



USAID
FROM THE AMERICAN PEOPLE

EVALUATION OF *AMAZONÍA LEE* READING INTERVENTION IN PERU: FINAL REPORT



January 2018

This report was produced for review by the United States Agency for International Development under task order AID-OAA-M-12-00020. It was prepared by Larissa Campuzano, Camila Fernández, Julieta Lugo-Gil, Steve Glazerman, Nancy Murray, and Ivonne Padilla of Mathematica Policy Research.

EVALUATION OF *AMAZONÍA LEE* READING INTERVENTION IN PERU: FINAL REPORT

DISCLAIMER

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

ABSTRACT

This study estimated the impacts of *Amazonía Lee*, an early grade reading program in Peru delivering in-service teacher training, coaching, and materials as well as leadership training and non-monetary incentives to improve reading instruction. We examined impacts on teaching practices and literacy skills of participating students. We randomly assigned schools to a program group that implemented *Amazonía Lee* or a control group that did not. The study was conducted in two areas (departments) of the Amazon region. In one department, Ucayali, control schools received the usual services provided by the Ministry of Education. In the other department, San Martín, most control schools received a program known as *Soporte Pedagógico*, a Ministry initiative that offered teacher training and coaching, pedagogical materials, and remedial tutoring.

In both regions, we found a few impacts on teacher instructional practices, including teachers' increased explanation of different types of texts and greater vocabulary development. The impacts on students differed by region. In Ucayali, *Amazonía Lee* had positive and substantial impacts on reading outcomes. For example, students in *Amazonía Lee* on average could read one additional familiar word (7 versus 6, an impact equivalent to 0.24 standard deviations), 3.5 additional invented words (16.8 versus 13.4, an impact on decoding of 0.24 standard deviations), and scored higher on a reading comprehension test (impact of 0.27 standard deviations), with no gender differences. In San Martín, where *Soporte Pedagógico* was prevalent, we found no impacts on students' reading outcomes relative to the control schools.

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

ACKNOWLEDGMENTS

The authors gratefully acknowledge numerous people and organizations that contributed to the implementation of all evaluation activities in Peru and to the production of this report.

The staff at the regional government of the departments of San Martín and Ucayali and their local education administrations, Dirección Regional de Educación (DRE) provided essential institutional support for the evaluation. Pilar Saavedra, Wilson Guerrero, Ricardo Quevedo, Esteban Ruíz, Ernesto Morales Fasanando, in San Martín and Wagner Sanchez, Victor García, Zoila García and Angela Villacorta in Ucayali provided important support in coordinating evaluation activities, continuous assistance during our data collection efforts, and useful qualitative information on program implementation.

Universidad Peruana Cayetano Heredia (UPCH), *Amazonía Lee* project director Raquel Villaseca provided us key information on program implementation and valuable input on data collection instruments and evaluation findings. Roxana Atoche Mondragón, Blanca Franco Díaz and the rest of UPCH's technical and administrative project teams in San Martín and Ucayali also provided valuable information and assistance in the coordination of the evaluation activities.

USAID/Washington and USAID/Peru staff have provided guidance and support to the evaluation design and its implementation. They include, from USAID/Washington, Barbara Knox-Seith, Senior Education Advisor Bureau for Latin America and the Caribbean, and while at USAID/Peru, Clifford Brown, Deputy Chief, Governance & Institutional Strengthening; Natalia Petroni, Monitoring and Evaluation Specialist; Sobeida Gonzalez, Project Coordinator Governance and Institutional Strengthening; Miriam Choy, Development Assistance Specialist; Evelyn Rodriguez-Perez, Chief, Health and Education Office; Marisol Pérez, Office Director, Governance and Institutional Strengthening; Jo Jean Elenes, Acting Chief, Health and Education Office; and Patricia Perez, Education Advisor, Health and Education Office.

We thank our local evaluation partner, Grupo de Análisis para el Desarrollo (GRADE), and its staff for high quality data collection in the study schools and communities. Miguel Jaramillo ensured careful planning of evaluation activities in Peru, and provided keen oversight of all data collection activities. Elsa Bardalez developed a collaborative relationship with regional staff at the DREs and has been instrumental in obtaining and maintaining local support for the study in the Amazonian region. Juan José Díaz provided overall project management support, oversaw the production of the datasets, and ensured consistency of data programming. Gabriela Guerrero assisted in various study activities including the design of the follow up reading assessments and enumerator training. María Balarín provided input on the approach for qualitative data collection on program implementation. Carla Pajares, Varinia Arévalo, Licet Infantas Díaz, Julio Almonacid, Luciana de la Flor and Emilio Legonia coordinated and supervised data collection and data entry activities.

We are also indebted to Sylvia Linan-Thompson for her thorough and incisive input at different stages of the evaluation. Her expertise in models of reading interventions and broad experience with measurement of early grade reading skills were very useful in designing the evaluation and the reading assessment instruments.

At Mathematica Policy Research, Galina Lapadatova, Sarah Chan, and Colleen Fitts provided critical support in data cleaning, programming, and presentation.

We are enormously appreciative of all the school directors and teachers for opening the doors of their classrooms and allowing us to observe the valuable work they do with students. Last, but not least, we thank the students who provided us with the important information we are able to share in this report.

CONTENTS

ABSTRACT	iii
ACKNOWLEDGMENTS.....	v
EXECUTIVE SUMMARY	ES-1
A. Introduction	ES-1
B. Evaluation questions and design.....	ES-2
C. Implementation findings.....	ES-3
D. Ucayali impact findings.....	ES-4
1. Impacts on use of services	ES-5
2. Impacts on teacher practices and use of time	ES-5
3. Impacts on students' reading skills	ES-6
E. San Martín impact findings	ES-7
1. Impacts on use of services	ES-7
2. Impacts on teacher practices and use of time	ES-8
3. Impacts on students' reading skills	ES-9
F. Overall Conclusions, Lessons Learned, and Recommendations.....	ES-9
1. Conclusions.....	ES-9
2. Lessons Learned.....	ES-10
3. Recommendations	ES-10
I INTRODUCTION.....	11
A. The <i>Amazonía Lee</i> intervention.....	12
B. Research questions, evaluation design and limitations	15
C. Data sources and indicators	19
1. Data collection instruments and procedures.....	19
2. Timing of data collection	23
3. Qualitative data collection on program implementation	23
II IMPLEMENTATION OF <i>AMAZONÍA LEE</i>	25
A. Rollout of program components in Ucayali and San Martín	25
B. Services provided to control schools.....	30
C. Principal and teacher perceptions of <i>Amazonia Lee</i>	33
III FINDINGS FOR UCAYALI	35
A. Impacts on school participation in support programs	35

1. Teacher-reported participation in professional development.....	36
2. Literacy resources and materials in the classroom.....	40
B. Impacts on teaching activities and practices	41
1. Impacts on classroom practices and time use	41
2. Impacts on differentiated instruction and support for underperforming students	45
3. Impacts on the use of formative assessments to monitor students' learning	46
C. Impacts on students' reading skills.....	47
IV FINDINGS FOR SAN MARTÍN	53
A. Impacts on school participation in support programs	53
1. Teacher-reported participation in professional development.....	55
2. Literacy resources and materials in the classroom.....	59
B. Impacts on teaching activities and practices	60
1. Impacts on classroom practices and time use	60
2. Impacts on differentiated instruction and support for underperforming students	65
3. Impacts on the use of formative assessments to monitor students' learning	65
C. Impacts on students' reading skills.....	65
V CONCLUSIONS, LESSONS LEARNED, AND RECOMMENDATIONS	69
A. Conclusions	69
1. What was the impact of <i>Amazonía Lee</i> programs on early-grade reading and other intermediate outcomes relative to other prevailing practices and programs?	69
2. How were <i>Amazonía Lee</i> intervention components, activities, and processes implemented in Ucayali and San Martín?	69
3. What services were provided to control schools in Ucayali and San Martín?	70
4. How were <i>Amazonía Lee</i> intervention components, activities, and impacts perceived by principals and teachers?.....	70
5. What were the facilitators and barriers to the implementation of <i>Amazonía Lee</i> ?	71
B. Lessons learned	72
C. Recommendations.....	72
REFERENCES.....	75
APPENDIX A EVALUATION SAMPLE AND IMPACT ESTIMATIONA	
APPENDIX B IMPLEMENTATION	
APPENDIX C ADDITIONAL TABLES AND FIGURES FOR UCAYALI	
APPENDIX D ADDITIONAL TABLES AND FIGURES FOR SAN MARTÍN	
APPENDIX E DEVELOPMENT OF READING ASSESSMENT INSTRUMENTS AND CLASSROOM OBSERVATION PROTOCOL	

TABLES

I.1	Summary of data sources and outcome indicators.....	20
I.2	Description of the endline reading assessment instrument	21
I.3	Timing of data collection	23
II.1	Program features of <i>Amazonía Lee</i> and <i>Soporte Pedagógico</i>	32
III.1	Treatment schools were more likely to participate in <i>Amazonía Lee</i>	36
III.2	Treatment group teachers reported participating in reading instruction professional development at much higher rates than control group teachers	37
III.3	More teachers in the treatment group than in the control reported receiving in-class coaching	40
III.4	Teachers were observed performing activities related to teaching reading at similar rates in both groups, with one exception	42
III.5	<i>Amazonía Lee</i> led to positive impacts of substantive magnitude on reading familiar words, decoding, fluency, and reading comprehension.....	48
III.6	Positive impacts for boys and for girls in decoding and reading comprehension and only for boys in reading familiar words and fluency.....	51
IV.1	Schools in both groups participate in various support programs.....	54
IV.2	Participation in reading instruction professional development activities is similar between groups	56
IV.3	Except for demonstration sessions, in-class coaching activities are similar between groups	59
IV.4	Teachers were observed performing activities related to teaching reading at similar rates in both groups, with two exceptions	61
IV.5	Students reading outcomes were similar between groups	66
IV.6	Students reading outcomes were similar between groups	68
A.1	Student sample and attrition rates	A-4
A.2	Baseline characteristics included in the regressions for students reading outcomes	A-6
A.3	Analysis and imputed sample	A-8
A.4	Transfers and dropouts in student analysis sample	A-9
B.1	Qualitative data collection (2016): respondents, methods, and topics of inquiry	B-5
C.1	Schools at baseline had similar infrastructure characteristics with some exceptions (percentage, unless otherwise indicated)	C-3
C.2	Students had similar characteristics at endline with the exception of being overage for first grade (percentage, unless otherwise indicated)	C-4

C.3	Children in both groups had similar early reading scores at baseline, except for phonemic awareness	C-4
C.4	Treatment group teachers are more likely to report receiving textbook, narrative books, workbooks, guides, and audiovisual resources from <i>Amazonía Lee</i>	C-7
C.5	Teachers in the treatment group were more likely to receive materials from <i>Amazonía Lee</i> and having a classroom library	C-8
C.6	Control teachers were more likely to report as a barrier to teach lack of pedagogical resources and lack of training to teach literacy.....	C-8
C.7	Teachers in both groups reported doing literacy activities at similar rates	C-9
C.8	Teachers in both groups reported spending similar time in all activities, except for participating in reading activities	C-12
C.9	Teachers in both groups spent similar time on teaching practices not focused on reading	C-13
C.10	Teachers in both groups reported using most activities related to classroom climate at similar rates	C-13
C.11	Teachers in both groups reported using differentiated instruction practices at similar rates, except for planning to teach for different levels of learning	C-14
C.12	Teachers in the treatment report using differentiated instruction less often than control teachers	C-14
C.13	Treatment group teachers reported working extra hours with underperforming students in reading	C-15
C.14	Teachers in both groups report using formative assessments at similar rates except for assessed student achievement at the end of the school year	C-15
C.15	Students reading outcomes were similar between groups	C-16
C.16	<i>Amazonía Lee</i> impacts on reading familiar words, decoding, fluency, and reading comprehension, sample with non-missing baseline data	C-17
C.17	Students characteristics at endline and early reading scores at baseline, sample with non-missing baseline data	C-17
D.1	Schools at baseline had similar infrastructure characteristics with some exceptions (percentage, unless otherwise indicated)	D-3
D.2	Students in both groups have similar characteristics at endline but significant differences in overage and repetition of first grade at baseline (percentage, unless otherwise indicated).....	D-4
D.3	Children in both groups had similar early reading scores at baseline with two exceptions.....	D-4
D.4	More teachers in the treatment group report having a library in the classroom and receiving material from <i>Amazonía Lee</i>	D-5
D.5	Control teachers are more likely to report lack of training to teach literacy as a factor limiting their ability to teach students to read.....	D-7
D.6	Teachers reported receiving materials from MINEDU at similar rates	D-8

D.7	Teachers in both groups reported doing literacy activities with similar rates.....	D-9
D.8	Teachers in both groups were observed performing general teaching practices with similar frequency	D-12
D.9	Teachers reported using most activities related to classroom climate at similar rates.....	D-12
D.10	Teachers in both groups report spending similar hours in teaching related activities, except for monitoring students during recess class	D-13
D.11	Teachers reported using differentiated instruction practices at similar rates, except for planning to teach for different levels of learning more times per month	D-14
D.12	Teachers in both groups work extra hours with underperforming students in reading but control teachers are more likely to refer them to receive academic reinforcement	D-14
D.13	Teachers in both groups report using formative assessments at similar rates.....	D-15
D.14	Percentage of boys and girls that got a score of zero in reading assessment task were similar in both groups	D-16
D.15	<i>Amazonía Lee</i> impacts on reading familiar words, decoding, fluency, and reading comprehension, sample with non-missing baseline data	D-17
D.16	Students characteristics at endline and early reading scores at baseline, sample with non-missing baseline data	D-17
E.1	Endline EGRA for pilot 1 (Form B).....	E-6
E.2	<i>Amazonía Lee</i> Endline EGRA.....	E-8
E.3	Description of classroom and school observation instrument.....	E-10

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

FIGURES

I.1	Map of study locations in Peru	11
I.2	Percentage of 2nd-grade students who achieved the expected (satisfactory) grade-level reading comprehension score on the ECE test	12
I.3	<i>Amazonía Lee</i> logic model.....	13
I.4	Evaluation design: randomized controlled trial	17
II.1	Implementation timeline, 2015 and 2016	26
III.1	Total training hours in reading instruction are similar between groups	38
III.2	Treatment group teachers reported higher participation rates in training for all the topics related to reading instruction.....	39
III.3	Teachers in both groups distributed class time on literacy activities similarly	44
III.4	Histogram of correct answers in reading comprehension.....	49
III.5	Students in the treatment group got a score of zero at lower rates than the control group.....	50
IV.1	Total training hours reported by teachers are similar between groups.....	57
IV.2	Treatment group teachers reported higher participation rates in training for most topics related to reading instruction.....	58
IV.3	Treatment teachers spent more class time on oral language and written text and less time on reading comprehension activities.....	63
IV.4	Histogram of correct answers in reading comprehension.....	67
IV.5	Fewer students in the treatment group got a zero score in the fluency test.....	68
B.1	Organizational chart for Ucayali, 2016.....	B-4
B.2	Organizational chart for San Martín, 2016	B-4
C.1	Treatment group teachers were more likely to report being trained in other pedagogical strategies and classroom management.....	C-5
C.2	Literacy resources and materials observed in second grade classrooms	C-6
C.3	Both groups performed initial literacy activities daily at similar rates.....	C-10
C.4	Treatment group more likely to read for pleasure every day and control group more likely to ask about the topic of the reading every day	C-11
C.5	Skills demonstrated by at least 50 percent of students are similar between groups, except for two skills (reported by teachers).....	C-16
D.1	Teachers in the two groups report being trained in other pedagogical strategies at similar rates	D-5
D.2	Observed literacy resources and materials were similar in both groups	D-6
D.3	Initial literacy activities performed every day were similar for both groups.....	D-10

D.4	Treatment group more likely to develop vocabulary every day	D-11
D.5	Skills demonstrated by at least 50 percent of students are similar between groups (reported by teachers).....	D-16
E.1	Process for the development of the endline reading assessment	E-3

ACRONYMS

AprenDes	Innovations in Decentralization and Active Schools
DRE	Dirección Regional de Educación (Regional Education Office)
ECE	Evaluación Censal de Estudiantes
EGRA	Early Grade Reading Assessment
GRADE	Grupo de Análisis para el Desarrollo (Analysis Group for Development)
LANN	Liderando los Aprendizajes de Niñas y Niños (Leading the Learning of Girls and Boys) (Local name in Ucayali for <i>Amazonía Lee</i>)
PEEL	Programa Enseñar Es Liderar (To Teach Is to Lead) (Local name in San Martín for <i>Amazonía Lee</i>)
PELA	Programa Estratégico Logros de Aprendizaje
MINEDU	Ministry of Education
REDES	Programa Nacional de Redes Educativas Rurales
SPI	Soporte Pedagógico Intercultural
UGEL	Unidad de gestión educativa local (Provincial Education Unit)
UNESCO	United Nations Educational, Scientific and Cultural
USAID	United States Agency for International Development
UPCH	Universidad Peruana Cayetano Heredia (Cayetano University of Peru)

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

EXECUTIVE SUMMARY

A. Introduction

This report presents results of the impact evaluation of *Amazonía Lee* (Amazon Reads), an early-grade reading intervention funded by the U.S. Agency for International Development (USAID). *Amazonía Lee*'s goal is to improve the reading skills of children in two administrative departments in the Amazon region of Peru: Ucayali and San Martín. Ucayali and, to a lesser extent, San Martín lag behind the nation and the metropolitan area of Lima in important student reading achievement indicators.

Amazonía Lee was implemented by the Department of Regional Education Offices (DREs) of Ucayali and San Martín starting in 2015. In San Martín, the program is known as “*Programa Enseñar Es Liderar*” (PEEL). In Ucayali the program is called “*Liderando los Aprendizajes de Niñas y Niños*” (LANN). The two *Amazonía Lee* programs share four main components, described below, and the same technical assistance provider, Universidad Peruana Cayetano Heredia (UPCH).

1. Capacity building. UPCH provided technical assistance to both departments for the implementation of five capacity-building subcomponents: teacher training, coaching, teacher study groups, more and varied materials to support reading, and school leadership training. The main subcomponent is the teacher training where teachers are trained on *Amazonía Lee*'s *balanced approach*, which combines evidenced-based best practice pedagogical skills such as alphabet knowledge, phonological awareness, reading fluency, comprehension, and writing with the *communicative textual* approach to early grade reading typically supported by the Ministry of Education. Teachers are also trained on the design and use of formative assessments, on positive classroom climate strategies, and on community engagement strategies in support of students' literacy. Training is 120 hours long delivered in 40-hour blocks throughout the year.

2. Community engagement. This component involves book fairs, family storytelling, shared book reading, and community workshops, and other reading promotion activities outside of school.

3. Regional assessment system. This component aims to introduce regional assessments at the beginning and end of the school year and enhance local capacity to design, implement, score, and interpret these assessments.

4. Regional incentive system for teachers. This component includes non-monetary strategies to encourage, recognize, and reward teacher performance.

The program logic is that these components should result in improved teacher training for reading instruction, more relevant reading materials, better school management, increased support of parents and the community for reading. These in turn should lead to students' greater acquisition of literacy skills—the program's ultimate goal.

B. Evaluation questions and design

The evaluation was structured to address the following central research question:

- What was the impact of *Amazonía Lee* programs on early-grade reading and other intermediate outcomes relative to other prevailing practices and programs?

In addition, to further advance our understanding of the implementation of the program in the context of the Amazon region of Peru and to help interpret the impacts, the evaluation sought to address a set of additional questions related to implementation:

- How were *Amazonía Lee* intervention components, activities, and processes implemented in Ucayali and San Martín?
- What services were provided to control schools in Ucayali and San Martín?
- How were *Amazonía Lee* intervention components, activities, and impacts perceived by principals and teachers?
- What were the facilitators and barriers to the implementation of *Amazonía Lee*?

To answer these questions, we conducted an impact evaluation and implementation study.

Impact evaluation. The study 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 this program, but could have participated in other programs. The randomized evaluation relies on comparing outcomes between these two groups of schools in each department to estimate the impacts of *Amazonía Lee* activities on the outcomes of interest.

Because the Ministry of Education was providing different services to the schools assigned to the control group in each of the two departments at the time of the evaluation, we conducted separate analyses for each department. In Ucayali, control schools received the usual services the Ministry of Education provides in the department. In San Martín, a majority of the control schools received additional assistance in the form of a primary education support program known as *Soporte Pedagógico* (Pedagogical Support), a Ministry of Education initiative implemented in San Martín starting in 2015.

It is important to understand *Soporte Pedagógico*, given that it is largely the basis for comparison for *Amazonía Lee* in San Martín. *Soporte Pedagógico* aims to improve student learning in the areas of mathematics, communication, science, and social skills in urban elementary schools by strengthening skills of teachers and administrators, providing after-school tutoring for students not meeting the expected learning standards, delivering educational materials and resources, and improving the school and local management capacity. There are three main differences between *Amazonía Lee* and *Soporte Pedagógico*. First, *Amazonía Lee* focuses on reading and communication, whereas *Soporte Pedagógico* focuses on several subjects, including communication, math, and science. Second, the approach to reading instruction of *Amazonía Lee* (balanced approach) integrates instruction on foundational reading skills with the national *communicative-textual* approach, whereas *Soporte Pedagógico* uses primarily the national *communicative-textual* approach. Third, *Soporte Pedagógico* provides

small-group tutoring to underperforming students after school with teachers specialized in academic reinforcement while *Amazonía Lee* does not provide this service.

Data sources and indicators

Quantitative data. The main data sources for the impact evaluation are student reading assessments, classroom observations, a teacher survey, and a school infrastructure survey. These instruments were used to collect baseline and endline data. The endline child assessment was a version of the Early Grade Reading Assessment that tested oral comprehension, familiar word reading, decoding, reading fluency, and reading comprehension. The classroom observation measured teachers' use of class time, their pedagogical practices, and the classroom literacy environment.

Implementation study data. We adopted a mixed-methods approach to understand the implementation of *Amazonía Lee*, integrating qualitative and survey data and a desk review of program materials. We conducted interviews with UPCH's technical assistance team, DRE project coordinators, *Amazonía Lee* trainers, school directors, and teachers. We also reviewed program documents including UPCH's quarterly reports, DRE's implementation reports, UPCH's design reports of intervention components, and UPCH's teacher training materials.

C. Implementation findings

The DREs of Ucayali and San Martín implemented *Amazonía Lee* with technical assistance from UPCH. Implementation started in 2015 in both departments with the rollout of the capacity-building component. The family and community engagement, teacher incentives, and reading assessment system components were rolled out more gradually and were only partially implemented.

The teacher training workshops and coaching sessions were implemented successfully especially after adjustments were made in the second year. UPCH trainers trained and coached LANN and PEEL trainers on the methodology for conducting teacher workshops, coaching sessions, demonstrative classes, and study group meetings. Subsequently, LANN and PEEL trainers trained teachers in *Amazonía Lee* schools. However, two challenges identified in the first year required adjustments for the second year of implementation. First, the number of UPCH trainers was not sufficient to support the cadre of local LANN and PEEL teacher trainers with the expected frequency and level of supervision needed. To address that challenge, UPCH hired more lead trainers for the second year of implementation. In San Martín, coaching visits were on average, every other month whereas in Ucayali they were monthly. The second challenge was that specific educational and work experience requirements for the recruitment and selection of teacher trainers were not well defined or enforced the first year. Consequently, LANN and PEEL trainers were not adequately prepared to carry out the teacher training and coaching activities with the level of proficiency that UPCH expected. To improve the recruitment and selection of teacher trainers for the second year of implementation, UPCH, in collaboration with the DREs in both departments, set stricter requirements.

Teachers reported being satisfied with materials provided by *Amazonía Lee* despite delays in delivery of students' materials. There were substantial delays in the delivery of the workbooks and classroom libraries in both departments, which resulted in a considerable lag

between the beginning of training activities and the availability of pedagogical materials for reading instruction and reading for pleasure. In 2015 teacher guides and student workbooks were delivered towards the end of the school year. In 2016, these pedagogical materials for reading instruction were delivered approximately six months after classes started. Similarly, classroom libraries to promote reading for pleasure, were only delivered in the last quarter of the second year of implementation. In spite of the delays, teachers were very satisfied with *Amazonía Lee* materials and felt that they had advantages over other pedagogical materials the teachers had used previously. Teachers perceived that the grade-level materials they receive from the Ministry of Education are generally too difficult for their students and are culturally biased to the coastal and Andean regions. In contrast, teachers felt *Amazonía Lee* teacher guides and student workbooks are well targeted for the skill level of their students and emphasize culturally relevant themes and activities of the Amazon region.

Teachers and school directors perceived that community engagement activities had promoted parental school engagement. In the first year of implementation, only teachers from Ucayali implemented parent workshops, but 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. Teachers and school directors noticed more frequent parental attendance to school meetings and greater efforts at home to support students' reading and school work.

In both departments, efforts to implement the reading assessment system were focused on Amazonía Lee schools only; however, the system was fully implemented in Ucayali but not in San Martín. UPOCH provided technical assistance, developed the reading assessments, and provided the administration, scoring, and interpretation guidelines to the regional implementation teams. In Ucayali, LANN trainers administered the reading assessments to students in 2015 and 2016. Teachers received the assessment results, but in our interviews, we learned that most teachers did not rely on them to plan differentiated instruction. In San Martín, the implementation team administered the assessments in 2016, but experienced operational problems with the optic readers; therefore, results were not able to be provided for teachers' use.

During the study period, only some activities related to the teacher incentive system were implemented in Amazonía Lee schools. The nonmonetary teacher incentive system was designed to recognize effective teaching and motivate teachers to improve instruction. Nonmonetary incentives were "distinction awards" conferred to high-performing teachers. In Ucayali, the implementation of this component was still in the planning stage in November 2016. In San Martín, the implementation of the incentive system was further along: 100 teachers had been selected as finalists, and the distinction awards were conferred in December 2016.

D. Ucayali impact findings

In Ucayali, most schools in the control group did not receive additional support beyond what the Ministry of Education offers. Therefore, the treatment-control contrast can be interpreted as the offer of *Amazonía Lee* versus business as usual services.

1. Impacts on use of services

***Amazonía Lee* led to more professional development in reading instruction.** The treatment group had higher rates of participation in reading professional development programs and they covered more topics than those reported by the control group. Nearly all the teachers in the treatment group reported participating in reading instruction professional development, whereas less than half of the control teachers did. Furthermore, the treatment group was more likely to report being trained by *Amazonía Lee* and the control group by other institutions. Teachers in the treatment group were more likely to be trained in topics related to reading instruction and in other pedagogical strategies and classroom management topics.

***Amazonía Lee* led to more in-class coaching with demonstration and observation opportunities for teachers.** Almost all teachers in the treatment group reported receiving in-class coaching, but only half of teachers in the control group did. Treatment teachers were more likely to be coached by *Amazonía Lee* and more likely to be coached in all the reading-specific topics we inquired about. The largest differences between *Amazonía Lee* teachers and their control group counterparts were in the extent to which they participated in demonstration sessions and received feedback on their teaching practices. *Amazonía Lee* teachers were also more likely to receive coaching on how to identify students' learning needs, reinforce key concepts, provide extra assistance for at risk students, conduct assessments to track students' performance in class, and design and prepare materials.

***Amazonía Lee* provided more books and materials for teaching reading.** Treatment teachers were more likely than control teachers to report having a classroom library. In the treatment group, by the end of the second year of the intervention 91 percent of teachers reported having a classroom library as compared to only 6 percent in the control group.

2. Impacts on teacher practices and use of time

Despite the positive impacts described above on training, coaching, and materials, teaching practices were similar across groups. However, the few impacts found are consistent with treatment teachers implementing *Amazonía Lee*'s balanced approach. Although teachers in both groups were observed using several practices related to teaching reading at similar rates, treatment teachers were more likely to be observed explaining the use of different types of text than control group teachers. Teachers in both groups also reported performing activities related to *emerging literacy and fluency* at similar rates and frequencies, with one exception—treatment teachers were less likely than control teachers to report spending time on daily dictation. Teachers in both groups were also asked if they used any of a number of recommended activities related to *reading comprehension*, and we found only one difference across the study groups. Teachers in the treatment group were more likely than those in the control group to report teaching the students how to make graphic organizers of the text, a learning tool to help students capture the meaning of a text. Furthermore, teachers in the treatment group were more likely than control group teachers to report spending time daily in class on reading for pleasure and less likely to report asking questions about topic or main ideas every day.

No impacts were found on observed allocation of class time among key literacy activities or on reported class duration. We found no statistically significant differences in the proportion of

time teachers spent in reading activities. In both groups, we estimated that teachers spent almost 30 percent of their time on reading comprehension, 27 percent of their time reading text, and around 19 percent of their class time on writing tasks.

***Amazonía Lee* teachers reported spending more time on activities to promote reading in the community after school time than control teachers.** We asked teachers to report the time spent per week on 15 activities related to their job such as time spent on administrative tasks, teaching, monitoring students, etc. We found only one statistically significant difference between groups—treatment teachers reported spending on average half an hour more per week than control teachers participating in activities to promote reading in the community after school time. This difference is consistent with the successful implementation of the community engagement component of *Amazonía Lee* by the teachers in the treatment group.

Treatment teachers reported using differentiated instruction strategies less frequently than control teachers. As part of the training workshop teachers were taught differentiated instruction strategies. Treatment group teachers reported using three of the six strategies related to differentiated instruction less frequently than control teachers. We found statistically significant differences in favor of the control group on the number of times (during the last month) that treatment teachers used some type of assessment to identify students' learning needs, and selected targeted content and planned teaching according to students' skill levels.

***Amazonía Lee* teachers reported higher rates of support for underperforming students.** In the treatment group, 86 percent of teachers reported working with underperforming students outside of class time, whereas only 57 percent of control teachers did. Furthermore, 31 percent of the treatment group reported referring underperforming students to academic reinforcement,¹ whereas only 11 percent of the control group did. This is consistent with *Amazonía Lee*'s focus on identifying the level at which students are performing and providing personalized support to meet children's learning needs.

