



**USAID**



# Long-Term Impact Evaluation of the Amazonía Lee Reading Intervention in Peru Final Report

**January 4, 2021**

This report was prepared for the United States Agency for International Development, under Contract/Cooperative Agreement Order Number AID-OAA-A-16-00025. It was prepared independently by Larissa Campuzano, Camila Fernández, Ivonne Padilla, and Nancy Murray.



**PULTE INSTITUTE**  
FOR GLOBAL DEVELOPMENT



**Mathematica**

---

## Acknowledgments

Undertaken by Mathematica Policy Research, this report was commissioned by USAID’s Latin America and Caribbean Bureau (LAC) Office of Regional Sustainable Development (RSD) Education Team and the USAID Global Development Lab (USAID/LAB) Office of Evaluation and Impact Assessment through the Pulte Institute for Global Development at the University of Notre Dame.

This publication was produced for review by the United States Agency for International Development. It was prepared by Mathematica Policy Research, with support from the Notre Dame Pulte Institute for Global Development. This report was authored by Larissa Campuzano, Camila Fernández, Ivonne Padilla, and Nancy Murray.

Funding for the report was provided by USAID LAC/RSD through ERIE-MERL.

### Expanding the Reach of Impact Evaluation

Expanding the Reach of Impact Evaluation (ERIE) is an initiative of the Monitoring, Evaluation, Research, and Learning Innovations (MERLIN) Program through the U.S. Global Development Lab (USAID/LAB) in partnership with the Bureau for Policy, Planning, and Learning. ERIE is an approach to conducting retrospective long-term impact evaluations of completed USAID interventions. These evaluations will leverage and build on existing program data to either assess if the observed short-term impacts are sustained, or to investigate results which might only be expected to emerge over a long-term horizon. This approach to program design and implementation for USAID and its partners is an operational innovation championed and pioneered by the Lab and being tested throughout Operating Units within the Agency. ERIE is led by the Notre Dame Pulte Institute for Global Development in partnership with AidData, the Geo-Spatial Impact Evaluation group, the Center for Effective Global Action, and Mathematica Policy Research.

### Suggested Citation

Campuzano, L., C. Fernandez, I. Padilla and N. Murray. "Long-Term Impact Evaluation of the *Amazonía Lee* Reading Intervention in Peru: Final Report." Submitted to the U.S. Agency for International Development. Washington, DC: Mathematica Policy Research and the Pulte Institute for Global Development, January 2021.

---

## Contents

Acknowledgments.....	ii
Executive Summary .....	v
A. Purpose and background .....	v
B. LTIE aims, design and methods .....	v
C. Summary of findings and conclusions .....	vi
D. Lessons learned.....	vii
I. Introduction .....	1
II. Background and context on the evaluation of <i>Amazonía Lee</i> .....	3
A. <i>Amazonía Lee</i> impact evaluation .....	5
B. Summary of findings from the impact evaluation of <i>Amazonía Lee</i> .....	7
1. Implementation and intermediate outcome findings.....	7
2. Impacts on student-level outcomes .....	8
III. Long-Term Impact Evaluation Design .....	9
A. Design and evaluation questions.....	9
B. Data sources, reading measures, and analysis sample.....	9
C. Analytic strategy.....	11
D. Limitations.....	13
IV. Findings .....	15
A. Short-term impacts.....	16
B. Long-term impacts .....	18
V. Conclusions, lessons learned, and recommendations.....	21
A. Conclusions .....	21
B. Lessons learned from the <i>Amazonía Lee</i> LTIE.....	22
C. Recommendations from the <i>Amazonía Lee</i> LTIE .....	23
References .....	25

---

## Tables

III.1	Data available for the LTIE analysis with ECE data .....	11
-------	--	----

## Figures

II.1	<i>Amazonía Lee</i> logic model .....	5
II.2	Evaluation design: randomized controlled trial.....	6
II.3	Timeline of key milestones in <i>Amazonía Lee</i> 's Implementation .....	7
III.1	Available data used for <i>Amazonía Lee</i> LTIE analyses .....	12
IV.1	Trends on ECE reading scores, by study groups and region .....	16
IV.2	Positive impact of <i>Amazonía Lee</i> on adjusted means second-grade ECE reading score by region.....	17
IV.3	Short-term impacts on EGRA reading comprehension and on ECE reading scores measured in effect sizes, by region .....	18
IV.4	Adjusted means on fourth-grade ECE reading scores in 2018 by region .....	19

---

## Executive Summary

### A. Purpose and background

The purpose of this long-term impact evaluation (LTIE) is to examine (1) the longer-term impacts of the United States Agency for International Development (USAID) investment in reading in Peru through the *Amazonía Lee* intervention; and (2) the factors that helped or hindered its uptake, sustainability, and longer-term impact, thus enhancing USAID’s understanding of the conditions and assistance needed to support countries’ self-reliance. This report presents findings of the long-term impact evaluation using the national reading scores (known as Evaluación Censal Educativa [ECE]) to explore the extent to which gains in students’ achievement were sustained one year after USAID stopped funding the implementation of *Amazonía Lee*’s capacity-building activities in the region.

The *Amazonía Lee* program was part of the USAID/Latin America and the Caribbean (LAC) Reads initiative to strengthen reading capabilities by generating and using information on successful approaches to improving early-grade reading. In line with this goal, *Amazonía Lee* was designed to improve early-grade reading performance and address the learning gaps of two Amazonian regions in Peru (San Martín and Ucayali) and targeted children in primary school (grades 1 to 3). The project was implemented by the Regional Education Directorates (Direcciones Regionales de Educación) from 2015 to 2017, with funding from USAID. Universidad Peruana Cayetano Heredia (UPCH) provided technical assistance in both regions, and Mathematica conducted an impact evaluation between 2015 and 2016. [This impact evaluation of \*Amazonía Lee\*](#), commissioned by USAID, focused on reading outcomes measured in 2016, about two years after implementation started in Ucayali and San Martín. Findings showed promising short-term improvements in children’s early-grade reading outcomes in Ucayali when compared to control schools providing the normal Ministry of Education (MINED) reading curriculum, and reading outcomes comparable to those achieved by the MINED’s program to improve education quality, known as *Soporte Pedagógico*, implemented in control schools in San Martín.

### B. LTIE aims, design and methods

The purpose of this LTIE is to examine longer term impacts of the USAID investment in reading in Peru through *Amazonía Lee*, and the factors that helped or hindered its uptake, sustainability, and longer-term impact, thus enhancing USAID understanding of the conditions and support needed to support countries’ self-reliance. Although USAID no longer funds early-grade reading programs in Peru, the long-term impact evaluation also supports USAID/Peru’s Learning Plan. The LTIE has two main goals: to determine (1) whether *Amazonía Lee* had a discernible impact on students’ scores on Peru’s national education achievement test (ECE) in second grade (in Ucayali and San Martín); and (2) whether improvements in students’ reading skills were sustained over time as children progressed to higher grades (grade 4).

We used a quantitative approach to examine whether the *Amazonía Lee* impacts on reading outcomes found for second-graders translated into impacts on their scores on the national reading test (ECE) for the same year. We then compared the impacts on the Early Grade Reading Assessment scores (EGRA) found in the initial impact evaluation to the impacts on ECE scores on second graders. We also estimated the long-term impact of *Amazonía Lee* on fourth-graders’ ECE reading scores in 2018, two years after the short-term impacts were measured.

To estimate short-term impacts, we used the baseline information we collected in 2015 to control for initial small differences that arose by chance between the treatment and control schools, despite random assignment. Because the ECE data does not include individual student identifiers, and we cannot match baseline EGRA with ECE scores at the student level, we therefore used school level data for the analysis. We estimated the impacts on ECE scores in 2016 adjusting for baseline school characteristics in 2014 and baseline average school-level EGRA reading scores in 2015.

Similarly, we used fourth-graders' ECE reading scores in 2018 as endline outcomes to estimate long-term impacts of *Amazonía Lee* and assess whether the short-term impacts found for second graders in 2016 were sustained when the students reached fourth grade in 2018—one year after the implementation funded by USAID ended.

### C. Summary of findings and conclusions

Overall, *Amazonía Lee*'s short-term impacts on student reading outcomes measured with the ECE test are consistent with the impact evaluation findings for second-graders using the EGRA test (Campuzano et al. 2018).

**In Ucayali, *Amazonía Lee* had a positive impact on second-grade ECE reading scores.** Adjusting for initial chance school-level differences between the treatment and control groups, *Amazonía Lee* second-graders had on average higher ECE reading scores than students in the control group in 2016, the second year of *Amazonía Lee* implementation. The ECE reading scores adjusted for baseline differences were on average 546 for *Amazonía Lee* schools and 520 for control schools in Ucayali. Thus, *Amazonía Lee* schools scored 26 points higher on the ECE test than the control group, an impact equivalent to 0.39 standard deviations (SD).

**In San Martín, second-grade students in both study groups had similar reading ECE scores in 2016.** *Amazonía Lee* had no impact on ECE scores in 2016 in San Martín. ECE reading scores for second-graders in 2016, adjusted for baseline differences, were on average 581 for *Amazonía Lee* and control group students.

The long-term impacts of *Amazonía Lee* also varied by region:

**In Ucayali, the 2018 ECE reading scores of fourth-grade students in *Amazonía Lee* schools were higher than scores of students in the control schools, but differences were not statistically significant.** The adjusted mean for *Amazonía Lee* fourth-graders on the ECE reading test was 430 points, 12 points higher than the adjusted mean score for control group students.

**In San Martín, students in control schools had higher adjusted means on ECE fourth-grade reading scores in 2018 than *Amazonía Lee* students, and the difference was statistically significant.** Fourth-graders had on average an adjusted ECE reading score of 477 in control schools, 11 points higher than the adjusted mean scores for *Amazonía Lee* schools. This merits investigation to gain a deeper understanding of the factors that may have helped *Soporte Pedagógico* schools improve their performance compared to *Amazonía Lee* schools.

**The short-term impacts proved to be consistent across the two types of reading assessments.** In Ucayali, *Amazonía Lee*'s positive impacts on the ECE scores of second-grade students were consistent with the positive impacts on EGRA scores. In San Martín, findings on ECE scores for second-graders were consistent with the EGRA findings. In *Amazonía Lee* schools students had, on average, similar ECE

reading scores than students in the control schools in 2016—most of which received services from *Soporte Pedagógico*.

