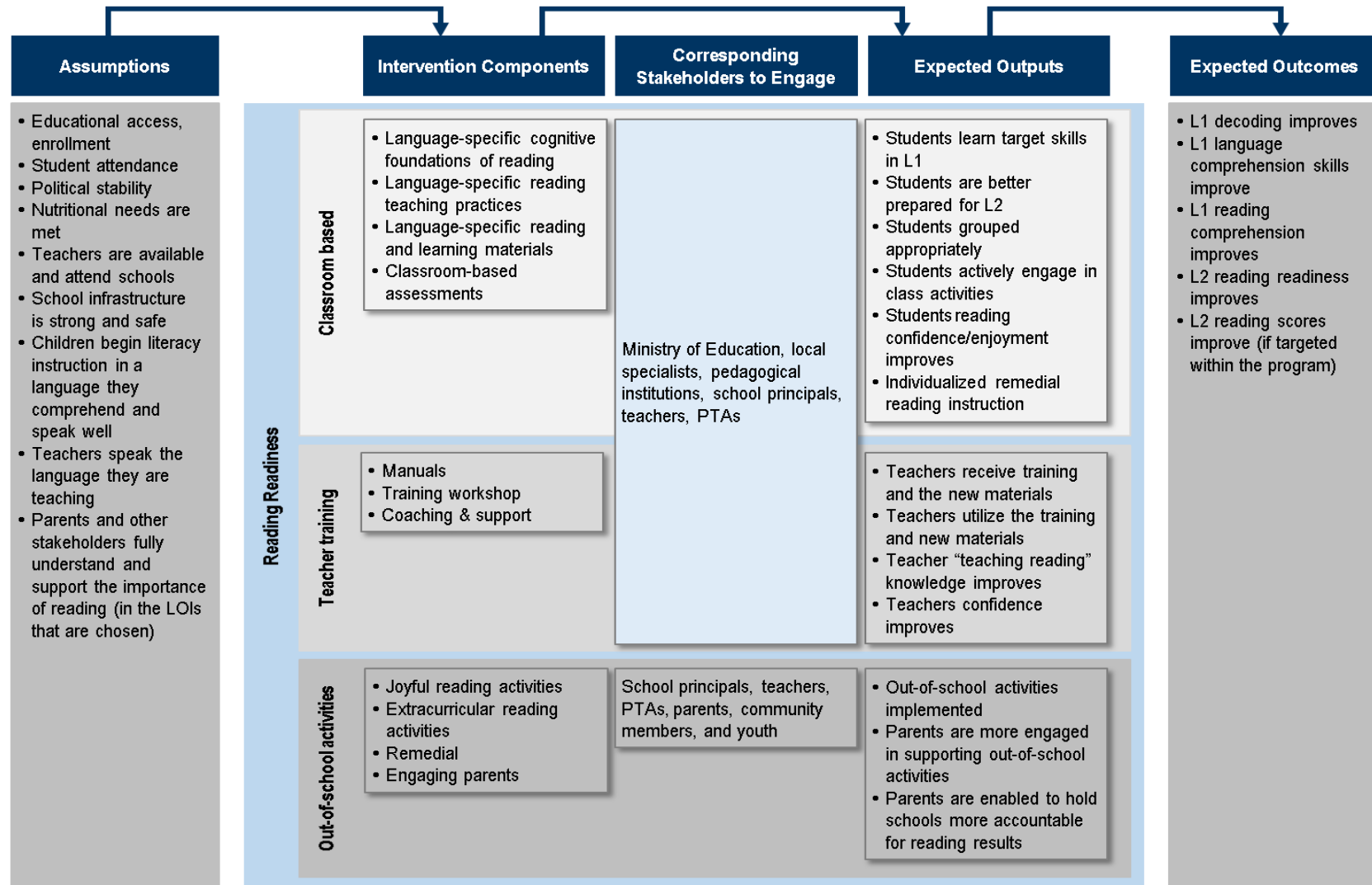
Theory of Change

Based on the literature on how children learn to read in multiple languages, an effective model for teaching reading is one that targets both decoding and oral language comprehension skills. Such a model should also be able to adapt to the relative importance of each of the subskills needed, depending on the language and writing system. An effective program will be taught in a language the children understand and speak—as determined by an objective assessment of oral language skills—and will help to develop “transferrable” skills, introducing a second language only after appropriate thresholds of readiness have been reached (if that information is known). Finally, and importantly, the program will also support teachers to use formative, classroom-based assessments and other methods to make instructional decisions, which will in turn enable them to determine how to adapt and adjust their teaching to improve student reading scores.

Figure 4 depicts a conceptual model for a proposed reading intervention. We first lay out the conditions that are required for enabling “reading readiness.” We hypothesize that these conditions do not have a direct impact on improving reading outcomes but are necessary for a program to have the potential to be effective. Reading program designers in any country can examine whether localities have the necessary enabling conditions, and if not, determine ways to meet the challenges. Next, we show the classroom-based activities, teacher training and support activities, and out-of-class activities, along with the corresponding stakeholders who need to be engaged in each set of activities. If these activities are implemented with fidelity, and with the involvement and engagement of stakeholders, we would expect to see an increase in more effective practices, behaviors, and material usage by teachers and students. This, in turn, will lead to the final outcomes of improved decoding, language comprehension skills, and

reading comprehension in the first language (L1), as well as improved reading readiness and reading scores in the second language (L2).

Figure 4. Theory of Change



Research Questions

We developed a set of research questions to assess **the literacy-related interventions** in Phase II of the Learning for Life project. This research will assess the impact of the various reading programs on reading ability—including all pertinent reading subskills—to ascertain whether the program effects are concentrated in any particular area of reading ability. Through two clustered RCTs, we can address the following three research questions related to impact.

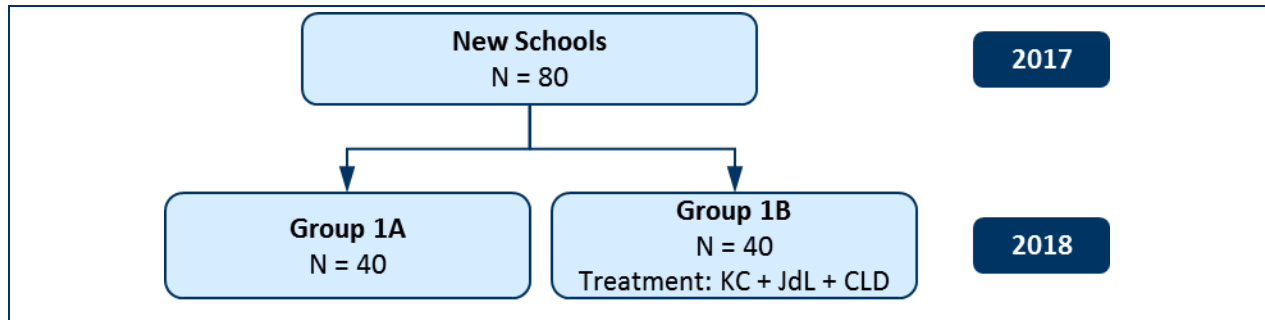
1. Does the package of school feeding (SF), Kemom Ch'ab'al (KC), Jardín de Letras (JdL), and the AIR-CRS Assessment Toolkit (CLD) impact student reading outcomes?
2. What is the added benefit of Jardín de Letras (JdL)?
3. What is the added benefit of the AIR-CRS Assessment Toolkit?

Study Design

The study design is a clustered RCT which included 106 schools that had previously only received the SF program with no other interventions (Group 1) in the municipalities of Santa Maria Chiqimula and Totonicapán; and 225 existing schools that had previous exposure to the SF program plus either the KC program, the JdL program, or both KC & JdL (Group 2) in the municipalities of Momostenango, San Andres Xecul, San Bartolo Aguas Calientes, and Santa Lucia La Reforma. Data collection was conducted at 80 randomly selected new schools from Group 1 and 99 randomly selected existing schools from Group 2. Within these groups, there are several treatment arms that are structured to provide insight into the added value of the various components of the intervention package. In the remainder of this section, we report the number of schools in the different treatment arms for which data will be collected.