3. Impacts on students' reading skills

We looked at impacts on reading comprehension and other important building blocks such as oral language comprehension, reading familiar words, decoding, and reading fluency.

***Amazonía Lee* students had better reading outcomes than students in the control group.** We found positive impacts on four of the five skills measured. Furthermore, the magnitude of the impacts was substantial (0.15 to 0.27 standard deviations) and similar to what has been found by other rigorous studies (USAID, 2016).

- *Amazonía Lee* had a positive impact on **reading familiar words**. Students in the treatment group correctly read on average 7 out 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 standard deviations.

¹ Referring underperforming students includes a broad set of options—for example, the teacher could refer them to additional help in school if it exists, they could also notify the parents that the students is underperforming, they could notify the principal, and so on.

- *Amazonía Lee* had a positive impact on **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 more words that is statistically significant. This is equivalent to 0.24 standard deviations.
- *Amazonía Lee* had a marginal positive impact on **fluency**. Students in the treatment group read on average 34 words correctly in one minute, whereas students in the control group read 30 words correctly, this difference of 4 words per minute is statistically significant at the 10 percent level. The difference is equivalent to 0.15 standard deviations. It is worth noting, however, that both groups have a fluency level that is still below the standard of 60 words per minute for 2nd graders suggested for Peru by the World Bank.
- *Amazonía Lee* had positive impacts on **reading comprehension**. Students in the treatment groups were able to answer on average three of the six comprehension questions correctly; students in the control group were able to answer two questions on average. The difference of one correct answer is statistically significant and equivalent to 0.27 standard deviations.

The results of the gender subgroup analysis are similar to what we found in the full sample. We conducted subgroup analysis to assess whether the impacts for boys were different than the impacts for girls and we found no statistically significant differences. Results for girls are similar to the results for boys with the exception of fluency where we find positive impacts for boys but not for girls.

E. San Martín impact findings

Unlike control schools in Ucayali, in San Martín, most schools in the control group were offered *Soporte Pedagógico*, resulting in a weaker treatment-control contrast. This is reflected in the implementation findings.

1. Impacts on use of services

Although we found no differences in participation in professional development between the groups, we found differences in the training provider and topics covered. Most teachers in both groups reported participating in reading instruction professional development activities. But teachers in the treatment group were more likely to report receiving training from *Amazonía Lee* and teachers in the control group from *Soporte Pedagógico*, *Programa Estratégico Logros de Aprendizaje* (PELA) and other institutions. Unlike in Ucayali, in San Martín the evaluation is comparing two groups that received professional development focused on reading instruction; rather than whether teachers received professional development, the groups differ on the topics covered by the training and who provided it.

Both groups report similar in-class coaching activities with the exception of demonstration sessions. In-class teacher coaching is an important feature of *Amazonía Lee*'s capacity building component. We asked teachers if they had received coaching, who provided it, and what topics were covered during the coaching sessions. Almost 96 percent of the teachers in both groups reported receiving in-class coaching. However, treatment teachers were more likely to have received coaching by *Amazonía Lee* and control teachers from *Soporte Pedagógico* and PELA. Furthermore, the teacher coaching used a similar approach in both groups, with the exception of demonstration sessions that were more likely to occur in the treatment group.

***Amazonía Lee* teachers were more likely to report having a classroom library. However, there were no differences in the range of reading materials available in the observed classrooms.** One goal of *Amazonía Lee* was to improve the availability, variety and relevance of materials to promote reading. We used the teacher survey to inquire about the availability of classroom libraries and materials. Sixty-two percent of teachers in the treatment group reported having a classroom library, and only 2 percent in the control group reported having one. However, we did not find differences across the study groups when we asked whether a classroom library was actually seen during the classroom observation visit. We found that groups did not differ substantially in the type or amount of reading materials available in classrooms. The only difference was that more classrooms in the treatment group had 11 or more printed books than in the control group.

2. Impacts on teacher practices and use of time

Teachers in the *Amazonía Lee* treatment group used literacy instructional practices similar to those used in the control group in San Martín. However, the few impacts found are consistent with the treatment group implementing strategies promoted by *Amazonía Lee*'s balanced approach. In our observation of 15 activities related to teaching reading, we found only two significant differences. Treatment teachers were more likely to be observed providing instruction on grammar, punctuation, or spelling and explaining the use of different types of text than control group teachers. Teachers in both groups reported performing most activities related to *emergent literacy and fluency* at similar rates and frequencies, with two exceptions. Treatment group teachers were more likely than the control teachers to report practicing graphomotor skills and teaching punctuation every day. Similarly, teachers in both groups reported performing 10 activities related to *reading comprehension* at similar rates and frequencies, with one exception. The treatment group was more likely to report developing students' vocabulary every day than the control group. In sum, the treatment group was more likely to report using some practices consistent with the balanced approach of *Amazonía Lee* such as practicing graphomotor skills; teaching punctuation, grammar and spelling; explaining the use of different texts; and developing vocabulary.

***Amazonía Lee* led to positive impacts on the reported duration of the class and a different use of class time that is consistent with *Amazonía Lee*'s balanced approach.** We asked teachers to report the typical duration of the Reading and Communication class. Treatment teachers reported on average a three-hour class and control teachers a two-and-a-half hour class, a statistically significant difference of almost a half of an hour. We estimated the proportion of time teachers spent in seven categories of activities during the class observation. Treatment teachers spent a greater proportion of class time on oral language (3 percentage points) and written text activities (6 percentage points), whereas control teachers spent a greater proportion of class time on reading comprehension activities (6 percentage points). This is consistent with treatment teachers implementing *Amazonía Lee*'s balanced approach, which emphasizes oral language development and written skills.

Teachers in both groups used similar differentiated instruction strategies, with one exception. More than 90 percent of the teachers reported using differentiated instruction for students underperforming in reading. The only difference is that teachers in the treatment group were more likely to report planning teaching for the different levels of student reading learning in the last month than the control group.

Although teachers in both groups reported working with underperforming students at similar rates, control teachers were more likely to refer these students for additional support. On average, 75 percent of teachers in both groups reported working with underperforming students for almost four hours per week. Although only 38 percent of the treatment group teachers reported referring students underperforming in reading to academic reinforcement, 59 percent of the control group did—a statistically significant difference of 21 percentage points. This is consistent with the fact that *Soporte Pedagógico* is implemented in tandem with the program *Docentes Fortaleza*, which provides academic reinforcement sessions to underperforming students after school hours.

3. Impacts on students' reading skills

Students in the treatment group had similar reading outcomes to students in the control group. We found no differences between groups in oral comprehension, words read correctly, fluency, decoding, and reading comprehension scores. Students in both groups correctly read on average 8 out of 10 familiar words presented. Students in both groups had similar fluency results: on average students read correctly almost 44 words per minute and 20 invented or pseudo-words per minute. Furthermore, students in both groups were able to answer three of the six questions correctly. We should note, however, that both groups have a fluency level that is still below the standard of 60 words per minute suggested by the World Bank for 2nd graders in Peru.

F. Overall Conclusions, Lessons Learned, and Recommendations

1. Conclusions

When compared to prevailing practice, *Amazonía Lee* achieved substantive effects of magnitudes similar to those found in other rigorous evaluations of early reading programs. In Ucayali, where *Amazonía Lee* was compared to a control group that did not receive additional support beyond what the Ministry of Education typically provides, we found that students in *Amazonía Lee* had better reading outcomes than the control group. The magnitude of the impacts ranged from 0.15 standard deviations to 0.27, which are within the range of what other studies have found.

When compared to a program that also provided teacher reading instruction but different from *Amazonía Lee*, students in both groups had similar reading outcomes. In San Martín, the control group received similar rates of professional development training and in-class coaching as the treatment group. Furthermore, *Soporte Pedagógico* commanded additional resources, such as tutors for underachieving students and provided monthly coaching visits as opposed to bi-monthly visits from *Amazonía Lee*. Despite *Amazonía Lee* teachers' reports of longer reading classes and evidence of teachers implementing the balanced approach, we found no differences in student outcomes.

***Amazonía Lee* services reached the treatment groups in both departments.** Our findings confirm that training and in-class coaching were delivered to the treatment groups in Ucayali and San Martín. Although we found some impacts of the program on the provision of pedagogical materials in Ucayali, we found no impacts in San Martín. Furthermore, we found delays in the delivery of materials.

Teachers in both departments seem to be implementing some aspects of *Amazonía Lee*. In Ucayali, *Amazonía Lee* had positive impacts on the number of teachers who explain the use of different kinds of text, emphasize reading daily for pleasure, and teach the use graphic organizers—practices consistent with the balanced approach of *Amazonía Lee*. Furthermore, *Amazonía Lee* teachers reported spending more time supporting underperforming students and promoting reading in the community—consistent with the community engagement component of the program. In San Martín, *Amazonía Lee* had positive impacts on practicing graphomotor skills; teaching punctuation, grammar, and spelling; explaining the use of different texts, and developing vocabulary—also practices consistent with the balanced approach of *Amazonía Lee*. Furthermore, we found that *Amazonía Lee* teachers reported teaching longer Reading and Communication classes and using class time differently than control teachers.

2. Lessons Learned

Supporting teachers explicitly in best teaching practices in reading is a promising approach to improving students' early grade reading. In the one department where there were no significant differences in reading outcomes at endline, teachers in both groups had been receiving support on reading instruction. The differences in the nature of the support did not translate into impacts. In contrast, in the department where such support was lacking in the control group, the impacts were significant and meaningful. This suggests that in-service training for teachers as well as individualized teacher coaching, are useful for teachers to improve their reading instruction and manage their classrooms effectively.

High teacher turnover hinders the potential reach of teacher training efforts. The commitment of school directors and regional authorities to retaining trained teachers in the early grades is important for the gradual and continuous improvement of teachers' reading instruction over time.

When conducting or sponsoring rigorous evaluations with a counterfactual, the local context and regional conditions matter for the implementation of impact evaluations and type of questions that can be answered with rigor. Although our original evaluation design planned comparisons with control groups that were not exposed to similar interventions, this was only possible in one of the departments.

3. Recommendations

Policymakers in Peru should continue to support *Amazonía Lee* in Ucayali. Based on the positive results found there, and given that the department has gained experience and capacity to implement *Amazonía Lee*, continuing the intervention there may be a way to sustain the progress made on early grade reading.

Policymakers in Peru should continue to support early grade reading in San Martín, but should examine the costs of different approaches before selecting the one to continue. Students in *Amazonía Lee* schools had similar reading performance as those in schools with other programs, primarily *Soprote Pedagógico*. Since *Soprote Pedagógico* has only been implemented in urban schools and *Amazonía Lee* was implemented in urban, peri-urban, and rural schools, continuing the intervention may be a way to sustain achievement in all types of schools.

I. INTRODUCTION

In the past 15 years, countries in Latin America and the Caribbean have made considerable progress in increasing primary school enrollment and completion among the very poor (United Nations Educational, Scientific and Cultural Organization [UNESCO] 2015). However, challenges in improving reading persist. Reading is essential for academic success. Without reading and comprehension skills, one cannot understand text information, write correctly, or study independently (Gove and Cvelich 2011). Recent studies have identified that training and coaching teachers with a focus on reading instruction can be effective in improving reading outcomes for children (American Institutes for Research [AIR] 2017, Friedlander and Goldenberg 2016). However, the Landscape Report on Early Reading Literacy (U.S. Agency for International Development [USAID] 2016) does not report any studies based on a rigorous impact evaluation design implemented in Latin America and the Caribbean. Moreover, validating teacher training models in different contexts will contribute to our understanding on the conditions needed for these program to work.

This report presents results of the impact evaluation of *Amazonía Lee* (Amazon Reads), an early-grade reading intervention funded by USAID. *Amazonía Lee*'s goal is to improve the reading skills of children in two departments in the Amazon region of Peru: Ucayali and San Martín.

Figure I.1. Map of study locations in Peru



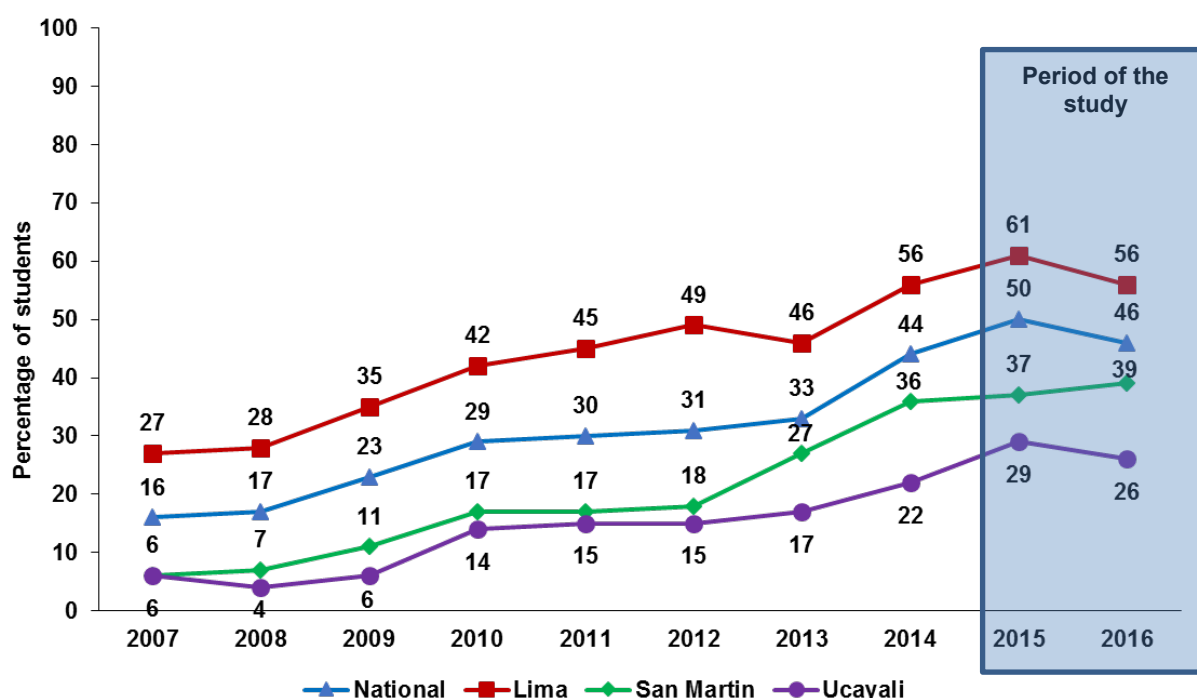
The geography of the two departments is shown in Figure I.1. San Martín lies in the northern part of the Peruvian Amazon forest; Ucayali is located south of San Martín, bordering Brazil. These departments are primarily Spanish speaking, with a small percentage of the population speaking Quechua, Aymara, or other indigenous languages. In terms of area, Ucayali is nearly twice as large as San Martín, but San Martín is more densely populated. San Martín has approximately 1,302 active primary public education institutions, whereas Ucayali has approximately 796.²

Within Peru, the departments of Ucayali and, to a lesser extent, San Martín lag behind the nation and the metropolitan area of Lima on important student reading achievement indicators. According to Peru's national standardized achievement test (*Evaluación Censal de Estudiantes*, ECE), in both Ucayali and San Martín, the percentage of 2nd-grade students achieving the expected grade-level reading comprehension has historically been lower compared with the percentage of students at

² Active public primary schools in 2015. Source: ESCALE—*Estadísticas de la Calidad Educativa*. Ministerio de Educación Perú. Available at <http://escale.minedu.gob.pe/web/inicio/padron-de-ieee>.

the national level and Lima (Figure I.2). Ucayali has maintained the same pattern as the capital and the nation, although with a persistent gap, whereas San Martín has been catching up to Lima since 2012, a trend that slowed in 2013 and 2014. In 2016, 56 percent of 2nd-grade students in Lima achieved the expected grade level for reading comprehension compared to 39 percent of students in San Martín and 26 percent of students in Ucayali in 2016.

Figure I.2. Percentage of 2nd-grade students who achieved the expected (satisfactory) grade-level reading comprehension score on the ECE test



Source: Reading comprehension results from the *Evaluación Censal de Estudiantes*, ECE. Ministry of Education. Office measuring the quality of learning.

Note: 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. The *Amazonía Lee* intervention

To address these issues USAID funded the *Amazonía Lee* intervention. We start below with the description of the intervention and then turn to the evaluation questions and design.

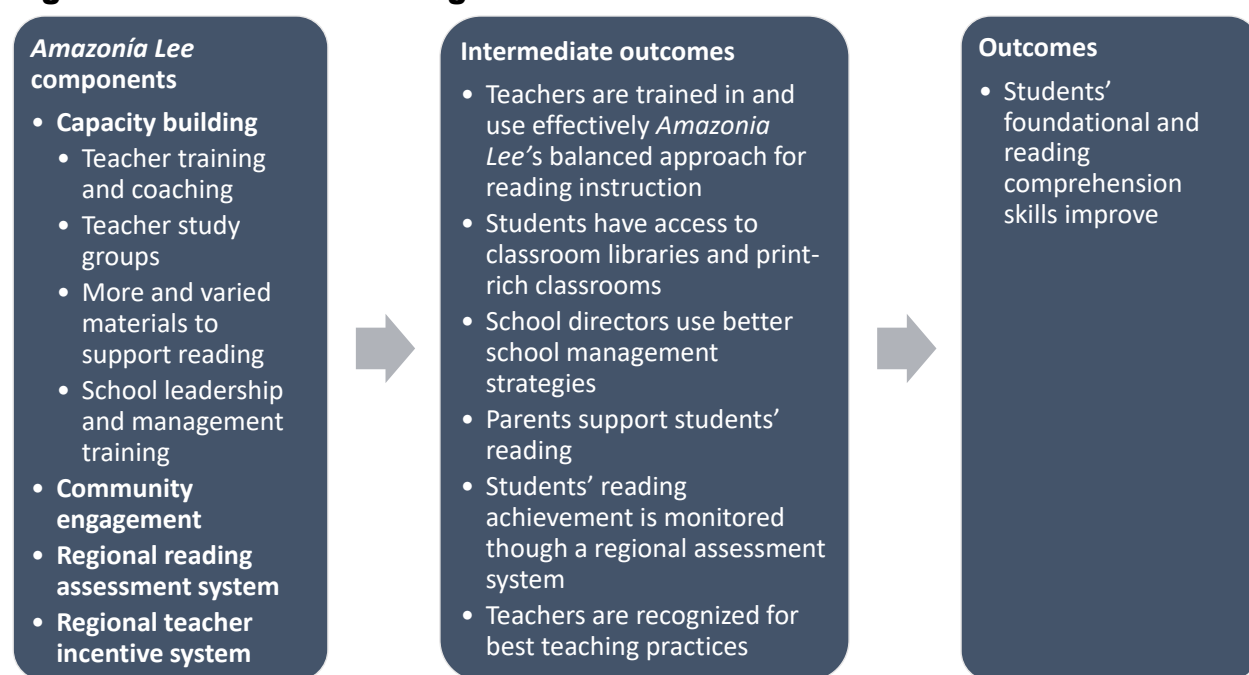
The *Amazonía Lee* intervention aims to address the learning gaps of San Martín and Ucayali by improving the quality of reading instruction in these departments.³ *Amazonía Lee* was implemented with funding from USAID by the Department of Regional Education Offices (DREs) of Ucayali and San Martín with technical assistance of the Universidad Peruana Cayetano Heredia (UPCH), beginning in 2015. In Ucayali, the program is known as “*Liderando*

³ In this report, we also refer to these departments as regions to be consistent with other related usage—for example, the Regional Educational Offices, which pertain to each department. We should note that *region* is also used for broader regions such as the Latin America and the Caribbean region.

los Aprendizajes de Niñas y Niños” (LANN), which means Leading the Learning of Girls and Boys. In San Martín, the program is known as “*Programa Enseñar Es Liderar*” (PEEL), which means To Teach Is to Lead.

The two *Amazonía Lee* programs share the main components: capacity building of teachers, school directors, and specialists and other components such as a regional assessment system to monitor students’ achievement, a community engagement strategy to support student learning, and a system of incentives to recognize and motivate teachers’ performance. In Figure I.3 we provide a high-level logic model for the program, summarizing the main components of the intervention and the expected intermediate outcomes through which the intended changes in foundational reading skills are expected to be achieved.

Figure I.3. *Amazonía Lee* logic model



Source: Authors' high level logic model depiction of *Amazonía Lee* program

Capacity building. The aim of this component is to strengthen pedagogical skills and instructional capabilities of teachers, school directors, and regional specialists. At the beginning of 2015, UPOCH began providing technical assistance to the DREs both regions for the implementation of the following five capacity-building mechanisms (subcomponents):

1. **Teacher training workshops based on *Amazonía Lee*'s balanced approach.** *Amazonía Lee*'s balanced approach to reading instruction integrates evidence-based foundational reading skills instruction with the Ministry of Education's existing *communicative-textual* approach. Training workshops for 1st- to 3rd-grade teachers aim to improve teachers' approach to reading instruction, class and curricular planning, differentiated instruction based on students' learning needs, and personalized attention for underperforming students. The training is structured in five modules: (1) print environment; (2) understanding the alphabetic principle (alphabet knowledge, phonological awareness, and fluency); (3) reading comprehension; (4) writing and production of texts; and (5) leadership and interpersonal

relations. Also, teachers are trained on the design and use of formative assessments to monitor students' learning, on positive classroom climate strategies, and on community engagement strategies in support of students' literacy. The full set of workshops offer 120 training hours, delivered in three blocks of 40-hour blocks during the school year.

2. **Teacher coaching.** Specialists conduct classroom visits to coach teachers on site and help them improve content area knowledge and pedagogical strategies for literacy instruction, differentiated instruction, assessment and monitoring of student learning, and effective classroom management. A primary way in which the specialists build teacher capacity is through individual demonstration and observation sessions. In demonstration sessions, the coach, known as the *specialist*, serves as a role model for effective classroom practice, and demonstrates the teaching strategies introduced during the training workshops in the teacher's actual classroom. Demonstrations are typically followed by a discussion of what the teacher learned while observing the coach deliver the lesson. Observation sessions are also part of the teacher coaching strategies and complement demonstration sessions. In observation sessions, the coach observes the teacher deliver a live class and offers feedback on areas for improvement. Teachers receive about six to eight individual coaching visits per year, but the number of visits is adjusted based on teachers' learning needs.
3. **Teacher study groups.** Teachers gather in small groups of 8 to 16 participants to strengthen their learning, review pedagogical materials, share teaching strategies, and plan for classes with support from peer teachers. Study groups do not have a defined set of contents; rather, they are meant to be flexible spaces for collective feedback. Trainers set the agenda for the study groups in collaboration with teachers. The content of the study groups is intended to respond to teachers' needs and interests, and to the skill-building needs trainers identify during the coaching visits. The goal is for the teacher study groups to become autonomous over time, and the groups are intended to promote teachers' pursuit of continuous professional development in collaboration with their peers. Teacher study groups are expected to meet once a month for approximately four hours. Teacher groups receive about six coaching visits a year, but the frequency of visits may vary depending on teachers' needs.
4. **More and more varied pedagogical materials to support reading.** The goal of this subcomponent is to improve quality and regional pertinence of pedagogical materials for reading and writing instruction, and to increase the availability of classroom book libraries in primary grades. To this end, *Amazonia Lee* classroom materials and teacher guides integrated evidence-based foundational reading skills instruction into the national pedagogical communicative approach of the Ministry of Education in Peru. The materials also emphasized culturally relevant stories and activities for students in the Amazon region. *Amazonia 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 were provided during the teacher training workshops and offered a conceptual basis on 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 activity guides and student workbooks included learning activities based on *Amazonia 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.

5. **School leadership and management training.** School directors receive training on pedagogical leadership and school management, and support from specialists on issues related to pedagogical matters and institutional management. Regional governments receive technical assistance to improve school planning and budgeting and to implement information systems on schools management indicators.

In addition to the five capacity-building mechanisms described above, *Amazonía Lee* included another three components that were originally envisioned to be implemented at the regional level in both Ucayali and San Martín. Therefore, at the evaluation planning stage, our assumption was that these regional components would be implemented in all schools (treatment, control, and non study schools). However, roll out of these components was complex, and they were implemented only in schools that were also receiving capacity building (treatment schools). Furthermore, some of these components were not fully implemented at the time of the evaluation. The description of these three components follows.

Community engagement. The community engagement component aims to promote the active engagement of the educational community in support of student learning, particularly in gaining reading proficiency during the first three grades of primary education. This component focused on engaging parents and the community at large in promoting reading opportunities for children in and outside the school context through a variety of activities such as book fairs, family oral storytelling, shared book reading, and community workshops. The implementation of this component mostly occurred through parent engagement in the *Amazonía Lee* schools.

Regional reading assessment system. The goals of this component were (1) to introduce regional assessments at the beginning and end of the school year to monitor student achievement (on reading comprehension and production of texts); and (2) to develop local capacity in the design, implementation, scoring, data production and interpretation of these assessments. The DREs of Ucayali and San Martín planned to implement the assessments and lead the dissemination of assessment results with stakeholders. It should be noted that although this component is intended to be implemented at the regional level, for the study period it was only implemented in the schools participating in *Amazonía Lee*.

Regional incentive system for teachers. This component includes strategies to motivate, encourage, recognize, and reward teacher performance—thereby engaging all stakeholders of the education system (teachers, specialists or facilitators, school directors, students, parents, and others) in efforts to improve student learning and academic achievement. Since 2013, the Ministry of Education in Peru organizes a national contest to recognize excellence in teaching. Based on the framework of the national contest, UPCH designed a regional system of nonmonetary incentives. During the evaluation period, the incentive system was implemented only in schools participating in *Amazonía Lee*.

B. Research questions, evaluation design and limitations

USAID/Washington and USAID/Peru in collaboration with the Ministry of Education in Peru and with the DREs of Ucayali and San Martín, sought to contribute to the evidence base on

what works to improve early-grade reading skills by funding and evaluating a program designed and implemented by the regional education authorities in the Amazon region of Peru, with support from a national academic institution. The evaluation was structured to address the following central research question:

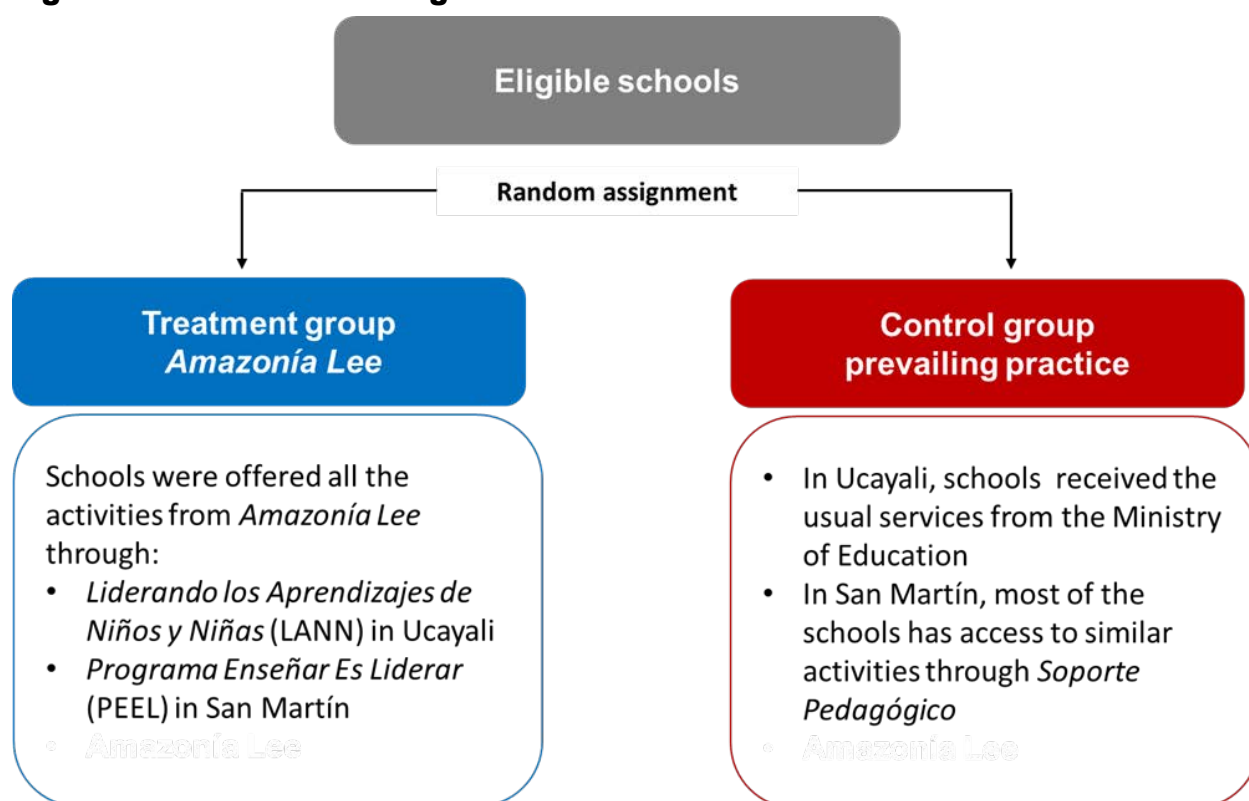
- What was the impact of *Amazonía Lee* programs on early-grade reading and other intermediate outcomes relative to other prevailing practices and programs?

In addition, to further advance our understanding of the implementation of the program in the context of the Amazon region of Peru and to help interpret the impacts, the evaluation sought to address a set of additional questions related to implementation:

- How were *Amazonía Lee* intervention components, activities, and processes implemented in Ucayali and San Martín?
- What services were provided to control schools in Ucayali and San Martín?
- How were *Amazonía Lee* components, activities, and impacts perceived by principals and teachers?
- What were the facilitators and barriers to the implementation of *Amazonía Lee* components?

To answer these questions, we conducted a rigorous impact evaluation, as well as an implementation study.

Impact evaluation. Mathematica partnered with the *Grupo de Análisis para el Desarrollo* (GRADE) as our local research and data collector partner to conduct a randomized control trial to rigorously estimate program 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 this program (Figure I.4). The randomized evaluation relies on comparing outcomes between these two groups of schools in each region to estimate the impacts of *Amazonía Lee* activities on the outcomes of interest. Because the Ministry of Education was providing different services to the schools assigned to the control group in each of the two regions at the time of the evaluation, we conducted separate analyses for each region..