**The short-term impacts were not sustained as students progressed to higher grades.** In Ucayali, *Amazonía Lee*'s positive short-term impacts on the ECE scores of second-graders were not sustained when students reached fourth grade. In San Martín, although second-graders had similar ECE reading scores in 2016, when students reached fourth grade in 2018, adjusted mean ECE scores for control schools were higher than adjusted mean ECE scores for *Amazonía Lee* schools.

### D. Lessons learned

**In-service teacher training on early-grade reading instruction, whether supported by local technical assistance providers or by the national government, contributes to better student reading skills in the primary grades in the short and the long term.** In Ucayali, in the short-term when *Amazonía Lee* schools were receiving in-service training and the control group was not, students in *Amazonía Lee* schools had better reading scores than students in the control group. In San Martín, where the control group was also receiving teacher training through *Soporte Pedagógico* students had similar scores in both groups. In the long-run, when *Amazonía Lee* schools were not offering teacher training, the positive impacts found in Ucayali, for second graders, were not sustained when students reached 4th grade. In San Martín, students in the control schools outperformed students in *Amazonía Lee* schools.

**Using national tests can be a cost-effective data source for future education evaluations.** The RCT design of the *Amazonía Lee* evaluation enabled us to measure long-term program impacts. Moreover, the *Amazonía Lee* LTIE offers a solid evidence base for future evaluations that wish to use the ECE as the main reading outcome measure without incurring the high costs of individually administered early-grade reading assessments given the robustness of the short-term impacts to these two reading assessments. The advantages of using national tests, when available, could include (1) better statistical power in cases where the national test is collected for all students (census) as the ECE was usually collected, (2) local and national authorities have more confidence in the metrics they historically use to measure students' progress in reading, and (3) the costs to obtain these data are low compared with tailored and individually administered EGRA tests. Further research should confirm that national tests in other countries can also serve as a good alternative.

---

## I. Introduction

The Expanding the Reach of Impact Evaluation Program (ERIE) helps international development projects understand whether they achieved their intended long-term results or sustained their shorter-term results over time. A long-term impact evaluation (LTIE) allows projects to measure outcomes over longer time horizons and to explore whether impacts grow over time, are sustained, or attenuate. Because LTIEs usually occur after program implementation is complete, they afford an opportunity to measure additional outcomes that would not have been manifest in the typical evaluation or program life cycle ([USAID 2018](#)).

Early-grade reading interventions in developing countries have often included a combination of various components—typically, training teachers to improve reading instruction using evidence-based approaches; in-class teacher coaching; and provision of instructional guidelines, instructional materials, and tools for student assessment. The effectiveness of such reading interventions has been tested in a variety of contexts in developing countries. For example, a study of 18 reading interventions implemented mostly in African and Middle Eastern countries found that early-grade interventions were effective, showing significant impacts on at least one reading subtask. In many cases, the effect sizes were equivalent to over half a year of schooling. However, most of the interventions seem to be only a partial solution, as they did not raise students’ oral reading fluency in a meaningful way (Graham and Kelly 2018).

In a recent meta-analysis, Kim, Lee, and Zuilkowski (2019) reviewed 68 impact evaluations of literacy interventions in 32 low- and middle-income countries and estimated their effects on children’s reading skills. Their results showed an overall average effect size of 0.30 standard deviations (SD) across various literacy outcomes; the highest effects were in emergent literacy skills (0.40 SD), and the smallest were in reading comprehension (0.25 SD). Much of the research in low- and middle-income countries consists of evaluations of multicomponent interventions collectively showing positive effects on literacy outcomes, but there are large variations in the sizes of the effects depending on specific literacy outcome measures, ranging from no effect to a large effect on reading comprehension. Further, most of the evidence to date comes from short-term evaluations, which results in a notable gap in the evidence on the potential for longer-term impacts that reading interventions might have in low- and middle-income countries.

The purpose of this LTIE is to examine longer-term impacts of the USAID investment in reading in Peru through the *Amazonía Lee* intervention, and the factors that helped or hindered its uptake, sustainability, and longer-term impact, thus enhancing USAID’s understanding of what is needed to best support countries’ journeys towards self-reliance, particularly in building effective early grade reading programs. The *Amazonía Lee* LTIE focused on exploring the extent to which gains in students’ reading achievements were sustained one year after USAID stopped funding the implementation of *Amazonía Lee*’s capacity-building activities in the region. This report presents the long-term impacts using Peru’s national reading achievement scores, known as Evaluación Censal Educativa (ECE), as the main reading outcome.

*Amazonía Lee* was a reading intervention intended for use with students in first through third grade. The intervention sought to improve the quality of reading instruction in two regions of the Peruvian Amazon—San Martín and Ucayali—which had historically showed marked gaps in early-grade reading achievement. It was implemented in the Peruvian Amazon region between 2015 and 2017. *Amazonía Lee* was part of USAID’s Latin America and the Caribbean Read (LAC Reads) project to increase the availability of, demand for, and capacity to use evidence-based, cost-effective practices by ministries of education and other education stakeholders to improve early-grade reading achievement in the LAC



region. The purpose of this LTIE is to examine longer term impacts of the USAID investment in reading in Peru through *Amazonia Lee*, thus enhancing USAID understanding of the programs that can support countries' self-reliance. Although USAID no longer funds early-grade reading programs in Peru, the long-term impact evaluation also supports USAID/Peru's Learning Plan. Specifically, the evaluation will help address the following cross-cutting learning question: "Under what conditions will the Government of Peru deliver their services in a more transparent, responsive, and efficient way?"

*Amazonia Lee's* approach to improving the reading skills of children in the early grades focused on a teacher training and coaching program that blended best practices from the National Reading Panel with the communicative textual approach espoused by Peru's Ministry of Education (MINEDU) (Campuzano et al. 2018). [The impact evaluation of Amazonia Lee](#), commissioned by USAID, focused on reading outcomes collected in 2016, about two years after implementation started in Ucayali and San Martín. Findings from the impact evaluation showed promising short-term improvements in children's early-grade reading outcomes in Ucayali as compared to a control group receiving typical reading instruction reflecting prevailing practice in most public schools, and reading outcomes comparable to those achieved by the MINEDU's program to improve education quality, known as *Soporte Pedagógico*, implemented in control schools in San Martín.

The ERIE team felt that *Amazonia Lee* would be a good candidate for an LTIE for these reasons (USAID 2018):

- *Amazonia Lee* has a theory of change that outlines the logic of how the intervention could realize long-term impacts. Because the program focused on improving teaching practices, it is expected that improved teaching practices would continue even after the teacher training and coaching period ended (Campuzano et al. 2018).
- The original evaluation design for *Amazonia Lee* intended to assess impacts on children's reading outcomes in third grade (in 2017). However, the evaluation period was shortened when USAID/Peru phased out support to basic education in Peru by early 2018. Because stakeholders were interested in obtaining results before the end of *Amazonia Lee* implementation, we modified the endline period for the impact evaluation and looked at impacts on reading outcomes in second grade (2016), one school year earlier than originally planned. An LTIE offers a unique opportunity to explore whether short-term improvements in early-grade reading were sustained after USAID funding ended.
- The *Amazonia Lee* evaluation randomly assigned schools to treatment and control groups, so there is a credible comparison group that can be used to conduct an LTIE. MINEDU annually measures children's reading comprehension skills through the ECE scores, which are a credible country-specific data source for reading achievement over time in treatment and control schools.

This report presents the findings from the *Amazonia Lee* LTIE. In Section II, we provide an overview of the *Amazonia Lee* intervention, its implementation, and the key findings from the impact evaluation. In Section III, we summarize the design for the LTIE of *Amazonia Lee*, define the research questions, and describe the analytic strategy and data sources. In Section IV we present findings on the short- and long-term impacts of *Amazonia Lee* using Peru's ECE as the reading outcome measure. The first set of findings shows the short-term impacts of *Amazonia Lee* on student reading using the ECE scores. We also compare the short-term impacts using ECE to the impacts previously found using EGRA; this allows to check robustness of the results across two different types of reading measures. The second set of findings show the long-term impacts of *Amazonia Lee* on fourth-grade ECE scores to determine whether the short-term impacts were sustained one year after the implementation funded by USAID ended. In Section V, we discuss the conclusions, lessons learned, and recommendations.

---

## II. Background and context on the evaluation of *Amazonía Lee*

The USAID Regional Sustainable Development Office of the Bureau for LAC contracted the services of Mathematica for an independent impact evaluation and cost-effectiveness analysis of reading and education access interventions in the LAC region (USAID/LAC Reads). Five impact evaluations of reading programs were completed under the contract. Reading interventions to be evaluated were selected based on showing promise to improve early-grade reading performance, and on their potential to contribute to the emerging evidence base in education.

The *Amazonía Lee* project was part of the USAID/LAC Reads effort to strengthen reading capabilities in Latin America and the Caribbean by generating and using information on successful approaches to improving early-grade reading. In line with this goal, *Amazonía Lee* was designed to improve early-grade reading performance and address the learning gaps of two Amazonian regions in Peru: San Martín and Ucayali. Intended for children in early primary school (grades 1–3), the project was implemented by the Regional Education Directorates (Direcciones Regionales de Educación [DREs]) from 2015 to 2017 with funding from USAID. Universidad Peruana Cayetano Heredia (UPCH) provided technical assistance in both regions, and Mathematica conducted the impact evaluation between 2015 and 2016.

*Amazonía Lee* was implemented in schools with at least six teachers and a minimum of one per classroom (known as “polidocente completo”), and in schools with between 2 and 5 teachers, with at least two grades grouped in one classroom (known as “polidocente multigrado”). Although the evaluation was originally designed to follow students from first through third grade, we were able to measure the effects of exposure to the program over only about one and a half school years.<sup>1</sup> A detailed description of the evaluation is in [Campuzano et al. \(2018\)](#).

*Amazonía Lee* had four components: (1) capacity building, (2) community engagement, (3) a regional assessment system, and (4) a regional teacher incentive system. However, the only component that was fully implemented in both regions during the evaluation period was **capacity-building**, which aimed to strengthen the pedagogical skills and instructional capabilities of teachers, school directors, and regional specialists. Next, we describe this component, which was the focus of the initial impact evaluation:

- **Teacher training workshops.** Workshops were based on *Amazonía Lee*’s balanced approach to reading instruction, which integrated evidence-based foundational reading skills instruction with MINEDU’s existing communicative-textual approach.<sup>2</sup> Training workshops sought to improve teachers’ approach to reading instruction, class and curriculum planning, differentiated instruction, and personalized attention for underperforming students. The training covered five core dimensions of reading instruction: (1) print environment; (2) understanding the alphabetic principle (alphabet

---

<sup>1</sup> USAID/Peru phased out support to basic education in Peru by 2018 and discontinued most funding for *Amazonía Lee*’s capacity-building activities at the end of 2017. To obtain results by the end of 2017, we could not measure the effects of the full intended dosage of the program. As a result, the impacts were measured for 13 to 14 months (May or June 2015 to September or October 2016) in second grade, rather than in third grade as originally envisioned.