Group 1: In 2017, all Group 1 schools only received the SF program. Starting in 2018, half of these schools (Group 1B; N = 40) were randomly selected to receive a package of interventions that includes the KC program, the JdL program, and the AIR-CRS Assessment Toolkit (CLD), as well as the SF program. The remaining schools (Group 1A; N = 40) continued to only receive the SF program in 2018. Figure 5 depicts the study design for Group 1 schools.

Figure 5. Study Design: New Schools

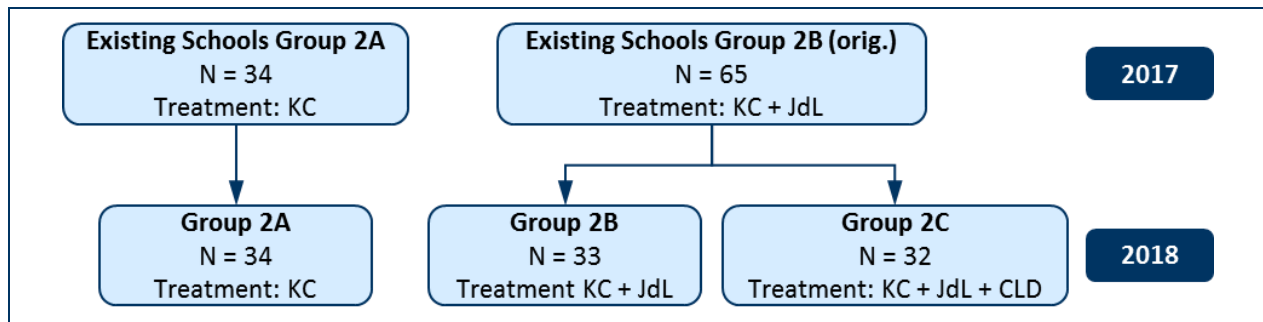


Note: All schools receive the SF program.

Group 2: In 2017, there were 99 Group 2 schools. Of these, 34 were randomly selected into Group 2A, which received the SF and KC programs. The remaining 65 schools in Group 2B (original) received the SF, KC, and JdL programs. Starting in 2018, 32 schools from Group 2B (orig.) were randomly selected to form Group 2C, which will additionally receive the AIR-CRS Assessment Toolkit. The remaining schools in Group 2B (orig.) became Group 2B and continued to get the SF, KC, and JdL programs. Group 2A continued to receive the SF and KC programs. Figure 6 depicts the study design for Group 2 schools.

Note that 15 schools spread across the three groups (2A, 2B, and 2C) received an extra-curricular program called Space to Grow (StG) in 2017, and another set of schools will receive it in 2018 (these were not determined at the time of data collection). This assignment was not/will not be done randomly, which complicates the impact analysis because the treatment status of a school may influence its take-up of the StG program. Our analysis would therefore incorrectly attribute the impact of the StG program to the treatment being evaluated. To address this complication, we will estimate impacts for all schools in our sample and compare the results to a sample that excludes the schools that received StG in 2017 and 2018. This will demonstrate whether the overall impact results are affected by this non-randomly assigned program.

Figure 6. Study Design: Existing Schools



Note: All schools received the SF program.

Analysis Plan

Based on our study design, the impact analysis will compare outcomes between the following groups:

- Group 2A versus Group 2B using (impact of the JdL program).*
- Group 2B versus Group 2C using (impact of the AIR-CRS Assessment Toolkit).
- Group 1A versus Group 1B using (impact of the KC and JdL programs and the AIR-CRS Assessment Toolkit, combined).

The first comparison will enable us to estimate the impact of the JdL program by comparing the difference in average outcomes between Group 2A and Group 2B. The second comparison will enable us to estimate the impact of the AIR-CRS Assessment Toolkit by comparing students in Group 2B and Group 2C schools. The third comparison will enable us to estimate the impact of the full package of interventions—the KC and JdL programs and the AIR-CRS Assessment Toolkit—by comparing Groups 1A and 1B. We will conduct subgroup analysis by estimating the interactions of the key demographic and baseline variables of students.

Baseline Data Collection

The study involves survey data and a reading assessment in Spanish and K’iche’ of an intended sample of 10 randomly-selected Grade 1 students from each of the sampled schools (80 in Group 1A/1B and 99 in Group 2A/2B/2C) for a total sample of 800 in Group 1A/1B and 990 in Group 2A/2B/2C. The baseline data collection took place in January 2018 and resulted in an

* Note that the impact of the JdL program can also be estimated by comparing Group 2A with Group 2B (orig.) using 2017 outcomes. We focus on using Round 2 outcomes to maximize statistical power since it has both 2017 and 2018 data.

actual sample of 1,786 total students separated in treatment arms such that 822 students were sampled in Group 1A/1B schools, and 964 students were sampled in Group 2A/2B/2C schools. The tools used collected data on several outcome variables and consisted of the following tests: oral directions, concept of print, oral vocabulary knowledge, phonological awareness, oral reading fluency, and reading comprehension. More details about the results from the baseline data collection are presented in the baseline results section below.

Instruments

Two assessments were used to collect data on the baseline literacy levels of Grade 1 students: the Evaluación de Lectura para Grados Iniciales (ELGI) in Spanish and the (EESNAJ) in K'iche'. These tests were administered to all students in the study. The tests are a Guatemalan adaptation of the Early Grade Reading Assessment (EGRA). The EGRA is an adaptation of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), created in the United States to assess the level of early reading subskills, such as fluency and reading comprehension.

The ELGI and EESNAJ were adapted to the Guatemalan context through the Educational Reform program in the classroom (REAULA). The tests are designed to be administered orally on an individual basis, with a duration of approximately 30 to 45 minutes (according to the original design). The ELGI and EESNAJ test eight skills with multiple subtests. Table 1 lists the various skills measured, along with how they are measured.

Table 1. ELGI and EESNAJ Skills

Reading Skill Measured	Construct Definitions	Measure
1. Oral Language	- Comprehension of oral instructions ✓ K'iche' ✓ Spanish	Understanding of oral instructions such as, "Touch your head, raise your hands."
	- Comprehension of an oral passage ✓ K'iche' ✓ Spanish	Ability to understand questions about a story a student listens to
	Oral vocabulary knowledge ✓ K'iche'	Ability to name pictures in 60 seconds
2. Alphabetic Principle	- Knowledge of letter names ✓ K'iche' ✓ Spanish	Ability to state the names of listed letters
	- Knowledge of letter sounds ✓ K'iche' ✓ Spanish	Ability to sound out listed letters
3. Decoding	- Sounding out short words ✓ K'iche' ✓ Spanish	Speed of reading short words over 60 seconds

Reading Skill Measured	Construct Definitions	Measure
	- Speed of reading nonsense words ✓ K'iche' ✓ Spanish	Speed of reading words without meaning
4. Phonological Awareness	- Knowledge of phonemes ✓ K'iche' ✓ Spanish	Ability to identify the initial phonemes in a word
	- Separation of phonemes ✓ K'iche' ✓ Spanish	Ability to separate phonemes in a word
5. Fluent Letter/Sound Recognition	- Speed of naming letters ✓ Spanish	How many letters a child can name in 60 seconds
	- Speed of naming the sounds of letters ✓ Spanish	How many letters a child can sound out in 60 seconds
	- Speed of reading familiar words ✓ K'iche' ✓ Spanish	Speed of reading familiar words over 60 seconds
	- Fluent reading of a passage ✓ K'iche' ✓ Spanish	Ability to read a passage with fluency
7. Reading Comprehension	- Reading and understanding a passage ✓ K'iche' ✓ Spanish	Ability to read and answer questions about a passage
8. Writing	- Dictation ✓ K'iche' ✓ Spanish	Reading text aloud

In addition to these eight subtests, AIR designed two additional subtests for both K'iche' and Spanish, which were also administered to all students in the sample. The first subtest measures concept of print, which is an understanding of how print functions. To measure concept of print, we provided students with a small story book and asked them to identify the title, a word, a picture, the direction in which the text should be read, and so on. The second additional subtest measured oral vocabulary knowledge. This was done using a version of the Peabody Picture Vocabulary Test, in which students are provided with a set of four pictures and a word said aloud by the test administrator; the students must then identify the picture that corresponds to the word. These two subtests were added to get a measure of children's concept of print as well as their oral vocabulary knowledge in both languages at baseline. The CLD program focuses on these skills as fundamental to pre-literacy and early literacy instruction therefore it was important to include them in the baseline measure.