Figure I.4. Evaluation design: randomized controlled trial

In Ucayali, the evaluation sample includes 70 schools, 35 of which were randomly assigned to the treatment group and 35 of which were randomly assigned to the control group in September 2014. In San Martín, the evaluation sample includes 200 schools, 100 of which were randomly assigned to the treatment group and 100 of which were randomly assigned to the control group in February 2014.⁴

To define the meaning of “prevailing practices and programs” in each region, in this report we describe the types of supports received by students and school staff in the two control groups. In Ucayali, control schools received the usual services the Ministry of Education provides in the department. In San Martín, a majority of the control schools received additional assistance in the form of a primary education support program known as *Soporte Pedagógico* (Pedagogical Support), a Ministry of Education initiative implemented in selected departments. San Martín was one of the departments that started implementing this new initiative in 2015 at the same time as *Amazonía Lee* was planned to start. In order to preserve the random assignment of schools for the evaluation of *Amazonía Lee* in San Martín and to take advantage of comparing the impacts of *Amazonía Lee* to those of the Ministry’s *Soporte Pedagógico* program, *Soporte Pedagógico* was offered to control schools only and thus became incorporated into the study’s research question

⁴ Key characteristics were geographic distribution within the department, school size, school structure (separate or combined grades), bilingual population, implementation of the Strategic Learning Achievements Program (PELA), and ECE reading scores. We created pairs of schools with similar characteristics and randomly assigned one school of the pair to treatment and the other to control. We provided full details on random assignment in our evaluation plan (Murray et al. 2016).

as the alternative to *Amazonía Lee*. *Soporte Pedagógico* aims to improve student learning in the areas of mathematics, communication, science, and social skills in urban elementary schools by strengthening skills of teachers and administrators, providing after-school tutoring for students not meeting the expected learning standards, delivering educational materials and resources, and improving the school and local management capacity. Thus, for San Martín, we are comparing outcomes from schools receiving the *Amazonía Lee* intervention to schools receiving *Soporte Pedagógico*. *Amazonía Lee* and *Soporte Pedagógico* differ in three main ways. First, *Amazonía Lee* focuses on Reading and Communication, whereas *Soporte Pedagógico* focuses on several curricular areas, including Reading and Communication. Second, the approach to reading instruction of *Amazonía Lee*—a *balanced* approach—integrates instruction on foundational reading skills with the national *communicative-textual* approach, whereas *Soporte Pedagógico* primarily uses the national *communicative-textual* approach. Third, *Soporte Pedagógico* provides small-group tutoring to underperforming students after school with teachers specialized in academic reinforcement.

Impact estimation. *Amazonía Lee* is designed to be implemented beginning in the 1st grade, when students begin acquiring foundational literacy skills, and to continue through the 3rd grade when students should have acquired basic reading skills. While the evaluation was originally designed to follow a cohort of students that were in 1st grade in 2014 for three years, implementation delays required us to follow the 2015 cohort of 1st graders instead. Subsequently, when USAID/Peru phased out support to basic education in Peru by 2018, it was decided that we'd measure impacts of the program at the end of 2016, when the students should have been in second grade, so that we could share the results of the evaluation in 2017. Hence, a limitation of this study is that we are not measuring the effects of the full intended dosage of the program. We were only able to measure the effects of exposure to the program during two school years for approximately 13-14 months (May or June 2015 to September or October 2016), rather than the three school years originally envisioned.

Student impacts. In each region, we estimated the impacts of *Amazonía Lee* activities on student reading outcomes at the end of 2nd grade (in 2016) by comparing the average outcomes of students in the treatment schools to those in the control schools. The impact estimates account for students' baseline performance on an emergent literacy assessment administered when students were in 1st grade (in 2015), before any treatment effects could reasonably have affected students' performance. We also controlled for student background characteristics such as gender, age, whether they were over age for grade, and whether they were repeating 1st grade. We also accounted for school characteristics, such as the schools' physical infrastructure and the schools' average emergent literacy scores, which we measured in 2014 for an earlier cohort of 1st-grade students. The models also account for the stratification used in random assignment. Details on the impact estimation models are presented in Appendix A.

Teacher impacts. Because the evaluation was only able to follow students from first to second grade, at follow-up we focus on second grade teachers. Impacts on teacher outcomes were estimated by comparing 2nd-grade teachers in the treatment and control groups. The teacher sample is a cross sectional survey of 2nd-grade teachers in the study schools at endline, so the teachers present at endline are not necessarily those who were in the school at baseline. The regression model for the analysis of impacts at the teacher level accounts for the

stratification used in the random assignment, but it does not control for any baseline characteristics. Details on the impact estimation models are presented in Appendix A.

This report focuses on results significant at the 5 percent level. We also mention results that are statistically significant at the 10 percent level, and the tables present all outcomes regardless of statistical significance. An important issue to keep in mind is that in some cases, especially for the teacher outcomes, we conducted many tests, corresponding to a list of related outcomes. When many tests are conducted there is a higher chance of finding statistically significant impacts even when there are none (false positives). While we do not do statistical adjustments for multiple comparisons, we provide information on the number of tests conducted to provide the reader the appropriate context.

2. Implementation study

We adopted a mixed-methods approach to understand the implementation of *Amazonía Lee* in Ucayali and San Martín, integrating qualitative and survey data and a desk review of program materials. In particular, qualitative implementation data helped us understand how the different components of the *Amazonía Lee* program were implemented in each region, how services were received by local stakeholders, principals, and teachers; teachers' perceptions of *Amazonía Lee* program outcomes; and the challenges the implementation teams in Ucayali and San Martín encountered during the course of the *Amazonía Lee* implementation. We also gathered qualitative information on control schools to understand their usual approach to reading instruction and the services they typically receive from the Ministry of Education.

We recruited a purposive sample of program implementers, technical assistance providers, and program beneficiaries for the qualitative data collection effort in collaboration with GRADE, the regional education authorities of Ucayali and San Martín, and UPCH. We used thematic analysis to summarize and interpret the results of focus groups, key informant interviews, and program document reviews.⁵

C. Data sources and indicators

Following the logic model presented in Figure I.3, our goal was to collect data that would allow us to assess impacts on intermediate outcomes such as teacher practices and availability of materials in the classrooms, as well as on final student reading outcomes. Here we describe, in the first section, the sources of data used for constructing these indicators and the timing of data collection. In the second section, we describe the data collected for the implementation study.

1. Data collection instruments and procedures

The main data sources for the impact evaluation are student reading assessments, classroom observations, a teacher survey, and a school infrastructure survey. These instruments yielded rich information on indicators such as student reading performance, teachers' characteristics and instructional practices, time on literacy tasks, classroom literacy environment, and school characteristics. Table I.1 presents a summary of the indicators, outcomes, and data sources. Below we describe each of the instruments and the outcome indicators constructed with each of

⁵ We used inductive content analysis to identify patterns in the narrative data gathered through interviews and focus groups, gain insight, and synthesize findings.

them. Additional findings on context variables and other outcomes can be found in the appendices.

Table I.1. Summary of data sources and outcome indicators

Outcome indicators	Data sources		
	Reading assessment	Teacher survey	Classroom observation
Student reading outcomes			
Decoding accuracy (correct pseudo-words per minute)	X		
Reading fluency (correct words per minute)	X		
Familiar word reading (familiar words read correctly)	X		
Reading comprehension (number of correct answers)	X		
Main teacher outcomes			
Participation in teacher training (amount and duration)		X	
Topics learned in teacher training		X	
Observed time spent on literacy-related activities (time on literacy tasks)			X
Observed use of foundational reading instruction practices			X
<i>Sample items:</i>			
Uses reading fluency exercises			
Discusses main characters			
Explains word meanings			
Other outcomes			
Classroom organization			X
General pedagogical practices			X
<i>Sample items:</i>			
Gives feedback			
Praises students			
Treats students with respect			
Literacy environment in the classroom			X
<i>Sample items:</i>			
Reading corner available			
Number of reading books accessible to students			
Alphabet visible to students			
Labeled objects			
Teaching practices and activities for reading instruction		X	X
<i>Sample items:</i>			
Dictation			
Fill in the blank			
Silent reading			
Demonstration reading			
Proportion of students that met foundational reading milestones		X	
Use of formative assessments		X	
Use of differentiated instruction		X	
Promotes positive classroom climate		X	X

Individual reading assessment. The individual assessment was collected on a longitudinal sample of students at two points in time. At baseline in 2015, the student sample was selected as follows. In each study school, we randomly selected one 1st grade classroom. Based on the

classroom roster for the selected classroom, we randomly selected, on average, 8 first graders in San Martín and 10 in Ucayali (evenly divided by gender when possible). Students who were identified as having a physical or severe cognitive disability⁶ or as no longer attending the school were considered ineligible for the study and were therefore not included in the sample. Students who were selected for the study sample (that is, students who were enrolled in the school and did not have a severe disability) but were absent on the days enumerators visited the school to conduct the literacy assessments were still included in the study sample. These absent students were considered non-respondents at baseline but were kept in the study sample to be followed at endline. At endline, we tracked the study students from the 2015 1st-grade cohort to evaluate their reading skills at the end of the second year of the intervention, when they were expected to be in 2nd grade. We attempted to locate and assess all sample students eligible at baseline as long as they were still residing in Ucayali and San Martín, including those that were repeating 1st grade, had transferred to other schools, were no longer attending school, or had not responded the baseline reading assessment. The analysis sample includes all the students who had follow-up assessments; the baseline assessment was imputed for the students who did not take it (see Appendix A for details on estimation models and imputations).

In collaboration with GRADE, our data collection partner in Peru, we design and piloted an endline Early Grade Reading Assessment (EGRA) consisting of five tasks: oral comprehension, familiar word reading, decoding, reading fluency, and reading comprehension. The reading assessment tasks are described in more detail in Table 1.2 below, and the assessment design framework and pilot test are presented in Appendix E. Reading assessments were administered individually by a trained enumerator. Most students were assessed at school during regular school hours either in their baseline school or in another school if they had transferred. However, some students that were absent from study schools during the visits for data collection or had dropped out but were located were assessed in their homes (See Appendix Table A.4). The final student reading outcomes were as follows: decoding accuracy (correct pseudo-words per minute), reading fluency (correct words per minute), familiar word reading (number of familiar words read correctly), and reading comprehension (number of correct answers).

Table 1.2. Description of the endline reading assessment instrument

Task	Description	Score
Oral comprehension	Assessors read children a short story on a familiar topic and then ask children six comprehension questions about the story they heard.	0–6
Familiar word reading	Children read a list of words they likely had seen before and were expected to read at their grade level. Stop rule: The task was discontinued if the child failed to read the first five words correctly.	0–10
Decoding	Children read out loud as many pseudo-words as they could in 60 seconds. Pseudo-words are vowel-consonant combinations that follow Spanish phonological and spelling rules but are not actual meaningful words. Stop rule: The task was discontinued if the child failed to read the first five pseudo-words.	0–50

⁶ Down syndrome and autism were considered severe cognitive disabilities, and blindness, deafness, and muteness were considered physical disabilities.

Task	Description	Score
Reading fluency	Children were asked to read a short story out loud, quickly but carefully. The story had 77 words (six sentences). To obtain a measure of children's ability to read text quickly and accurately, this task had a 60-second time limit. Stop rule: The task was discontinued after 60 seconds. To reduce burden on lower performing children, those that were unable to read the first line of this story (12 words) did not take subsequent reading comprehension task.	Number of correct words per minute
Reading comprehension	Children read a short story out loud, then were asked six comprehension questions (literal and inferential) that could be answered only by having read the passage. Lookbacks were not allowed.	0–6

Source: Adapted from RTI International. 2015. Early Grade Reading Assessment (EGRA) Toolkit, Second Edition. Washington, DC: United States Agency for International Development.

Classroom observation. We developed a classroom observation instrument to be fielded in the Reading and Communication class.⁷ The instrument focused on measuring teachers' use of class time on different types of literacy-related activities, teachers' literacy instruction (for example, book reading, fluency, vocabulary development, reading comprehension), teachers use of general pedagogical practices (for example, engage students in class activities, check for understanding, give feedback, and promote positive classroom climate). In addition, the observation protocol measured the quality of the classroom literacy environment (for example, availability of books and print materials, variety and overall quality of books and materials, exposure to print in classroom walls with labeled objects, familiar words, student's names, bulletin boards) and availability of resources for writing. The classroom observation was collected at two different points in time. At baseline, a random sample of 1st-grade classrooms (one per school) was observed; similarly, a random sample of 2nd-grade classrooms at endline was observed. The main teacher outcomes measured at endline through direct observation of a Reading and Communication class were as follows: time spent on literacy-related activities (time on literacy tasks), use of foundational reading instruction practices, and availability of a print-rich classroom environment.

Teacher survey. We also administered a teacher survey to all teachers whose classrooms were observed. At baseline, we have a random sample of one 1st-grade teacher per school. Similarly, at endline, we have a random sample of one 2nd-grade teacher per school. Both teacher surveys gathered self-reported use of instructional practices, proportion of students in the classroom that achieved milestones in reading acquisition, use of formative assessments, participation in professional development or in-service training, education and experience, occupational needs, career expectations, and background characteristics. The main teacher outcomes measured in the teacher survey at endline were as follows: participation in teacher training (amount and duration) and topics learned during teacher training.

School infrastructure checklist. The school infrastructure 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

⁷ In Peru, they refer to this class as Language and Communication.

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.

2. Timing of data collection

Table I.3 shows the timing of data collection activities during the study period. At endline in 2016 we collected student reading assessments, classroom observations, a teacher survey, and a school infrastructure survey. All instruments except the classroom observation and school infrastructure survey were also applied at baseline in 2015 for the target cohort of the impact evaluation. Data were also collected in 2014 using these same instruments for an earlier cohort of students who were attending 1st grade in (2014 cohort), as we originally expected the 2014 cohort to receive the program. We were able to use these data to control for baseline school-level characteristics in our impact estimations.

Table I.3. Timing of data collection

Year	2014		2015		2016	
Semester	1st	2nd	1st	2nd	1st	2nd
			Baseline (1st grade)	Baseline (1st grade)	Endline (2nd grade)	Endline (2nd grade)
Individual reading assessment	X		X			X
Classroom observation	X					X
Teacher survey	X		X			X
School infrastructure survey	X					X
Implementation data collection				X		X

Note: For reasons explained earlier, endline data collection originally was planned for the end of 2017 (when most of the 2015 cohort of 1st-grade students would have been in 3rd grade). Given the timing of the end of USAID support to basic education in Peru and the desire to share information from the study as early in 2017 as possible, USAID/Washington, USAID/Peru, and Mathematica subsequently agreed to focus the impact evaluation on 2nd grade outcomes. Therefore, endline data collection took place at the end of the 2016 school year.

3. Qualitative data collection on program implementation

We gathered qualitative information on the implementation of *Amazonia Lee* in the first and second years of implementation. In the first year, we interviewed the director of UPCH's technical assistance team and conducted a group interview with the LANN-DRE project coordinator for Ucayali and the PEEL-DRE project coordinator for San Martín. At the end of the second year of implementation (November 2016), we visited Ucayali and San Martín to collect qualitative information from UPCH's key staff, *Amazonia Lee* trainers, DRE project coordinators, school directors, and teachers. Table B.1. in Appendix B summarizes the range of qualitative data collected in 2016, at the end of the second year of the study, including respondents, method of data collection, and topics addressed in the qualitative study. Focus groups and interviews were digitally recorded and transcribed verbatim.

In addition to collecting primary qualitative data, we reviewed program documents including UPCH's quarterly reports, DRE's project work plans, DRE's implementation reports, UPCH's design reports of intervention components, and UPCH's teacher training materials to understand how the program was conceptualized and rolled out in each department.

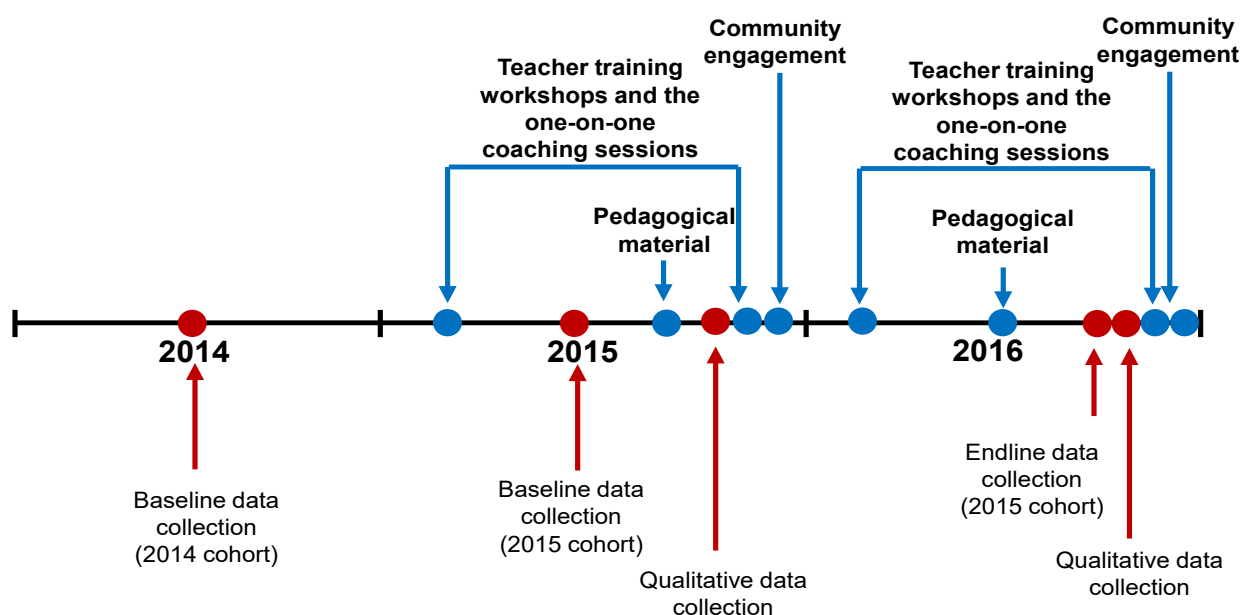
This page has been left blank for double-sided copying.

II. IMPLEMENTATION OF *AMAZONÍA LEE*

Ucayali and San Martín are both located in the Amazon region; however, they differ in their geography, students' historical performance on national reading assessments, and in their DREs' management capacity. We gathered interview and focus group data from program stakeholders to better understand the context in which the impact evaluation took place, and to identify potential variations in the implementation of *Amazonía Lee* across regions. In this chapter, we present findings from our qualitative data. First, we discuss what we learned about the way program components were rolled out and the challenges experienced in implementation of program activities in Ucayali and San Martín, emphasizing conditions that delayed or hindered the implementation of program components. Second, we describe how *Amazonía Lee* components and activities were received by principals and teachers, and their perceptions about the program. Finally, we describe what we learned about the prevailing practice in Ucayali and the services that control schools received through *Soporte Pedagógico* in San Martín from our document review and from focus groups with teachers that did not receive *Amazonía Lee* services.

A. Rollout of program components in Ucayali and San Martín

The implementation of *Amazonía Lee* was led by the regional technical coordinators in the DREs of Ucayali and San Martín, with technical assistance from UPCH. (The structure of the regional implementation teams and their organizational charts are available in Appendix B.) Core program activities started in both regions with the rollout of the capacity building component, which was implemented during 2015 and 2016. The training workshops and coaching subcomponents of capacity building were successfully implemented, however the pedagogical materials subcomponent was delayed. In fact, classroom libraries were not delivered to any of the *Amazonía Lee* schools in either region until November of 2016, the second year of the intervention. In the first year of implementation, teacher guides and student reading workbooks developed specifically by the project for the Amazonian region arrived during the last quarter of the 2015 school year. In the second year of implementation schools did not receive these materials until July of 2016. Other program components—the regional assessments, the family and community engagement, and the teacher incentives—were not fully implemented during the study period (see Figure II.1 for the implementation timeline). For example, while the reading assessments were administered successfully in Ucayali, and reports were provided to teachers and used by *Amazonía Lee* staff, in San Martín the assessments were administered, but the results were not fed back into training or teaching activities. In the case of the family and community activities, Ucayali implemented parent workshops and learning fair activities as planned each year. San Martín only implemented parent workshops in 2016, but conducted learning fairs each year.

Figure II.1. Implementation timeline, 2015 and 2016

Source: Author development based on qualitative data collected for the study.

Note: The school in Peru starts in March and ends in December.

Capacity building

Amazonía Lee trainers conducted the teacher training workshops and coaching visits, but they faced some challenges consolidating the teacher trainer cadres. UPCH lead trainers trained and coached *Amazonía Lee* trainers on the methodology of conducting teacher workshops, coaching sessions, and study group meetings.⁸ Subsequently, *Amazonía Lee* trainers trained teachers in treatment schools. The first training workshop for trainers took place during the first quarter of 2015. Training workshops for teachers took place the month after the LANN (Ucayali) and PEEL (San Martín) trainer workshops. In-class coaching activities started soon after the 2015 school year started in both regions, but the demonstration sessions which were part of the teacher coaching started somewhat later the first year.⁹ In 2016, teacher coaching activities started about one month after the beginning of classes in both departments. However, the implementation team faced some challenges with the train-the-trainer approach, which required adjustments for the second year of implementation.

First, the number of UPCH lead trainers in Ucayali and San Martín was not sufficient to support the cadre of local *Amazonía Lee* teacher trainers with the expected frequency and level of supervision needed. To address that challenge, UPCH hired more lead trainers for the second year that could support training and coaching activities in all treatment schools. Second, specific educational and work experience requirements for the recruitment and selection of *Amazonía Lee* teacher trainers were not well defined or enforced the first year. Consequently, the first cadre of teacher trainers was not adequately prepared to train and coach teachers with the level of

⁸ UPCH lead trainers are UPCH staff based in the regions.

⁹ Coaching sessions started in May in Ucayali and in October in San Martín. Not all San Martín (PEEL) trainers conducted the expected number of coaching sessions with teaching demonstrations.

proficiency that UPCH expected. UPCH staff had to work closely with the less skilled Ucayali (LANN) trainers during the implementation of training workshops and coaching visits. To improve the recruitment and selection of teacher trainers for the second year of implementation, UPCH in collaboration with the program coordinators at the DREs in both departments developed a profile of educational background and experience required to join the team of *Amazonía Lee* trainers. The stricter recruitment and selection requirements resulted in a more qualified cadre of teacher trainers for the second year of implementation and decreased the burden on UPCH lead trainers. Teacher trainers were then able to implement training workshops and coaching sessions on their own with less on-site support from UPCH staff. Furthermore, due to the number of schools and program staff available for each region, the coaching component was implemented with less intensity in San Martín. Stakeholders reported monthly coaching visits on average for teachers in Ucayali and bi-monthly visits in San Martín.

An insufficient number of qualified teachers that meet the criteria to become *Amazonía Lee* trainers was a barrier to implementation, as was the high turnover of regional technical team members throughout the project. This situation was due in part to weaknesses in teachers' pre-service training, and the scarcity of highly qualified teachers created competition between *Soporte Pedagógico* and *Amazonía Lee* in San Martín. Initially, *Amazonía Lee* was at a disadvantage because *Soporte*'s teacher trainers were paid more; therefore, more-qualified teachers preferred to join *Soporte*'s training teams. The DREs and program staff agreed to adjust pay scales and to decrease the pay differential between *Amazonía Lee* and *Soporte*, but the perception that *Soporte Pedagógico* had more resources remained.

As with any teacher training program, stakeholder interviews revealed that there was high turnover of trained teachers within and across schools during the implementation of *Amazonía Lee*. These qualitative findings on teacher turnover are consistent with findings from the teacher survey. For example, in Ucayali only 16 out of the 35 teachers surveyed in 2016 reported teaching first, second or third grade at the *Amazonía Lee* schools in 2015. In San Martín, 55 out of 100 teachers surveyed in 2016 reported teaching first, second or third grade at the ALEE school in 2015. Furthermore, substantial delays in the hiring process interrupted implementation of program activities in *Amazonía Lee* schools. Every time a teacher trained by *Amazonía Lee* left, a replacement teacher had to be trained. Continuity of teachers in the program is key for the interventions' success and sustainability. Teachers need consecutive years of practice to deepen their knowledge about literacy instruction and build their teaching skills over time with targeted feedback from the coaching sessions.

Amazonia Lee teachers had difficulty implementing differentiated instruction strategies. As part of the training workshop teachers were taught differentiated instruction strategies. *Amazonia Lee* teacher training aimed to promote teachers ability to identify students' learning needs and their level of social and emotional maturity, and provide instructional support for reading instruction accordingly. *Amazonia Lee*'s strategies for differentiated instruction included establishing a positive and inclusive classroom climate, grouping students intentionally and flexibly according to their skill level, and using diagnostic and formative assessment for lesson planning, monitoring students' progress, and adjusting instructional support. However, in focus groups, teachers mentioned that the strategies for differentiated instruction were difficult for them to adopt and use in their regular classroom practice. Teachers felt they needed to

continue improving their ability to tailor instructional support according to students' needs and to apply differentiated instruction strategies more regularly.

Pedagogical materials

UPCH provided culturally relevant student materials and teacher guides based on *Amazonía Lee*'s balanced approach to reading and writing instruction, but not until the second year of the program. During the two years of program implementation, UPCH provided training materials for *Amazonía Lee* trainers. Materials specific to the *Amazonía Lee* intervention were not available in the first year of implementation because the 2015 school year started shortly after UPCH's agreement was signed, leaving little time for design and production of materials specific to *Amazonía Lee* before the start of the year. Instead, in the first year (2015), UPCH provided student workbooks and teacher guides developed and validated by UPCH, with support from USAID under previous education projects in Peru. For the second year of implementation (2016), UPCH produced new pedagogical materials for *Amazonía Lee* trainers, teachers, and students. Compared with the materials used during the first year of implementation, the new *Amazonía Lee* materials were better targeted to the achievement level of children in the region, and were tailored to the Amazonian cultural context. The new materials were reviewed and approved by staff from the pedagogical management team and by primary education specialists at the Ministry of Education (MINEDU) to ensure they were aligned with MINEDU standards. UPCH validated the student workbooks and teacher guides with staff of DREs, and then distributed the materials to *Amazonía Lee* schools in June 2016.

Challenges with the delivery of *Amazonía Lee* program materials in both departments resulted in a considerable lag between the beginning of training activities and the availability of pedagogical materials for teacher trainers and teachers in both years.

Pedagogical materials for reading instruction and reading for pleasure were not delivered at the beginning of the school year in 2015 or 2016. In both regions, UPCH trainers delivered teacher guides and student workbooks toward the end of the school year in 2015 and midyear in 2016. Similarly, classroom libraries to promote reading for pleasure, which are central to the *Amazonía Lee* intervention, were only delivered in the last quarter of the second year of implementation when the study data collection was ending. Trainers and teachers expressed the need for grade-specific pedagogical materials. In response to those needs, UPCH provided grade-specific teacher guides for use in 2017 (outside the evaluation time frame).

In spite of the delays in the availability of *Amazonía Lee* materials, teachers were very satisfied with the quality of the materials when they did receive them, and felt that they had some advantages over other pedagogical materials they had used previously. For example, they perceived that the grade-level materials they generally receive from MINEDU are too difficult for their students and are culturally biased to the coastal and Andean regions. In contrast, teachers felt *Amazonía Lee* teacher guides and student workbooks were well targeted for the skill level of their students and emphasized culturally relevant themes and activities of the Amazon region. They also noted that materials were well structured and helpful for instruction following *Amazonía Lee*'s balanced approach. Word of mouth about the high quality of *Amazonía Lee* pedagogical materials prompted teachers from schools that were not receiving the program to seek *Amazonía Lee* materials. It was not uncommon for teachers in nontreatment schools to ask *Amazonía Lee* teachers to share their program materials. We learned that *Amazonía Lee* materials

and teaching practices may have been shared informally with some nontreatment school teachers.

Community engagement

Parent workshops and learning fairs were the main activities implemented under the family and community involvement component in 2015 and 2016. This component of *Amazonía Lee* aims to promote the active engagement of families and the educational community in support of student literacy. Parent workshops included activities such as informational sessions and demonstration sessions to acquaint parents with *Amazonía Lee*'s balanced approach to reading, to provide information on how children learn to read, and to offer parents ideas for how to engage actively in their children's journey to becoming readers.

As in the capacity-building component, the implementation of the parent workshop followed a train-the-trainer approach, where UPCH lead trainers provided technical assistance to *Amazonía Lee* teacher trainers on how to conduct parent workshops. Subsequently, LANN and PEEL trainers assisted and coached teachers to conduct parent workshops in both departments. In Ucayali (LANN), at least two parent workshops took place each year (2015 and 2016) during the evaluation period after program implementation began. In San Martín, two parent workshops were implemented in 2016, but none during the first year of implementation (2015).

Learning fairs were community-wide events to showcase reading, storytelling, and literacy-related activities and achievements. Various members of the community participated in the learning fairs, including municipal authorities, civil society organizations, teachers, students, and parents. In Ucayali, learning fairs were held during both years of program implementation towards the end of the school year. In San Martín, learning fairs were held at the end of the 2015 school year, and were planned for December 2016, the second year of implementation.

Even though some parental and community engagement activities were implemented in each region during the study period, *Amazonía Lee* trainers highlighted that parental and community engagement efforts could be improved the following year. For example, San Martín DRE staff highlighted that the program invested the greatest effort in the capacity-building component, particularly the teacher trainings. All other program components were still under development or had not yet been consolidated by the end of the study period. Even though program staff and teachers in *Amazonía Lee* schools had noticed an increase in parental involvement, they stated that those activities were not sufficient to enable parents to support children's reading at home.