<sup>2</sup> MINEDU promotes the *communicative and textual* approach (*enfoque comunicativo y textual*) to reading instruction. This approach focuses on teaching reading comprehension and supporting the development of communicative competencies in meaningful contexts by presenting images and text to students, explaining their meaning, and helping students communicate that meaning through oral and written language. The *balanced* approach to teaching reading integrates the national *communicative and textual* approach with instruction on foundational reading skills such as alphabet knowledge, phonological awareness, decoding, vocabulary, fluency, and comprehension (National Institute of Child Health and Human Development 2000).

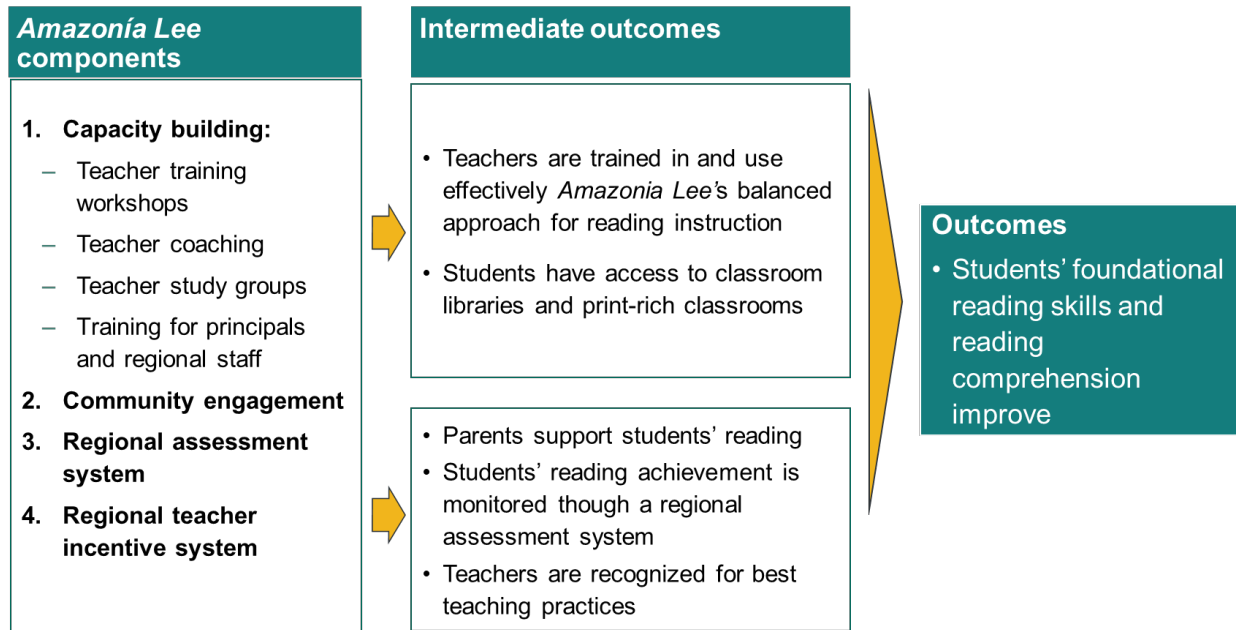
knowledge, phonological awareness, and fluency); (3) reading comprehension; (4) writing and production of texts; and (5) leadership and interpersonal relations. It also covered the use of formative assessments to monitor students' learning, positive classroom climate strategies, and community engagement strategies to support student literacy. The training was planned for 120 hours to be delivered in 40-hour blocks throughout the year.

- **Appropriate materials.** In addition to teacher workshops, schools received more and more varied pedagogical materials to support reading, to improve the quality and cultural relevance of pedagogical materials for reading and writing instruction, and to increase the availability of classroom book libraries in primary grades. *Amazonía Lee* provided three broad types of materials to support reading instruction: (1) the UPOCH teacher training modules, (2) teacher activity guides and student workbooks, and (3) classroom libraries. The training modules offered a conceptual basis for early-grade teaching and pedagogical guidelines for reading comprehension, production of texts, creating an enriched literacy classroom environment, and engaging parents and the community in students' reading. The teacher guides and student workbooks included learning activities based on *Amazonía Lee*'s balanced approach to foundational reading skills, reading comprehension, and production of texts. The program also delivered "reading for pleasure" classroom libraries with a variety of books purposefully selected to be grade-level appropriate and culturally relevant.
- **Teacher coaching.** Program specialists conducted visits to coach teachers in their classrooms and help them improve content area knowledge and pedagogical strategies for literacy instruction. This capacity-building component included demonstration and observation sessions. In demonstration lessons, specialists served as role models for effective classroom practice. In observation sessions, trainers observed teachers delivering instruction to a class and provided feedback. Teachers received six to eight coaching visits per year, depending on their learning needs.
- **Teacher study groups.** Teachers gathered in groups of 8 to 16 to strengthen their learning, review pedagogical materials, share teaching strategies, and plan for classes with support from peer teachers. Study groups (known as "Grupos de Interaprendizaje") were intended as flexible spaces for collective feedback and to address the skill-building needs trainers identified during the coaching visits. Teacher study groups were expected to meet once a month for about four hours.
- **Training for principals and regional staff.** This involved school leadership and management training for school directors.

In addition to these teacher capacity-building activities, *Amazonía Lee* included a set of strategies to support students' early-grade reading achievement (for example, community engagement activities, reading assessments, and teacher incentives). These activities were not fully implemented on the same timeline as the capacity-building activities, and therefore the impact evaluation did not account for their effects.

In Figure II.1 we provide a high-level logic model of the program, summarizing the main components of *Amazonía Lee* intervention and the expected intermediate outcomes through which the intended changes in foundational reading skills are expected to be achieved.

Figure II.1. *Amazonía Lee* logic model

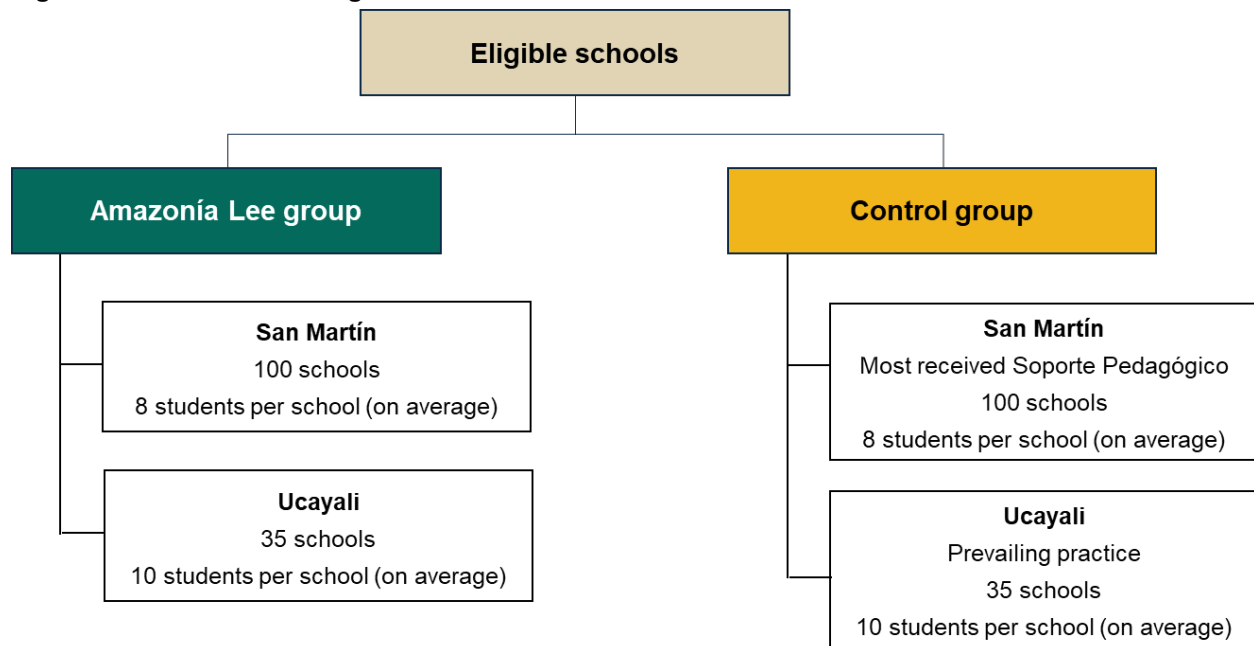


### A. *Amazonía Lee* impact evaluation

In this section, we summarize the impact evaluation; a more detailed description can be found in [Campuzano et al. \(2018\)](#). Mathematica partnered with Grupo de Análisis para el Desarrollo as our local research and data collection partner to conduct a randomized controlled trial to rigorously estimate *Amazonía Lee* impacts. We randomly assigned eligible schools in Ucayali and San Martín to one of two groups: the treatment group was offered all program components of *Amazonía Lee*, and the control group was not offered the capacity-building component of the program but may have received services offered at the regional level.

In Ucayali, the evaluation sample included 70 schools; 35 were randomly assigned to the treatment group and 35 were randomly assigned to the control group. In San Martín, the evaluation sample included 200 schools; 100 were randomly assigned to the treatment group and 100 were randomly assigned to the control group (Figure II.2). To achieve better balance on key characteristics, we stratified the random assignment using pairs of schools with similar key school-level characteristics. Random assignment took place in 2014, when the program was originally intended to be rolled out, and initial baseline data collection occurred in study schools for the 2014 first-grade cohort. However, the rollout was delayed, and the evaluation ultimately focused on the 2015 first-grade cohort. As mentioned above, although the original design planned to follow the 2015 first-grade cohort for the full three years of *Amazonía Lee* implementation, the phasing out of USAID/Peru's basic education program and the need to obtain results by the end of 2017 prevented us from following the cohort past second grade (2016).