For the sections of the test in which the student is asked to say the name or sound of letters, or words or nonsense words in a fixed amount of time, results are obtained in the following manner: The speed is determined considering all those items that the student has read

correctly or incorrectly in a maximum time of 60 seconds. To establish accuracy, only the items that were correctly named are taken into consideration.

A rule of suspension is used in some sections of the test, which indicates to the test administrator the moment when he/she must discontinue administering the section. This rule is followed when the student does not respond correctly to certain items, which indicates that the student has not yet developed the necessary skills to complete that section of the test.

Data Collection

Juarez & Associates trained thirty-six enumerators from the external evaluator, ADOC, over 10 days. This training included practice sessions in schools outside the project’s target area. The enumerators applied the tests according to the manuals, under the supervision of Juarez and Associates[†] and DIGEDUCA[‡], during the training period. Thirty-six test administrators conducted the data collection, with four enumerators per school, organized into nine teams. Nine schools were covered during each day of the data collection period. During the baseline data collection, the four enumerators each administered the tests to one student at a time (and a maximum of four students during the day) at each school.

Baseline Results

This section describes the findings by assessment and treatment group to better understand the sample composition and the baseline reading scores for all students in the sample before the start of the intervention(s).

Sample Characteristics

In this section, we describe the sample for the study by presenting child-level characteristics such as sex, age, language spoken at home, reading behaviors, and pre-primary education. Each of these indicators is an important influencer of a child’s literacy. Table 2 presents the baseline values for these factors.

Table 2. Household Roster

	Mean	SD	Obs
Boy students	0.52	0.50	1,796

[†] Juárez & Associates is a research and consulting firm which provides targeted services to governmental agencies, institutions, community organizations, business and private sector clients. They provided training for the local firm who conducted the data collection.

[‡] DIGEDUCA stands for the Dirección General de Evaluación e Investigación Educativa de Guatemala. DIGEDUCA was responsible for adapting the EGRA into the ELGI and ESNAAJ tests.

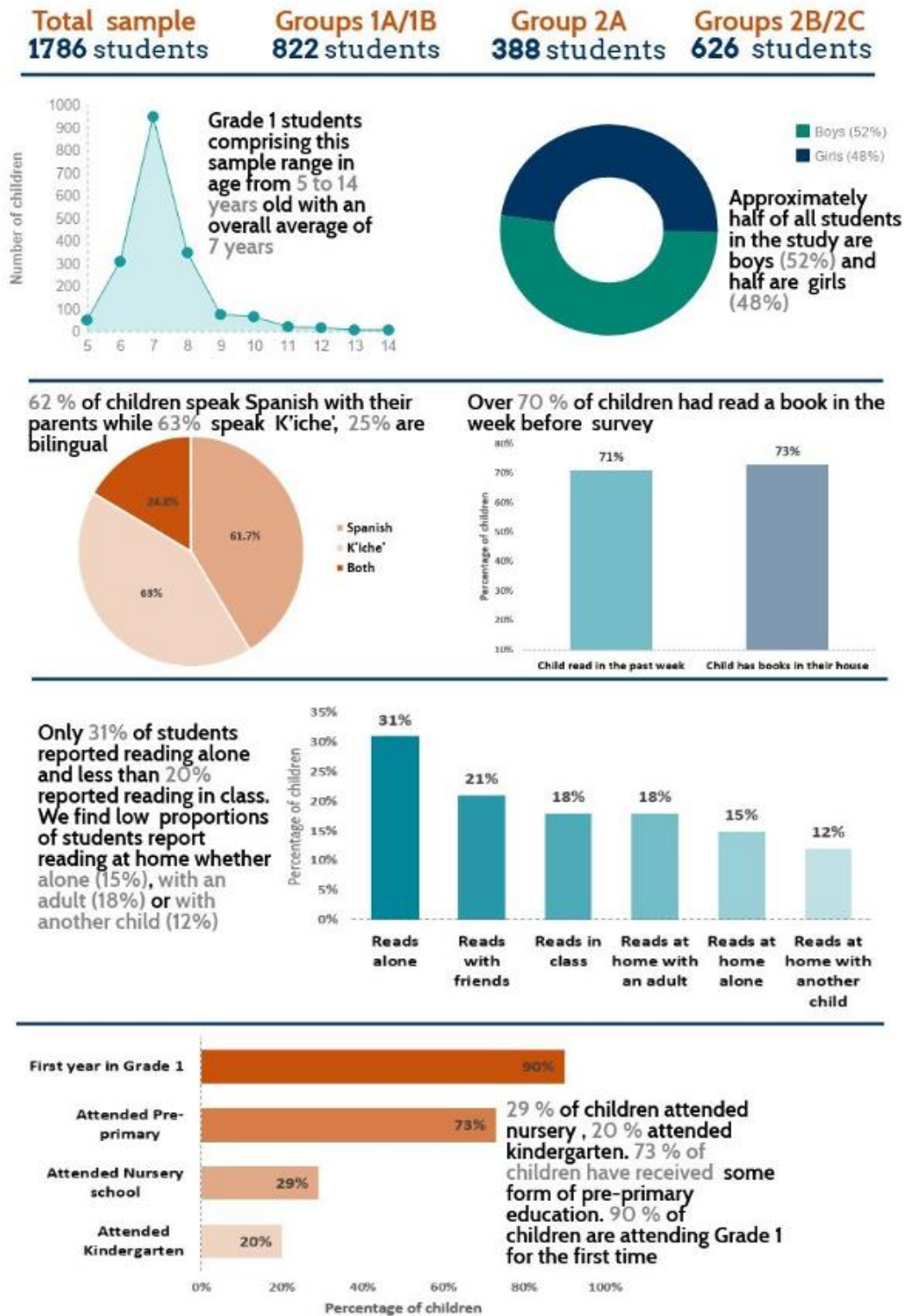
	Mean	SD	Obs
Girl students	0.48	0.50	1,796
Age	7.23	1.12	1,786
Speaks Spanish with parents	0.62	0.49	1,786
Speaks K'iche' with parents	0.63	0.48	1,786
Reads alone	0.31	0.46	1,786
Reads with friends	0.21	0.41	1,786
Reads in class	0.18	0.39	1,786
Reads at home alone	0.15	0.36	1,786
Reads at home with an adult	0.18	0.38	1,786
Reads at home with another child	0.12	0.33	1,786
Read in the past week	0.71	0.45	1,786
Books in the house	0.73	0.45	1,786
Attended kindergarten	0.20	0.40	1,786
Attended nursery school	0.29	0.45	1,786
Attended pre-primary	0.73	0.44	1,786
First year in Grade 1	0.90	0.31	1,786
Repeated a grade	0.18	0.38	1,786

The sample consists of 1,786 students: 822 in Groups 1A/1B, 338 in Group 2A, and 626 in Groups 2B/2C. Approximately half of all students in the study are boys (52%) and half are girls (48%). The sample consists of Grade 1, ranging in age from 5 to 14 years old, with an average age of 7 years. We found that equal proportions of these students speak Spanish or K'iche' with their parents (62% and 63%, respectively) and that 25% are bilingual. Few students practice reading in any capacity; only 31% reported reading alone and less than 20% reported reading in class. Low proportions of students reported reading at home, whether alone (15%), with an adult (18%), or with another child (12%). However, 71% of students did report reading in the past week, and 73% mentioned having books in their home. Lastly, we found that almost three quarters of the students in our sample attended some pre-primary education, and only 18% had repeated a grade.

Taken together, this information suggests that children in our sample are mostly at the appropriate grade for their age, with the majority having attended pre-primary education. However, the findings on reading are mixed. 71% report reading a book in the past week and 73% say they have books in their home. However, only 31% report reading alone, and only 18% report reading at home with an adult. Perhaps most surprising is that only 18% of the sampled students report reading in class. A mix of languages is spoken in the homes of students in our

sample. Oral proficiency in a language predicts children's reading comprehension and overall literacy in that language and obtaining proficiency in one's mother tongue is essential for a successful transition to literacy in a second language, when there is little or no oral language proficiency in the second language.^{6, 7, 8, 9} Figure 7 highlights some of these baseline characteristics.