Reading assessments system

The main goals of the regional assessment system were: (1) to provide monitoring data on students' reading achievement at the regional level, and (2) to provide individual-level data on students' reading progress and inform reading instruction. UPCH provided technical assistance, developed the reading assessments, and provided the administration, scoring, and interpretation guidelines to the regional implementation teams in both departments. In Ucayali, LANN trainers administered the reading assessments to students in 2015 and 2016. Teachers received the assessment results, but in our interviews, we learned that most teachers did not rely on them to differentiate instruction. In San Martín, the implementation team administered the assessments in 2016, but experienced operational problems with the optic readers; therefore, results were not available for teachers' use.

Teacher nonmonetary incentive system

Some activities related to the nonmonetary incentive system were implemented in *Amazonía Lee* schools, but this component was not rolled out completely during the study period. UPCH designed a system of incentives using as a framework MINEDU's contest to recognize excellence in teaching. The incentive system was implemented only in schools participating in *Amazonía Lee*. To benefit from incentives, teachers needed to meet two requirements: they (1) had to be teaching 1st, 2nd or 3rd grade (by contract or tenured) and (2) had to be continuously engaged in the teacher training activities of *Amazonía Lee*. Criteria to select the best-performing teachers for the awards were based on the rubric UPCH designed to assess teachers' progress and on a teacher narrative illustrating his or her best practice experiences in reading and writing instruction. Nonmonetary incentives were in the form of "distinction awards" conferred to high-performing teachers.

In Ucayali, the implementation of this component was still in the planning stage in November 2016, when the study data collection had ended. LANN trainers assessed teachers' performance using UPCH's rubric, and teachers were scheduled to submit their narrative documents at the end of November 2016. The date for conferring the distinction awards had not yet been designated. In addition to the distinction awards, Ucayali regional authorities and program leadership were assessing the option of awarding laptops in addition to the distinction awards. The implementation of the incentive system was more advanced in San Martín—approximately 100 teachers (out of 852) had been selected as finalists, and the distinction awards were conferred in December 2016.

B. Services provided to control schools

While it is important to describe the services provided by *Amazonía Lee* as we did in the previous section, we also need to understand what services were provided to the schools assigned to the control groups in Ucayali and San Martín. Understanding the difference in services between the study groups will allow us to better interpret the evaluation findings. In this section we describe findings from focus groups and interviews conducted in control group schools.

Ucayali

Control schools in Ucayali did not participate in special training programs or professional development initiatives; teachers reported having little access to in-service training and faced challenges with MINEDU's pedagogical materials. Most teachers stated they received little or no in-service training at their current school. Some teachers attended sporadic training workshops at the UGEL, whereas others had been proactive in seeking professional development trainings during vacation periods and paying out of pocket for those trainings. Often, teachers looked for resources online and relied on their peer teachers to collaborate on planning lessons based on MINEDU's pedagogical materials for the national curriculum *Rutas del Aprendizaje*.

Training by the state, never. I look for trainings on my own and go during my vacations . . . I pay with my own money and that's how I get training. I've looked for training in reading comprehension . . . it costs 90 soles [equivalent to \$28 USD]. . . and lasts almost two months.
 —Ucayali, prevailing practice teacher focus groups

Some control teachers reported lacking pedagogical materials for reading instruction or books to encourage reading for pleasure. Teachers expressed that MINEDU's student workbooks are not adequate for children in the Amazon region because they are culturally biased to the Andean and coastal regions and seem to be above grade level for children in Ucayali. Most teachers in the focus groups shared that they were unable to use MINEDU's workbooks at the beginning of the year because children's literacy skills were not yet at the level expected for workbook activities. The lack of in-service training and the challenges teachers face with pedagogical materials are consistent with how teachers described their approach to reading instruction. Teachers stated they usually pick and choose strategies and activities from various sources, not necessarily adhering to a specific pedagogical approach. For example, some 1st-grade teachers stated they do phonologic awareness exercises, and others the syllabic approach (*silabeo*).

I develop my own material because I do not have text books this year. My colleague has, but I don't. I made a formal request to the DRE, but the text books have not arrived yet . . . I learn from my colleague the units that we are going to teach
 —Ucayali, prevailing practice teacher focus group

San Martín

As described in chapter one, approximately three-quarters of the schools assigned to the control group in San Martín received *Soporte Pedagógico* services. *Soporte Pedagógico* is the flagship primary education support program from MINEDU that aims to improve student learning in the areas of mathematics, reading and communication, science, and social skills in urban elementary schools through four main lines of action: (1) strengthening skills of teachers and administrators with training workshops, peer teacher study groups, and virtual educational support for teachers and principals; (2) providing after-school tutoring for students not meeting the expected learning standards delivered by teachers specialized in remedial support; (3) providing educational materials and resources; and (4) improving the school and local management capacity. *Soporte Pedagógico* operates in 18 of the 25 regions in the country, but doesn't necessarily serve all schools in the regions in which it operates, since it targets mostly urban schools.

Soporte Pedagógico and *Amazonía Lee* have similar goals, and to some extent, offer comparable services to teachers and school principals; yet, each program has some unique features. *Soporte Pedagógico* and *Amazonía Lee* offer in-service training and coaching to teachers and school principals, and provide educational materials to teachers and students. *Soporte Pedagógico* targets several content areas and social skills, whereas *Amazonía Lee* focuses on communication and reading. *Soporte Pedagógico* offers after-school tutoring to underperforming students by specialized personnel; whereas, *Amazonía Lee* trains teachers to use differentiated instruction during regular school hours. *Soporte Pedagógico* is a centralized strategy designed and managed by MINEDU; on the other hand, *Amazonía Lee* is led and implemented by regional governments with technical assistance from UPCH and addresses the local needs and culture of the Amazon region. *Amazonía Lee* has a nonmonetary teacher incentive system to promote effective teaching practices among trained teachers. *Soporte Pedagógico* does not have a teacher incentive system, but teachers are free to participate in the national contest from MINEDU. Table II.1 compares in more detail services offered by both programs.

Table II.1. Program features of *Amazonía Lee* and *Soporte Pedagógico*

Program features	<i>Amazonía Lee</i>	<i>Soporte Pedagógico</i>
Teacher supports		
Teacher training workshops	X	X
Teacher coaching (one-on-one support)	X	X
Demonstration sessions	X	X
Peer teacher study groups	X	X
School supports		
Program-specific pedagogical materials	X	X
Leadership and administrative training (school principals)	X	
Teacher incentive system (other than Bono Escuela ^a)	X	
Standardized student assessments	X ^b	ECE ^c
Remedial tutoring for underperforming students (after school)		X
Pedagogical approach		
Foundational reading skills (phonemic awareness, alphabetic knowledge, vocabulary, fluency, reading comprehension)	X	
Intercultural bilingual education		SPI ^d
Individualized/diversified instruction based on students' skill level	X	X
Subject areas		
Focus on language and communication	X	X
Focus on math		X
Focus on social skills		X
Which teachers participate		
Main classroom teachers	X	X
After-school teachers		X
Parent and community supports		
Parent workshops	X	X
Learning fairs and community events	X	X

^a*Bono Escuela* (2015–2016) is a monetary incentive program for high-performing public school principals and teachers. The awards for primary schools range between 1,000 and 1,500 nuevos soles.

^b The reading assessment component of *Amazonía Lee* schools was implemented in treatment schools in addition to the ECE. It follows a similar format to the ECE but is administered at the beginning and end of the school year to students in 1st through 3rd grades.

^c *Evaluación Censal de Estudiantes* administered to all public schools in 2nd grade (except bilingual schools in which it is administered in 4th grade).

^d *Soporte Pedagógico Intercultural* (SPI). *Soporte Pedagógico* and SPI aim to improve the quality of education and student achievement; however, SPI emphasizes the implementation of the Ministry of Education's bilingual intercultural education model for literacy acquisition in rural areas (in mother tongue and in Spanish). *Soporte Pedagógico* targets schools in urban areas.

Teachers in control schools in San Martín reported being trained on and using the communicative-textual approach for reading instruction, but they seem to face challenges with MINEDU's pedagogical materials. Teachers reported that subject matter textbooks and workbooks often don't arrive to their school on time or arrive incomplete. They also said that MINEDU's books (known as *Rutas del Aprendizaje*) are pertinent for the coastal region but are not culturally relevant for students in the Amazon region. Teachers often have to adapt materials to make them more culturally relevant to their students. Further, teachers use other pedagogical resources for reading instruction that they find more helpful than MINEDU's materials, such as the *Coquito* guides and workbooks from the program *PeruEduca*.

The books reflect the reality of children from the coastal region, and we adapt it to our reality . . . to our region.
—San Martín, prevailing practice teacher focus groups

First- and 2nd-grade teachers in control schools also received monthly coaching sessions and participated in teacher study groups. Teachers in schools receiving services from *Soporte Pedagógico* participate in coaching sessions, where a coach monitors their teaching practice and provides feedback on their strengths and improvement areas. Furthermore, *Soporte Pedagógico* offers monthly visits, while in San Martín, *Amazonía Lee* offered bi-monthly visits due to the large number of schools served by the PEEL trainers.

Underperforming students in schools receiving *Soporte Pedagógico* are eligible for after-school tutoring, but teachers have noticed marked absenteeism during tutoring sessions. The after-school component for academic reinforcement of *Soporte Pedagógico* is implemented with teachers specialized in individualized instruction for underperforming students (known as *Docentes Fortaleza*). *Docentes Fortaleza* visit schools about three times per week, after the regular school schedule, to provide tutoring or remedial support to students that have been identified at risk for academic failure. Some teachers feel that the duration and frequency of the tutoring sessions may be insufficient to really help underperforming students catch up. They've also noticed poor attendance at tutoring sessions, usually by students in greater need for remedial support.

C. Principal and teacher perceptions of *Amazonia Lee*

Some teachers and principals were initially resistant to *Amazonía Lee* teacher training activities but over time came to recognize the value of *Amazonía Lee*'s balanced approach to reading instruction. Earlier educational initiatives implemented in San Martín, such as SUMA or Aprender, were eventually discontinued. As a result, some teachers and principals were somewhat skeptical about the potential reach and sustainability of another educational program in the region. Further, older and more experienced teachers were initially resistant to a new pedagogical approach for reading instruction. Over time, most teachers became engaged and motivated by the training activities in which they participated. Teachers acknowledged the value of *Amazonía Lee*'s balanced approach and reported they have changed the way they teach reading significantly. School principals reported they liked this balanced approach because it integrates two complementary approaches to reading instruction, and emphasizes the oral and written aspects of literacy. Previously, there was greater emphasis on text production and writing instruction. Early on teachers struggled with *Amazonía Lee*'s focus on oral language skills, but recognized that it actually helped students in their classrooms. In addition, teachers felt they were successful at creating an enriched literacy environment in the classroom and improved their classroom management skills.

What I liked is that [the program] has left nothing aside . . . they have taken what's good from this approach, from the other and from the other and have created the balanced [approach]. Teachers have been empowered with this approach, and have adopted didactic processes and strategies.
 —San Martín, school principal

Teachers found *Amazonía Lee* training activities highly demanding in terms of time and effort and sometimes felt they were unable to keep up with what was expected from the program. Program staff had trouble finding suitable periods of time to conduct the teacher trainings that would not interrupt classes, compromise student learning time, or use teachers' vacation periods. Training activities added to teachers' burden and required that they use

personal time to participate in the *Amazonía Lee* training activities. Also, teachers that lived far away from the training locations experienced challenges related to travel and lodging cost reimbursement.

School principals believe teachers have improved their reading instruction practices and have noticed widespread desire among teachers for continuous improvement through participation in teacher study groups, coaching sessions with teaching demonstrations, and the use of culturally relevant pedagogical materials. Most teachers showed ownership of *Amazonía Lee*'s balanced approach to reading instruction, were highly motivated by the *Amazonía Lee* training activities, and seemed committed to continue improving their pedagogical skills. School principals acknowledged that several primary teachers in their schools lacked effective teaching skills, but they showed improvements and were moving toward “teaching reading for comprehension” as opposed to “teaching reading for the sake of reading.” Teachers shared that there were some aspects of reading instruction that have been more challenging for them to learn. For example, working with inferential reading comprehension questions, using formative evaluations, and working with students of different levels of achievement were still areas in which they felt they need to improve.

Teachers found the *Amazonía Lee* balanced approach effective and noticed their students were achieving reading milestones earlier in the school year. Teachers noticed their students were gaining reading skills sooner in the school year with the *Amazonía Lee* balanced approach, compared to previous years. Also, they observed increased student engagement and more active class participation. Teachers expressed that they feel pressure to demonstrate improvement in ECE scores. However, a regional project coordinator emphasized that the scope and reach of *Amazonía Lee* extends beyond raising ECE scores and strives for holistic improvements in students reading and learning. We also found evidence of this effort to focus on student learning beyond ECE indicators from the teacher survey in San Martín—teachers in *Amazonía Lee* schools tend to spend less time preparing students for the ECE test than teachers in control schools.

Teachers say that children now talk more, are more self-assured . . . and are improving in reading and writing in less time. In June and the vast majority of children were reading and writing . . . they were understanding what they read, especially in second and third grade
—San Martín, regional program staff

III. FINDINGS FOR UCAYALI

The *Amazonía Lee* intervention was intended to provide capacity building services to teachers in order to improve their reading instruction strategies and ultimately improve student reading outcomes. This chapter presents the findings for Ucayali. Because this evaluation uses an intent-to-treat design and analysis, we begin by looking at what services the study groups actually received. We first look at capacity building services provided to the schools and teachers in the study groups. Overall, we find that *Amazonía Lee* led to improvements in professional development support for teachers and potentially greater resources available to support student reading. Next, we look at impacts of *Amazonía Lee* on teacher practices. We found mixed results in actual teaching practices with no impacts on the majority of teaching practices and use of time. However, the impacts on teaching practices that were identified are consistent with teachers in the treatment group implementing the *Amazonía Lee's balanced* approach. Finally, we look at impacts on student reading outcomes. Consistent with the implementation of the intervention, we found that *Amazonía Lee* led to better student reading outcomes compared to a control group of students whose teachers were not receiving the same level of services in support of reading instruction as *Amazonía Lee* provided.

As described in Chapter I, we estimated *Amazonía Lee* impacts on teacher outcomes by comparing teachers across the treatment and control groups in 2016. The impacts on student outcomes used a longitudinal sample of students who were assessed in 2016, when most students were in second grade, and control for baseline characteristics collected in 2015 when students were in first grade.

Before turning to the findings, we briefly discuss baseline equivalence. Because we used a random assignment design, we expect that the treatment and control groups have similar characteristics before the intervention starts. Our baseline report (Campuzano et al. 2016) discussed the equivalence of the study groups in more detail. Tables C.1 through C.3 in Appendix C present baseline equivalence for the analysis sample in Ucayali. While the two groups are similar in almost all school and student level baseline characteristics, there are a few differences that are accounted for in the regression used for estimating impacts student outcomes (see Appendix A for a detailed description of the impact estimation).

A. Impacts on school participation in support programs

As noted above, this study used an intent-to-treat design and analysis. Schools in the treatment group were offered *Amazonía Lee*, while those in the control group received the prevailing educational practices. However, schools and teachers in the treatment group were not required to participate in the *Amazonía Lee* services. Furthermore, schools and teachers in either group were not prevented from participating in alternative services. In order to interpret the program impacts, we need to understand what services were actually received by each group. In this section, we describe the principal- and teacher- reported participation in training and coaching and materials received. The findings are based on data from principal and teacher surveys and classrooms observations administered in 2016, the second year of the intervention.

Schools in both study groups participated in various support programs. Many schools participate in support programs that provide school nutrition, education technology, and have other inputs to learning that could affect the reading outcomes that are the focus of this study. Most of the differences between groups were not statistically significant (see Table III.1). One difference that is significant, as expected, is that the treatment group was more likely to report participating in *Amazonía Lee*. In the treatment group, 97 percent of the principals reported participating in *Amazonía Lee*, and 14 percent of the control principals reported participating. This is equivalent to one principal in the treatment group reporting the school did not participate in *Amazonía Lee* and six control principals reporting their school did participate. We suspect these differences are due to respondents' lack of recall or knowledge of the names of particular programs implemented in their schools, given that administrative data provided to us by the DRE and UPOCH consistently reported that all schools in the treatment group participated in *Amazonía Lee* and none of the control schools did.

Table III.1. Treatment schools were more likely to participate in *Amazonía Lee*

Program	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
<i>Amazonía Lee</i> , <i>Liderando los Aprendizajes de Niñas y Niños</i> (LAAN)	97.1	14.3	82.9*	0.000
<i>Soporte Pedagógico</i> (MINEDU)	14.3	14.3	0.0	1.000
<i>Docentes Fortaleza</i>	11.4	8.6	2.9	0.711
<i>Programa Estratégico Logros de Aprendizaje</i> (PELA)	31.4	31.4	-0.0	1.000
<i>Programa Nacional de Redes Educativas Rurales</i> (REDES)	8.6	14.3	-5.7	0.487
Education materials provided by MINEDU	94.3	97.1	-2.9	0.571
One Laptop per Child	45.7	57.1	-11.4	0.402
Computer lab/technology room	51.4	28.6	22.9*	0.044
Information and communication technology				
Huascarán Program	5.7	5.7	0.0	1.000
National Children's Nutrition Program Qali Warma	100.0	100.0	0.0	
Reduction of vulnerability	77.1	80.0	-2.9	0.768
Other program	20.0	31.4	-11.4	0.292
Number of schools	35	35		

Source: *Amazonía Lee* School Infrastructure—Final Follow-Up 2016.

Note: Adjusted means were predicted values from an ordinary least squares regression that included covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

MINEDU = Ministry of Education.

1. Teacher-reported participation in professional development

Because one important subcomponent of the *Amazonía Lee* program is the teacher training workshops, we assessed whether the groups had participated in training for literacy instruction, who provided the training, the duration of the training, and which topics were covered.

Assignment to *Amazonía Lee* resulted in more teachers receiving professional development in reading instruction. As shown in Table III.2, in the treatment group, nearly all (97 percent) teachers reported having participated in reading instruction professional development activities, whereas less than half (43 percent) of teachers in the control group did. The difference of 54 percentage points is statistically significant. Treatment teachers' training came from *Amazonía Lee*, whereas control teachers' training came from other sources. We also found that the 11 percent of teachers in the control group reported being trained by other institutions and that none of the treatment teachers did. This finding is important because it confirms that in Ucayali, most of the treatment group was receiving reading instruction training from *Amazonía Lee*, whereas this did not happen in the control group.

Table III.2. Treatment group teachers reported participating in reading instruction professional development at much higher rates than control group teachers

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Teachers who participated in reading instruction professional development activities (percentage)	97.1	42.9	54.3*	0.000
Percentage of teachers who reported receiving reading instruction professional development from:				
<i>Amazonía Lee, Liderando los Aprendizajes de Niñas y Niños</i> (LAAN)	97.1	2.9	94.3*	0.000
<i>Soporte Pedagógico</i> (MINEDU)	14.3	5.7	8.6	0.183
<i>Programa Estratégico Logros de Aprendizaje</i> (PELA)	25.7	20.0	5.7	0.487
UGEL specialist	14.3	17.1	-2.9	0.711
Other institution	0.0	11.4	-11.4*	0.044
Number of teachers	35	35		

Source: *Amazonía Lee* Teacher Survey—Final Follow-Up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

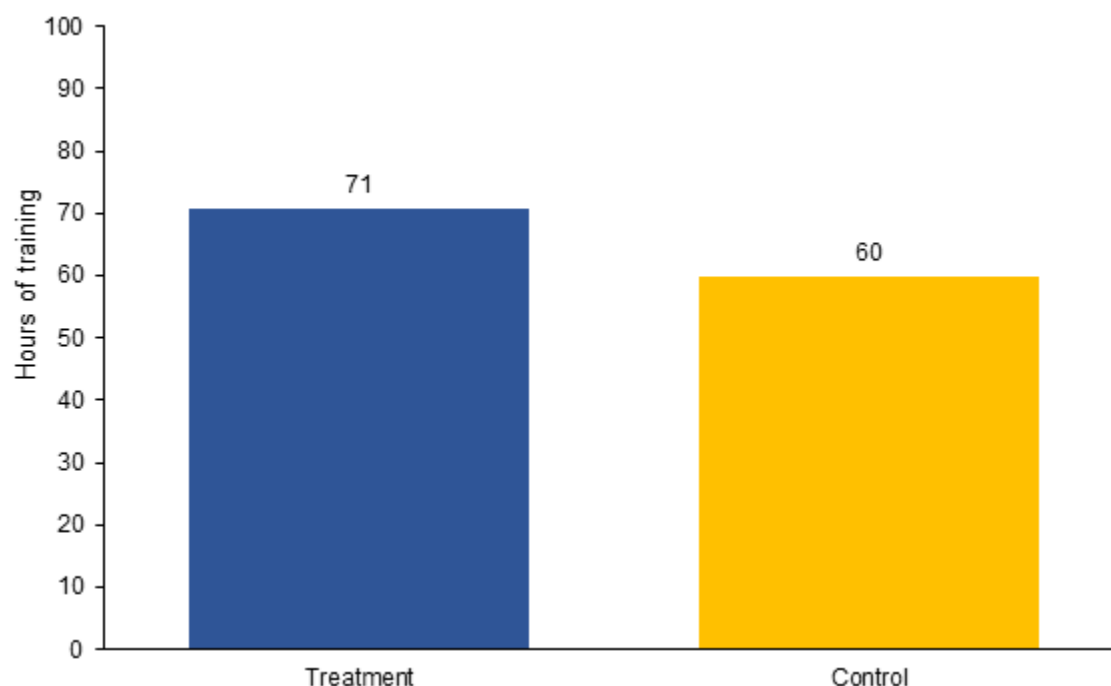
* Difference in group means is statistically significant at the 0.05 level.

MINEDU = Ministry of Education; UGEL = Unidad de gestión educativa local.

Second-grade teachers in both groups reported over 60 hours of training on average. As expected, teachers in the treatment group reported more hours of training from *Amazonía Lee*. Teachers could have received reading instruction training from more than one program, so we added the hours of reading instruction training across all programs. Teachers in the treatment group reported participating 71 hours of training across all programs on average and teachers in the control group reported participating 60 hours across all programs on average, but the difference is not statistically significant (see Figure III.1). However, as expected, when we look at training hours by program, we found significant differences. Teachers in the treatment group received on average, 45 more hours of training from *Amazonía Lee* than teachers in the control group, a statistically significant difference (results not presented). We should note that the averages reported include teachers that did not receive any training, so it masks differences in the training duration for those teachers who did receive some training. If we take into account only the teachers who participated in reading instruction training, we find that treatment teachers

who reported participating in training (34 out of 35) received an average of 73 hours across all training providers and control teachers who participated in training (15 out of 35) received an average of 139 hours across all training providers.¹⁰ So although only half of the control teachers received training (Table III.2), the average training duration reported was almost twice as long as average training duration reported by treatment teachers.

Figure III.1. Total training hours in reading instruction are similar between groups

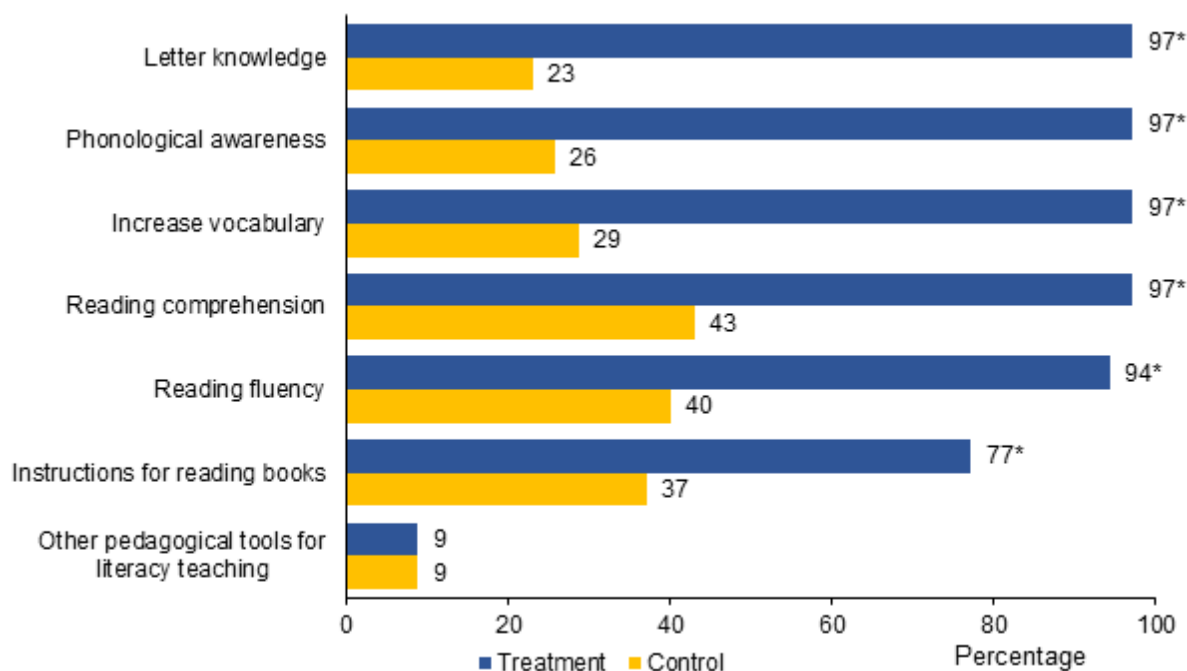


Source: *Amazonía Lee* Teacher Survey—Final Follow-Up 2016.

***Amazonía Lee* teachers reported more training than did control teachers in all the topics related to reading instruction.** As shown in Figure III.2, more teachers in the treatment group reported being trained on topics related to reading instruction. The differences are statistically significant and the magnitude ranges from 40 to 74 percentage points. Teachers in the treatment group are 2 to 3 times more likely to report receiving training on most topics related to reading instruction.

¹⁰ Three teachers in the control group reported very large duration of training. These outliers are driving the large average duration of training in the control group.

Figure III.2. Treatment group teachers reported higher participation rates in training for all the topics related to reading instruction



Source: *Amazonía Lee* Teacher Survey—Final Follow-Up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

***Amazonía Lee* had positive impacts on training in general pedagogical strategies and classroom management.** More treatment teachers than control teachers reported being trained on the following topics: class planning, classroom climate, differentiated instruction, classroom management, and use of regular evaluations. These differences range from 34 percentage points to 57 percentage points (see Figure C.1 in Appendix C).

***Amazonía Lee* led to more in-class coaching, in particular demonstration sessions.** In-class teacher coaching is an important subcomponent of the *Amazonía Lee* program, so we asked teachers if they had received coaching, who provided it, and what topics were covered. Table III.3 summarizes the main findings. Almost all teachers in the treatment group (94 percent) reported receiving in-class coaching, but only half (54 percent) of teachers in the control group did. The difference of 40 percentage points was statistically significant. When we asked about the activities teachers were coached on, we found positive impacts from receiving coaching for virtually all the activities listed.: the largest difference was related to demonstration and observation coaching sessions (77 percentage points difference), followed by professional development coaching meetings, identifying students learning needs, reinforcement for students, helping at-risk students, assessing students' performance during class, designing and preparing materials, class observation, and diagnosing teaching practices. Furthermore, in the focus groups conducted with teachers, they mentioned the demonstration and observation sessions, which are a key coaching activity in *Amazonía Lee*, as one of the features of coaching that was most useful to them.

Table III.3. More teachers in the treatment group than in the control reported receiving in-class coaching

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Teachers who reported receiving in-class coaching	94.3	54.3	40.0*	0.000
Teachers who reported receiving in-class coaching from:				
<i>Amazonía Lee, Liderando los Aprendizajes de Niñas y Niños (LANN)</i>	94.3	2.9	91.4*	0.000
<i>Soporte Pedagógico (MINEDU)</i>	-0.0	8.6	-8.6	0.083
<i>Programa Estratégico Logros de Aprendizaje (PELA)</i>	25.7	28.6	-2.9	0.711
Unidad de gestión educativa local (UGEL)	2.9	5.7	-2.9	0.571
Other	3.0	23.0	-20.0*	0.017
Activities carried out during in-class coaching:				
Demonstration sessions	94.3	17.1	77.1*	0.000
Professional development coaching meetings (individual or in groups)	94.3	42.9	51.4*	0.000
Identify students' learning needs	88.6	37.1	51.4*	0.000
Reinforcement for students	71.4	20.0	51.4*	0.000
Assessment of students' performance during class	88.6	40.0	48.6*	0.000
Help to students at risk	74.3	25.7	48.6*	0.000
Design and preparation of materials	68.6	22.9	45.7*	0.000
Observation sessions	94.3	51.4	42.9*	0.000
Support in the implementation of the safety system	80.0	37.1	42.9*	0.000
Diagnose teaching practices	82.9	42.9	40.0*	0.000
Number of teachers	35	35		

Source: *Amazonía Lee* Teacher Survey—Final Follow-Up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

MINEDU = Ministry of Education; UGEL =Unidad de gestión educativa local

2. Literacy resources and materials in the classroom

One goal of *Amazonía Lee* was to improve the availability of materials that facilitate reading. In particular, the program aimed to provide classroom libraries for all participating schools. We used the teacher survey to inquire about the availability of classroom libraries and materials. In addition, we observed classrooms to assess whether they had materials available and the types of available materials.

The program provided more physical resources for teaching reading. More treatment classrooms than control group classrooms had a space dedicated for reading at endline. An equipped reading area was observed in 69 percent of treatment classrooms and in only 23 percent of control classrooms, a statistically significant difference of 46 percentage points. Furthermore, treatment classrooms were more likely than control classrooms to have materials that facilitate reading instruction. In the treatment group, 71 percent of observed classrooms had 11 or more materials written by the teacher, whereas only 37 percent of control classrooms did, a statistically

significant difference of 34 percentage points. Treatment classrooms were more likely to have picture books with text (69 percent versus 17 percent) and books with stories (83 percent versus 23 percent). See Appendix C, Figure C.2 for literacy resources and materials observed in 2nd-grade classrooms.