Figure II.2. Evaluation design: randomized controlled trial



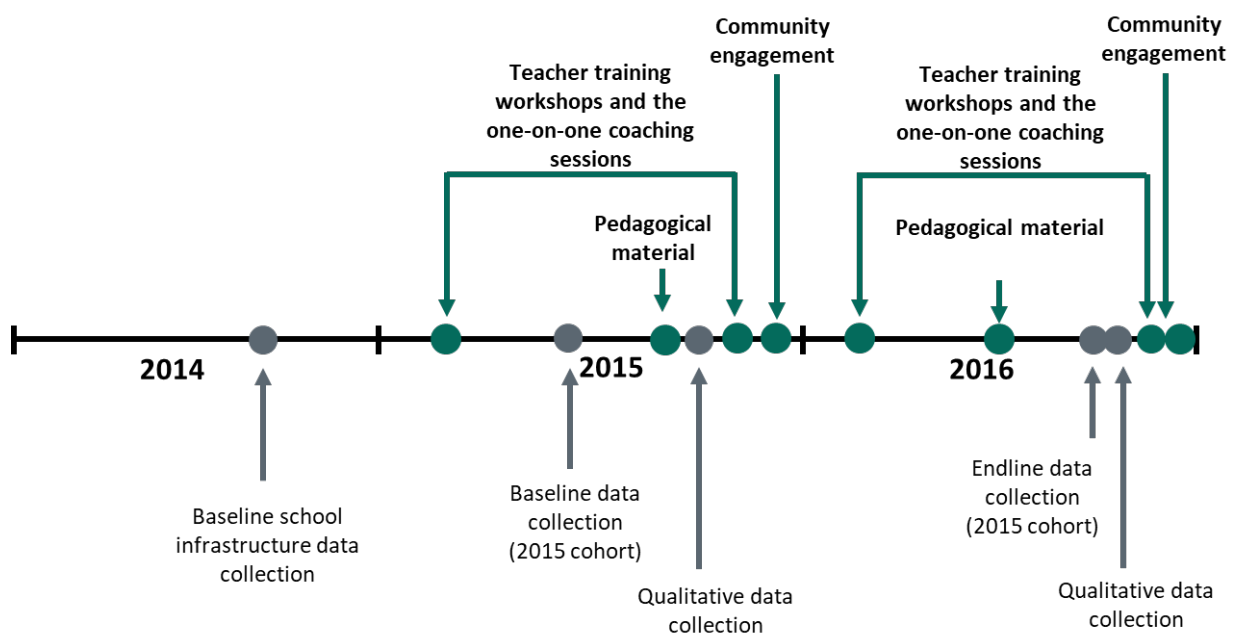
The control group schools were receiving different services in the two regions at the time of the evaluation, so we conducted separate analyses for each. In Ucayali, control schools received the usual services MINEDU provides in the department. In San Martín, most of the control schools received additional assistance in the form of a primary education support program known as *Soporte Pedagógico* (Pedagogical Support), which Peru’s flagship education quality improvement program rolled out in San Martín starting in 2015. *Soporte Pedagógico* seeks to improve student learning by strengthening the skills of teachers and administrators, providing after-school tutoring for students who do not meet the expected learning standards, delivering educational materials and resources, and improving the school and local management capacity. It included (1) training activities to strengthen first- through sixth-grade teachers’ reading, math, and social studies instruction; (2) remedial academic support for underperforming students after regular school hours (*Docentes Fortaleza*); and (3) other related services to improve teachers’ performance and students’ academic outcomes. There are three main differences between *Amazonía Lee* and *Soporte Pedagógico*. First, *Amazonía Lee* focused on reading and communication, whereas *Soporte Pedagógico* focused on mathematics, communication, science, and social skills in urban elementary schools. Second, the approach to reading instruction of *Amazonía Lee* (balanced approach) integrated instruction on foundational reading skills with the national *communicative-textual* approach, whereas *Soporte Pedagógico* used primarily the national *communicative-textual* approach. Third, *Soporte Pedagógico* was a more comprehensive and resource-intensive program than *Amazonía Lee*. In contrast to *Amazonía Lee*, *Soporte Pedagógico* provided small-group tutoring to underperforming students after school with specialists in academic reinforcement known as *Docentes Fortaleza*. *Soporte Pedagógico* had a much lower program specialist to teacher ratio (1:15 for *Docentes Fortaleza*; 1:30 for *Amazonía Lee*).

The main data sources for the impact evaluation were student reading assessments, classroom observations, a teacher survey, and a school infrastructure survey. The child assessments were a version of the Early Grade Reading Assessment (EGRA), which tested early reading skills; we used these assessments to estimate impacts on students’ reading skills, the primary measure for the impact

evaluation. The classroom observation measured teachers’ use of class time, their pedagogical practices, and the classroom literacy environment. We also collected qualitative data at two points in time.

The implementation of *Amazonía Lee* began in 2015 and ended in 2017. The impact evaluation focused on the first two years of implementation. Baseline student assessments for the 2015 first-grade cohort were conducted between May and June 2015, and endline student assessments were conducted with the same students between September and October 2016, when most were in second grade. Therefore, the impact evaluation measured the effects of 13 to 14 months of exposure to the program in first and second grades, equivalent to about 1.5 academic school years (Figure II.3 shows a high-level timeline of key milestones in *Amazonía Lee* implementation).

**Figure II.3. Timeline of key milestones in *Amazonía Lee*’s Implementation**



Source: Developed by authors, based on qualitative data collected for the impact evaluation (Campuzano et al. 2018).

Note: The school year in Peru starts in March and ends in December. The community engagement component was intended to support the school capacity-building activities; however, there were delays in implementation, and the community engagement activities lagged those of capacity building. In the first year of *Amazonía Lee*, Ucayali implemented parent workshops. In the second year, both departments did. Parent workshops focused on providing parents with information on how children learn to read and how to provide them opportunities to practice reading. Once each year, *Amazonía Lee* schools held community-wide learning fairs to showcase reading, storytelling, and literacy-related activities and achievements.

## B. Summary of findings from the impact evaluation of *Amazonía Lee*

### 1. Implementation and intermediate outcome findings

*Amazonía Lee* led to more professional development in reading instruction, in-class coaching, demonstration, observation of teachers, and availability of books and materials for teaching reading (Campuzano et al. 2018). In both regions, we found some impacts on teacher instructional practices.

These impacts were consistent with treatment teachers implementing *Amazonía Lee*'s balanced approach. For example, compared to control group teachers, treatment group teachers were (1) more likely to be observed explaining the use of different types of text, (2) less likely to report spending time on daily dictation, and (3) more likely to report teaching students how to make graphic organizers to describe the meaning of a text (Campuzano et al. 2018). A summary of impacts on intermediate outcomes is in the evaluation plan (see Fernandez et al. 2020 for a detailed description).

## 2. Impacts on student-level outcomes

The impacts on students' reading outcomes differed by region. In Ucayali, where impacts of *Amazonía Lee* were compared to outcomes among a control group that received only the support that MINEDU typically provides, students receiving *Amazonía Lee* had better reading outcomes than students in the control group. In San Martín, where *Amazonía Lee* was compared to a control group that was receiving *Soporte Pedagógico*, a more comprehensive and resource-intensive program, students receiving *Amazonía Lee* had outcomes similar to those of the control group. Below we describe the findings for each region in more detail.

**Ucayali.** We looked at impacts of *Amazonía Lee*'s capacity-building activities on reading comprehension and other important reading building blocks, such as familiar-word reading, decoding, and reading fluency. These reading skills were measured with the EGRA test. In Ucayali, *Amazonía Lee* students had better reading outcomes than students in the control group; we found positive and statistically significant impacts on four of the five skills measured. Furthermore, the magnitude of the impacts was substantial (0.15 to 0.27 SD) and comparable to what has been found by rigorous studies of other educational programs implemented internationally ([USAID 2016](#)). *Amazonía Lee* had positive impacts on:

- **Familiar-word reading.** Students in the treatment group correctly read on average 7 of 10 familiar words presented, whereas students in the control group read on average 6 words, a statistically significant difference of 1 word. This difference is equivalent to 0.24 SD.
- **Decoding skills.** Students in the treatment group correctly read on average 17 invented or pseudo-words in one minute, whereas students in the control group read on average 13 invented words, a difference of almost 4 words that is statistically significant. This difference is equivalent to 0.24 SD.
- **Fluency (marginal).** Students in the treatment group read on average 35 words correctly in one minute, whereas students in the control group read 30 words correctly. This difference of 5 words per minute is statistically significant at the 10 percent level and is equivalent to 0.15 SD. However, the fluency level of both groups was below 60 words per minute for second-graders, the standard suggested for Peru (World Bank, 2006).
- **Reading comprehension.** Students in the treatment group were able to answer on average 3 of the 6 comprehension questions correctly; students in the control group were able to answer 2 questions on average. The difference of 1 correct answer is statistically significant and equivalent to 0.27 SD.

**San Martín.** In San Martín, students in the treatment group had reading outcomes similar to those of students in the control group. We found no differences between groups on oral comprehension, familiar-word reading, fluency, decoding, and reading comprehension scores. Students in both groups correctly read on average 8 of 10 familiar words presented. Students in both groups had similar fluency results: on average they read correctly almost 44 words per minute, and 20 invented or pseudo-words per minute. Furthermore, students in both groups were able to answer 3 of the 6 reading comprehension questions correctly. However, as in Ucayali, both groups showed fluency scores below the standard of 60 words per minute suggested by the World Bank for second-graders in Peru.

---

### III. Long-Term Impact Evaluation Design

In this chapter, we briefly discuss the *Amazonía Lee* LTIE research questions, the data sources, and the analysis we conducted to answer these questions. A more detailed description of the LTIE design is in Fernández et al. (2020).

#### A. Design and evaluation questions

The aims of this LTIE are to (1) reexamine the short-term impacts of *Amazonía Lee* on student reading outcomes measured with Peru's national reading test (ECE); (2) assess long-term impacts to determine whether they were sustained after the implementation funded by USAID ended; (3) understand the factors that helped or hindered *Amazonía Lee*'s uptake, sustainability, and longer-term results. We also compare the short-term impacts from this LTIE to those of the original evaluation that used EGRA to measure reading skills, which will allow us to assess whether short-term impacts are robust to the type of reading assessment used. The ECE is a contextually relevant metric for Peru, so impacts based on this metric can inform or serve as a benchmark for future evaluation efforts.

The LTIE examines whether the *Amazonía Lee* impacts on reading outcomes found in the impact evaluation for second-graders translated into short-term impacts on their scores on the national reading test (ECE) for the same year. We then estimate the long-term impact of *Amazonía Lee* on fourth-graders' ECE reading scores in 2018, two years after the short-term impacts were measured. Together, this enabled us to assess whether the short-term ECE impacts found for second-graders in 2016 were strengthened or sustained when the students reached fourth grade in 2018, one year after the program ended, and after the students had been exposed to the program for almost three years (grades 1 to 3). The **main research questions** for the LTIE are:

1. Did *Amazonía Lee* have a discernible impact on students' scores on Peru's national education achievement test (Evaluación Censal Educativa, ECE) in second grade (in Ucayali and San Martín)?
2. Were improvements in students' reading skills sustained as children progressed to higher grades (grade 4)?