Figure 7. Baseline Characteristics



Assessments

Each of the students in our sample was administered a one-on-one language assessment in both Spanish and K'iche' covering multiple reading subskills. Table 3 presents summary statistics for all students on the Spanish language assessment. Students exhibited a relatively high degree of comprehension of the oral instructions but performed poorly on most of the other subskills. Overall, on average children scored 25%[§] or below on both timed and untimed letter name recognition and letter sound knowledge modules, which suggests they are unfamiliar with the alphabetic structure of the language. Scores for initial phoneme identification were slightly higher (with an average score of 32%) but, unsurprisingly, separation of phonemes proved more difficult for students at Grade 1 (with an average score of just 16%).

Reading Spanish words was the most difficult task for students, with average scores for short word reading, familiar word reading, and nonsense word speed falling well below 15% on average. Student performance on listening comprehension was slightly better, with students scoring 33% on average, which suggests that students have a higher oral language competency. Students scored 5% on fluent passage reading and 11% on dictation, indicating that students struggle with both reading and writing in Spanish. For the subset of students who took the reading comprehension portion of the assessment,** the average score was 2%.

Students scored high on the adapted Peabody Picture Vocabulary Test (PPVT) of oral vocabulary knowledge with an average score of 82%. However, they scored low on the concept of print items with an average score of 46%, which reflects their poor preparedness to begin print instruction. The low concept of print scores could stem from several sources. Although the data indicate that 73% of students have books at home this does not mean that they are age and/or language appropriate nor does it mean that they are reading them with an adult who can show them how. For example, only 18% report reading at home with an adult and only 18% report reading in class. Concept of print typically measures a student's familiarity with print concepts which usually stems from exposure to print in various forms. These baseline results identify considerable room for improvement for the students in our sample.

[§] All scores are standardized to be out of 100%, based on the highest possible score for each subskill. This means that scores range from 0.0 to 1.0 for each subskill, regardless of the scoring method for the particular module. Subskills showing a maximum score of less than 1.0 indicate that no students answered all questions on that module correctly; the maximum is the highest score for the students in our sample.

** Students who failed the listening comprehension section of the test were unable to progress to the reading comprehension section. Of the total sample, only 138 students performed well enough to advance to the reading comprehension section of the test.

Table 3. Spanish Assessment

	Mean	SD	Min	Max	Obs
Oral Instruction Comprehension	0.67	0.24	0.00	0.92	1,784
Letter Name Recognition, Untimed	0.16	0.18	0.00	1.00	1,786
Letter Name Recognition, Timed	0.10	0.12	0.00	0.90	1,786
Letter Sound Knowledge, Untimed	0.25	0.27	0.00	1.00	1,786
Letter Sound Knowledge, Timed	0.15	0.17	0.00	0.90	1,786
Initial Phoneme Identification	0.32	0.27	0.00	1.00	1,786
Separation of Phonemes	0.16	0.18	0.05	0.90	1,786
Short Word Reading	0.14	0.29	0.00	1.00	1,786
Familiar Word Reading	0.02	0.08	0.00	0.79	1,786
Nonsense Word Speed	0.02	0.06	0.00	0.76	1,786
Listening Comprehension	0.33	0.28	0.00	1.00	1,786
Fluent Passage Reading	0.05	0.17	0.00	0.90	1,786
Reading Comprehension	0.02	0.12	0.00	1.00	138
Dictation	0.11	0.19	0.00	1.00	1,786
Oral Vocabulary Knowledge	0.82	0.25	0.00	1.00	1,786
Concept of Print	0.46	0.30	0.00	1.00	1,786

The results from the K'iche' assessment are presented in Table 4. Overall, scores resembled those from the Spanish language assessment, except for separation of phonemes, for which students scored 5% on average; and initial phoneme identification, for which students scored 47%. Again, students scored relatively poorly on letter name (16%) and sound knowledge (21%), as well as oral vocabulary knowledge (12%). Similarly, K'iche' reading skills were relatively poor, with average scores of 8% for short word reading, 1% for familiar word reading, and 1% for nonsense word speed. As with the Spanish language assessment, scores were slightly higher for the listening comprehension module (27%) but fell again for fluent passage reading (1%) and dictation (10%). Lastly, the small cohort of students who performed well enough on the listening comprehension module to progress to the reading comprehension module fared poorly. The average score was 0%, though scores ranged from 0% to 64% among those students. Compared to the Spanish oral vocabulary knowledge, students performed worse on the K'iche' oral vocabulary knowledge. The average score was 57%. Again, these results identify areas ripe for improvement, and we see that students' literacy skills are similarly undeveloped for both languages.

Table 4. K'iche' Assessment

	Mean	SD	Min	Max	Obs
Oral Instruction Comprehension	0.54	0.23	0.07	0.79	1,786
Letter Name Recognition	0.16	0.17	0.00	1.00	1,786
Letter Sound Knowledge	0.21	0.22	0.00	1.00	1,786
Oral Vocabulary Knowledge, Timed	0.12	0.15	0.00	0.82	1,786
Initial Phoneme Identification	0.47	0.24	0.05	0.95	1,786
Separation of Phonemes	0.05	0.02	0.00	0.09	1,786
Short Word Reading	0.08	0.20	0.00	0.90	1,786
Familiar Word Reading	0.01	0.06	0.00	0.69	1,786
Nonsense Word Speed	0.01	0.06	0.00	0.56	1,786
Listening Comprehension	0.27	0.24	0.00	1.00	1,786
Fluent Passage Reading	0.01	0.07	0.00	0.81	1,786
Reading Comprehension	0.00	0.03	0.00	0.64	77
Dictation	0.10	0.18	0.00	1.00	1,786
Oral Vocabulary Knowledge	0.57	0.26	0.00	1.00	1,786

We conducted a series of *t*-tests to determine whether the differences in scores between the Spanish and K'iche' assessments were statistically significant. Table 5 presents the results of this analysis. Overall, we found that—with the exception of the letter name recognition and nonsense words per minute subtasks—scores in the Spanish language assessment were statistically significantly higher than scores in the K'iche' assessment.

Table 5. Spanish Language versus K'iche' Language Comparison

Variable	Spanish		K'iche'		Balance Test			
	Mean	N	Mean	N	Diff	SE	p-value	Effect Size
Oral Instruction Comprehension	0.67	1784	0.54	1786	-0.14	0.01	0.00	-0.14
Letter Name Recognition	0.16	1786	0.16	1786	0.00	0.01	0.90	0.00
Letter Sound Knowledge	0.25	1786	0.21	1786	-0.04	0.01	0.00	-0.04
Initial Phoneme Identification	0.32	1786	0.47	1786	0.14	0.01	0.00	0.14
Separation of Phonemes	0.16	1786	0.05	1786	-0.11	0.00	0.00	-0.11
Short Word Reading	0.14	1786	0.08	1786	-0.07	0.01	0.00	-0.07
Familiar Word Reading	0.02	1786	0.01	1786	-0.01	0.00	0.00	-0.01
Nonsense Word Speed	0.02	1786	0.01	1786	0.00	0.00	0.09	0.00
Listening Comprehension	0.33	1786	0.27	1786	-0.06	0.01	0.00	-0.06
Fluent Passage Reading	0.05	1786	0.01	1786	-0.03	0.00	0.00	-0.03
Reading Comprehension	0.02	138	0.00	77	-0.02	0.00	0.00	-0.02

Variable	Spanish		K'iche'		Balance Test			
	Mean	N	Mean	N	Diff	SE	p-value	Effect Size
Dictation	0.11	1786	0.10	1786	-0.01	0.01	0.01	-0.01
Oral Vocabulary Knowledge	0.82	1786	0.57	1786	-0.24	0.01	0.00	-0.24

We further examine the nuance of these results by running univariate regressions to assess the correlation between student characteristics and their literacy skills. Specifically, we test for correlations by gender, municipality, pre-school education, and having breakfast in their home (a proxy for household economic status). Tables 10 and 11 in Annexes 1 and 2 present the results for the Spanish and K'iche' language assessments, respectively. Overall, we find gender has no bearing on students' literacy skills while their locality (based on their municipality) significantly affects performance on most literacy skills and having breakfast at home and one's pre-primary education influence literacy skills to some degree though less consistently than locality. These trends are similar for both Spanish and K'iche' literacy skills. The largest correlation appears between municipality of residence and students' OVK and concept of print for K'iche' literacy while Table 11 shows that municipality is highly correlated with a broader range of Spanish literacy skills including oral instructions, untimed letter sound knowledge, listening comprehension, and concept of print.