Teachers in the treatment group were more likely to report receiving materials from *Amazonía Lee* and having a classroom library. The treatment group was more likely than the control group to receive the following materials from *Amazonía Lee*, workbooks (74 percentage point difference), textbooks (43 percentage point difference), guide books (26 percentage point difference), posters and laminates (29 percentage point difference), storybooks (83 percentage point difference), and audiovisual resources (14 percentage point difference) (see Appendix C, Table C.4). Furthermore, in the treatment group, 91 percent of teachers reported having a classroom library, whereas only 6 percent of teachers in the control reported having a library. The difference of 86 percentage points is statistically significant (see Appendix C, Table C.5). We found no differences between groups in the likelihood that they reported having received materials from MINEDU: around 90 percent of teachers in both groups reported receiving MINEDU materials. See Appendix C, Table C.4 for more information on types of materials provided by MINEDU.

We also asked teachers about their perceptions of barriers to teaching reading. Teachers in the control group were more likely than treatment teachers to report a lack of pedagogical resources and materials and lack of training in methods to teach literacy. In the treatment group, 49 percent of teachers mentioned a lack of training methods for teaching literacy, and 80 percent of control teachers did, a statistically significant difference of 31 percentage points. (See Appendix C, Table C.6).

B. Impacts on teaching activities and practices

In the previous section, we discussed the services that the study groups reported receiving. Following the logic model of the program, we expect those services to first affect teacher practices and use of time and, through these changes in teacher practices, to ultimately affect student literacy outcomes. In this section, we discuss the impacts of *Amazonía Lee* on teacher activities and use of time. We obtained data on teacher activities and practices as part of our classroom observations as well as through teacher reports in surveys we conducted with teachers. We focus on the findings from the teacher observation conducted at endline in 2016, to minimize potential teacher reporting bias that may arise in the teacher self-reports. However, some outcomes, such as use of assessments or differentiated instruction, can only be analyzed with teacher-reported data. We therefore also discuss findings from teacher survey data collected in 2016, but when both sources of data are available, we emphasize classroom observation findings.

1. Impacts on classroom practices and time use

Despite having positive impacts on teacher training and coaching, *Amazonía Lee* had no detectable impact on most measures of literacy instructional practices. However, the few impacts we found are consistent with the treatment group implementing strategies promoted by *Amazonía Lee*'s balanced approach. Based on classrooms observations of the Reading and Communication class conducted in 2nd grade classrooms in 2016, observers reported whether the teacher had done 15 activities related to teaching reading during the

observation period (see Table III.4). Teachers in both groups were observed conducting similar activities; we found only one statistically significant difference between groups. In the treatment group, 20 percent of teachers were observed explaining the use of different type of texts, whereas only 6 percent in the control group did, a statistically significant difference of 14 percentage points. In addition, we found two differences significant at the 10 percent level. In the treatment group, 75 percent of teachers used pre-reading activities, whereas 54 percent did in the control group, a difference of 20 percentage points. Also, in the treatment group, 26 percent of teachers encouraged students to make inferences, whereas only 9 percent of control teachers did. It is possible that these differences are due to the treatment teachers implementing strategies promoted by *Amazonia Lee*'s balanced approach, such as explaining use of texts, using pre-reading activities, and encouraging students to make inferences based on the texts.

Table III.4. Teachers were observed performing activities related to teaching reading at similar rates in both groups, with one exception

Observed teacher practice	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Phonological and phonemic awareness				
Encouraged students to recognize letters, identify the names of letter	11.4	17.1	-5.7	0.535
Introduced or rehearsed vocabulary words and explained word meaning	31.4	37.1	-5.7	0.600
Provided explicit instruction on phonological awareness	31.4	17.1	14.3	0.169
Prompted students to segment words into specific units	20.0	8.6	11.4	0.211
Provided instruction on grammar, punctuation, or spelling	34.3	37.1	-2.9	0.800
Provided instruction on word writing	77.1	82.9	-5.7	0.571
Fluency				
Teacher engaged students in fluency-building activities	94.3	97.1	-2.9	0.571
Book reading				
Teacher used pre-reading or book preview	74.3	54.3	20.0	0.051
Teacher encouraged students to make inferences	25.7	8.6	17.1	0.057
Teacher explored children interest in the story and facilitated connections with it	42.9	25.7	17.1	0.110
Teacher explained the use of different types of text	20.0	5.7	14.3*	0.023
Vocabulary and reading comprehension				
Encouraged students to use context clues for reading comprehension	11.4	8.6	2.9	0.711
Discussed text structure	14.3	5.7	8.6	0.263
Discussed the characters in the text	51.4	34.3	17.1	0.160
Guided students in the use of reading comprehension strategies	91.4	80.0	11.4	0.160
Number of classrooms	35	35		

Source: *Amazonia Lee* Classroom Observation—Final Follow-Up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Similarly, findings from the teacher survey are mixed but the positive impacts found were consistent with the treatment group implementing strategies from the *Amazonia Lee*'s balanced approach. The teacher survey asked teachers whether they did 15 activities related to **emergent literacy and fluency**.¹¹ Most teachers reported using these activities, and we found no

¹¹ Teachers reported on the following emergent literacy and fluency activities: rhymes or games with the sounds of words, asking students to make predictions about the readings, practice to complete incomplete words, practice to complete incomplete sentences, quick reading exercise on familiar words, pre-reading exercises, demonstrating reading, silent reading, reading exercises in pairs or small groups, group reading exercises in chorus, practice to

statistically significant differences between groups (see Table C.7 in Appendix C). However, more teachers in the control group reported doing dictation every day, 14 percent in the treatment group and 37 percent in the control group. The difference is statistically significant (see Figure C.3 in Appendix C). This could be related to control teachers using a more “traditional” approach to teaching spelling. Second, the teacher survey asked whether the teacher used 10 activities related to **reading comprehension**.¹² We found one statistically significant difference. In the treatment group, 97 percent of teachers reported teaching students how to make a graphic organizer of the text, whereas only 74 percent of control teachers did. This strategy in particular is something that *Amazonia Lee* emphasizes as a way to help students improve their reading comprehension. Furthermore, when we asked about daily activities, we found two differences. In the treatment group, 69 percent of teachers reported doing reading activities for pleasure or entertainment every day, whereas 40 percent of teachers in the control group did, a statistically significant difference of 29 percentage points. Reading for pleasure is something promoted by *Amazonia Lee*. We also found that 17 percent of teachers in the treatment group reported asking students questions about topic or main ideas every day, whereas 51 percent of the control teachers did—a 34 percentage points statistically significant difference (see Appendix C, Figure C.4). This could be related to control group teachers using a more traditional approach to reading comprehension, such as asking questions after reading. In sum, we found only a few differences in teacher practices. But the positive impacts on teacher reported practices we found are also consistent with the treatment group implementing *Amazonia Lee*’s balanced approach.

Teachers’ allocation of class time among key literacy activities was similar for treatment and control teachers. Both groups were observed in the Reading and Communication class for 39 minutes on average. We estimated the proportion of class time teachers spent in seven categories (six categories related to literacy and one not related to literacy, see list in Figure III.3), and assessed whether there are differences between groups.¹³ We found no statistically significant difference in the proportion of time teachers spent in these activities. The distribution of time among activities is shown in Figure III.3. Although classroom observations for the evaluation were typically less than one hour, the total amount of time that teachers reported spending on a typical Reading and Communication class (period) was almost

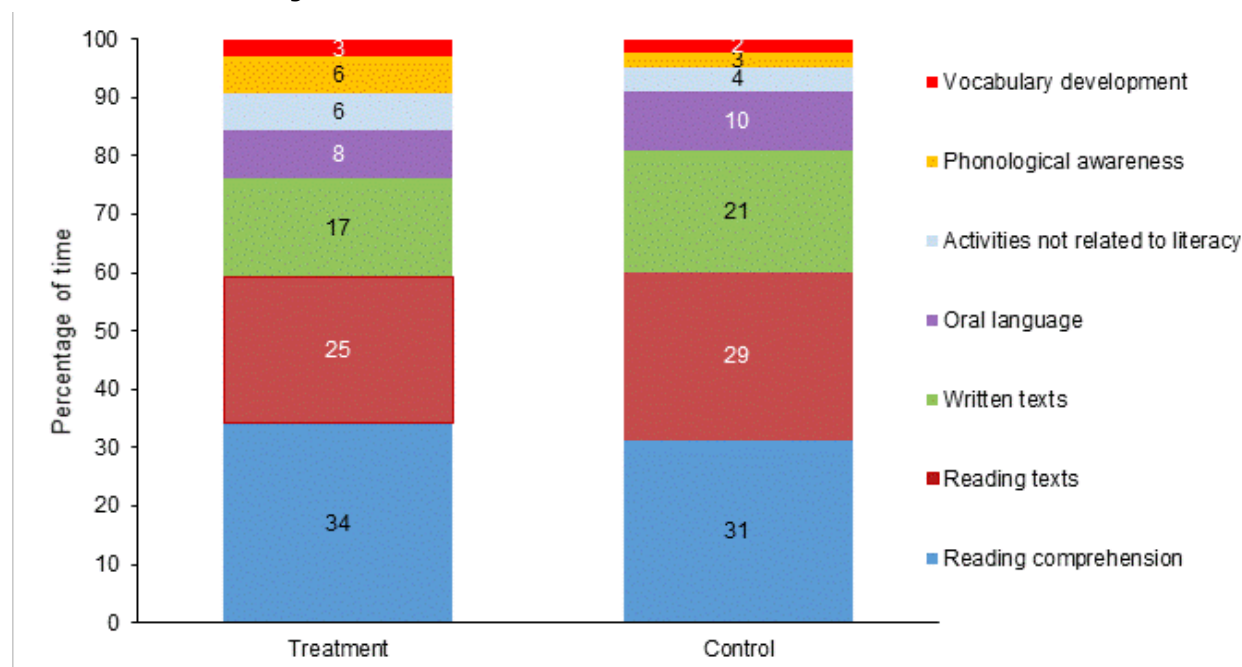
develop graphomotor skills, copying words or sentences, practicing punctuation, dictation, and practicing spelling rules or accents.

¹² Teachers reported on the following reading comprehension activities: defining the purpose of the reading, conducting pre-reading exercises, asking students questions to activate previous knowledge, practicing predicting or anticipating what a text is about, asking students questions about the topic or characters, practicing summarizing a text, teaching students how to make a graphic organizer of the text, developing vocabulary or learning new words, reading activities for pleasure or entertainment, and reading activities to look for information.

¹³ To estimate teachers’ time use we used the following process. In the classroom observation, observers coded whether the teacher had engaged in activities in any of 7 predefined categories (see list in Figure III.3) in a 5 minute observation period. The observers marked all the activities the teacher engaged in during the 5 minute period. There were on average 8 observation periods per classroom observation. First we counted the number of activities the teacher had engaged in the 5 minute period. Then, we divided 5 by the number of activities engaged in, this process assumes that the teacher spent equal time those activities. To estimate the total time spent in each activity per class, we added the time spent in that activity during all the observations periods. Finally, to obtain the percentage of time a teacher spent in each activity, we divided the total time spent in each activity by the total observation time for that classroom observation.

two-and-a-half hours per day for both groups.¹⁴ We also asked the teachers to report the time spent per week teaching reading, and both groups reported spending on average eight hours per week teaching reading (Appendix C, Table C.8).

Figure III.3. Teachers in both groups distributed class time on literacy activities similarly



Source: *Amazonía Lee* Classroom observation—Final Follow-Up 2016.

Note: Difference in group means is not statistically significant.

Looking beyond literacy instruction, there was no pattern of other impacts of *Amazonía Lee* on teachers' general pedagogical practices and classroom climate. As part of the classroom observation, we also looked at general pedagogical practices such as verifying that students understood instruction, verifying that students had relevant materials, providing feedback, etc. (see Appendix C, Table C.9). Teachers in both groups were observed using general pedagogical practices with similar frequencies. In both groups, during the class, teachers were observed having informal conversations or socializing with students on average five times, speaking respectfully to students on average three times, praising students sincerely on average four times, providing feedback to students on average three times, and offering individualized help to students a little over two times on average.

We found no impacts on teacher-reported activities related to classroom climate, with one exception. Teachers in both groups reported always using the following activities at similar rates: praise students, show affection or appreciation to students, display jobs that students have done, and remind students the rules of coexistence in the classroom. However, teachers in the treatment

¹⁴ We inquired about the typical duration of a Reading and Communication class (period), but we do not know how many times a week it was taught.

group were less likely to report always giving students responsibilities in the classroom (49 percent versus 71 percent) (see Table C.10 in Appendix C).

Teachers in both groups were observed spending most of their time teaching the whole class and almost no time teaching small groups. During classroom observations, teachers in both groups spent just more than two-thirds of the time observed teaching the entire class; in most of the remaining time, they worked with individual students. Teachers spent little time working with small groups (results not presented).

***Amazonía Lee* had a positive impact on the time teachers reported spending per week on activities to promote reading in the community after school time.** Teachers in both groups reported spending similar amounts of time in 14 of the 15 general activities we asked about, with one exception.¹⁵ Teachers in the treatment group reported spending almost one hour per week participating in activities to promote reading in the community after school time, whereas control group teachers reported spending on average only a few minutes, a statistically significant difference (Appendix C, Table C.8). This could be related to the implementation of *Amazonía Lee*'s community engagement component.

2. Impacts on differentiated instruction and support for underperforming students

Although we found no impacts on the use of differentiated instruction strategies, the treatment group reported using them less frequently than control teachers. *Amazonía Lee*'s approach to differentiated instruction emphasized student-centered instruction, and promotes teachers' ability to support students according to their skill level though the use of several strategies like diagnostic and formative assessments, peer groups, and personalized student attention. It does not focus only on underperforming students; it is inclusive and directed to support the full range of student skills, motivation and level of maturity. In both groups, around 92 percent of the teachers reported using differentiated instruction for underperforming students. We also asked if teachers used six differentiated instruction strategies: use assessments to identify student underperforming students, organize students to teach based on their needs, plan teaching for different levels of learning, select differentiated content students for different student levels, use differentiated activities for different learning levels, and monitoring individual progress. We found that teachers in both groups reported using these strategies at similar rates. The only statistically significant difference was that 74 percent of treatment group teachers reported planning teaching for different levels of student learning, whereas 91 percent of control group teachers did, a 17 percentage point difference that is statistically significant (see Appendix C, Table C.11).

We also asked to teachers to report the number of times they used these strategies in the last month. Treatment group teachers reported using three of the six strategies related to differential instruction less frequently than control teachers. We found a negative and statistically significant

¹⁵ Teachers reported the time they spent per week on the following activities: administrative tasks, preparing the class, meeting with the principal or other teachers, meeting with students' parents, monitoring students during recess and free time, teaching reading techniques, teaching math, teaching social science, teaching science, teaching art, planning or managing or correcting exams or homework, providing support to students with low academic achievement, preparing students for national examinations, and participating in activities to promote reading in the community after school time.

difference in the number of times treatment group teachers used the following strategies in the last month compared to control teachers: using some type of assessment to identify student learning needs (a difference of –3 times), planning for teaching for different levels of students learning (a difference of –5 times), and selecting differentiated contents for the different levels of student’s learning (a difference of –4 times) (see Appendix C, Table C.12). This is consistent with qualitative findings showing that *Amazonia Lee* teachers experienced difficulties using the differentiated instruction strategies they learned in training during regular classroom practice. Teachers who participated in focus groups recognized that they were not implementing differentiated instruction strategies as the program intended and felt they needed additional practice and support to do it well. It is possible that in the teacher survey treatment group, teachers reported using such strategies less frequently than control teachers as a result of exposure to *Amazonia Lee*’s teacher training and the program’s higher expectations for differentiated instruction. *Amazonia Lee* teachers were probably more conscious of their actual use of differentiated instruction and therefore mean differences across groups reflect reference bias for the treatment group.

***Amazonia Lee* had positive impacts on teacher reported support for underperforming students.** Teachers in both groups reported that on average almost six students out of a typical class of 25 students underperformed in developing reading skills. In the treatment group, 86 percent of teachers reported providing support to these underperforming students outside regular class time, whereas only 57 percent of control teachers reported doing so, a statistically significant difference of almost 29 percentage points. In the treatment group, teachers reported providing additional support to underperforming students on average four hours per week, whereas control teachers reported three hours on average; the difference of one hour is not statistically significant at the 5 percent level but is at the 10 percent level. Furthermore, 31 percent of the treatment group reported referring underperforming students to academic reinforcement, whereas only 11 percent of the control group did, a difference of 20 percentage points that is not statistically significant at the 5 percent level but is at the 10 percent level.¹⁶ (See Appendix C, Table C.13). It is possible that these additional supports to underachieving students are helping drive some of the positive impacts we found on student achievement.

3. Impacts on the use of formative assessments to monitor students’ learning

We found no impacts in the use of formative assessments with two exceptions. Teachers in both groups reported using assessments for making initial diagnostics to plan classes and to organize learning groups at similar rates. They also reported using formative assessments to determine progress in literacy and provide feedback to students at similar rates. However, 77 percent of treatment group teachers reported performing an end-of-year assessment for literacy, whereas all teachers in the control group did, a statistically significant difference of 23 percentage points. Table C.14 in Appendix C, provides information on the reported use of formative assessments, and Figure D.5 reports teachers’ perceptions on students’ performance. We also asked teachers whether they assessed alphabet knowledge, phonological awareness, oral comprehension, reading fluency, reading comprehension, vocabulary, and writing daily, and we only found one statistically significant difference. In the treatment group 29 percent of teachers

¹⁶ Referring underperforming students includes a broad set of options for example the teacher could refer them to additional help in school if it exists, they could also notify the parents that the students is underperforming, they could notify the principal, and so on.

reported assessing oral comprehension daily, whereas 57 percent of the control group did (results not presented).

C. Impacts on students' reading skills

To assess students' reading skills, we looked at impacts on key foundational skills as well as on the main goal, improved reading comprehension (Table III.5). In the first subtest, we assessed oral language comprehension. This provides information on whether students can comprehend a short story that is read to them. The second subtest assessed emergent reading skills by focusing on reading familiar words. The third subtest measured decoding skills. In this subtest, students were asked to read a list of invented words and were given one minute. This subtest allows us to assess whether the student has learned the correspondence between letters and sounds. Reading fluency is associated with reading comprehension. Therefore, the fourth subtest assessed how many words per minute the students were able to read correctly. The last subtest assessed reading comprehension skills.

***Amazonía Lee* had no impacts on oral comprehension skills.** Students in both groups demonstrated emerging oral comprehension skills, responding correctly to 40 percent of the questions. The assessor read a four-line passage to the student and asked the student six oral comprehension questions about the story. At endline, students in both groups responded to two-and-a-half questions correctly, on average, out of six questions.

***Amazonía Lee* had a positive impact on reading familiar words.** Students in the treatment group correctly read on average 7 out of 10 familiar words presented, whereas students in the control group read on average 6 out of 10 familiar words, a statistically significant difference of 1 word. This difference is equivalent to an impact of 0.24 standard deviations. A recent systematic review of educational interventions finds evaluations of structured pedagogy programs, similar to *Amazonía Lee*, have had impacts that are on average 0.23 standard deviations of language arts outcomes, with a 95 percent confidence interval of (0.13, 0.34). *Amazonía Lee*'s impact is in the middle of the range of impacts found by this review (Snilstveit et al. 2016).

***Amazonía Lee* had positive impacts on decoding skills equivalent to 0.24 standard deviations.** Students in the treatment group were able to read correctly almost four more invented words in one minute than students in the control group. 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 statistically significant difference of almost four more words. This is equivalent to 0.24 standard deviations. A recent review of early literacy interventions (USAID 2016) reports that programs have found positive impacts on fluency outcomes as small as 0.14 standard deviations and as large as 0.73 standard deviations. *Amazonía Lee*'s impact is in the middle of the range of impacts found for fluency outcomes by this recent review.

***Amazonía Lee* had marginally positive impacts on fluency equivalent to 0.15 standard deviations.** Students in the treatment group were able to correctly read almost five more words in one minute than students in the control group. To assess fluency, students were given a short story with 77 familiar words and were asked to read the story for one minute. Students in the

treatment group read on average almost 34 words correctly in one minute, and students in the control group read 30 words correctly in one minute. The difference of 4 words per minute is statistically significant at the 10 percent level. This is equivalent to 0.15 standard deviations, which is smaller than the average impact of similar programs but still within the range of what other studies found for fluency (USAID 2016). We should note, however, that both groups have a fluency level that is still below the standard of 60 words per minute that Peru has set for 2nd graders (Banco Mundial 2006). Second-grade students in Ucayali are barely over the standard set for fluency in first grade, 30 words per minute.

***Amazonía Lee* had positive impacts on reading comprehension equivalent to 0.27 standard deviations.** Students were given a simple seven-line story to read in silence, and then the assessor asked six questions related to the story. Students in the treatment groups were able to answer almost three of the six questions correctly, and students in the control group were able to answer on average two questions, a statistically significant difference of almost one question. This is equivalent to 0.27 standard deviations, which is of similar magnitude to what other studies of similar programs have found. Despite the positive impacts, students in both groups still struggle to answer reading comprehension questions correctly, only answering on average 50 percent of the questions correctly.

Table III.5. *Amazonía Lee* led to positive impacts of substantive magnitude on reading familiar words, decoding, fluency, and reading comprehension

	Treatment adjusted mean	Control adjusted mean	Difference	p-value	Effect sizes
Number of correct oral comprehension responses	2.7	2.6	0.1	0.358	0.08
Number of familiar words read correctly	7.0	6.0	1.0*	0.019	0.24
Number of pseudo-words read correctly in one minute (decoding)	16.8	13.4	3.4*	0.002	0.24
Number of words read correctly in one minute (fluency)	34.6	30.2	4.3†	0.060	0.15
Number of correct answers in reading comprehension	2.5	2.0	0.5*	0.001	0.27
Number of students	380	360			

Source: *Amazonía Lee* Child Assessment—Final Follow-Up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables). We also include baseline scores of students and some students' characteristics (2015) and baseline infrastructure characteristics of schools (2014). Effect sizes were calculated before rounding the adjusted difference.

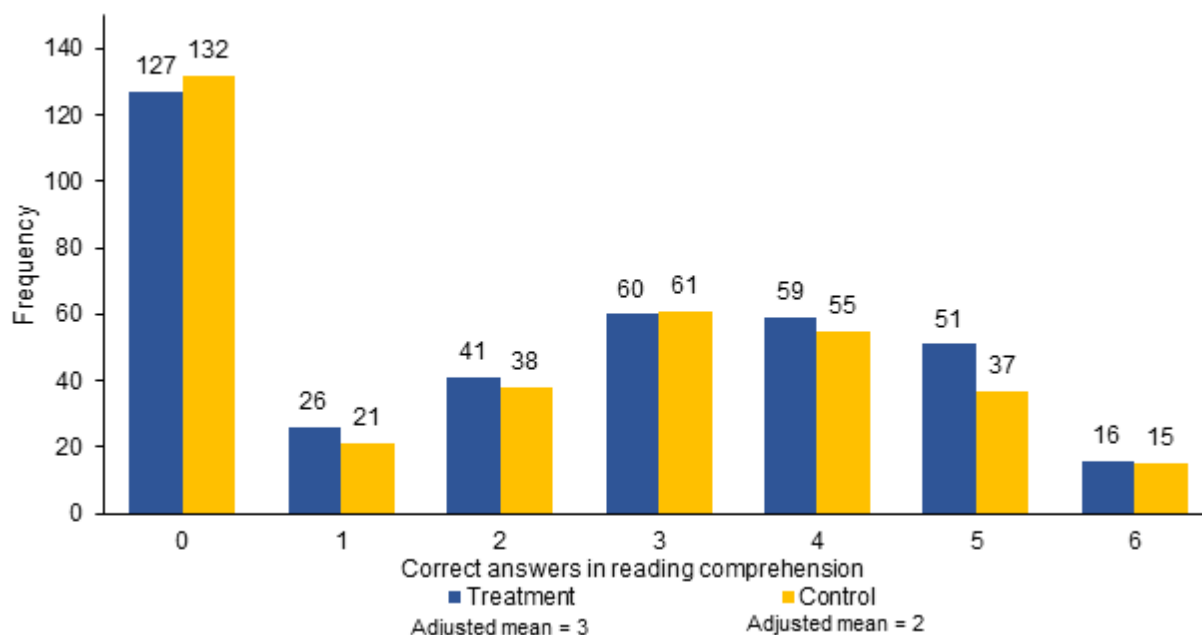
* Difference in group means is statistically significant at the 0.05 level.

† Difference in group means is statistically significant at the 0.10 level.

Many students in both groups are still unable to correctly answer at least one reading comprehension question. Figure III.4 presents the number of students who got each of the possible scores on the reading comprehension test (the lowest is zero correct answers and the largest is six correct answers). The first two columns show that 127 students in the treatment group got a score of zero in the reading comprehension test while 132 students in the control group did. Another way of interpreting this data is that 31 percent of students in the treatment group and 39 percent of students in the control group were unable to answer even one reading comprehension question correctly. Therefore, there were still more than 30 percent of second

grade students (almost 40 percent in the control group) that were unable to read and comprehend. A significant percentage (90 percent) of the students who were scored zero on the reading comprehension test were students that were unable to demonstrate basic reading fluency skills and therefore were not tested on comprehension (See Table I.2).

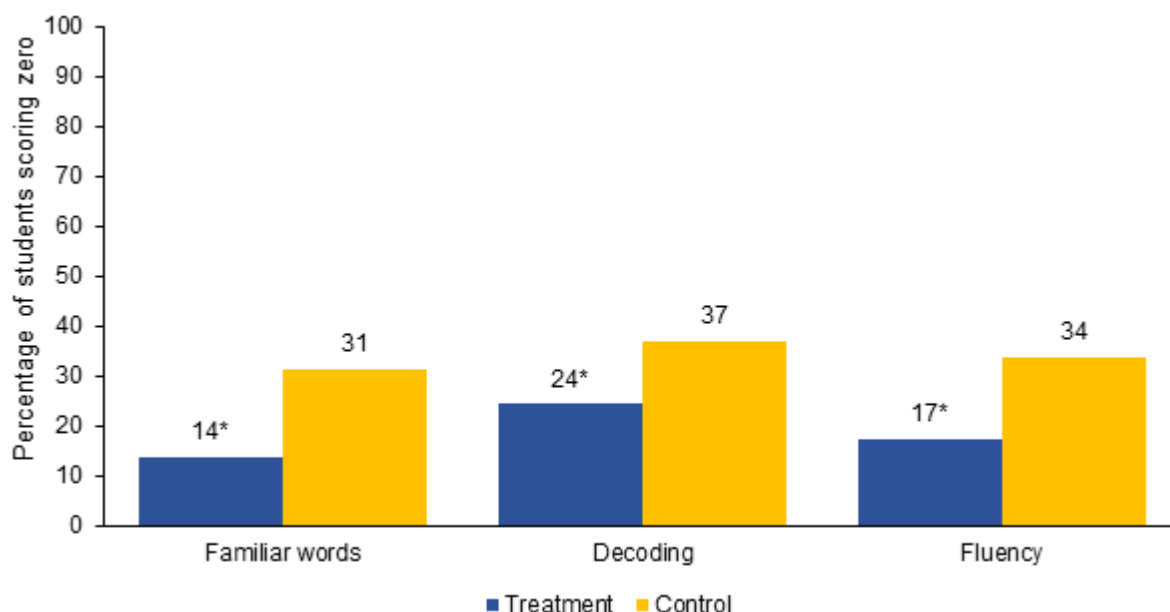
Figure III.4. Histogram of correct answers in reading comprehension



Source: *Amazonía Lee* Child Assessment—Final Follow-Up 2016.

***Amazonía Lee* led to less students scoring zero in the reading familiar words subtest, the decoding subtest, and the fluency subtest.** Students that took the reading assessment were asked to read 10 familiar words. (Figure III.5). In the treatment group, 14 percent of students had a score of zero in the reading familiar words subtest, whereas 31 percent of the control group students had a score of zero, a statistically significant difference. Second, the students attempted the decoding test in which they were given one minute to read aloud invented words. In the treatment group, 24 percent of students had a score of zero in the decoding test, and 37 percent of control group student had a score of zero. The difference of 13 percentage points is statistically significant. For the fluency test, in the treatment group, 17 percent of students had a score of zero, and 34 percent of control group students had a score of zero, a statistically significant difference.

Figure III.5. Students in the treatment group got a score of zero at lower rates than the control group



Source: *Amazonía Lee* Child Assessment—Final Follow-Up 2016.

* Difference in group means is statistically significant at the 0.05 level.

Gender analysis

The results of the gender subgroup analysis were similar to what we found in the full sample. Table III.6 summarizes the findings. We found no differences between groups on the number of correct oral comprehension responses for boys and for girls. We also found that boys in the treatment group were able to read one more familiar word than the control group. For girls, the result is similar but is not significant at the 5 percent level—only at the 10 percent level. As with the full sample, we found that boys in the treatment group were able to read three more pseudo-words in one minute than boys in the control group. We found the same result for girls. Boys in the treatment group were able to answer almost one more reading comprehension correctly than boys in the control group. The results were similar for girls.

The one exception was fluency, where we found a marginal difference for boys, boys in the treatment group were able to read five more words in a minute than boys in the control group, but we found no statistically significant difference for girls. We also looked at differences in the percentage of students who got a score of zero in the reading familiar words, decoding, and fluency tests. While we found that the differences between treatment group and control group boys are greater than the differences between treatment group and control group girls they are not statistically significant different. (See Appendix C, Table C.15).

Table III.6. Positive impacts for boys and for girls in decoding and reading comprehension and only for boys in reading familiar words and fluency

	Impact for girls	p-value	Impact for boys	p-value	Difference in impacts by gender (p-value)
Number of correct oral comprehension responses	0.0	0.986	0.2	0.146	0.240
Number of familiar words read correctly	0.8 [†]	0.081	1.2*	0.018	0.403
Number of pseudo-words read correctly in one minute (decoding)	3.3*	0.018	3.5*	0.007	0.855
Number of words read correctly in one minute (fluency)	3.5	0.204	5.1 [†]	0.053	0.587
Number of correct answers in reading comprehension	0.4*	0.029	0.6*	0.002	0.406
Number of students	386		354		

Source: *Amazonía Lee* Child Assessment—Final Follow-Up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables). We also include baseline scores of students and some students' characteristics (2015) and baseline infrastructure characteristics of schools (2014). Effect sizes were calculated before rounding the adjusted difference.