#### B. Data sources, reading measures, and analysis sample

The two main data sources we use for the *Amazonía Lee* LTIE are (1) data collected for the *Amazonía Lee* impact evaluation, and (2) ECE data provided by Peru's MINEDU. Next, we summarize the data sources and the reading outcome measures (see Fernández et al. 2020 for a detailed description).

**Data from the *Amazonía Lee*'s impact evaluation.** Grupo de Análisis para el Desarrollo, our in-country partner, collected survey and reading assessment data for the impact evaluation. To increase the precision of the LTIE impacts, we used baseline data from the impact evaluation in the LTIE analysis. We also used school infrastructure data collected in 2014 and EGRA student reading assessment results collected in 2015 in Ucayali and San Martín as baseline data.

- **School infrastructure in 2014.** This checklist gathered information on general characteristics of the schools' infrastructure, such as the number of classrooms, the distance between the school and the district capital, and the availability of electricity and potable water. To complete the infrastructure checklist, enumerators conducted a visual scan of the school premises and buildings and asked a few questions of the school principals.



- **Amazonía Lee early-grade reading assessment for first grade in 2015.** We collected baseline data for all schools in the evaluation sample, 70 schools in Ucayali and 200 in San Martín. We randomly selected a sample of first grade students in each school (8 students per classroom in San Martín and 10 students in Ucayali) to assess early grade reading skills at baseline. The 2015 cohort was assessed between May and July 2015 in San Martín and in June and July 2015 in Ucayali. The assessment included three components: (1) the oral language screener (PreLas) to assess students' oral language skills in Spanish and determine if the assessment should be administered in Spanish; (2) the Evaluation of Emergent Reading and Writing (LEE, for its name in Spanish) to assess alphabet knowledge, simple word reading, and emergent writing; and (3) the Evaluation of Early Grade Reading (EGRA or ELGI, for its name in Spanish) to assess phonemic awareness, decoding, and oral comprehension. A description of the baseline and follow-up EGRA assessment sources, tasks, and item examples are available in Fernandez et al. 2020.

**The national education achievement test (ECE).** For this analysis, we used second-grade reading ECE scores in 2016 as endline outcomes for the short-term impact analysis and fourth-grade reading ECE scores in 2018 as endline outcomes for the long-term impact analysis. The ECE test is a group-administered, diagnostic assessment of reading comprehension that does not evaluate individual students' academic standing. It is administered in second and fourth grade.<sup>3</sup> The ECE assesses students' ability to (1) read sentences (match a sentence to a drawing), (2) identify literal information (distinguish explicit facts and recognize sequences of events), and (3) make inferences (deduce cause-and-effect relations, word meaning, characters' traits, main themes, moral of a story, and purpose of a text). Although the ECE is intended as a census of second- and fourth-graders, MINEDU did not administer the ECE test in schools that had fewer than five enrolled students. Therefore, some schools in the impact evaluation sample have missing ECE data. We excluded those schools from the LTIE analysis and from the corresponding schools in the treatment-control matched pair used for random assignment. Excluding the school pairs allows us to maintain the balance we sought to achieve with the random assignment.

We used two waves of ECE reading scores for the LTIE analysis:

- **ECE reading scores for second-graders in 2016.** MINEDU provided ECE reading scores for students enrolled in second grade in 2016 in the study schools. In Ucayali, five schools have missing ECE data for 2016, representing two complete pairs and one partial pair of our evaluation sample. We excluded these three school pairs from the analysis sample, for a total of six excluded schools. This is 8.6 percent overall school-level attrition and zero percent differential attrition. According to WWC standards, this level of attrition at the school level is low, preserving the strong causal validity of the findings. In Ucayali, the analysis sample for question 1 consists of 64 schools, 32 in the treatment group and 32 in the control group. In San Martín, the analysis sample for question 1 includes all 200 schools in the evaluation sample, or 100 in each of the treatment and control groups (see Table III.1).
- **ECE reading scores for fourth-graders in 2018.** MINEDU provided ECE scores for students enrolled in fourth grade in 2018 in the study schools. In Ucayali, two schools have missing ECE data for 2018. We excluded those school pairs, for a total of four excluded schools. In Ucayali, the

---

<sup>3</sup> In 2006, the national census evaluations in Peru started to collect information on reading comprehension skills for second-grade students. ECE was collected annually for all second-graders until 2016. In 2016, the census was collected also for fourth-grade students. In 2017, a teacher strike and inclement weather resulted in school closures, so MINEDU suspended the ECE in 2017, and ECE data are not available for that year. In 2018, ECE was collected as a census for fourth grade but not for second grade; ECE was collected only for a sample of schools and students in second grade.

analysis sample for question 2 consists of 66 schools, 33 in the treatment group and 33 in the control group. In San Martín, the ECE was not collected in two schools in 2018, so we also excluded those two and their corresponding pairs from the analysis. The analysis sample for question 2 in San Martín is 196 schools, 98 in treatment and 98 in control (see Table III.1).

**Table III.1. Data available for the LTIE analysis with ECE data**

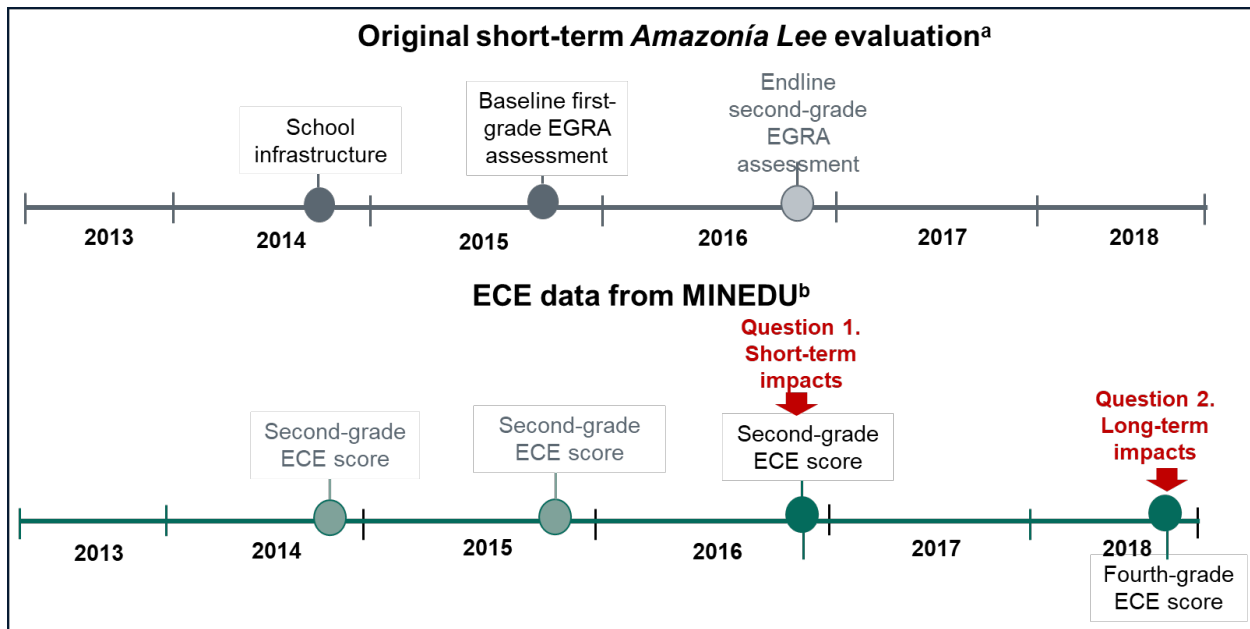
	2014 school infrastructure		2015 1st-grade EGRA scores		2016 2nd-grade ECE scores		2018 4th-grade ECE scores	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
<b>Ucayali</b>								
Students	n.a.	n.a.	348	326	2,020	1,911	1,892	1,814
Schools	35	35	35	35	32	32	33	33
<b>San Martín</b>								
Students	n.a.	n.a.	790	789	4,130	4,090	4,021	4,030
Schools	100	100	100	100	100	100	98	98

n.a. = not available

### C. Analytic strategy

Figure III.1 depicts the data available for the *Amazonía Lee* LTIE and how it was used to answer the research questions. We used second-graders’ ECE reading scores in 2016 as endline outcomes to estimate the short-term impacts of *Amazonía Lee* (question 1). We used fourth-graders’ ECE reading scores in 2018 as endline outcomes to estimate the long-term impacts of *Amazonía Lee* one year after the implementation funded by USAID ended (question 2). As shown in Figure III.1, we used baseline data from the 2014 infrastructure school and 2015 student data from *Amazonía Lee* evaluation to control for initial differences and increase the precision of the short- and long-term impact estimates.

Figure III.1. Available data used for *Amazonía Lee* LTIE analyses



<sup>a</sup> The *Amazonía Lee* evaluation collected school infrastructure data in 2014 and EGRA student assessments in 2015 (baseline) and EGRA in 2016 (endline). The LTIE analysis uses the 2014 and 2015 data as baseline information.

<sup>b</sup> MINEDU's data sets include student second-grade ECE reading scores for school years 2013, 2014, 2015, 2016, and fourth-grade ECE reading scores for 2016 and 2018. The LTIE analysis uses second-grade ECE scores in 2016 as endline reading outcomes to estimate short-term impacts (question 1) and fourth-grade ECE scores in 2018 as endline reading outcomes to estimate long-term impacts (question 2).

Next, we describe the analytic approach to respond to the two research questions.

- 1. Short-term impacts.** To answer question 1, we estimated the impacts of *Amazonía Lee* on second-grade ECE reading scores in 2016. As explained above, we used the baseline information from 2015 to control for initial small differences between the treatment and control schools that arose by chance, despite random assignment. Because we cannot match EGRA with ECE scores at the student level, we used average EGRA score at the school level. The regression specification for estimating impacts on ECE scores in 2016 controls for school baseline characteristics in 2014 and school-average baseline EGRA reading scores in 2015. For each region, we estimated the impacts on students with the following regression:

$$(1) Y_{is} = a + \beta X_s + \lambda T_s + \varepsilon_{is}$$

where  $Y_{is}$  is the outcome of interest (second-grade ECE reading scores) for student  $i$  in school  $s$  at follow-up. The vector  $X_s$  represents the baseline characteristics of school  $s$ , which include average baseline reading measures at the school level and school characteristics such as access to services. The variable  $T_s$  is an indicator equal to one for students in schools assigned to receive the *Amazonía Lee* intervention and zero for those in the control group.  $\varepsilon_{is}$  is a random error term for student  $i$  in school  $s$ . We included cluster correction for school pairs used for randomization (de Chaisemartin 2020). The parameter  $\lambda$  is our main interest—the impact of the *Amazonía Lee* intervention on the outcome of interest.