Baseline Equivalence

In this section, we present the results of testing for any imbalances across treatment groups. In a randomized experiment, different groups are expected to be comparable—i.e., balanced across observed and unobserved characteristics—so that the average differences in outcomes between the two groups at the end of the study can be attributed to the intervention. However, there is a chance (albeit small, if the sample size is large) that the groups are not balanced. We conducted tests of differences between observed characteristics (before treatment) corresponding to the group comparisons presented above. Even though allocation of schools to treatment conditions is randomly assigned, we wanted to test that the randomization successfully produced samples that were comparable across groups at baseline.

Table 6 describes the results of the Spanish assessment broken down by skill and treatment status for Group 1A schools (i.e., schools that continued with business as usual and will receive the SF program in 2018) and Group 1B schools (i.e., schools that will receive the full package of interventions in 2018: the SF, KC, and JdL programs and the CSR-AIR Assessment Toolkit). Across all assessed skills, we found no evidence of a significant difference between Group 1A and Group 1B students as the p-values of the difference were all well above 0.05 (column 7). These results suggest the randomization was successful in creating balance across these domains for the Group 1 schools.

In other words, all children assessed in Group 1 schools scored similarly on each of the Spanish language skills; students uniformly struggled with the timed modules and the word and passage reading modules. This is important as it shows us that the groups were not different from each other in terms of children’s reading abilities before the start of the program, allowing us to attribute any endline differences between the two groups to the program.

Table 6. Group 1A versus Group 1B Spanish Assessment

Variables	Group 1A		Group 1B		Balance Test		
	Mean	N1	Mean	N2	Diff	SE	p-value
Oral Instruction Comprehension	0.65	403	0.64	419	-0.01	0.03	0.78
Letter Name Recognition, Untimed	0.16	403	0.13	419	-0.03	0.02	0.16
Letter Name Recognition, Timed	0.11	403	0.09	419	-0.02	0.01	0.20
Letter Sound Knowledge, Untimed	0.23	403	0.20	419	-0.03	0.03	0.33
Letter Sound Knowledge, Timed	0.14	403	0.12	419	-0.02	0.02	0.30
Initial Phoneme Identification	0.29	403	0.31	419	0.02	0.02	0.32
Separation of Phonemes	0.14	403	0.16	419	0.01	0.02	0.50
Short Word Reading	0.12	403	0.09	419	-0.03	0.03	0.32
Familiar Word Reading	0.02	403	0.01	419	-0.01	0.01	0.18
Nonsense Word Speed	0.02	403	0.01	419	-0.01	0.01	0.21
Listening Comprehension	0.29	403	0.27	419	-0.02	0.04	0.67
Fluent Passage Reading	0.04	403	0.05	419	0.01	0.02	0.72
Reading Comprehension	0.03	403	0.01	419	-0.01	0.01	0.17
Dictation	0.10	403	0.09	419	-0.00	0.02	0.95
Oral Vocabulary Knowledge	0.79	403	0.81	419	0.02	0.03	0.49
Concept of Print	0.43	403	0.41	419	-0.02	0.04	0.63

Notes: Bold denotes significance at the 0.05 level. Standard errors are clustered at the school level.

Table 7 presents the average scores from the baseline Spanish assessment by skill and treatment status for Group 2A, 2B and 2C schools. Schools in Group 2A will receive the SF and KC programs in 2018, 2B will receive the SF and KC programs plus the JdL program, and Group 2C schools will receive the same programs in addition to the AIR-CRS Assessment Toolkit. We find evidence that the three groups are not balanced across more than half of all assessed skills. However, while these differences are statistically significant, the majority are minuscule in practical significance. Seven of the ten significant skills differ across the three groups by less than 5 percentage points. The groups exhibit a larger divergence on untimed letter name recognition, untimed letter sound knowledge, short word reading, and listening comprehension skills, though each of these

differences are within 10 percentage points. Estimation techniques in future rounds can be used to correct for these initial imbalances across groups at balance.

Table 7. Group 2A versus Groups 2B and 2C Spanish Assessment

Variables	Group 2A		Group 2B		Group 2C		Balance Test
	Mean	N1	Mean	N2	Mean	N3	p-value
Oral Instruction Comprehension	0.70	338	0.70	287	0.71	339	0.86
Letter Name Recognition, Untimed	0.20	338	0.15	287	0.14	339	0.00
Letter Name Recognition, Timed	0.12	338	0.10	287	0.09	339	0.00
Letter Sound Knowledge, Untimed	0.32	338	0.25	287	0.23	339	0.00
Letter Sound Knowledge, Timed	0.17	338	0.15	287	0.15	339	0.12
Initial Phoneme Identification	0.35	338	0.33	287	0.33	339	0.45
Separation of Phonemes	0.19	338	0.17	287	0.16	339	0.12
Short Word Reading	0.23	338	0.14	287	0.14	339	0.00
Familiar Word Reading	0.04	338	0.03	287	0.02	339	0.02
Nonsense Word Speed	0.03	338	0.02	287	0.01	339	0.01
Listening Comprehension	0.40	338	0.32	287	0.37	339	0.00
Fluent Passage Reading	0.07	338	0.03	287	0.04	339	0.02
Reading Comprehension	0.05	338	0.02	287	0.02	339	0.00
Dictation	0.15	338	0.11	287	0.12	339	0.01
Oral Vocabulary Knowledge	0.82	338	0.84	287	0.83	339	0.53
Concept of Print	0.50	338	0.47	287	0.49	339	0.55

Notes: Bold denotes significance at the 0.05 level. Standard errors are clustered at the school level.

Results for the K'iche' assessment for Group 1 schools are presented in Table 8. We found evidence of a significant difference in only one subskill: oral vocabulary knowledge (timed). In the timed exercise the average oral vocabulary knowledge score was higher for children in Group 1B (18%) than for Group 1A children (11%), indicating that children in Group 1A struggled slightly more with orally associating images and words in K'iche'. The difference is also reflected in the PPVT oral vocabulary knowledge test. The average for Group 1A children is 61% compared to 65% for Group 1B children. This difference is significant at the 10% level. We did not detect evidence of significant differences for any of the other assessed skills.

Table 8. Group 1A versus Group 1B K’iche’ Assessment

Variables	Group 1A		Group 1B		Balance Test		
	Mean	N1	Mean	N2	Diff	SE	p-value
Oral Instruction Comprehension	0.57	403	0.59	419	0.03	0.03	0.33
Letter Name Recognition	0.16	403	0.14	419	-0.03	0.02	0.11
Letter Sound Knowledge	0.21	403	0.18	419	-0.03	0.03	0.31
Oral Vocabulary Knowledge, Timed	0.11	403	0.18	419	0.06	0.02	0.00
Initial Phoneme Identification	0.46	403	0.45	419	-0.01	0.03	0.75
Separation of Phonemes	0.05	403	0.05	419	0.00	0.00	0.95
Short Word Reading	0.07	403	0.06	419	-0.01	0.02	0.56
Familiar Word Reading	0.01	403	0.01	419	-0.01	0.01	0.26
Nonsense Word Speed	0.01	403	0.01	419	-0.01	0.01	0.28
Listening Comprehension	0.30	403	0.34	419	0.04	0.03	0.17
Fluent Passage Reading	0.01	403	0.00	419	-0.01	0.01	0.31
Reading Comprehension	0.00	403	0.00	419	-0.00	0.00	0.68
Dictation	0.08	403	0.08	419	-0.00	0.02	0.98
Oral Vocabulary Knowledge	0.61	403	0.65	419	0.04	0.03	0.10

Notes: Bold denotes significance at the 0.05 level. Standard errors are clustered at the school level.