* Difference in group means is statistically significant at the 0.05 level.

† Difference in group means is statistically significant at the 0.10 level.

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

IV. FINDINGS FOR SAN MARTÍN

The *Amazonía Lee* intervention was intended to provide capacity building services to teachers in order to improve their reading instruction strategies and ultimately improve student reading outcomes. This chapter presents the findings for San Martín. Because this evaluation uses an intent-to-treat design and analysis, we begin by looking at what services the study groups actually received. We first look at capacity building services provided to the schools and teachers in the study groups. Overall, we find that in San Martín *Amazonía Lee* provided similar professional development support for teachers and resources for reading compared to what the control group received, mostly from *Soporte Pedagógico*. Next, we look at impacts of *Amazonía Lee* on teacher practices. We find mixed results on actual teaching practices with no impacts on the majority of them. Where we do find impacts on teaching practices and use of time, they are consistent with teachers in the *Amazonía Lee* group implementing the program's balanced approach. Last, we look at impacts on student reading outcomes. Consistent with the data showing that both groups received substantial training and support in reading instruction, we found no differences in reading outcomes of *Amazonía Lee* students compared to control students, the majority of whom had teachers receiving *Soporte Pedagógico*. As described in Chapter I, we estimated *Amazonía Lee* impacts on teacher outcomes by comparing teachers across the treatment and control groups in 2016. The impacts on student outcomes used a longitudinal sample of students who were assessed in 2016, when most students were in second grade, and control for baseline characteristics collected in 2015 when students were in first grade.

Before turning to the findings, we briefly discuss baseline equivalence. Because we used a random assignment design, we expect that the treatment and control groups have similar characteristics before the intervention starts. Our baseline report (Campuzano et al. 2016) discussed the equivalence of the study groups in more detail. Tables D.1 to D.3 in Appendix D present baseline equivalence for the analysis sample in San Martín. While the two groups are similar in almost all school and student level baseline characteristics, we found a few differences that are accounted for in the regression used for estimating impacts student outcomes (see Appendix A for a detailed description of the impact estimation).

A. Impacts on school participation in support programs

As noted above, this study used an intent-to-treat design and analysis. Schools in the treatment group were offered *Amazonía Lee*, while those in the control group received the prevailing educational practices. However, schools and teachers in the treatment group were not required to participate in the *Amazonía Lee* services. Furthermore, schools and teachers in the treatment or control group were not prevented from participating in alternative services. Therefore, in order to interpret these impacts we need to understand what services were actually received in both groups. In this section, we describe the principals' and teachers' reports on participation in training and coaching, and on the materials they received.

Schools in both study groups participated in various support programs. Many schools participate in support programs that provide school nutrition, education technology, and other inputs to learning that could affect the reading outcomes that are the focus of this study. Most of the differences in participation in these programs were not statistically significant across the two

groups. However, as expected, schools that were assigned to the treatment group were more likely to report participating in the *Amazonía Lee* program (88 percent of treatment group schools and 7 percent of control group schools), and the schools assigned to the control group were more likely to report participating in *Soporte Pedagógico* (5 percent of treatment group schools and 74 percent of control group schools).¹⁷ Schools in the control group were also more likely to report participating in the *Docentes Fortaleza* program (6 percent in the treatment group and 67 percent in the control group), which is consistent with the fact that *Soporte Pedagógico* and *Docentes Fortaleza* are normally implemented together. Furthermore, both groups reported receiving other support programs from the Ministry of Education such as PELA¹⁸ and REDES¹⁹ at rates (16 percent in the treatment group and 22 in the control group reported participating in PELA and 14 percent and 12 percent, respectively, participating in REDES). See Table IV.1. Principals sometimes gave incorrect information regarding the name of the programs; therefore, we also confirmed their participation in *Amazonía Lee* with UPOCH's administrative data.

Table IV.1. Schools in both groups participate in various support programs

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
School participation in the following programs:				
<i>Amazonía Lee</i> , Programa Enseñar Es Liderar (PEEL)	88.0	7.0	81.0*	0.000
<i>Soporte Pedagógico</i> (MINEDU)	5.0	74.0	-69.0*	0.000
<i>Docentes Fortaleza</i>	6.0	67.0	-61.0*	0.000
Programa Estratégico Logros de Aprendizaje (PELA)	16.0	22.0	-6.0	0.222
Programa Nacional de Redes Educativas Rurales (REDES)	14.0	12.0	2.0	0.672
Education materials provided by MINEDU	98.0	98.0	0.0	1.000
One Laptop per Child	59.0	49.0	10.0	0.123
Computer lab/technology room	44.0	43.0	1.0	0.882
Information and communication technology Huascarán Program	3.0	4.0	-1.0	0.657
National Children's Nutrition Program Qali Warma	99.0	98.0	1.0	0.566
Reduction of vulnerability	63.0	61.0	2.0	0.774
Other program	24.0	22.0	2.0	0.741
Number of schools	100	100		

Source: *Amazonía Lee* School Infrastructure—Final Follow-Up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

MINEDU = Ministry of Education.

¹⁷ Table D.1 in the appendix shows that in both groups approximately 63 percent of the schools are urban and the rest were rural. Because *Soporte* targets mostly urban schools, we looked at how many of the 74 *Soporte Pedagógico* schools in the control group are urban and found that 55 are urban and 19 are rural. Of the 26 control schools that did not report participating in *Soporte* 18 are rural and 8 are urban.

¹⁸ PELA is a MINEDU program that is designed to improve educational management and administration, teacher training and coaching, school infrastructure, availability of pedagogical materials, and access to quality education. PELA focuses on reading and communication, math, and social skills and it focuses on 1st and 2nd grades.

¹⁹ REDES is a MINEDU program that creates clusters of neighboring schools who work together to strengthen management and administration and improve teacher training and coaching.

While not all schools assigned to the treatment group participated in the *Amazonía Lee* program, analysis comparing intent to treat and treatment-of-the-treated impacts yielded similar results. In our study sample, 100 schools were assigned to the treatment group and were supposed to participate in *Amazonía Lee*. However, in 2016 when we collected follow-up data, the San Martín DRE informed us that 11 schools that had been assigned to the treatment group had preferred to receive PELA instead of *Amazonía Lee*.²⁰ We confirmed with UPCH's administrative data that 89 of the 100 schools in our treatment group were participating in *Amazonía Lee* and that the 11 treatment schools that preferred PELA were not receiving *Amazonía Lee*. We also confirmed with UPCH data that none of the schools in the control group had received *Amazonía Lee*. Our main analysis compares all the schools assigned to the treatment group to the schools in the control group regardless of whether they participated in *Amazonía Lee*. The main analysis therefore, generates intent-to-treat impacts, and should be interpreted as the effect of the *offer* to participate in *Amazonía Lee*. In addition, we used UPCH's administrative data to construct an indicator variable of whether the school participated in *Amazonía Lee* or not. We conducted additional analysis to estimate the treatment-on-the-treated impacts. The treatment-on-the-treated impacts were estimated using an instrumental variable approach where random assignment to the treatment or control group is used as an instrument for participation in the *Amazonía Lee* program. The treatment-on-the-treated results are similar to the results with the intent-to-treat approach; therefore, they are not presented in this report.

1. Teacher-reported participation in professional development

One important subcomponent of the *Amazonía Lee* program is the teacher training workshops. Therefore, we assessed whether the groups had participated in training for literacy instruction, who had provided the training, how long the training was, and which topics were covered.

Both groups reported receiving teacher professional development in reading instruction, the treatment group from *Amazonía Lee*, and the control group from *Soporte Pedagógico*, PELA, and other institutions. As shown in Table IV.2, around 88 percent of teachers in both groups reported participating in professional development in reading instruction in 2016. As expected, treatment teachers were more likely than control teachers to report receiving training from *Amazonía Lee* (85 percent versus 3 percent) and less likely to be trained by *Soporte Pedagógico* (6 percent versus 64 percent), by PELA (10 percent versus 18 percent) and by other institutions (0 versus 4 percent). All four differences are statistically significant.²¹

²⁰ While *Amazonía Lee* focuses on reading and communication, PELA provides teacher training and coaching and the content areas include reading and communication, math, and social skills. So it is likely that schools needed support in other areas and preferred PELA.

²¹ Teachers' reports on program participation were more consistent with UPCH data than what principals reported, but we still found some inconsistencies. Three teachers in the control group reported being trained by *Amazonía Lee*. It is possible that these teachers either misreported the name of the program that trained them or were trained by *Amazonía Lee* in a treatment school and moved to a control school recently. With the available data, it was not possible to identify whether the teachers had been in a treatment school in these particular cases. In addition, only 85 treatment teachers reported receiving training from *Amazonía Lee*, but we would have expected 89 of them to have participated. We found that the 4 treatment teachers that were in *Amazonía Lee* schools but did not report participating in training had either been teaching a different grade in the same school that school year or had been

Table IV.2. Participation in reading instruction professional development activities is similar between groups

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Teachers who participated in reading instruction professional development activities (percentage)	90.0	86.0	4.0	0.348
Percentage of teachers who reported receiving reading instruction professional development from:				
<i>Amazonía Lee, Programa Enseñar Es Liderar (PEEL)</i>	85.0	3.0	82.0*	0.000
<i>Soporte Pedagógico (MINEDU)</i>	6.0	64.0	-58.0*	0.000
UGEL specialist	17.0	25.0	-8.0	0.145
<i>Programa Estratégico Logros de Aprendizaje (PELA)</i>	10.0	18.0	-8.0*	0.045
Other institution	0.0	4.0	-4.0*	0.045
Number of teachers	100	100		

Source: *Amazonía Lee* Teacher Survey—Final Follow-Up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

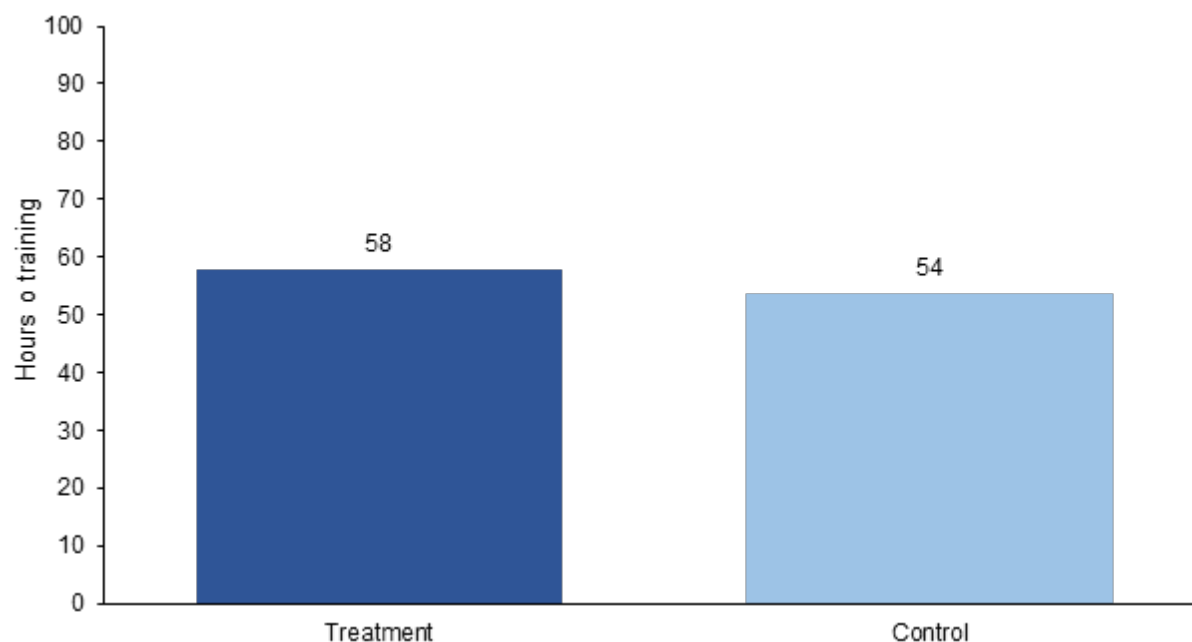
MINEDU = Ministry of Education; UGEL = Unidad de gestión educativa local.

Second-grade teachers in both groups reported having approximately 56 hours of training in reading instruction. But, as expected, treatment teachers received more training hours from *Amazonía Lee* and control teachers from *Soporte Pedagógico* and *PELA*.

Teachers could have received reading instruction training by more than one program, so we added the number of hours reported being trained across all programs. We did not find statistically significant differences between the groups on the hours of reading instruction training received across all programs. On average, teachers in both groups reported being trained in reading instruction for 56 hours across all programs (see Figure IV.1). However, as expected, we found differences when we looked at hours of training by program. The treatment group reported receiving, on average, 43 more training hours from *Amazonía Lee* than the control group, and the control group received 29 more training hours from *Soporte Pedagógico* than the treatment group and 4 more hours from *PELA*. These differences are statistically significant. We should note that the data reported includes teachers that did not receive training. If we take into account only the teachers who participated in each program, we find that teachers that participated in *Amazonía Lee* were trained on average 51 hours, teachers that participated in *Soporte Pedagógico* were trained on average 50 hours, teachers that participated in *PELA* were trained 53 hours, and teachers participating in UGEL training received 36 hours (results not shown). So all programs provided somewhat similar hours of training.

teaching at a different school that school year. This could explain why they had not been trained by *Amazonía Lee* even though their school was participating in the program.

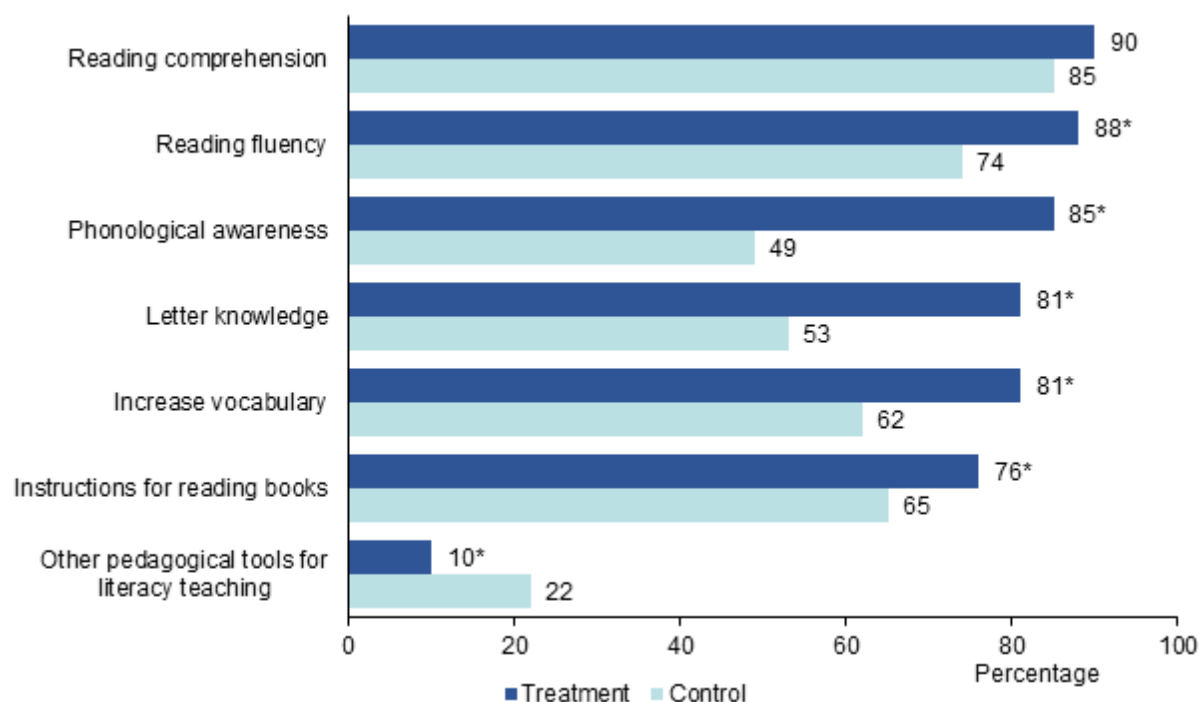
Figure IV.1. Total training hours reported by teachers are similar between groups



Source: *Amazonía Lee* Teacher Survey—Final Follow-Up 2016.

***Amazonía Lee* teachers reported more training than control teachers in letter knowledge, phonological awareness, reading fluency, and increasing vocabulary.** Although most teachers received training in reading instruction, we found some differences in the topics covered between groups. As shown in Figure IV.2, more teachers in the treatment group than in the control group reported receiving training in the following topics: letter knowledge (28 percentage point difference), phonological awareness (36 percentage point difference), reading fluency (14 percentage point difference), and increasing vocabulary (19 percentage point difference). We also found a difference in favor of the control group: the treatment group was less likely than the control group to report receiving training on other pedagogical tools of literacy teaching (12 percentage point difference). The range of other pedagogical tools for literacy reported by the teachers varied—some examples are related to use of materials, pedagogical or didactical processes for communication, strategies for learning, and strategies for literacy. These differences across groups are consistent with *Amazonía Lee*'s balanced approach which incorporates a special emphasis on foundational reading skills. Figure D.1 in the Appendix D presents differences in topics related to general instruction and classroom management; we did not find any differences in those topics between the groups.

Figure IV.2. Treatment group teachers reported higher participation rates in training for most topics related to reading instruction



Source: *Amazonía Lee* Teacher Survey—Final Follow-Up 2016.

* Difference in group means is statistically significant at the 0.05 level.

Both groups reported similar in-class coaching activities with the exception of demonstration and observation sessions that were more likely in the treatment group. In-class teacher coaching is an important subcomponent of the *Amazonía Lee* program, so we asked teachers if they had received coaching, who provided it, and what topics were covered. Most teachers (96 percent) in both groups reported receiving coaching in class. As with training, teachers could receive coaching from more than one program. We found statistically significant differences in coaching providers. More teachers in the control group reported receiving coaching from *Soporte Pedagógico* (a difference of 64 percentage points) and from PELA (a difference of 11 percentage points), and more teachers in the treatment group reported receiving coaching from *Amazonía Lee* (a difference of 82 percentage points). When we asked about the activities carried out in coaching, we found one statistically significant difference: 85 percent of the teachers in the treatment group reported receiving demonstration sessions, whereas 48 percent of control group teachers did (see Table IV.3). The demonstration sessions are a key aspect of *Amazonía Lee*'s teacher coaching approach, and as we learned through teacher focus groups, the component of coaching that was most useful to them.

Table IV.3. Except for demonstration sessions, in-class coaching activities are similar between groups

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Teachers who reported receipt of in-class coaching	95.0	97.0	-2.0	0.417
Teachers who reported receipt of in-class coaching from:				
<i>Amazonía Lee, Programa Enseñar Es Liderar</i> (PEEL)	85.0	3.0	82.0*	0.000
<i>Soporte Pedagógico</i> (MINEDU)	4.0	68.0	-64.0*	0.000
UGEL	31.0	35.0	-4.0	0.540
<i>Programa Estratégico Logros de Aprendizaje</i> (PELA)	11.0	22.0	-11.0*	0.021
Other	5.0	14.0	-9.0*	0.038
Activities carried out during in-class coaching:				
Demonstration sessions	85.0	48.0	37.0*	0.000
Design and preparation of materials	49.0	58.0	-9.0	0.209
Help to students at risk	49.0	57.0	-8.0	0.269
Identify students' learning needs	82.0	86.0	-4.0	0.452
Professional development coaching meetings (individual or in groups)	78.0	82.0	-4.0	0.495
Assessment of students' performance during class	82.0	78.9	3.1	0.593
Observation sessions	94.0	97.0	-3.0	0.259
Diagnose teaching practices	83.0	85.0	-2.0	0.697
Support in the implementation of the safety system	77.0	79.0	-2.0	0.717
Reinforcement for students	44.0	46.0	-2.0	0.770
Number of teachers	100	100		

Source: *Amazonía Lee* Teacher Survey—Final Follow-Up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

2. Literacy resources and materials in the classroom

One goal of *Amazonía Lee* was to improve the availability of materials that facilitate reading. In particular, the program aimed to provide classroom libraries for all participating schools. We used the teacher survey to inquire about the availability of classroom libraries and materials. In addition, we observed classrooms to assess whether they had materials available and the types of available materials.

***Amazonía Lee* teachers were more likely to have a classroom library; however, there were no differences in the types of books available.** We found a significant difference in having a classroom library: 62 percent of teachers in the treatment group reported having a classroom library, and only 2 percent in the control group did (see Appendix D, Table D.4). However, according to classroom observation data, an equipped reading area was observed in 37 percent of the classrooms, in both treatment and control groups. It is possible that the classroom libraries were not displayed in the treatment classrooms, and consequently the observers did not

code the classroom as having a reading area (something we heard was quite possible in our focus groups with teachers).

The observers also coded whether classrooms had materials that facilitated reading such as having more than 10 printed materials, more than 10 reading materials developed by the teacher, etc. (see Appendix D, Figure D.2 for a full list). We found one difference in favor of the treatment group regarding reading materials: In the treatment group, 41 percent of observed classrooms had more than 10 printed materials, whereas only 26 percent of control classrooms did. However, observers also coded the types of materials available in the classroom, and we did not find differences between groups. This is confirmed by teachers' perceptions (see Appendix D, Table D.5). In both treatment and control groups half of the teachers felt they lacked materials to support the teaching of reading.

We did not find statistically significant differences between groups on the type of materials teachers reported receiving from MINEDU. As expected, we found significant differences in materials provided by *Amazonía Lee*. The treatment group was more likely than the control group to report receiving the following materials from *Amazonía Lee*: workbooks, textbooks, guide books, posters and laminates, storybooks, and school supplies (see Table D.6 in Appendix D). These findings confirm that the treatment group received materials from *Amazonía Lee*, but because we did not ask for materials from all other possible sources, the classroom observation provides more reliable data to compare availability of materials in the classrooms.

B. Impacts on teaching activities and practices

In the previous section, we discussed the services that the study groups had received. Following the logic model of the program, we expect those services to first affect teacher practices and use of time and, through these changes in teacher practices, ultimately affect student literacy outcomes. In this section, we discuss the impacts of *Amazonía Lee* on teacher activities and use of time. Our discussion focuses on the findings from the teacher observation conducted at endline in 2016. This allows us to avoid issues with potential teacher reporting bias. However, some outcomes such as use of assessments or differentiated instruction could not be observed and thus are only available from teacher reported data. We, therefore, also discuss findings from teacher survey data collected in 2016, but when both sources of data are available, we emphasize classroom observational data over teacher reports.

1. Impacts on classroom practices and time use

***Amazonía Lee* teachers and control group teachers had similar literacy instructional practices in San Martín. The few differences found are consistent with the treatment group implementing strategies promoted by *Amazonía Lee*'s balanced approach.** Based on the classroom observations of the Reading and Communication class, observers reported whether the teacher had implemented any of 15 activities related to teaching reading during the observation period (see Table IV.4). We found only two statistically significant differences between groups of the 15 observed. In the treatment group, 46 percent of teachers were observed providing instruction on grammar, punctuation, or spelling, whereas only 26 percent of teachers did so in the control group, a statistically significant difference of 20 percentage points. This is consistent with our findings that treatment teachers were more likely to report teaching punctuation every day than control teachers, and with the fact that *Amazonía Lee* has a training module on writing

instruction. In the treatment group, 7 percent of teachers were observed explaining the use of different types of text, whereas only 1 percent in the control group did (a statistically significant difference of 6 percentage points). We also found that 20 percent of teachers in the treatment group and 31 percent of teachers in the control group were observed encouraging students to use context clues for reading comprehension. This difference of –11 percentage points is not statistically significant at the 5 percent level, but is significant at the 10 percent level.

Table IV.4. Teachers were observed performing activities related to teaching reading at similar rates in both groups, with two exceptions

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Phonological and phonemic awareness				
Encouraged students to recognize letters, identify the names of letter	9.0	8.0	1.0	0.810
Introduced or rehearsed vocabulary words and explained word meaning	33.0	29.0	4.0	0.549
Provided explicit instruction on phonological awareness	10.0	7.0	3.0	0.408
Prompted students to segment words into specific units	7.0	3.0	4.0	0.208
Provided instruction on grammar, punctuation, or spelling	46.0	26.0	20.0*	0.001
Provided instruction on word writing	78.0	74.0	4.0	0.519
Fluency				
Teacher engaged students in fluency-building activities	95.0	97.0	–2.0	0.482
Book reading				
Teacher used pre-reading or book-preview	66.0	72.0	–6.0	0.368
Teacher encouraged predictions	22.0	15.0	7.0	0.179
Teacher explored children interest in the story and facilitated connections with it	37.0	33.0	4.0	0.558
Teacher explain the use of different types of text	7.0	1.0	6.0*	0.033
Vocabulary and reading comprehension				
Encouraged students to use context clues for reading comprehension	20.0	31.0	–11.0	0.063
Discussed text structure	16.0	13.0	3.0	0.551
Discussed the characters in the text	41.0	41.0	0.0	1.000
Guided students in the use of reading comprehension strategies	92.0	94.0	–2.0	0.596
Number of classrooms	100	100		

Source: *Amazonía Lee* Classroom Observation—Final Follow-Up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Findings from the teacher surveys similarly showed few differences in literacy practices between the control and treatment groups, each of which was consistent with the emphases of *Amazonia Lee*. For example, more treatment group teachers than control group teachers reported practicing daily graphomotor skills, punctuation, and vocabulary development. First, the teacher survey asked teachers whether teachers performed 15 activities related to literacy and fluency.²² Most teachers in both groups reported using these activities (see Appendix D, Table D.7). Regarding how often these activities were performed, we found only two statistically significant differences in emergent literacy-related teaching activities performed every day (see Appendix D, Figure D.3). In the treatment group, 8 percent of teachers reported practicing how to develop graphomotor skills every day, and 2 percent of the control group reported doing it daily. Additionally, 36 percent of the treatment group teachers reported practicing punctuation every day, whereas 23 percent of the control group reported doing so. These differences seem to be consistent with some of the emphases of *Amazonia Lee*, given that the training emphasizes writing (punctuation marks, coherence of texts, and so on). Second, the survey asked teachers to report whether they used 10 activities related to reading comprehension.²³ We found no statistical differences between groups on the rates at which teachers reported using these activities. However, regarding how often the activities were conducted, we found one significant difference (see Appendix D, Figure D.4). In the treatment group, 30 percent of teachers reported developing vocabulary or learning new words every day, whereas only 15 percent of teachers in the control group did, a statistically significant difference of 15 percentage points. We also found that teachers in the treatment group were more likely to teach using graphic organizers of text two or three times a week (40 percent versus 28 percent, results not shown). This difference is not significant at the 5 percent level but is at the 10 percent level. The use of graphic organizers is a strategy emphasized by *Amazonia Lee* to improve reading comprehension.

***Amazonia Lee* has positive impacts on the reported duration of the Reading and Communication class, and it led to a different distribution of class time that is consistent with *Amazonia Lee*'s balanced approach.** We also asked teachers to report the typical duration of the Reading and Communication class. Teachers in the treatment group reported on average a three-hour class and the control group a two-and-a-half hour class, a statistically significant difference of almost a half hour.²⁴

Treatment teachers spent a greater proportion of class time on oral language and written text activities, whereas control teachers spent a greater proportion of class time on reading

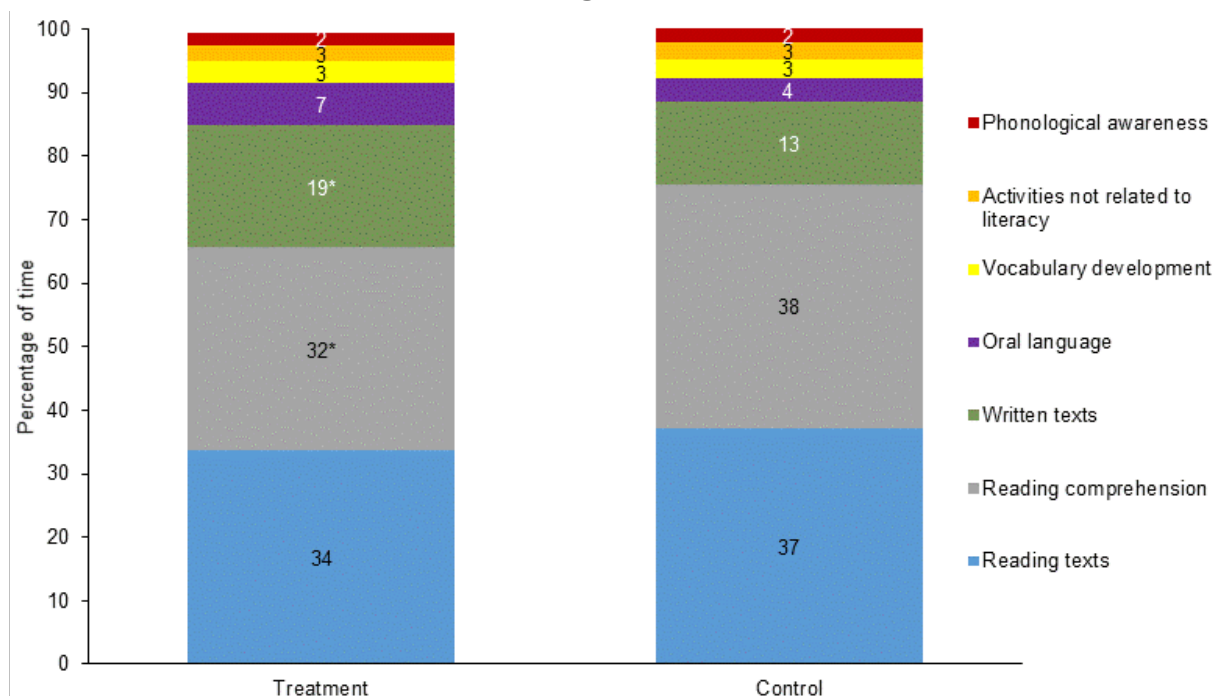
²² The 15 literacy and fluency activities were as follows: rhymes or games with the sounds of words, ask students to make predictions about the readings, practice to complete incomplete words, practice to complete incomplete sentences, quick reading exercise on familiar words, pre-reading exercises, demonstrate reading, silent reading, reading exercises in pairs or small groups, group reading exercises in chorus, practice to develop graphomotor skills, copy words or sentence, practice punctuation, dictation, and practice spelling rules or accents.