2. **Long-term impacts.** To answer question 2, we estimate impacts of *Amazonía Lee* on fourth-grade ECE scores from 2018. For each region, we assessed the impacts on fourth-grade ECE reading scores using the same specification as in question 1 (equation 1, above). The only difference is that  $Y_{is}$  is the fourth-grade ECE reading score for student  $i$  in school  $s$ .

#### D. Limitations

Below we describe a set of potential limitations worth highlighting:

- **We cannot assess whether the impacts of *Amazonía Lee* were sustained by other cohorts of students.** *Amazonía Lee*'s theory of change implies long-term impacts on student reading achievement as a result of improved teaching practices during and after the teacher training and coaching period. However, the only available ECE scores for 2018 pertain to the cohort of fourth-graders. In *Amazonía Lee* schools, this cohort of students was exposed to the program in first (2015), second (2016), and third grade (2017). We assessed long-term impacts with these data. We would have needed ECE scores on a cohort of younger students in 2018 or 2019 to assess whether students in post-implementation cohorts in the *Amazonía Lee* schools had better ECE reading scores relative to the control schools.
- **When comparing the magnitude of effects using EGRA and ECE tests, we need to consider differences in time of exposure to the intervention, skills measured, and sample.** Because of four key differences, the magnitude of the effects estimated in the *Amazonía Lee* impact evaluation using the EGRA test cannot be directly compared to the magnitude estimated with the ECE test. Those assessed with EGRA were exposed to the intervention for 13 to 14 months, and students assessed with ECE had a few more months of exposure. Second, EGRA was administered to a sample of second-grade students, and ECE is a census test, so we have more observations and hence greater statistical power to detect impacts for the latter. Third, the ECE and EGRA do not measure the same reading skills. For reading comprehension, we used EGRA, which better matches the skills measured in ECE, but the tests are different (see Fernández et al). Fourth, because the sample is different, the impact estimate may differ, and we did not conduct statistical tests to compare the magnitudes.
- **Distinguishing school-level improvements due to the program from school-level shocks is not feasible.** An important consideration is that the method we are using to identify variation in impacts across schools, as described in the appendix, could confound the effects of the intervention with other potential unobserved school factors. While we can identify a group of schools that sustained impacts, we cannot rule out that an external school-specific shock contributed to the apparent sustained impacts.

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

---

## IV. Findings

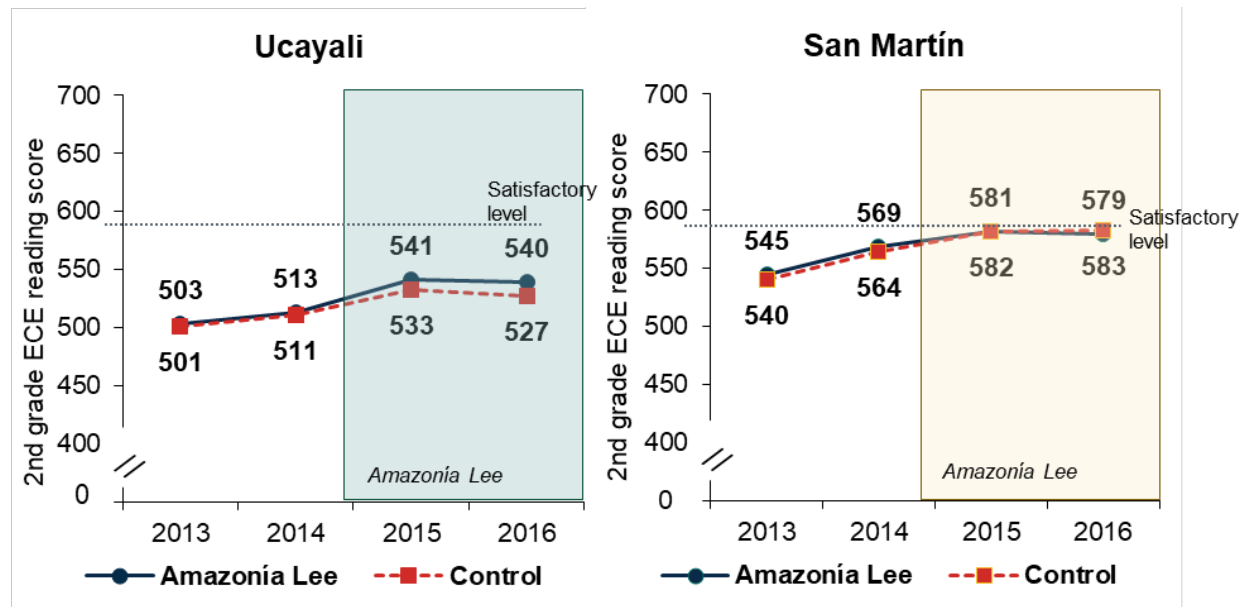
In this section, we first provide an overview of how the schools in the sample performed on the ECE from 2013 to 2016. Our goal is to give the reader contextual information on how the study schools scored on the ECE, and most important, to show that before the intervention started in 2015, the treatment and control schools in each region had similar scores using this metric. Second, we discuss the findings for short-term impacts on second-graders' ECE test in 2016 (question 1) and compare them to those found in the impact evaluation, which were measured with EGRA. Then we discuss the findings for long-term impacts on fourth-grade reading ECE scores (question 2) to determine whether *Amazonía Lee*'s impacts were sustained one year after the program ended.

To facilitate interpretation of the ECE results, we present the historical reading achievement level for schools in the evaluation sample. Figure IV.1 presents the trends for second-grade average ECE reading scores for schools in each of the two study groups in Ucayali and San Martín between 2013 and 2016.<sup>4</sup> In Ucayali and San Martín, the two study groups had similar reading scores before the implementation of *Amazonía Lee* in 2013 and 2014. In the two years of *Amazonía Lee* implementation (2015 and 2016), the treatment group in Ucayali seems to have higher scores than the control group. In San Martín, the ECE scores of treatment and control schools appear similar. However, a majority of schools in the control group in San Martín were implementing *Soporte Pedagógico*. The next section discusses the short-term impacts in 2016 found for each region.

---

<sup>4</sup> *Amazonía Lee* was implemented in both regions during 2017, but ECE data are not available for that year because MINEDU suspended the test.

Figure IV.1. Trends on ECE reading scores, by study groups and region



Source: Scores for 2nd-grade ECE reading test 2013–2016, Ministry of Education.

Note: *Amazonía Lee* implementation period is shown in blue shadow in Ucayali and yellow shadow in San Martín. ECE data are not available for 2017. A teacher strike and inclement weather caused school closures that affected the school calendar and students' learning, so MINEDU suspended the ECE in 2017. In 2018, ECE for 2nd grade was not collected as a census (in all schools) but only in a sample of schools which do not fully overlap with the evaluation sample.

The Ministry of Education reports three levels of achievement for the ECE scores: unsatisfactory, in process, and satisfactory. The definition of *satisfactory* is that the student achieved the expected learning at the end of 2nd grade and is prepared for the next grade. This figure refers to the satisfactory level as being at grade level for 2nd grade.

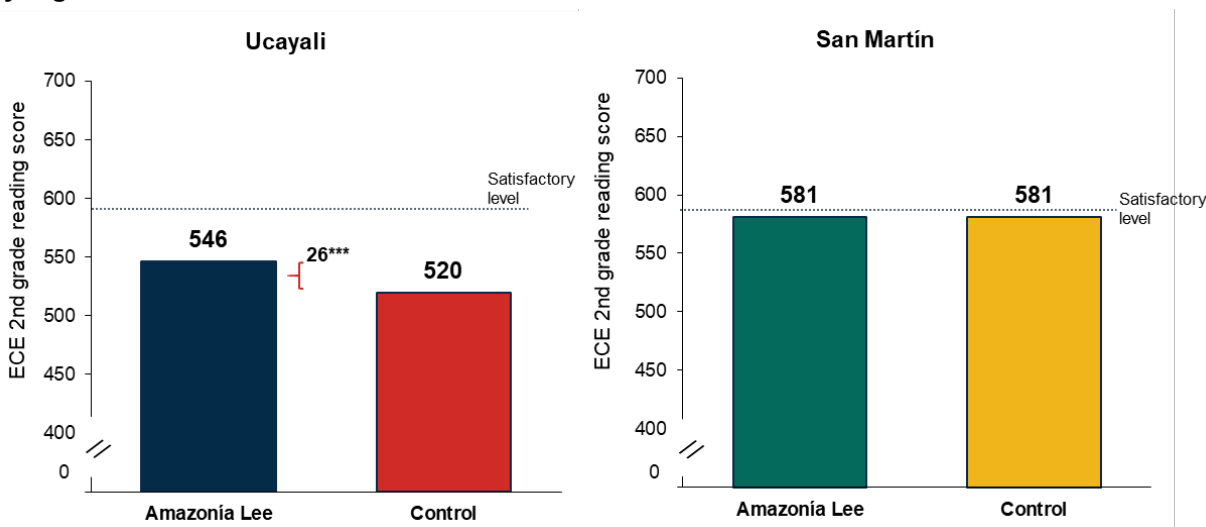
## A. Short-term impacts

To answer the first research question, we estimated the impacts of *Amazonía Lee* on second-grade ECE reading scores in 2016. We conducted separate analysis for each region.

**In Ucayali, *Amazonía Lee* had a positive impact on second-grade ECE reading scores.** Adjusting for initial chance school-level differences between the treatment and control groups, *Amazonía Lee* second-graders had on average higher ECE reading scores than students in the control group in 2016, the second year of *Amazonía Lee* implementation. As shown in Figure IV.2, the difference of 26 points is statistically significant. The magnitude of the impact is equivalent to 0.39 SD.

**In San Martín, second-grade students in both study groups had similar reading ECE scores in 2016.** *Amazonía Lee* had no impact on ECE scores in 2016 in San Martín. ECE reading scores for second-graders in 2016 adjusted for baseline differences were on average 581 for *Amazonía Lee* and control group students (see Figure IV.2).

**Figure IV.2. Positive impact of *Amazonía Lee* on adjusted means second-grade ECE reading score by region**



Source: Scores on the ECE national reading test in 2016, Ministry of Education.