We also detected a significant difference in letter name recognition, letter sound knowledge, and oral vocabulary knowledge (timed) scores between children in Groups 2A, 2B and 2C (Table 9). Based on the averages reported in columns 1 and 3, children in Groups 2A and 2C struggled more with orally associating images and words in K’iche’ than their counterparts in Group 2B while Groups 2B and 2C performed worse than students in Group 2A on letter name and letter sound tasks. These differences, however, are not reflected by the PPVT oral vocabulary knowledge, phoneme or word reading tests. We did not detect evidence of significant differences between the three groups in any other assessed skills.

Table 9. Group 2A versus Group 2B and 2C K’iche’ Assessment

Variables	Group 2A		Group 2B		Group 2C		Balance Test
	Mean	N1	Mean	N2	Mean	N3	p-value
Oral Instruction Comprehension	0.49	338	0.51	287	0.51	339	0.43
Letter Name Recognition	0.21	338	0.14	287	0.15	339	0.00
Letter Sound Knowledge	0.26	338	0.19	287	0.21	339	0.00
Oral Vocabulary Knowledge, Timed	0.08	338	0.12	287	0.08	339	0.00

Variables	Group 2A		Group 2B		Group 2C		Balance Test
	Mean	N1	Mean	N2	Mean	N3	p-value
Initial Phoneme Identification	0.48	338	0.48	287	0.48	339	0.96
Separation of Phonemes	0.05	338	0.05	287	0.05	339	0.34
Short Word Reading	0.10	338	0.08	287	0.08	339	0.59
Familiar Word Reading	0.02	338	0.02	287	0.01	339	0.14
Nonsense Word Speed	0.02	338	0.02	287	0.01	339	0.09
Listening Comprehension	0.22	338	0.23	287	0.24	339	0.63
Fluent Passage Reading	0.02	338	0.02	287	0.01	339	0.32
Reading Comprehension	0.01	338	0.00	287	0.00	339	0.72
Dictation	0.12	338	0.10	287	0.13	339	0.23
Oral Vocabulary Knowledge	0.52	338	0.53	287	0.53	339	0.95
Concept of Print	0.50	338	0.47	287	0.49	339	0.55

Notes: Bold denotes significance at the 0.05 level. Standard errors are clustered at the school level.

Discussion

Baseline results show that randomization of the schools was successful in creating balanced subgroups throughout the sample. This sets the stage for a randomized controlled trial, in which we can attribute the differences between the groups – if any – at baseline to the program, which in turn allows us to test the impact of the program. That said, we did find evidence of a significant difference in oral vocabulary knowledge scores on K’iche’ assessments in both Group 1 and Group 2 treatment groups. However, these statistical differences represent less than 5% of all the compared baseline scores, indicating that randomization led to statistically equivalent samples and did not introduce bias into our analyses. Since baseline balance was achieved, any statistically significant differences observed between treatment arms at endline can likely be attributed to the program.

Overall, the baseline results show that children in all groups had low reading scores. Students’ oral language skills were much more developed than their letter and sound knowledge and reading abilities in both languages, even though books seem to be readily accessible for most of these students (although it is unclear in which languages they are available). These results suggest there is ample room for improvement for these students, which may be achieved through some combination of the interventions that will be evaluated in this study. The AIR-CRS toolkits aim to specifically address the pre-reading skills (concept of print and oral language) which are clearly still being developed at the start of the school year but then to support

development of increasingly difficult skills such as decoding and reading comprehension. We would expect to see scores on these constructs, which are extremely low in both languages at baseline, improve dramatically by endline if teachers are implementing the toolkits (including remedial activities) with fidelity. However, fidelity of implementation reports show low levels of fidelity thus this will have an effect on size of the reading outcomes.

In general, one might expect scores on oral language proficiency subskills to be higher for K'iche' than Spanish, however, our results suggest that students have higher oral vocabulary knowledge in Spanish than in K'iche'. In Totonicapán it is more and more common to find schools with more Spanish monolingual students than K'iche' speaking students. The language and literacy landscape has been changing in recent years due to multiple factors one of which is the lack of true bilingual programming in schools, as well as more and more parents not teaching their children K'iche' because they are placing a higher value on Spanish¹⁰.

Baseline data also show that children scored relatively poorly on concept of print. This is also to be expected given that relatively few children reported reading at home (18% with an adult, 15% alone, or 12% with another child) or in school (20%). Although 73% of children report having books in their home, it is unclear if the books are level or context appropriate for grade one children or if they are in K'iche' or Spanish. The concept of print finding is somewhat surprising, however, considering the high number of students (73%) that attended some form of pre-primary education. One would expect that children would be exposed to books in various forms during pre-primary education, but further study is needed to determine to what extent books are being used in the pre-primary classroom.

The baseline results followed expectations in terms of skills we would expect first grade students to do better on and those we would expect them to struggle with. In general, we would expect to see higher scores on the oral language related skills and lower scores on the more advanced skills of decoding and reading comprehension. Phonological awareness and letter name scores are also relatively low which also suggests that the preprimary curriculum (if indeed a curriculum is being followed) does not seem to address the necessary early reading skills needed to prepare children for entry into first grade.

Limitations

Overall, the study randomization delivered a balanced baseline sample. Nonetheless, these statistics should be interpreted with some caution. Ideally, baseline data collection occurs prior to program implementation to better understand how children perform before being targeted with any intervention. Unfortunately, some treatment schools in our sample began

implementing the program before students were assessed. This means that the balance we see across our sample could be a result of the intervention bringing up students' scores in the treatment group, rather than actual balance. However, the program began only 1 month prior to the assessments, and we do not believe the interventions could have influenced student reading outcomes in this short amount of time. For this reason, we believe the fidelity of randomization was not compromised, and that our results show actual baseline values balanced across groups.

Additionally, the extra-curricular StG program—which was implemented in 15 schools in 2017 and will be operating in an additional unknown number of schools in 2018—may pose a threat to the validity of our impact results. Since schools were purposively assigned to this program (rather than randomly), a school's treatment status may influence take-up of the StG program. In other words, treatment schools may be less inclined to add another program to their repertoire, while control schools implementing fewer or no additional programs may more readily agree to participate. This may mean that we underestimate the true impact of the reading interventions underpinning this study if the StG program improves student outcomes but is not factored into our analysis. Conversely, treatment schools may be more inclined to add on another program than control schools, leading us to overestimate the impact of the reading programs on student literacy (if the StG positively influences students' reading skills). To address this potential complication, we will estimate impacts for all schools in our sample and compare the results to a sample that excludes the schools that received StG in 2017 and 2018. This will demonstrate whether the overall impact results are affected by this non-randomly assigned program.

Conclusion

The primary purpose of this baseline report is to describe the student sample prior to receiving the programs, and to demonstrate the equivalence of the treatment and comparison groups. Describing the sample at baseline helps stakeholders assess if they have accurately targeted the students they want to benefit from the program. It also helps stakeholders understand where students need the most assistance, and how best to design the program to meet beneficiary needs.

Students in the study reported low levels of reading behavior both in school and at home, although the majority mentioned having access to books at home and three quarters of students reported reading in the past week. These results suggest that children have been exposed to books and written text but may not be getting enough instruction on how to

improve their reading skills. Teachers at the pre-primary level may lack the necessary skills to properly teach the necessary foundational reading skills as this is not something that is directly covered in the pre- or in-service teacher training curriculum. In addition, they often have no way of identifying which reading skills students are struggling with, and even if they do identify a challenge, they often don't know a variety of different ways to teach a skill. Moreover, teachers receive little to no coaching support from the Ministry of Education or other entities to teach reading in a bilingual environment, which requires a higher level of understanding of the nuances of multilingual reading instruction. The literacy interventions measured in this study aim to help children develop early bilingual literacy skills in both Spanish and K'iche'. The CLD toolkits aim to address many of these issues through training, coaching support, and toolkits that help teachers determine which specific skills students need additional support with, followed by remedial activities that help teachers provide targeted instruction. If these toolkits are implemented with fidelity, we would expect to see strong improvements in students' baseline scores in both K'iche' and Spanish.