²³ The 10 reading comprehension activities were as follows: defining the purpose of the reading, conducting pre-reading exercises, asking students questions to activate previous knowledge, practicing predicting or anticipating what a text is about, asking students questions about the topic or characters, practicing summarizing a text, teaching students how to make a graphic organizer of the text, developing vocabulary or learning new words, reading activities for pleasure or entertainment, and reading activities to look for information.

²⁴ The survey asked for the typical duration of a Reading and Communication class, but we did not ask for how many classes per week the teachers performed.

comprehension activities. Both groups were observed in the Reading and Communication class for 38 minutes on average. We estimated the proportion of class time teachers spent in seven categories (six categories related to literacy and one not related to literacy) and assessed whether there were differences between groups.²⁵ We estimated that the treatment group spent 7 percent of class time on oral language activities, whereas the control group spent only 4 percent of the time on this activity, a statistically significant difference of 3 percentage points. See Figure IV.3. This is consistent with *Amazonía Lee*'s balanced approach that emphasizes oral language development in addition to writing and reading comprehension. We also estimated that treatment group teachers spent more time on written texts than control group teachers, 19 percent versus 13 percent, a statistically significant difference of 6 percentage points. This is also consistent with *Amazonía Lee*'s balanced approach which also focuses on writing skills. Finally, we estimated that treatment teachers spent 32 percent of class time on reading comprehension activities, whereas teachers in the control group spent 38 percent of their time, a statistically significant difference of 6 percentage points.

Figure IV.3. Treatment teachers spent more class time on oral language and written text and less time on reading comprehension activities



Source: *Amazonía Lee* Classroom Observation—Final Follow-Up 2016.

* Difference in group means is statistically significant at the 0.05 level.

²⁵ To estimate teachers' time use we used the following process. In the classroom observation, observers coded whether the teacher had engaged in activities in any of 7 predefined categories (see list in Figure IV.3) in a 5 minute observation period. The observers marked all the activities the teacher engaged in during the 5 minute period. There were on average 8 observation periods per classroom observation. First we counted the number of activities the teacher had engaged in the 5 minute period. Then, we divided 5 by the number of activities engaged in, this process assumes that the teacher spent equal time those activities. To estimate the total time spent in each activity per class, we added the time spent in that activity during all the observations periods. Finally, to obtain the percentage of time a teacher spent in each activity, we divided the total time spent in each activity by the total observation time for that classroom observation.

Looking beyond literacy instruction, *Amazonía Lee* teachers used similar general pedagogical practices and established a similar classroom climate as did control group teachers. As part of the classroom observation, we also looked at general pedagogical practices such as verifying that students understood instruction, verifying that students had relevant materials, providing feedback, etc. (See Appendix D, Table D.8). In both groups, during the class, teachers were observed having informal conversations or socializing with students on average five times, speaking respectfully to students on average five times, praising students sincerely on average four times, providing feedback to students on average three times, and offering individualized help to students a little over two times on average.

We found no impacts on teacher-reported activities related to classroom climate, with two exceptions. Teachers in both groups reported always using the following activities at similar rates: praise students, show affection or appreciation to students, display jobs that students have done, and remind students the rules of coexistence in the classroom. However, teachers in the treatment group reported using two activities at lower rates than control teachers: always giving students specific responsibilities in the classroom (56 percent versus 70 percent) and always giving opportunities to students to cooperate with their peers (62 percent versus 76 percent). Furthermore, 69 percent of treatment teachers reported always teaching students to speak positively about themselves, whereas 81 percent of control teachers did, a difference of 12 percentage points that is not statistically significant at the 5 percent level but is at the 10 percent level (see Appendix D, Table D.9).

Teachers in both groups spent just more than two-thirds of the time observed teaching the entire class; in most of the remaining time, they worked with individual students. Teachers spent little time working with small groups (results not presented).

***Amazonía Lee* teachers reported spending a similar amount of time per week in several activities as control group teachers, with one exception.** Teachers in the treatment group reported spending less time per week monitoring students during recess or free time than control group teachers did. Teachers in both groups reported spending similar amounts of time on 14 of the 15 general activities we asked about.²⁶ Teachers in the treatment group reported spending a little over one hour per week monitoring students during recess and free time, whereas control teachers reported spending under two hours, a statistically significant difference of less than half an hour. We also found that teachers in the treatment group reported spending about one more hour a week teaching reading than control teachers (although not significant at the 5 percent level, this is significant at the 10 percent level). It is possible that this is related to a previous finding where treatment teachers reported having a longer Reading and Communication class than control teachers. Furthermore, teachers in the treatment group reported spending almost one less hour per week on preparing students for national exams than control teachers, a statistically significant difference at the 10 percent level. This is consistent with findings from the focus

²⁶ The 15 activities were administrative tasks, preparing the class, meeting with principal or other teachers, meeting with students' parents, monitoring students during recess and free time, teaching reading techniques, teaching math, teaching social science, teaching science, teaching art, planning or managing or correcting exams or homework, providing support to students with low academic achievement, preparing students for national examinations, and participating in reading activities.

groups, where stakeholders reported that control schools gave priority to the national ECE exam (see Appendix D, Table D.10).

2. Impacts on differentiated instruction and support for underperforming students

Teachers in both groups reported similar use of differentiated instruction, with one exception. We found only one area of impact on the use of differentiated instruction strategies. In both groups, more than 90 percent of the teachers reported using differentiated instruction for students underperforming in reading. Teachers in both groups also reported using differentiated instruction strategies such as using assessments to identify students' needs, organizing students based on their learning needs, and planning for different levels of reading learning at similar rates. We identified only one statistically significant difference between the groups: Almost 3 percent more teachers in the treatment group reported planning teaching for the different levels of student reading learning in the last month than in the control group (see Appendix D, Table D.11).

Although teachers in both groups reported working with underperforming students at similar rates, control teachers were more likely to refer these students for additional support. Teachers in both groups reported that on average four students out the 25 students in a typical class were underperforming in developing reading skills. On average, 75 percent of teachers in both groups reported working with underperforming students. Both groups reported working with students underperforming in reading for almost four hours per week. However, although only 38 percent of the treatment group reported referring underperforming students in reading to academic reinforcement, 59 percent of the control group did—a statistically significant difference of 21 percentage points. This is consistent with the fact that *Soporte Pedagógico* is implemented in tandem with *Docentes Fortaleza*, a program that provides reinforcement for underperforming students after school hours (see Appendix D, Table D.12).

3. Impacts on the use of formative assessments to monitor students' learning

Teachers in both groups reported using formative assessments for various purposes at similar rates. Teachers in both groups reported using assessments for making initial diagnostics to plan for class and to organize learning groups at similar rates. They also reported using formative assessments to determine progress in literacy, providing feedback to students, and performing an end-of-year assessment for literacy. We also asked about the frequency of assessing alphabet knowledge, phonological awareness, oral comprehension, and other skills related to early reading skills, and we found that teachers in both groups reported similar frequency of assessments. Table D.13 in Appendix D provides information on the reported use of formative assessments, and Figure D.5 reports teachers' perceptions on students' performance.

C. Impacts on students' reading skills

To assess students' reading skills, we looked at impacts on key foundational literacy skills as well as on reading comprehension. In the first subtest, we assessed oral language comprehension (Table IV.5). This provides information on whether students can comprehend a short story that is read to them. To assess emergent reading skills, the second subtest focused on reading familiar words. The third subtest measured decoding skills. In this subtest, students were asked to read a list of invented words and were given one minute. This subtest allowed us to assess whether the student had learned the correspondence between letters and sounds. Furthermore, being able to

read fluently is a necessary for reading comprehension. Therefore, the fourth subtest assessed how many words per minute the students were able to read correctly. The last subtest assessed reading comprehension skills.

Students in both groups had similar oral comprehension skills. The assessor read a four-line passage to the student and asked the student six oral comprehension questions about the story. On average, students in both groups responded to roughly three questions correctly.

Students in both groups had similar scores in reading familiar words, decoding, and fluency. Students in both groups correctly read on average 8 out of 10 familiar words presented. Furthermore, students in both groups correctly read on average 20 pseudo-words in one minute. To assess fluency, students were given a short story with 77 familiar words and were asked to read the story for one minute. Students in both groups had similar fluency results; on average students read almost 44 words correctly. We should note that this finding is not robust to other regression model specifications: model specifications that do not include controls for school infrastructure show significant negative results for Amazonia Lee as compared to the control group that was largely receiving *Soporte Pedagógico*; essentially suggesting that students do better with *Soporte*. However, because school characteristics are often related to factors that affect student achievement, such as community involvement or principal's effectiveness, our preferred model includes all school characteristics.

Students in both groups were able to answer correctly only half of the reading comprehension questions. Students were given a simple seven-line story to read in silence, and then the assessor asked six questions related to the story. Students in both groups were able to answer three of the six questions correctly.

Table IV.5. Students reading outcomes were similar between groups

	Treatment adjusted mean	Control adjusted mean	Difference	p-value	Effect sizes
Number of correct oral comprehension responses	2.9	2.9	0.1	0.267	0.05
Number of familiar words read correctly	8.1	8.2	-0.1	0.394	-0.03
Number of pseudo-words read correctly in one minute, decoding	19.6	20.2	-0.6	0.154	-0.05
Number of words read correctly in one minute, fluency	42.8	44.2	-1.3	0.117	-0.05
Number of correct answers to reading comprehension questions	3.0	3.0	0.0	0.768	0.01
Number of students	824	822			

Source: *Amazonia Lee* Child Assessment—Final Follow-Up 2016.

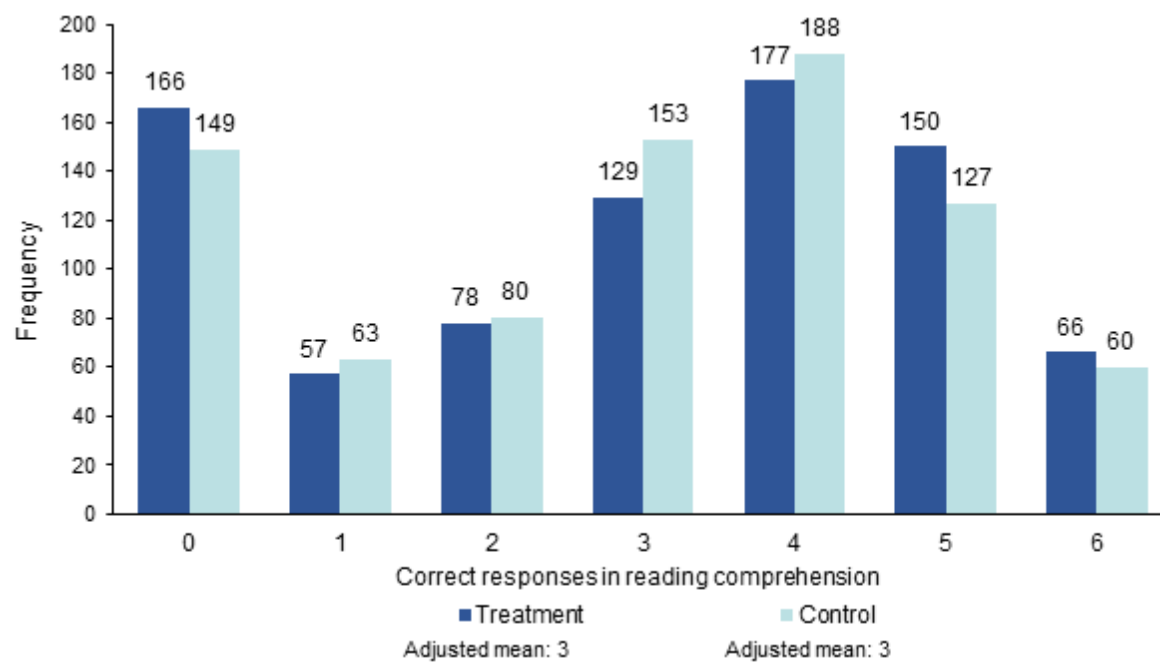
Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables). We also include baseline score of students and some students' characteristics (2015) and baseline infrastructure characteristics of schools (2014). Effect sizes were calculated before rounding the adjusted difference.

* Difference in group means is statistically significant at the 0.05 level.

To provide a more comprehensive perspective of student achievement in reading comprehension, Figure IV.4 presents the number of students per study group that answered correctly from zero to six questions. It is interesting to note that the greatest number of students

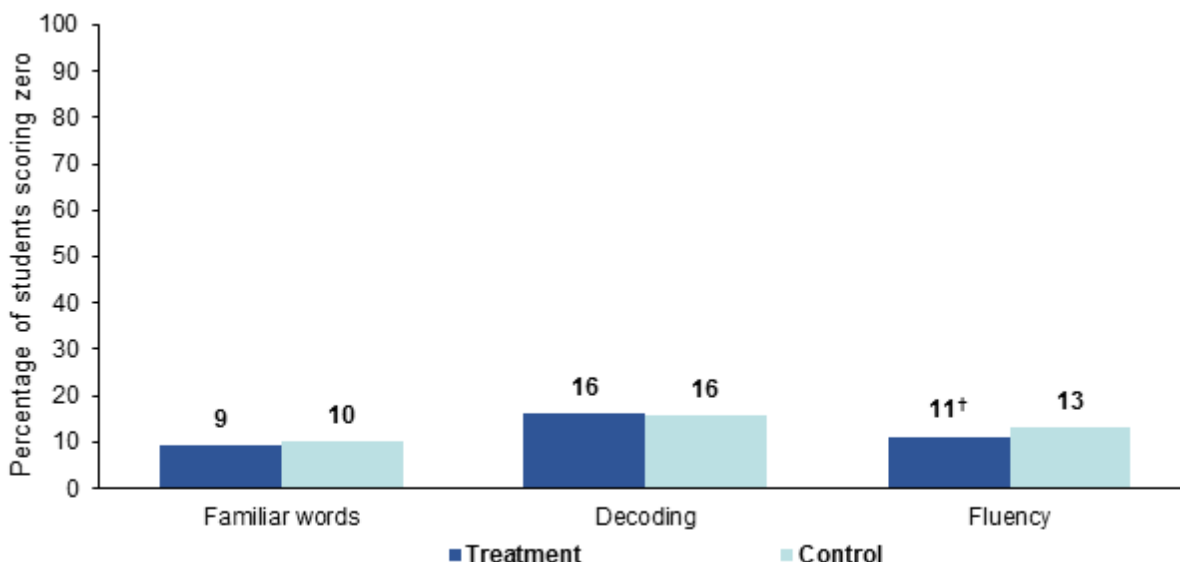
correctly answered four questions, but there were still many students – about one in five—who were not able to answer even one reading comprehension question correctly (bars in the zero correct responses). A significant percentage of the students (87 percent) who were scored zero on the reading comprehension test were students that were unable to demonstrate basic reading fluency skills and therefore were not tested in comprehension.

Figure IV.4. Histogram of correct answers in reading comprehension



Source: *Amazonía Lee* Child Assessment—Final Follow-Up 2016.

Fewer students in the treatment group got a score of zero on the fluency test than in the control group. In both groups, around 9 percent had a score of zero on the reading familiar words test, and 16 percent of students in both groups had a score of zero on the decoding test. On the fluency test, 2 percentage points fewer students in *Amazonía Lee* schools had a score of zero than those in the control schools. This difference is statistically significant at the 10 percent level and suggests that more students in *Amazonía Lee* schools were able to read at least 12 words correctly in one minute than students in the control group (Figure IV.5).

Figure IV.5. Fewer students in the treatment group got a zero score in the fluency test

Source: *Amazonía Lee* Child Assessment—Final Follow-Up 2016.

* Difference in group means is statistically significant at the 0.05 level.

† Difference in group means is statistically significant at the 0.10 level.

Gender analysis

The results of the gender subgroup analysis are similar to what we found in the full sample. We estimated impacts for girls and for boys, and we found no statistically significant difference in reading outcomes between treatment and control groups by gender (see Table IV.6). Furthermore, we found no differences in the magnitude of the impacts for boys compared to the impacts for girls. As shown in Appendix D, Table D.14, we found no statistically significant difference in the percentage of girls and boys who got a score of zero on reading subtests: familiar words, decoding, and fluency.

Table IV.6. Students reading outcomes were similar between groups

	Impact for girls	p-value	Impact for boys	p-value	Difference in impacts by gender (p- value)
Number of correct oral comprehension responses	-0.0	0.628	0.2†	0.061	0.116
Number of familiar words read correctly	-0.0	0.815	-0.2	0.372	0.658
Number of pseudo-words read correctly in one minute (decoding)	0.0	0.946	-1.3†	0.057	0.188
Number of words read correctly in one minute (fluency)	-0.7	0.597	-2.0	0.109	0.478
Number of correct answers in reading comprehension	0.0	0.944	0.0	0.718	0.846
Number of students	815		831		

Source: *Amazonía Lee* Child Assessment—Final Follow-Up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables). We also include baseline score of students and some student's characteristics (2015) and baseline infrastructure characteristics of schools (2014). Effect sizes were calculated before rounding the adjusted difference.

* Difference in group means is statistically significant at the 0.05 level.

† Difference in group means is statistically significant at the 0.10 level

V. CONCLUSIONS, LESSONS LEARNED, AND RECOMMENDATIONS

This study began with five questions. Here we list conclusions corresponding to each question and then discuss lessons learned and recommendations.

A. Conclusions

1. What was the impact of *Amazonía Lee* programs on early-grade reading and other intermediate outcomes relative to other prevailing practices and programs?

The impacts of *Amazonía Lee* programs depended critically on the context, meaning the “other prevailing practices and programs.” The study examined the programs in two departments of Peru, each representing a different point of comparison. In Ucayali, control schools received the usual services provided by the Ministry of Education. In the other department, San Martín, most control schools received a program known as *Soporte Pedagógico*, a Ministry initiative that offered teacher training and coaching, pedagogical materials, and remedial tutoring.

When compared to prevailing practice in a region where no new intervention was introduced, *Amazonía Lee* achieved substantive effects of magnitudes similar to those found in other rigorous evaluations. In Ucayali, we found that students in *Amazonía Lee* had better reading outcomes than the control group. The magnitude of the impacts ranged from 0.15 standard deviations to 0.27, which are within the range of what other studies have found.

When compared to a control group that received teacher reading instruction but through a different program than *Amazonía Lee*, students in both groups had similar reading outcomes. In San Martín, most teachers in the control group received training on reading instruction from *Soporte Pedagógico*. We found no significant differences in student outcomes, despite *Amazonía Lee* teachers’ report of longer classes and evidence of teachers implementing key components of *Amazonía Lee*’s balanced approach to reading. This finding did not support USAID’s original thinking, that *Amazonía Lee*’s balanced approach might lead to different student reading outcomes than *Soporte Pedagógico*. It is possible that the availability of *Soporte Pedagógico*’s tutors for underperforming students in the control group and more frequent coaching sessions for control teachers offset potential differences in student reading outcomes resulting from longer classes and instructional practices reported by *Amazonía Lee* teachers.

2. How were *Amazonía Lee* intervention components, activities, and processes implemented in Ucayali and San Martín?

Teachers in both departments seem to be implementing some aspects of *Amazonía Lee*’s balanced approach. In Ucayali, *Amazonía Lee* had positive impacts on explaining the use of different kinds of text, reading daily for pleasure, and teaching the use of graphic organizers—practices consistent with *Amazonía Lee*’s balanced approach. Furthermore, *Amazonía Lee* teachers reported spending more time supporting underperforming students and promoting reading in the community—consistent with the community engagement component of the program.

In San Martín, *Amazonía Lee* had positive impacts on practicing graphomotor skills; teaching punctuation, grammar, and spelling; explaining the use of different texts; and developing vocabulary—mostly practices consistent with *Amazonía Lee*'s balanced approach. Furthermore, we found that *Amazonía Lee* teachers reported teaching longer reading and communication classes and distributing their time differently than control teachers.

However, in both Ucayali and San Martín, some aspects of the program, such as differentiated instruction and use of formative assessments, did not generate impacts.

***Amazonía Lee* services reached the treatment groups in both departments.** The primary interventions of the program concentrated on teacher training, in-class coaching, and provision of materials during the study period (2015 and 2016). Our findings confirm that training and in-class coaching were delivered to the treatment groups in Ucayali and San Martín. We compared these implementation results to the control group to estimate impacts. Although we found some impacts on materials provision in Ucayali, we found no impacts in San Martín. Furthermore, findings from the implementation study indicated that *Amazonía Lee* materials were delivered with delays, later in program implementation.

3. What services were provided to control schools in Ucayali and San Martín?

In San Martín, the control group also received reading support services but from a different program than the treatment group. In San Martín, the control group had similar rates of professional development training and in-class coaching as the treatment group. Almost three fourths of schools in the control group received support from *Soporte Pedagógico*. Furthermore, *Soporte Pedagógico* a program that commanded additional resources, such as tutors for underachieving students, who may have helped underperforming students in the control group. *Soporte Pedagógico* also had monthly coaching sessions, whereas in San Martín, *Amazonía Lee* had bimonthly coaching sessions. In Ucayali, less than half of the teachers in the control group reported participating in reading instruction professional development activities the past year—those that did, typically received training from PELA, the UGEL, or sought training independently from other institutions. Survey and qualitative data suggested that teachers in control schools had far fewer opportunities for in-service training compared to *Amazonía Lee* teachers in that department.

4. How were *Amazonía Lee* intervention components, activities, and impacts perceived by principals and teachers?

Teachers perceived *Amazonía Lee*'s balanced approach to be effective. Teachers perceived that their students were gaining reading skills sooner in the school year with *Amazonía Lee*'s balanced approach, compared to previous years. Also, they observed increased student engagement and more active student participation in class. Teachers felt *Amazonía Lee*'s training and coaching helped them improve their reading instruction skills, their ability to create an enriched literacy environment in the classroom and their classroom management skills.

Teachers perceived that *Amazonía Lee* materials had some advantages over other pedagogical materials teachers had previously used. Teachers felt *Amazonía Lee* teacher guides and student workbooks were well targeted for the skill level of their students and emphasized culturally relevant themes and activities of the Amazon region. They also noted that

materials were well structured and helpful for instruction following *Amazonía Lee*'s balanced approach.

School principals perceived improvements in teachers' reading instruction practices and teachers' interest in continuous improvement. Such interest was expressed through participation in teacher study groups, coaching sessions, and use of culturally relevant pedagogical materials. Most teachers showed ownership of *Amazonía Lee*'s balanced approach to reading instruction, were highly motivated by the program's training activities, and seemed committed to continue improving their pedagogical skills. School principals also noticed improvement in primary teachers' reading instruction skills, and in their ability to apply pedagogical strategies that develop students' reading comprehension skills.

5. What were the facilitators and barriers to the implementation of *Amazonía Lee*?

The main barriers to the implementation of *Amazonía Lee* were teacher qualifications and staff turnover. Specifically, there were not enough qualified teachers that met the criteria to become *Amazonía Lee* trainers, and there was high turnover of regional technical team members throughout the project. Weaknesses in teachers' pre-service training, and *Soporte Pedagógico*'s concurrent demand for qualified personnel in San Martín contributed to the scarcity of qualified teacher trainers. Further, there was high turnover of trained teachers within across schools during the implementation of *Amazonía Lee* which limited the continuity of trained teachers in the focalized grades of treatment schools.

***Amazonía Lee*'s pedagogical materials did not reach teachers and students on time.** Pedagogical materials for reading instruction and reading for pleasure, which are central to the *Amazonía Lee* intervention, were not delivered at the beginning of the school year in 2015 or 2016. In both regions, treatment schools received teacher guides, student workbooks and classroom libraries with significant delays. These delays resulted in a considerable lag between the beginning of training activities and the availability of pedagogical materials.

Continuous improvement efforts of the UPOCH regional technical team, and buy-in from the regional authorities and school directors were critical for the successful implementation of *Amazonía Lee* activities. UPOCH routinely gathered feedback forms after from teachers after training workshops to identify strengths and weaknesses and improve the training workshops. In addition, UPOCH regional coordinators observed training workshops to identify areas for improvement and strengthen the skills of teachers' trainers. These continuous improvement efforts from UPOCH were likely to contribute to the quality of the teacher training workshops and activities.

Program buy-in from the regional authorities and school directors was another critical facilitator for the implementation of *Amazonía Lee*. The commitment of regional authorities and school directors to maintain trained teachers in treatment schools and the target primary grades enabled those teachers to deepen their knowledge and reading instruction skills over time.

B. Lessons learned

Supporting teachers explicitly in best teaching practices in reading is a promising approach to improving early grade reading. Teachers need and value comprehensive and sustained in-service training and coaching for literacy instruction in the early grades. In the one department where there were no significant differences at endline, teachers in both groups had been receiving such support. The differences in the nature of the support did not translate into impacts. In contrast, in the department where such support was lacking in the control group, the impacts were significant and meaningful. This suggests that teachers need to deepen the basic knowledge about literacy acquisition and instruction they receive in pre-service teacher education programs. Comprehensive professional development programs for in-service teachers are crucial to fill in gaps in teachers' knowledge about reading instruction and to help them improve their reading instruction practices. Group-based learning in the form of teacher workshops and teacher study groups, as well as individualized teacher coaching, are useful for teachers to improve their reading instruction, increase use of formative assessments, incorporate more and more varied reading materials and activities, and manage their classrooms effectively.

High teacher turnover hinders the potential reach of teacher training efforts. The program logic assumes that teachers' ability to implement reading instruction practices effectively grows over time. The commitment of school directors and regional authorities to retaining trained teachers in the early grades is important for the gradual and continuous improvement of teachers' reading instruction over time.

The local context and regional conditions matter for the implementation of impact evaluations and type of questions that can be answered with rigor. Although our original evaluation design planned comparisons with control groups that were not exposed to similar interventions, this was only possible in one of the regions. In Ucayali, we were able to assess the impact of *Amazonía Lee* with a control group that was not exposed to program services similar to those of *Amazonía Lee*. In San Martín, *Soporte Pedagógico* was implemented at the same time the *Amazonía Lee* evaluation started. The Ministry of Education agreed to keep the implementation of *Soporte Pedagógico* limited to control schools to preserve the evaluation in San Martín; however, a prevalent practice control group was no longer available in San Martín as it was in Ucayali. In Ucayali, the local context allowed us to assess the impacts of *Amazonía Lee* as compared to prevalent practice. In San Martín, we assessed the impacts of *Amazonía Lee* as compared to a competing program.

C. Recommendations

Policymakers in Peru should continue to support *Amazonía Lee* in Ucayali. Based on the positive results found there, and given that the department has gained experience and capacity to implement *Amazonía Lee*, continuing the intervention there may be a way to sustain the progress made on early grade reading.

Policymakers in Peru should continue to support early grade reading in San Martín, but should examine the costs of different approaches before selecting the one to continue. Students in schools offered *Amazonía Lee* had similar reading performance as those in schools that offered other programs, primarily *Soporte Pedagógico*. Since *Soporte Pedagógico* has only been implemented in urban schools and *Amazonía Lee* was implemented in urban, peri-urban,

and rural schools, continuing the intervention may be a way to sustain student achievement in all types of schools.

Funders interested in learning about the potential impacts of educational interventions throughout Latin America and the rest of the world should consider carefully the regional context and assess the local programmatic needs before investing in designing and implementing new programs. This would help to understand whether any proposed intervention has not only a sound internal logic but represents a compelling innovation relative to other approaches that have been tried or would be tried in its absence.

Similarly, sponsors of program evaluation and research on education interventions like the ones studied here should carefully consider the context before investing in an evaluation. A deep understanding of the often-neglected “prevailing practices” is critical for setting up a contrast that will illuminate important tradeoffs and generate lessons that are applicable well beyond the local context. The learning questions that can be answered with impact evaluations, to a great extent, depend on the contrasts chosen for comparison. USAID and Mathematica initially planned to assess the impact of *Amazonía Lee* as compared to the prevailing pedagogical approaches to teaching early grade reading, but this was only possible in Ucayali where the contrast allowed for it. In the case of San Martín, although we are limited to conclusions about the benefits of *Amazonía Lee* in relation to *Soporte Pedagógico*, the evaluation also offered valuable information about the potential reach of *Amazonía Lee*, an educational program implemented with regional leadership and oversight, as compared to *Soporte* which is a program implemented with central leadership. Donors and evaluators should be aware that there are often changes on the ground that provide cause for concern or could potentially threaten the original evaluation design, as in the case of what the prevailing practice ended up being for San Martín. However, such changes can provide an opportunity to consider how to think about the changes and incorporate them so that the evaluation can still generate policy relevant learning, as happened in this case.