Note: The bars present regression-adjusted means ordinary least squares. The specification includes school average baseline scores of students using the EGRA test in 2015 and baseline infrastructure characteristics of schools (2014).

The Ministry of Education reports three levels of achievement for the ECE scores: unsatisfactory, in process, and satisfactory. The definition of *satisfactory* is that the student achieved the expected learning at the end of second grade and is prepared for the next grade. This figure refers to the satisfactory level as being at grade level for second grade.

\*\*\* Difference in group-adjusted means is statistically significant at the 0.01 level ( $p$ -value=0.009).

**Comparison of *Amazonía Lee* short-term impacts measured with ECE vs. EGRA.** Overall, *Amazonía Lee*'s short-term impacts on student reading outcomes measured with the ECE test are consistent with the impact evaluation findings for second-graders using the EGRA test (Campuzano et al. 2018). Figure IV.3 shows *Amazonía Lee*'s short-term impacts on student reading tests comparing ECE and EGRA scores. We are comparing reading comprehension subtest from EGRA test with ECE reading scores only because the national test focuses primarily on reading comprehension<sup>5</sup>.

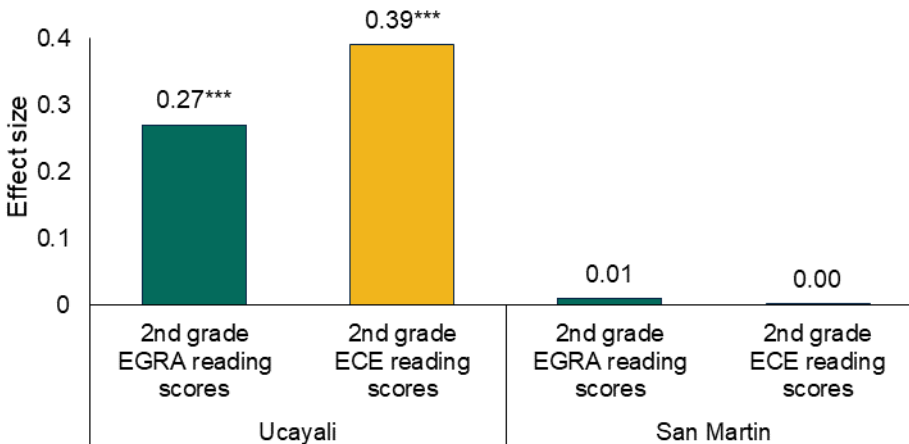
- In Ucayali, *Amazonía Lee* had positive impacts on reading comprehension measured with EGRA and ECE tests. Based on EGRA reading comprehension results, *Amazonía Lee* students were able to answer on average 3 of the 6 comprehension questions correctly; students in the control group were able to answer 2 questions correctly on average. The difference of 1 correct answer is statistically significant and equivalent to 0.27 SD. As mentioned above, we also found a positive impact of *Amazonía Lee* on students' ECE reading scores. The *Amazonía Lee* schools scored 26 points higher on the ECE test than the control group; and the impact is equivalent to 0.39 SD.
- In San Martín, *Amazonía Lee* students' reading skills were comparable to those of the control group in both metrics, EGRA and ECE scores. Students in both groups were able to answer 3 of the 6

<sup>5</sup> In the impact evaluation of *Amazonía Lee* program, we looked at impacts on these skills: (1) oral language comprehension, (2) familiar-word reading, (3) decoding skills, (4) oral reading fluency, and (5) reading comprehension.



reading comprehension questions correctly on the EGRA reading comprehension subtest. Both study groups also had similar ECE reading scores, 581 points on average.

**Figure IV.3. Short-term impacts on EGRA reading comprehension and on ECE reading scores measured in effect sizes, by region<sup>6</sup>**



Source: The effect size of 2nd-grade reading scores was estimated for the *Amazonía Lee* impact evaluation. The effect size of 2nd-grade ECE reading scores was estimated for LTIE study.

\*\*\* Impact is statistically significant at the 0.01 level.

## B. Long-term impacts

To answer the second research question, we estimated the impacts of *Amazonía Lee*, for both Ucayali and San Martín, on fourth-graders’ ECE scores in 2018. These children would have been exposed to the *Amazonía Lee* program in first, second, and third grade in 2015, 2016, and 2017 but not in 2018, when they were in fourth grade. We consider these estimates long-term impacts because they are measured one year after *Amazonía Lee* funding ended and one year after students had been exposed to the program from 2015 to 2017. As with research question 1, we estimated these impacts separately for Ucayali and San Martín.

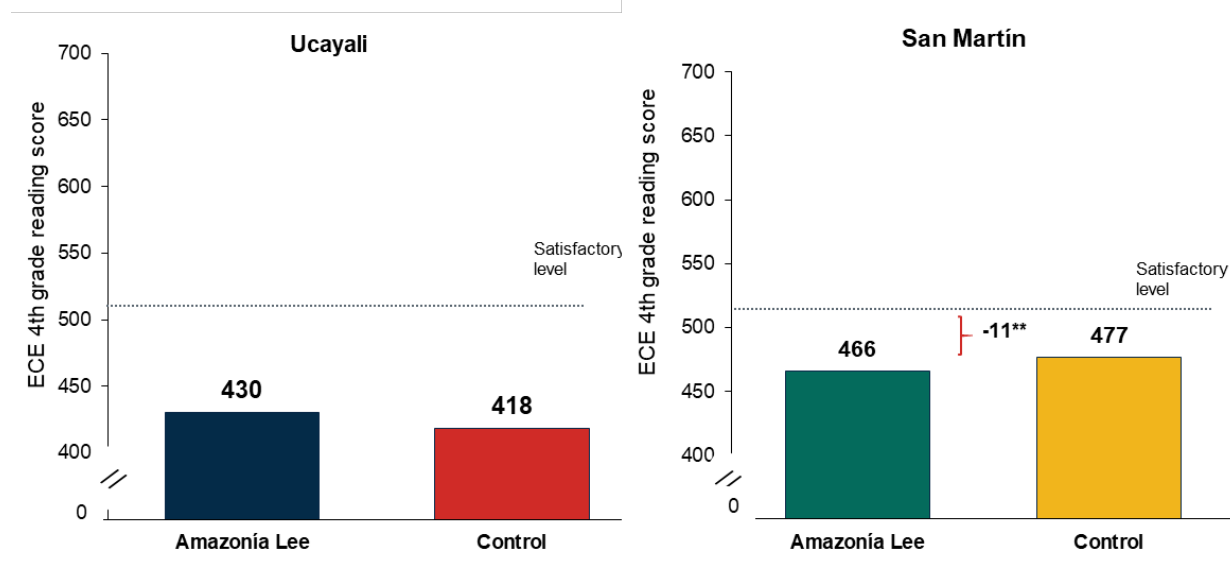
In Ucayali in 2018, the ECE reading scores of fourth grade-students in *Amazonía Lee* were higher than in the control schools, but differences were not statistically significant. The adjusted mean for *Amazonía Lee* fourth-grade students on the ECE reading test was 430 points. The adjusted mean for control group students on fourth-grade ECE reading score was 418. The 12-point difference favoring *Amazonía Lee* is not large enough to be considered statistically significant at the 5 percent level ( $p$ -value=0.102, see Figure IV.4). This difference is equivalent to 0.15 SD.

In San Martín, students in control schools had higher adjusted means on ECE fourth-grade reading scores in 2018 than *Amazonía Lee* students, and the difference is statistically significant. Fourth-graders had on average an adjusted ECE reading score of 477 in control schools and 466 in *Amazonía Lee* schools. The 11-point difference in favor of control students was statistically significant at the 5 percent level. This

<sup>6</sup> Effect sizes were estimated using the following formula for continuous variables: where  $w = [1 - 3 / (4N - 9)]$ ,  $N$  is total sample size,  $y_t - y_c$  is the difference in adjusted means,  $n_t$  is the sample size of treatment students and  $n_c$  is the sample size of control students,  $s_t^2$  is the squared standard deviation in the treatment group, and  $s_c^2$  is the squared standard deviation in the control group.

difference is equivalent to 0.12 SD. See Figure IV.4. Note that while in Ucayali, the difference in average scores was similar (12 points) it is not statistically significant because the standard error of the estimate was higher and the sample size was smaller.

**Figure IV.4. Adjusted means on fourth-grade ECE reading scores in 2018 by region**



Source: Scores on fourth-grade ECE reading test in 2018, Ministry of Education.

Note: Regression-adjusted means ordinary least squares. The specification includes school average baseline scores of students using EGRA test in 2015 and baseline infrastructure characteristics of schools (2014). The difference in means is not statistically significant  $p$ -value=0.102.

The Ministry of Education reports three levels of achievement for the ECE scores: unsatisfactory, in process, and satisfactory. The definition of *satisfactory* is that the student achieved the expected learning at the end of fourth grade and is prepared for the next grade. This figure refers to the satisfactory level as being at grade level for fourth grade.

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

---

## V. Conclusions, lessons learned, and recommendations

### A. Conclusions

**In Ucayali, *Amazonía Lee*'s positive impacts on second-grade students' ECE scores are consistent with the positive impacts on EGRA scores.** In Ucayali, second-grade students in *Amazonía Lee* schools had higher ECE reading scores than control group students in 2016, with a difference equivalent to 0.39 SD. Similarly, the previous impact evaluation had found that second-grade students in *Amazonía Lee* schools had higher reading comprehension EGRA scores in 2016 than control group students. The difference on EGRA scores is equivalent to 0.27 SD. The magnitude of these differences is within the range of what other studies have found for similar education interventions.<sup>7</sup> Using the ECE as another metric to identify the short-term impacts of *Amazonía Lee* was important in two ways: (1) short term-impacts are robust to the type of reading assessment used (for example, individual vs. group administered; evaluation-specific vs. high stakes); (2) the ECE is a contextually relevant metric for Peru for which there are historical data, and likely prospective data, which can inform or serve as a benchmark for future evaluation efforts. However, because of the differences in time of exposure, in skills measured, and in samples used to estimate impacts with each metric, the magnitude of the impacts obtained with EGRA and ECE should not be compared directly<sup>8</sup>.

**In San Martín, the failure to detect an impact of *Amazonía Lee* on second-grade students was consistent across the two metrics, ECE and EGRA scores.** In San Martín, second-grade students in *Amazonía Lee* schools had on average the same ECE reading scores of control group students in 2016, the difference being close to zero SD. Similarly, in the previous short-term evaluation we found that second-grade students in *Amazonía Lee* and control schools had similar reading comprehension EGRA scores in 2016, on average (close to zero SD difference). In San Martín, most teachers in the control group received training on reading instruction from *Soporte Pedagógico*. The findings suggest that *Amazonía Lee* and *Soporte Pedagógico* may have a similar effect on reading skills as measured by two assessments.