¹ American Institutes for Research. (2015). *Are Latin American children's reading skills improving? Highlights from the Second and Third Regional Comparative and Explanatory Studies (SERCE & TERCE)*. Washington, DC: American Institutes for Research.

² Deputies of the National Constituent Assembly. (1986). *The Constitution of the Republic of Guatemala of 1985*.

³ Evans, D. K., & Popova, A. (2016). What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews. *World Bank Research Observer*, 31(2), 242–70. Retrieved from <https://openknowledge.worldbank.org/handle/10986/29308>

⁴ World Bank. (2018). *World development report 2018: Learning to realize education's promise*. Washington, DC: Author. doi:10.1596/978-1-4648-1096-1.

⁵ Trudell, B. (2016). *The impact of language policy and practice on children's learning: Evidence from Eastern and Southern Africa*. Nairobi: UNICEF. Retrieved from [https://www.unicef.org/esaro/UNICEF\(2016\)LanguageandLearning-FullReport\(SingleView\).pdf](https://www.unicef.org/esaro/UNICEF(2016)LanguageandLearning-FullReport(SingleView).pdf).

⁶ Cummins, J. (1991). Interdependence of first-and second-language proficiency in bilingual children. *Language processing in bilingual children*, 70-89.

- ⁷ Koda, K. (2008). Impacts of prior literacy experience on second language learning to read. In K. Koda & A. M. Zehler (Eds.), *Learning to read across languages: Cross-linguistic relationships in first- and second language literacy development* (pp. 68–96). New York, NY: Routledge.
- ⁸ August, D., & Shanahan, T. (2006). Developing literacy in second-language learners. Report of the National Literacy Panel on Language-Minority Children and Youth. Mahwah, NJ: Lawrence Erlbaum Associates Publisher.
- ⁹ Alidou, H., Boly, A., Brock-Utne, B., Diallo, Y. S., Heugh, K., & Wolff, H. E. (2006). Optimizing learning and education in Africa—the language factor. *Paris: ADEA*.
- ¹⁰ Stone, R. (2016). *Guatemala Rapid Situational Analysis Report*. American Institutes for Research.

Annex 1

Table 10: Ki'iche Correlations Language Assessments

Variables	Oral Inst. Comp.	Letter Name Recog., Un-timed	Letter Name Recog., Time	Letter Sound, Un-timed	Letter Sound, Timed	Initial Phon.	Sep. of Phon.	Short Word	Familiar Words	Nonsense Words	Fluent Passage	Reading Comp.	Dictation	OVK	Concept of Print
Municipality															
Momostenango	-0.06* (0.01)	0.01 (0.01)	0.04*** (0.01)	-0.05*** (0.01)	0.02* (0.01)	0.00* (0.00)	0.04*** (0.01)	0.01*** (0.00)	0.01*** (0.00)	-0.04*** (0.01)	0.01*** (0.00)	0.00 (0.00)	0.06*** (0.01)	-0.09*** (0.01)	0.07*** (0.01)
San Andres Xecul	-0.18*** (0.03)	0.04 (0.03)	0.07* (0.04)	-0.03 (0.02)	0.12*** (0.03)	0.00 (0.00)	0.04 (0.03)	0.01 (0.00)	0.01 (0.01)	-0.17*** (0.03)	0.01 (0.10)	0.01 (0.01)	0.03 (0.02)	-0.09*** (0.03)	0.06 (0.04)
San Bartolo Aguas Calient	-0.07** (0.03)	0.04** (0.02)	0.02 (0.02)	-0.06*** (0.01)	0.03 (0.03)	0.00 (0.00)	-0.02 (0.02)	-0.01 (0.00)	-0.01 (0.01)	-0.04 (0.03)	-0.01 (0.00)	-0.00 (0.00)	-0.03** (0.01)	-0.09*** (0.03)	0.09*** (0.03)
Santa Lucia La Reforma	0.02 (0.02)	-0.02 (0.01)	-0.06*** (0.01)	0.02* (0.01)	-0.02 (0.02)	-0.00 (0.00)	-0.04*** (0.01)	-0.01*** (0.00)	-0.01* (0.00)	-0.04** (0.02)	-0.01*** (0.00)	-0.00*** (0.00)	-0.05*** (0.01)	0.01 (0.02)	-0.07*** (0.03)
Santa Maria Chiquimula	0.05*** (0.01)	-0.03*** (0.01)	-0.05*** (0.01)	0.03*** (0.01)	0.00 (0.01)	0.00 (0.00)	-0.04*** (0.01)	-0.01** (0.00)	-0.01*** (0.00)	0.01 (0.01)	-0.00 (0.00)	-0.00 (0.00)	-0.04*** (0.01)	0.07*** (0.01)	-0.12*** (0.02)
Totonicapan	0.08*** (0.01)	0.02* (0.01)	0.03** (0.01)	0.06*** (0.01)	0.01 (0.01)	-0.01*** (0.00)	0.01 (0.01)	-0.00 (0.00)	-0.00 (0.00)	0.13*** (0.02)	-0.01* (0.00)	-0.00 (0.00)	-0.00 (0.01)	0.09*** (0.01)	0.06*** (0.02)
Gender															
Male	-0.01 (0.01)	-0.00 (0.01)	-0.01 (0.01)	-0.01 (0.01)	0.02 (0.01)	-0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	-0.01 (0.01)	0.00 (0.00)	0.00* (0.00)	-0.01 (0.01)	-0.00 (0.01)	-0.00 (0.01)
Female	0.01 (0.01)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)	0.03 (0.01)	0.00 (0.00)	-0.00 (0.01)	-0.00 (0.00)	-0.00 (0.00)	0.01 (0.01)	-0.00 (0.00)	-0.00* (0.00)	0.01 (0.01)	0.00 (0.01)	0.00 (0.01)
Breakfast at home	0.03 (0.02)	0.04*** (0.01)	0.04** (0.02)	-0.01 (0.01)	0.04 (0.02)	-0.00 (0.00)	0.03* (0.02)	0.01* (0.00)	0.01 (0.00)	0.05** (0.02)	0.01** (0.00)*	0.00*** (0.00)	0.01 (0.02)	0.04* (0.02)	0.08*** (0.03)
Early Childhood Education															

Variables	Oral Inst. Comp.	Letter Name Recog., Un-timed	Letter Name Recog., Time	Letter Sound, Un-timed	Letter Sound, Timed	Initial Phon.	Sep. of Phon.	Short Word	Familiar Words	Nonsense Words	Fluent Passage	Reading Comp.	Dictation	OVK	Concept of Print
Attended Kindergarten	-0.02 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.05 (0.01)	-0.00*** (0.00)	-0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	-0.01 (0.01)	-0.00 (0.00)	-0.00 (0.00)	0.02** (0.01)	0.02 (0.01)	-0.03 (0.02)
Attended Parvulos	-0.02 (0.01)	0.02** (0.01)	0.01 (0.01)	-0.01 (0.01)	0.06 (0.01)	-0.00*** (0.00)	0.02 (0.01)	0.00 (0.00)	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	0.02** (0.01)	-0.00 (0.01)	0.01 (0.02)
Attended Pre-Primary	0.01 (0.01)	0.00 (0.01)	0.03*** (0.01)	-0.02 (0.01)	0.07 (0.01)	-0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	0.05*** (0.01)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.01)	-0.02 (0.01)	0.03 (0.02)
<i>Observations</i>	<i>1,784</i>	<i>1,786</i>	<i>1,786</i>	<i>1,786</i>	<i>1,786</i>	<i>1,786</i>	<i>1,786</i>	<i>1,786</i>	<i>1,786</i>	<i>1,786</i>	<i>1,786</i>	<i>1,786</i>	<i>138</i>	<i>1,786</i>	<i>1,786</i>

Annex 2

Table 11: Spanish Correlations Language Assessments

Variables	Oral Inst. Comp.	Letter Name Recog., Un-timed	Letter Name Recog., Time	Letter Sound, Un-timed	Letter Sound, Timed	Initial Phon.	Sep. of Phon.	Short Word	Familiar Words	Nonsense Words	Fluent Passage	Reading Comp.	Dictation	OVK	Concept of Print
Municipality															
Momostenango	0.08*** (0.01)	0.01 (0.01)	0.01 (0.01)	0.06*** (0.01)	0.03*** (0.01)	-0.00 (0.01)	0.01 (0.01)	0.07*** (0.02)	0.02*** (0.00)	0.01*** (0.00)	0.11*** (0.01)	0.02*** (0.01)	0.02*** (0.01)	0.05*** (0.01)	0.02** (0.01)
San Andres Xecul	0.06* (0.03)	0.07** (0.03)	0.01 (0.02)	0.20*** (0.04)	0.05** (0.02)	0.13*** (0.03)	0.04 (0.02)	0.08* (0.04)	0.00 (0.01)	0.01 (0.01)	0.09** (0.04)	0.01 (0.02)	0.04* (0.02)	0.03 (0.03)	0.08*** (0.01)
San Bartolo Aguas Calient	0.08*** (0.02)	0.02 (0.02)	0.01 (0.01)	0.03 (0.03)	0.02 (0.02)	0.03 (0.02)	0.01 (0.02)	0.06* (0.03)	-0.01 (0.01)	-0.01* (0.00)	0.13*** (0.03)	-0.01 (0.02)	-0.02** (0.01)	0.01 (0.02)	0.04** (0.01)
Santa Lucia La Reforma	-0.11*** (0.02)	-0.03** (0.01)	-0.02* (0.01)	-0.11*** (0.02)	-0.05*** (0.01)	0.02 (0.03)	0.01 (0.02)	-0.07*** (0.02)	-0.01*** (0.00)	-0.01** (0.00)	-0.16*** (0.02)	-0.04*** (0.01)	-0.02*** (0.01)	-0.05*** (0.01)	-0.05*** (0.02)
Santa Maria Chiquimula	-0.11*** (0.01)	-0.04*** (0.01)	-0.03*** (0.01)	-0.09*** (0.01)	-0.05*** (0.01)	-0.05*** (0.01)	-0.01 (0.01)	-0.11*** (0.01)	-0.01*** (0.00)	-0.01*** (0.00)	-0.16*** (0.01)	0.00 (0.01)	-0.02*** (0.01)	-0.06*** (0.01)	-0.06*** (0.01)
Totonicapan	0.06*** (0.01)	0.03*** (0.01)	0.03*** (0.01)	0.04** (0.02)	0.04*** (0.01)	0.02 (0.02)	-0.02** (0.01)	0.05*** (0.02)	0.00 (0.01)	0.00 (0.00)	0.08*** (0.02)	-0.02* (0.01)	0.00 (0.01)	0.02* (0.01)	0.06*** (0.01)
Gender															
Male	-0.01 (0.01)	-0.01 (0.01)	0.01** (0.01)	-0.02 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.01* (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	0.02 (0.01)	-0.01 (0.01)	-0.00 (0.01)	0.00 (0.01)	-0.01 (0.01)
Female	0.01 (0.01)	0.01 (0.01)	-0.01** (0.01)	0.02 (0.01)	0.00 (0.01)	0.00 (0.01)	0.01* (0.01)	-0.00 (0.01)	-0.00 (0.00)	-0.00 (0.00)	-0.02 (0.01)	0.01 (0.01)	0.00 (0.01)	-0.00 (0.01)	0.01 (0.01)
Breakfast at home	0.07*** (0.02)	0.03* (0.01)	0.03** (0.01)	0.05** (0.02)	0.02* (0.01)	0.02 (0.03)	0.01 (0.02)	0.06** (0.02)	0.01 (0.01)	0.01 (0.01)	0.04 (0.03)	0.02* (0.01)	0.01 (0.01)	0.02 (0.02)	0.03* (0.02)
Early Childhood Education															

Variables	Oral Inst. Comp.	Letter Name Recog., Un-timed	Letter Name Recog., Time	Letter Sound, Un-timed	Letter Sound, Timed	Initial Phon.	Sep. of Phon.	Short Word	Familiar Words	Nonsense Words	Fluent Passage	Reading Comp.	Dictation	OVK	Concept of Print
Attended Kindergarten	-0.03* (0.01)	0.01 (0.01)	0.00 (0.01)	-0.00 (0.02)	-0.01 (0.01)	-0.02 (0.02)	-0.01 (0.01)	-0.01 (0.02)	0.00 (0.01)	0.00 (0.00)	-0.04** (0.02)	-0.01 (0.01)	0.00 (0.01)	0.02* (0.01)	0.02 (0.01)
Attended Parvulos	0.01 (0.01)	0.02** (0.01)	0.02** (0.01)	0.03** (0.01)	0.02* (0.01)	-0.01 (0.01)	-0.00 (0.01)	0.04** (0.02)	0.01* (0.00)	0.01** (0.00)	0.01 (0.01)	-0.01 (0.01)	0.02** (0.01)	0.03*** (0.01)	0.02* (0.01)
Attended Pre-Primary	0.06*** (0.01)	0.00 (0.01)	0.01 (0.01)	0.05*** (0.01)	0.03*** (0.01)	-0.00 (0.01)	-0.01 (0.01)	0.05*** (0.01)	0.00 (0.00)	0.00 (0.00)	0.09*** (0.01)	-0.00 (0.01)	0.00 (0.01)	0.03*** (0.01)	0.02* (0.01)
Observations	1,784	1,786	1,786	1,786	1,786	1,786	1,786	1,786	1,786	1,786	1,786	1,786	138	1,786	1,786

References

1. * American Institutes for Research. (2015). *Are Latin American children's reading skills improving? Highlights from the Second and Third Regional Comparative and Explanatory Studies (SERCE & TERCE)*. Washington, DC: American Institutes for Research.
2. * Deputies of the National Constituent Assembly. (1986). *The Constitution of the Republic of Guatemala of 1985*.
3. * Evans, D. K., & Popova, A. (2016). What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews. *World Bank Research Observer*, 31(2), 242–70. Retrieved from <https://openknowledge.worldbank.org/handle/10986/29308>
4. * World Bank. (2018). *World development report 2018: Learning to realize education's promise*. Washington, DC: Author. doi:10.1596/978-1-4648-1096-1.
5. * Trudell, B. (2016). *The impact of language policy and practice on children's learning: Evidence from Eastern and Southern Africa*. Nairobi: UNICEF. Retrieved from [https://www.unicef.org/esaro/UNICEF\(2016\)LanguageandLearning-FullReport\(SingleView\).pdf](https://www.unicef.org/esaro/UNICEF(2016)LanguageandLearning-FullReport(SingleView).pdf).
6. * Cummins, J. (1991). Interdependence of first-and second-language proficiency in bilingual children. *Language processing in bilingual children*, 70-89.
7. * Koda, K. (2008). Impacts of prior literacy experience on second language learning to read. In K. Koda & A. M. Zehler (Eds.), *Learning to read across languages: Cross-linguistic relationships in first- and second language literacy development* (pp. 68–96). New York, NY: Routledge.
8. * August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners. Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum Associates Publisher.
9. * Alidou, H., Boly, A., Brock-Utne, B., Diallo, Y. S., Heugh, K., & Wolff, H. E. (2006). *Optimizing learning and education in Africa—the language factor*. Paris: ADEA.

10. * Stone, R. (2016). *Guatemala Rapid Situational Analysis Report*. American Institutes for Research.



Established in 1946, the American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research on important social issues and delivers technical assistance, both domestically and internationally, in the areas of education, health, and workforce productivity.

MAKING RESEARCH RELEVANT

AMERICAN INSTITUTES FOR RESEARCH
1000 Thomas Jefferson Street NW
Washington, DC 20007-3835 | 202.403.5000
www.air.org

LOCATIONS

Domestic: Washington, DC (HQ) | Monterey, Sacramento, and San Mateo, CA | Atlanta, GA | Honolulu, HI | Chicago and Naperville, IL | Indianapolis, IN | Metairie, LA | Waltham, MA | Frederick and Rockville, MD | Chapel Hill, NC | New York, NY | Columbus, OH | Cayce, SC | Austin, TX | Reston, VA

International: El Salvador | Ethiopia | Haiti | Honduras | Zambia