**In Ucayali, *Amazonía Lee*'s positive short-term impacts on ECE reading scores in second-graders were not found to be sustained when students reached fourth grade.** In 2016, *Amazonía Lee* students in second grade had higher ECE scores than control students; the difference was statistically and equivalent to 0.39 SD. By contrast, in 2018, *Amazonía Lee* fourth-grade students scored on average 430 points on the ECE test compared to 418 in the control group. Although *Amazonía Lee* students have higher scores than control students, the 12-point difference (equivalent to 0.15 SD) is not statistically significant at the 5 percent level.

**In San Martín, although second-graders had similar ECE reading scores in 2016, when students reached fourth grade in 2018, control group students had higher adjusted means on ECE reading scores compared to *Amazonía Lee* students.** In 2016, second-grade students in *Amazonía Lee* schools had reading ECE scores similar to those of students in the control group where the majority of schools implemented *Soporte Pedagógico* (a difference equivalent to close to zero SD). In contrast, in 2018, *Amazonía Lee* fourth-grade students scored on average 466 points on the ECE test compared to 477 in the

---

<sup>7</sup> A systematic review of education interventions (Snilstveit et al. 2016) finds that evaluations of structured pedagogy programs, similar to *Amazonía Lee*, have had impacts that are on average 0.23 SD of language arts outcomes, with a 95 percent confidence interval of (0.13, 0.34).

<sup>8</sup> The EGRA student sample in Ucayali was 740, and in San Martín it was 1646. The ECE student sample in Ucayali was 3931, and in San Martín it was 8220

comparison group. The 11-point difference in favor of the comparison group (equivalent to negative 0.12 SD) is statistically significant. To interpret the long-term differences found in San Martín, we need to understand what type of training services schools implemented after we conducted the follow-up for the initial *Amazonía Lee* impact evaluation. *Soporte Pedagógico* training was offered to teachers in all elementary grades from 2015 to 2017 school years, but we do not have reliable information on the services offered to teachers in 2018<sup>9</sup>.

**In both regions, the long-term impacts on ECE reading scores were smaller than the short-term impacts.** In San Martín, impacts became negative. In Ucayali, the short-term impact on 2016 second-grade ECE reading scores is equivalent to 0.39 SD and statistically significant, the long-term impact on 2018 fourth-grade ECE reading scores is equivalent to 0.15 SD, but the difference is not statistically significant. In San Martín, the short-term impact on 2016 second-grade ECE reading scores is close to zero SD and not statistically significant, but the long-term impact on fourth-grade ECE reading scores is equivalent to negative 0.12 SD and statistically significant. An important consideration is that *Amazonía Lee*'s teacher training focused on the first three grades of primary education. In contrast, *Soporte Pedagógico* offers training to teachers in all elementary grades. It is possible that because fourth grade teachers received *Soporte Pedagógico* training in the previous years, their students obtained better reading scores than students in *Amazonía Lee*, or that *Soporte Pedagógico* was offered in 2018 to control group teachers while *Amazonía Lee* was no longer implemented in the treatment group.

## B. Lessons learned from the *Amazonía Lee* LTIE

**In-service teacher training on early-grade reading instruction, whether supported by local technical assistance providers or by the national government, contributes to better student reading skills in the primary grades in the short and the long term.** In the short term, in Ucayali, *Amazonía Lee* schools had higher ECE reading scores than the control group which was not receiving in-service teacher training. In San Martín, where treatment and control school students had comparable scores, teachers from both groups received in-service training. Over the long term, when teacher training was no longer offered in *Amazonía Lee* schools, the difference with control schools was not large enough to be significant in Ucayali. But control schools in San Martín, which may have still been receiving in-service training through *Soporte Pedagógico*, outperformed *Amazonía Lee* schools over the long term.

**Using national tests can be a cost-effective data source for future education evaluations.** The RCT design of the *Amazonía Lee* evaluation enabled us to measure long-term program impacts and can continue to be used to explore other relevant questions related to early-grade reading in future cohorts of students in early primary grades. Moreover, the *Amazonía Lee* LTIE offers a solid evidence base for future evaluations that wish to use the ECE as the main reading outcome measure without incurring the high costs of individually administered early-grade reading assessments given the robustness of the short-term impacts to these two reading assessments. The advantages of using national tests, when available, could include (1) better statistical power in cases where the national test is collected for all students (census) as the ECE was collected, (2) local and national authorities have more confidence in the metrics they historically use to measure students' progress in reading, and (3) the costs to obtain these data are low compared with tailored and individually administered EGRA tests. However, some tradeoffs need to be considered. Because students cannot be identified in public data sets, the analysis may not attain the same level of precision as if we collected student pretest and posttest data. We are also limited on the type

---

<sup>9</sup> We designed a qualitative study to enrich learning gained through the LTIE. Due to COVID-19, it was not feasible to undertake qualitative data collection.

of skills assessed, most national data sets include reading comprehension, but fluency is not frequently included.

### C. Recommendations from the *Amazonía Lee* LTIE

**Continued in-service training improves students' early-grade reading in the short and long term and should be prioritized by policymakers, Ministries of Education, and the local DREs.** Based on the LTIE's main findings (1) regardless of the test used to measure reading skills (ECE national test or EGRA test), short-term impacts on second-graders are similar; and (2) long-term impacts were found for students attending schools where *Soporte Pedagógico* might still have been being implemented compared to schools that were no longer implementing a reading program. Policy makers should consider offering in-service teacher training on reading instruction to teachers in grades 1 through 4, whether it is delivered by a national education quality improvement program or by local technical assistance providers.

**Examining the costs of different approaches to support early-grade reading is critical for an equitable and efficient allocation of resources to schools in regions with different needs.** Both programs aiming to support teaching practices in reading, *Amazonía Lee* and *Soporte Pedagógico*, show promising results. Policymakers in Peru should assess which program is the most cost-effective to achieve impacts on students' early-grade reading—a critical building block of learning and further academic achievement. They should also consider the cost to scale up with fidelity and the regional capacity for program implementation to correctly size the in-service teacher training approach to maximize the use of limited resources.

**Adding to the evidence base on early grade reading through investing in rigorous LTIES provides additional value.** For the *Amazonía Lee* LTIE, we took advantage of the rigorous design of the initial *Amazonía Lee* evaluation to examine the robustness of our findings with different early-reading metrics, and to explore the sustainability of the impacts over time in a context of changing teacher training policies and investments at relatively low additional cost. Policymakers should invest in this type of LTIE, which takes advantage of rigorous designs used for short-term evaluations to get more value out of the evaluations they fund.

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

---

## References

- Bloom, H., L. Richburg-Hayes, and A. Black. "Using Covariates to Improve Precision for Studies That Randomize Schools to Evaluate Educational Interventions." *Educational Evaluation and Policy Analysis*, vol. 29, no. 1, 2007, pp. 30–59.
- Campuzano, L., C. Fernandez, N. Murray, J. Lugo-Gil, S. Glazerman, and I. Padilla. "Evaluation of Amazonía Lee Reading Intervention in Peru: Final Report." Submitted to the U.S. Agency for International Development. Washington, DC: Mathematica Policy Research, January 2018.
- de Chaisemartin, C., and J. Ramirez-Cuellar. "At What Level Should One Cluster Standard Errors in Paired Experiments, and in Stratified Experiments with Small Strata? July 22, 2020. Available at SSRN: <https://ssrn.com/abstract=3520820> or <http://dx.doi.org/10.2139/ssrn.3520820>.
- Deke, J., L. Dragoset, and R. Moore. "Precision Gains from Publically Available School Proficiency Measures Compared to Study-Collected Test Scores in Education Cluster-Randomized Trials." (NCEE 2010-4003.) Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2010.
- Dickinson, D.K., J.P. Brady, and J.A. Hirschler. "Evaluación Lectoescritura Emergente: Versión adaptada de la versión en inglés." Centro para el Desarrollo Educativo y Carolyn Chaney. Adaptation by Proyecto MEDIR (USAID / Juárez & Asociados), 2004.
- Kim, Y.-S. G., H.N. Boyle, S.S. Zuikowski, and P. Nakamura. "Landscape Report on Early Grade Literacy." Washington, DC: USAID, Global Reading Network, 2016.
- Murray, Nancy, Larissa Campuzano, Steve Glazerman, and Camila Fernandez. "Latin American and Caribbean Reading and Evaluation Services: Amazonía Reads Evaluation Plan." Submitted to the U.S. Agency for International Development. Washington, DC: Mathematica Policy Research, April 2016.
- National Institute of Child Health and Human Development. "Report of the National Reading Panel, Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction." NIH publication no. 00-4769, Washington, DC: U.S. Government Printing Office, 2000.
- Schochet, P. "Statistical Power for Random Assignment Evaluations of Education Programs." *Journal of Educational and Behavioral Statistics*, vol. 33, no. 1, 2008, pp. 62–87.
- Snilstveit, B., J. Stevenson, R. Menon, D. Phillips, E. Gallagher, M. Geleen, H. Jobse, T. Schmidt, and E. Jimenez. "The Impact of Education Programs on Learning and School Participation in Low- and Middle-Income Countries: A Systematic Review Summary Report." 3ie Systematic Review Summary 7. London: International Initiative for Impact Evaluation (3ie), 2016.
- USAID. "Guide for Planning Long-Term Impact Evaluations (LTIEs). Utilizing the Expertise of the Expanding the Reach of Impact Evaluation (ERIE) Program Consortium. USAID, September 2018.
- USAID/Reforma Educativa en el Aula. "Evaluación de Lectura para Grados Iniciales." Misión Guatemala (USAID/G), with Juárez y Asociados, 2013. Unpublished instrument.
- Policy Research Working Paper 8292. "How Effective Are Early Grade Reading Interventions? A Review of the Evidence." World Bank Group, January 2018.



## References

---

World Bank. “Por una educación de calidad para el Perú. Estándares, rendición de cuentas y fortalecimiento de capacidades”. Lima, Peru, 2006.

Young-Suk G. Kim; Hansol Lee; Stephanie S. Zuilkowski. “Impact of Literacy Interventions on Reading Skills in Low- and Middle-Income Countries: A Meta-Analysis.” *Child Development*. March/April 2020, Volume 91, Number 2, Pages 638–660



**USAID**



**PULTE INSTITUTE**  
FOR GLOBAL DEVELOPMENT



**Mathematica**  
Progress Together