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

REFERENCES

- Banco Mundial “Por una educación de calidad para el Perú. Estándares, rendición de cuentas y fortalecimiento de capacidades.” 2006. Available at <http://mapeal.cippec.org/wp-content/uploads/2014/05/Por-una-educaci%C3%B3n-de-calidad-para-el-Per%C3%BA-estandares-rendici%C3%B3n-de-cuentas-y-fortalecimient-de-las-capacidades.pdf>.
- Campuzano, L., Glazerman, S., Murray, N., Liuzzi, S. and Fernández, C. “Evaluation of Amazonía Lee reading intervention in Peru: First report. Report prepared for USAID. Washington, DC: Mathematica Policy Research. November 2016.
- Friedlander, E., y C. Goldenberg, Editores. “Literacy Boost in Rwanda: Impact Evaluation of a 2-year Randomized Control Trial.” Stanford, CA: Stanford University, 2016.
- Gove, A. and P. Cvelich. Early Reading: Igniting Education for All’ A report by the Early Grade Learning Community of Practice. Revised Edition. Research Triangle Park, NC, USA: Research Triangle Institute. 2011.
- Kim, Y.-S. G., Boyle, H. N., Zuilkowski, S. S., & Nakamura, P. (2016). Landscape Report on Early Grade Literacy. Washington, D.C.: USAID. Available at <https://globalreadingnetwork.net/publications-and-research/landscape-report-early-grade-literacy-skills>.
- Ministerio de Educación. “Resultados por Región/DRE de la Evaluación Censal Estudiantil; 2013,” 2016. Available at <http://umc.minedu.gob.pe/resultadosece2016/>.
- Snilstveit, B, Stevenson, J, Menon, R, Phillips, D, Gallagher, E, Geleen, M, Jobse, H, Schmidt, T and Jimenez, E. 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.
- United Nations Educational, Scientific and Cultural Organization. Latin America and the Caribbean. Education for all 2015 Regional Review. Available at <http://unesdoc.unesco.org/images/0023/002327/232701e.pdf>.
- USAID, Global Reading Network. Landscape Report on Early Grade Literacy. Kim, Y.-S. G., H. N. Boyle, S. S. Zuikowski, y P. Nakamura, Autores. Washington, D.C.: USAID, 2016.
- USAID and AIR, Programa de Capacidades LAC Reads. “Systematic Review of the Evidence on EGR from LAC.” Presentación por Rebecca Stone, Mayo 9, 2017.

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

APPENDIX A

EVALUATION SAMPLE AND IMPACT ESTIMATION

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

1. Evaluation sample

To evaluate the impacts of *Amazonía Lee*, we selected the evaluation sample schools in each region eligible to receive the intervention. In Ucayali, out of 76 eligible schools, 70 schools were part of the evaluation sample, 35 schools were randomly assigned to the treatment group and 35 to the control group. In San Martín, out of 314 eligible schools, 200 schools were part of the evaluation sample, 100 were randomly assigned to the treatment group and 100 to the control group.

Student sample and attrition rates

Table A.1 presents the study sample of students selected for baseline data collection, the number of students assessed at baseline, and the number of students assessed at follow-up, by study group and region. Because the study focus on a longitudinal sample of students, we followed all the students that were eligible for baseline collection independently of whether they had been assessed at baseline or not. Our analysis sample includes all the students who were assessed at follow-up given that we imputed baseline data for students who were assessed at endline but had missing baseline information. Attrition is, therefore, measured as the percent of students who were eligible for baseline data collection but did not have follow-up data and hence were not included in the analysis sample.

The number of children in the Ucayali sample that were eligible for baseline data collection in 2015 was 796 (416 in the treatment group and 380 in the control group). As we explained before, only students that had dropped out of school at the time of baseline data collection or students with severe disabilities were ineligible for baseline data collection. Among the students eligible for data collection, 740 (380 in treatment group and 360 in the control group) completed a reading skills assessment at follow-up in 2016. This reflects an attrition rate of 8.7 percent in the treatment group and 5.3 in the control group.

In San Martín, the number of children eligible for baseline data collection in the 2015 was 1,738 (870 in the treatment group and 868 in the control group). For the follow-up, the number of children assessed was 1,646 (824 and 822 students in the treatment and control group, respectively). The attrition rate in the two groups is 5.3 percent.

Attrition levels are low and within acceptable ranges for a randomized control trial, according to What Works Clearinghouse Standards.²⁷ We also tested for baseline equivalence of the analysis sample and found groups are balanced in most characteristics; these results are presented in Tables C.1 and C.2 for Ucayali and D.1 and D.3 for San Martín. In addition, our baseline report (Campuzano et al. 2016) presented baseline comparisons for the sample of students that was assessed at baseline.

²⁷ https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_standards_handbook_v4_draft.pdf.

Table A.1. Student sample and attrition rates

Region/ study group	Study eligible sample: Students eligible for data collection ¹	Students assessed at baseline	Analysis sample: Students assessed at follow-up ²	Attrition rate ³
Ucayali				
Treatment	416	350	380	8.7
Control	380	328	360	5.3
Total	796	678	740	7.0
San Martín				
Treatment	870	789	824	5.3
Control	868	790	822	5.3
Total	1738	1579	1646	5.3

Source: *Amazonía Lee* Emergent Literacy Skills Assessment—Baseline 2015, and Final Follow-up 2016.

¹ In each study school, the data collection team, randomly selected on average 10 students (8 in San Martín) enrolled in first grade at the time of the visit and who did not have disabilities for baseline data collection. The sample of eligible students for data collection includes students who were not present on the day of the assessment, students assessed, and students that refused to be assessed.

² For the follow-up we tracked all the students eligible for data collection independently of whether they had been assessed at baseline or not.

³ The attrition rate was estimated as the difference between the number of students eligible for data collection and the number of students assessed at follow-up, divided by the number of number of eligible students for data collection.

2. Impact estimation

Amazonía Lee impacts on students reading outcomes were assessed using longitudinal data from the 2015 student cohort. Endline data was collected at the end of the 2016 school year when most of the cohort's students were in 2nd grade and had been exposed to the intervention for almost two years.²⁸ To estimate the impacts on student reading scores, we conducted regression analyses that accounted for initial differences in student characteristics in 2015, including reading scores, and initial differences in school characteristics such as infrastructure and school average reading outcomes of the 2014 cohort as well as for the stratification used in random assignment. The regression model used to estimate impacts is the following:

$$Y_{is} = \alpha + \beta X_{is} + \gamma z_s + \lambda T_s + \varepsilon_{is}$$

where Y_{is} is the outcome of interest (such as the score on the reading test) for student i in school s at follow-up. The vector X_{is} represents the baseline characteristics of student i in school s , for example age, gender and baseline reading measures y (the list of student characteristics used in the estimations is presented in Table A.2). The vector z_s includes school baseline characteristics for example infrastructure information and also indicator variables for each stratum (matched school pairs) used in the random assignment design (the list of school variables used in the estimations is presented in Table A.2). The variable T_s is an indicator equal to one for students in schools assigned to the treatment group, and zero for those in the control group. ε_{is} is a random

²⁸ Data collection was fielded from September 21 to November 16 for San Martín and in October for Ucayali.

error term for student i in school s . The parameter λ is our main interest—the impact of *Amazonía Lee* on the student outcome. The regression is clustered at the school level to account for students clustered in schools.

The impact estimation for teacher and classroom observation outcomes does not include controls for baseline measures but does account for stratification. The regression model for teacher outcome is the following:

$$Y_{js} = \alpha + \gamma z_s + \lambda T_s + \varepsilon_{js}$$

where Y_{js} is the outcome of teacher j in school s , and the vector z_s includes indicator variables for each stratum used in the random assignment design (matched pairs). The variable T_s is an indicator equal to one for teachers in schools assigned to the treatment group, and zero for teachers in schools assigned to the control group. ε_{js} is a random error term for teacher j in school s . The parameter λ is our main interest—the impact of *Amazonía Lee* on the outcome.

We also examined the program effects for subgroups defined by gender to explore whether the impacts for girls and boys differed. The regression model used to estimate impacts for subgroups is the following:

$$Y_{is} = \alpha + b\theta_{is} + \beta X_{is} + \gamma z_s + \lambda T_s + \mu\theta_{is} * T_s + \varepsilon_{is}$$

where Y_{is} is the outcome of interest (such as the score on the reading test) for student i in school s at follow-up. The variable θ_{is} is an indicator that takes the value of one for female students and zero for male students. The term $\theta_{is} * T_s$ is the interaction of the gender variable and the treatment indicator. The other variables are the same as in the model explained above. The parameters of interest for the subgroup analysis are λ —the impact of *Amazonía Lee* on males and $\lambda + \mu$ the impact of *Amazonía Lee* on females. The regression is clustered at the school level to account for students clustered in schools.

Table A.2 below lists the control variables included in the regression models we used to estimate program impacts on the reading skills of the children in the study in 2016, when they should have been in 2nd grade. The table also indicates whether the variable is dichotomous or continuous, and the source of the variable. All the variables in this list are unlikely to be affected by the intervention, student level covariates were measured at baseline when the students were in first grade in 2015 and school level covariates were measured in 2014.

Table A.2. Baseline characteristics included in the regressions for students reading outcomes

Level	Variables	Type	Source
Student	• Child age at baseline assessment (in months) • Child gender • Child was overage at baseline • Child repeated first grade at baseline^{1\}	Continuous Binary	<i>ALEE</i> Child assessment, Baseline 2015
Student	• Baseline letter identification score • Baseline emergent reading score • Baseline emergent writing score • Baseline phonemic awareness score • Baseline number of pseudo-words read in one minute • Baseline passage comprehension score	Continuous	<i>ALEE</i> Emergent Literacy Skills Assessment, Baseline 2015
School	• Baseline letter identification (average score per school) • Baseline emergent reading (average score per school) • Baseline emergent writing (average score per school) • Baseline phonemic awareness (average score per school) • Baseline number of pseudo-words read in one minute (average number of pseudo-words per school) • Baseline passage comprehension (average score per school)	Continuous	<i>ALEE Emergent Literacy Skills Assessment, Baseline 2014</i>
School	• School has multigrade classrooms • School has library or resource room • School has computers for students • School has computers for teachers • School has internet connectivity • School has phone service • School has piped water supply • School has unpiped water supply • School has potable drinking water • School has electricity • School has permanent building • School has painted interior walls • School has painted exterior walls • School has classrooms have ceilings • School has perimeter fence • School has working restroom facilities for children • School has latrines or urinals	Binary	<i>ALEE</i> School Infrastructure Observation, Baseline 2014

Level	Variables	Type	Source
School	• School has functional toilets • School has separate facilities for boys and girls • School has hand-washing facilities • School has working waste disposal • School has plumbing waste disposal • School has septic or similar disposal • School has other waste disposal • School has kitchen • School has vegetable garden • School has teacher's lounge • School has outdoors recreational space • School has gym or sports facilities • School has infirmary	Binary	ALEE School Infrastructure Observation, Baseline 2014
School	• Number of physical hazards observed in school • Number of health hazards observed in school	Continuous	ALEE School Infrastructure Observation, Baseline 2014
School	• At baseline, school participated in the One Laptop per Child program • At baseline, school participated in the National Strategic Learning Achievement Program (PELA) • At baseline, school participated in the National Rural Education Networks Program (REDES) • At baseline, school received education materials from the Ministry of Education • At baseline, school participated in the National School Nutrition Program (Qali Warma)	Binary	ALEE School Infrastructure Observation, Baseline 2014

¹⁾ Repetition of first grade at baseline was included only in San Martin analysis because we find initial differences at baseline.

Imputations. Some of the baseline variables listed in Table A.2, in particular the baseline reading student scores, had missing values for some students. In order to keep all the observations with follow-up data in the analysis sample, we imputed the missing baseline values. For emergent reading baseline scores, we imputed missing values through the Stata command “impute.” To impute subtest scores of the baseline reading test, we used that variable as the dependent variable and baseline age and baseline age squared as the independent variables. We also imputed the missing values of school-level variables (school characteristics and resources) at baseline, with the mean value for all the schools with baseline data for that variable. We included an indicator variable in the regressions described above to account for imputed values. Table A.3 presents the number of students in the analysis sample (students with endline assessment), the number of students in the analysis sample who also have complete baseline information, and the number of students in the analysis for which we imputed at least one baseline covariate.

Table A.3. Analysis and imputed sample

Region/ study group	Analysis sample: Students assessed at follow-up	Students assessed at follow- up with non-missing baseline covariates	Imputed sample: Students assessed at follow- up with at least one missing baseline covariate
Ucayali			
Treatment	380	315	65
Control	360	314	46
San Martín			
Treatment	824	706	118
Control	822	640	182

By imputing baseline data, we were able to include in the analysis sample students who had follow-up data but who had been absent during the two school baseline data collection visits or who had refused to take the assessment. As a robustness check, we also estimated impacts on the sample that had endline assessment and complete baseline data, so no imputations were necessary. Impact results for this sample are presented in Table C.16 for Ucayali and Table D.15 for San Martín. The sample of students with both baseline and endline data, however, are more likely to have lower absenteeism than the analysis sample. We also compared baseline data for the non-imputed sample, these results are presented in Tables C.17 and D.18.

Student mobility and dropout. In Peru, most students attend the school that is nearest to their home. Although, students can attend any public school if there is space available. As described in the main text, at endline, we attempted to follow students eligible at baseline regardless of whether they were still attending the same school that they had attended at baseline. We also assessed students in their homes when they had been absent from school during the data collection visits or had dropped out from school. Consistent with an intent-to-treat approach, the treatment status of the students is the one they had at baseline regardless of whether they transferred school or dropped out. Below we provide information on the number of students who had transferred out of their schools or had dropped out from school at endline. Note that some of the students who transferred to another school could be attending another treatment school. But we do not have information on what school they transferred to, so we cannot estimate the percentage of transfers who was still exposed to the treatment by endline data collection.

Table A.4. Transfers and dropouts in student analysis sample

Region/ study group	Analysis sample: Students assessed at follow-up	Transfers: Students assessed attending a different school from baseline	Dropouts: Students assessed not attending school	Percentage of students assessed attending a different school or not attending school
Ucayali				
Treatment	380	34	6	10.5%
Control	360	28	5	9.2%
Total	740	62	11	9.9%
San Martin				
Treatment	824	65	4	8.4%
Control	822	52	3	6.7%
Total	1,646	117	7	7.5%

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

APPENDIX B

IMPLEMENTATION

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

Structure of the regional implementation teams in Ucayali and San Martín

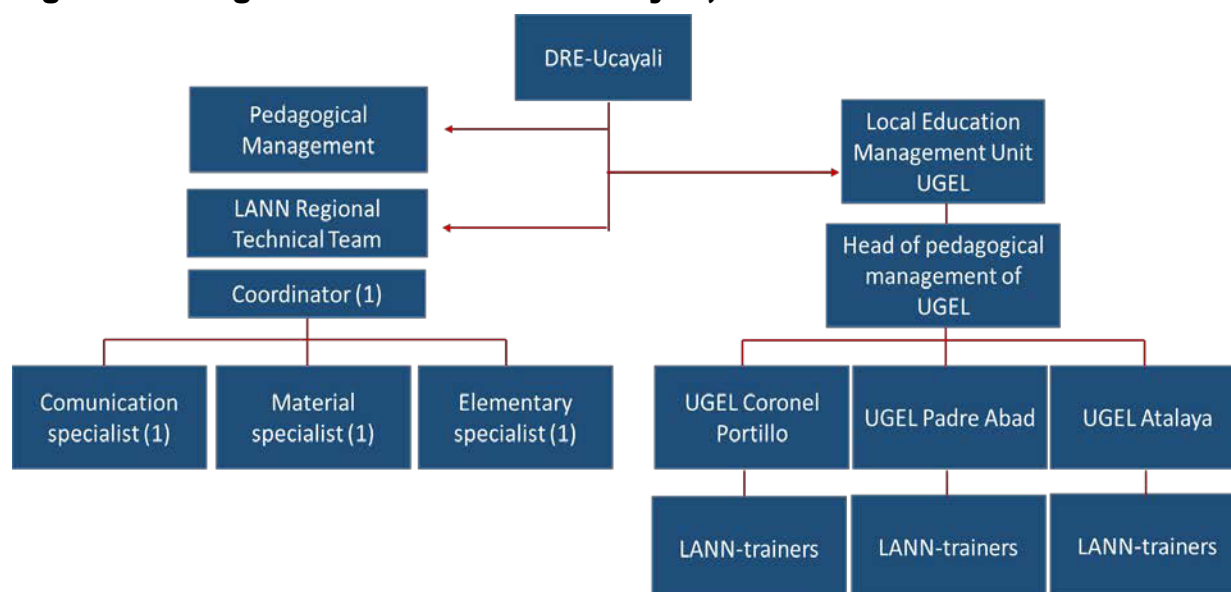
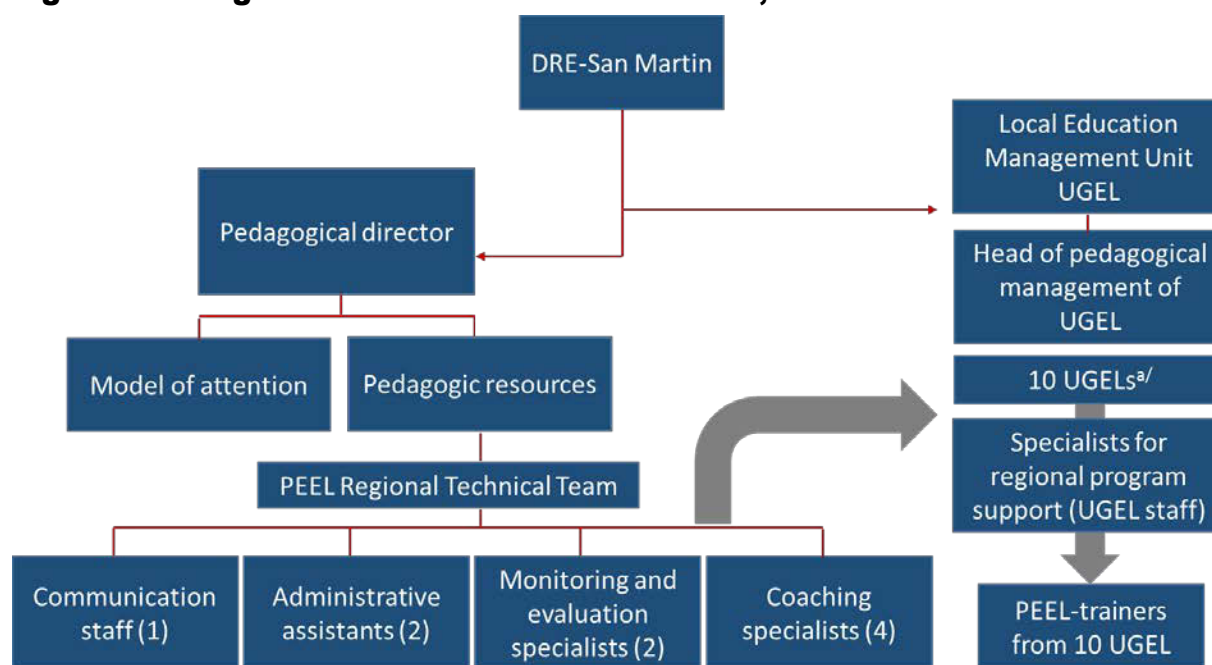
The organization of the regional implementation teams was similar in Ucayali and San Martín; however, San Martín had to increase the number of staff supporting the implementation of PEEL to deliver coaching sessions in all program schools. The Regional Education Offices of Ucayali and San Martín have a pedagogical branch and a local education management unit (the UGEL). The LANN regional technical team in Ucayali and the PEEL regional technical team in San Martín coordinate the implementation of program activities with the UGEL of their respective region. UPCHs provided technical assistance to the regional technical teams and training for the PEEL-trainers and LANN-trainers in Ucayali and San Martín, respectively.

In Ucayali, the regional technical team consists of three area specialists (communication, primary education, and materials) and two administrative staff (secretary and an accounting assistant). The Ucayali regional technical team is overseen by the pedagogical management director and focuses exclusively on the implementation of the LANN project. The regional technical team and UGEL staff coordinate training and coaching activities for PEEL trainers. UGEL staff supports the logistics of program activities, while the technical regional team assist LANN-trainers and monitor the delivery of the teacher training and coaching. Figure II.1 shows the organizational chart for the regional implementation team in Ucayali.

In San Martín, the organizational structure changed in the second year of implementation. During the first year, PEEL trainers faced constraints delivering coaching sessions as planned given the large number of intervened schools (89) spread over a large, and often hard-to-reach geographic area. As a result, an additional trainer joined UPCH staff and nine new posts for PEEL-trainers were opened for the second year. In 2016, a total of 22 PEEL-trainers joined the program—13 PEEL-trainers occupied existing vacant posts and filled 9 new posts.

To strengthen the implementation of PEEL program and carry out the monitoring and evaluation plan as envisioned, four coaching specialists, two monitoring and evaluation specialists, one communication specialist and two administrative assistants joined the PEEL regional technical team for the second year of implementation.²⁹ Figure II.2 shows the organizational chart for the regional implementation team in San Martín.

²⁹ Also, the cadre of UGEL specialists were integrated into the team with the intent that they would support PEEL activities, and to prepare to be liaisons for a sustainability phase later on.

Figure B.1. Organizational chart for Ucayali, 2016**Figure B.2. Organizational chart for San Martín, 2016**

Note: ^{a/}In San Martín, the PEEL program is being implemented in 10 provinces through the local education management units (UGEL). The UGELs are Bellavista, El Dorado, Huallaga, Lamas, Mariscal Cáceres, Moyobamba, Picota, Rioja, San Martín, and Tocache.

Table B.1. Qualitative data collection (2016): respondents, methods, and topics of inquiry

Respondents	Method	Topics of inquiry	Location
UPCH technical assistance director and central team	1 group interview (3 participants)	• Program components • Pedagogical approach • Program rollout • Fidelity of implementation • Facilitators and barriers to implementation • Perceived program outcomes • Sustainability	Lima
UPCH technical assistance director and central team	1 group interview	• UPCH technical assistance director and central team	1 group interview
UPCH program coordinators for Ucayali and San Martín	2 interviews	• Program rollout • Fidelity of implementation • Facilitators and barriers to implementation	Ucayali San Martín
LANN-DRE project coordinator for Ucayali	1 interview	• Program rollout • Fidelity of implementation • Institutional coordination • Facilitators and barriers to implementation	Ucayali
PEEL-DRE project coordinator for San Martín	1 interview	• Rollout of program components • Fidelity of implementation • Facilitators and barriers to implementation • Institutional coordination	San Martín
DRE pedagogical management director for Ucayali and San Martín	2 interviews	• Understanding of program goals • Facilitators and barriers to implementation • Institutional coordination	Ucayali San Martín
LANN trainers for Ucayali and PEEL trainers for San Martín	5 focus groups (2 Ucayali, 3 San Martín)	• Program components • Program participation and take-up • Perception of program services • Perception of program outcomes • Facilitators and barriers to implementation	Ucayali San Martín
Group A teachers (Amazonía Lee schools)	7 focus groups	• Program participation and take-up • Reading instruction practices • Perception of program services • Perception of program outcomes • Perception of other programs	Ucayali San Martín
Group B teachers (Prevailing practice)	6 focus groups (4 Ucayali, 2 San Martín)	• Reading instruction practices • In-service training received • Challenges with reading instruction • Challenges with students' reading achievement	Ucayali San Martín

Respondents	Method	Topics of inquiry	Location
Group A school directors (Amazonía Lee schools)	7 interviews	• Program participation and take-up • Leadership and management • Perception of program services • Perception of program outcomes • Perception of other programs	Ucayali and San Martín

UPCH = Universidad Peruana Cayetano Heredia; LANN = Liderando los Aprendizajes de Niñas y Niños; DRE = regional education office; PEEL = Programa Enseñar Es Liderar.

APPENDIX C

ADDITIONAL TABLES AND FIGURES FOR UCAYALI

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

Table C.1. Schools at baseline had similar infrastructure characteristics with some exceptions (percentage, unless otherwise indicated)

	Treatment adjusted mean	Control adjusted mean	Difference	p-value	Effect sizes
Programs implemented at baseline					
PELA	17.1	17.1	-0.0	1.000	-0.00
Redes	2.9	11.4	-8.6	0.183	-0.33
One Laptop per Child	65.7	85.7	-20.0*	0.017	-0.46
Computer lab/Technology room	28.6	37.1	-8.6	0.413	-0.18
National children's program Qali Warma	91.4	97.1	-5.7	0.324	-0.23
Number of risks observed in school (average)					
Physical hazards	1.3	1.3	0.1	0.786	0.06
Health risks	1.3	1.4	-0.1	0.607	-0.12
Percentage of schools with multigrade classrooms	42.9	37.1	5.7	0.571	0.11
Utilities and facilities					
Library or resource room	20.0	22.9	-2.9	0.768	-0.07
Piped water supply	65.7	77.1	-11.4	0.324	-0.24
Unpiped water supply	31.4	14.3	17.1	0.11	0.38
Potable drinking water supply	62.9	54.3	8.6	0.475	0.17
Electricity	88.6	82.9	5.7	0.535	0.16
Permanent building	91.4	97.1	-5.7	0.16	-0.25
Painted exterior walls	60.0	65.7	-5.7	0.571	-0.11
Perimeter fence	48.6	71.4	-22.9*	0.019	-0.46
Working restroom facilities for children	97.1	94.3	2.9	0.571	0.15
Latrines or urinals	71.4	88.6	-17.1	0.11	-0.44
Functional toilets	80.0	68.6	11.4	0.292	0.26
Separate facilities for boys and girls	77.1	74.3	2.9	0.768	0.07
Hand-washing facilities	71.4	74.3	-2.9	0.786	-0.06
Working waste disposal	91.4	91.4	0.0	1.000	0.00
Plumbing waste disposal	37.5	43.2	-5.7	0.601	-0.12
Septic (or similar) disposal	59.3	53.6	5.7	0.626	0.12
Other waste disposal (absorption bed or trench, river)	6.2	0.4	5.8	0.152	0.29
Kitchen	48.6	42.9	5.7	0.571	0.11
Vegetable garden	48.6	51.4	-2.9	0.768	-0.06
Teachers' lounge	22.9	14.3	8.6	0.324	0.22
Outdoors recreational space	91.4	94.3	-2.9	0.661	-0.10
Gymnasium or Sport Facilities	11.4	11.4	0.0	1.000	0.00
Infirmary	2.9	5.7	-2.9	0.571	-0.15
Computers for students	77.1	85.7	-8.6	0.324	-0.22
Computers for teachers	54.3	37.1	17.1	0.110	0.34
Internet connectivity	40.0	31.4	8.6	0.447	0.18
Phone service	14.3	8.6	5.7	0.422	0.18
Number of schools	35	35			

Source: *Amazonía Lee* Infrastructure observation—Baseline 2014.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table C.2. Students had similar characteristics at endline with the exception of being overage for first grade (percentage, unless otherwise indicated)

	Treatment adjusted mean	Control adjusted mean	Difference	p-value
Student is a boy	47.6	48.1	-0.5	0.789
Age in years at endline (average)	7.5	7.5	0.1	0.094
Current school is same as school in baseline	89.4	90.9	-1.6	0.336
Students not attending school at endline	1.6	1.3	0.3	0.514
Student attends second grade in endline school year	94.0	94.3	-0.3	0.858
Endline assessment took place in school	91.4	94.1	-2.7	0.119
Overage for first grade (percentage of students age 7 or older at baseline)	9.3	4.3	5.0*	0.004
Student repeated first grade	1.2	0.6	0.6	0.423
Number of students	380	360		

Source: *Amazonia Lee* Child assessment—Baseline 2015 and Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table C.3. Children in both groups had similar early reading scores at baseline, except for phonemic awareness

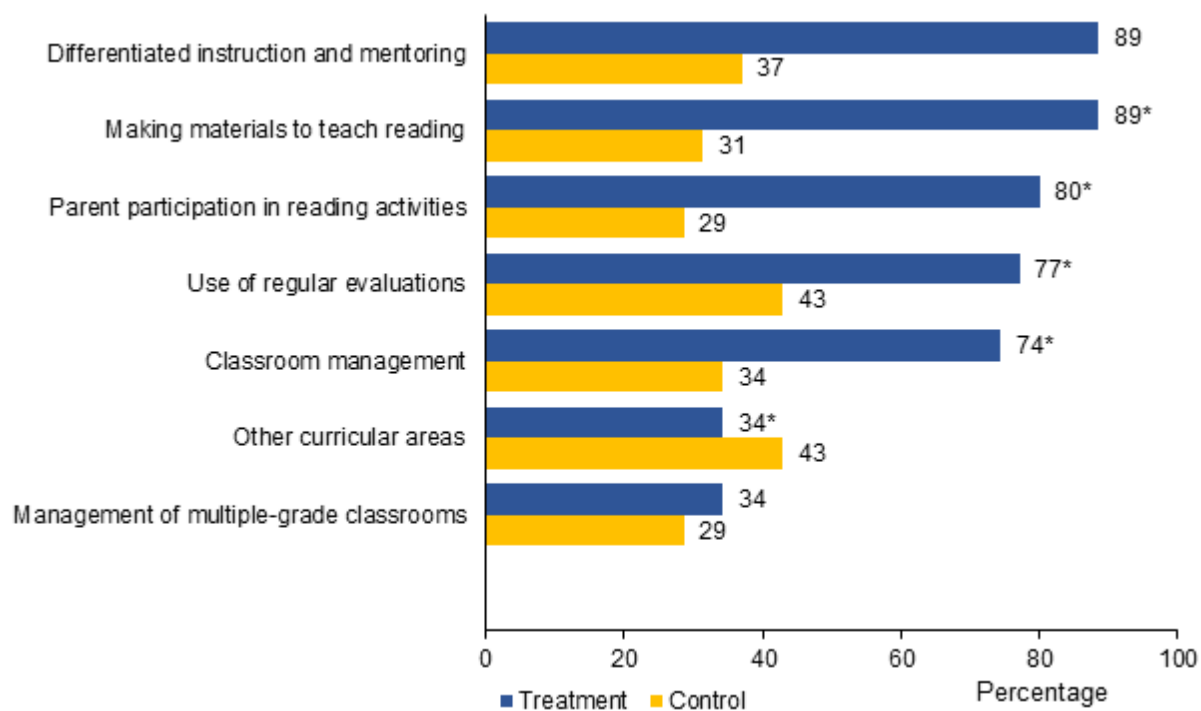
	Treatment adjusted mean	Control adjusted mean	Difference	p-value
Test scores of 1st-grade emergent literacy assessment at baseline				
Letter identification score	18.1	18.5	-0.5	0.212
Emergent writing score	10.5	10.8	-0.2	0.582
Fluency, number of correctly read words in one minute	2.1	2.1	-0.1	0.746
Phonemic awareness score	3.1	2.7	0.4*	0.003
Decoding, number of correctly read pseudo-words in one minute	1.9	2.0	-0.1	0.889
Oral comprehension score	2.2	2.1	0.1	0.335
Number of students	380	360		

Source: *Amazonia Lee* Child assessment—Baseline 2015.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

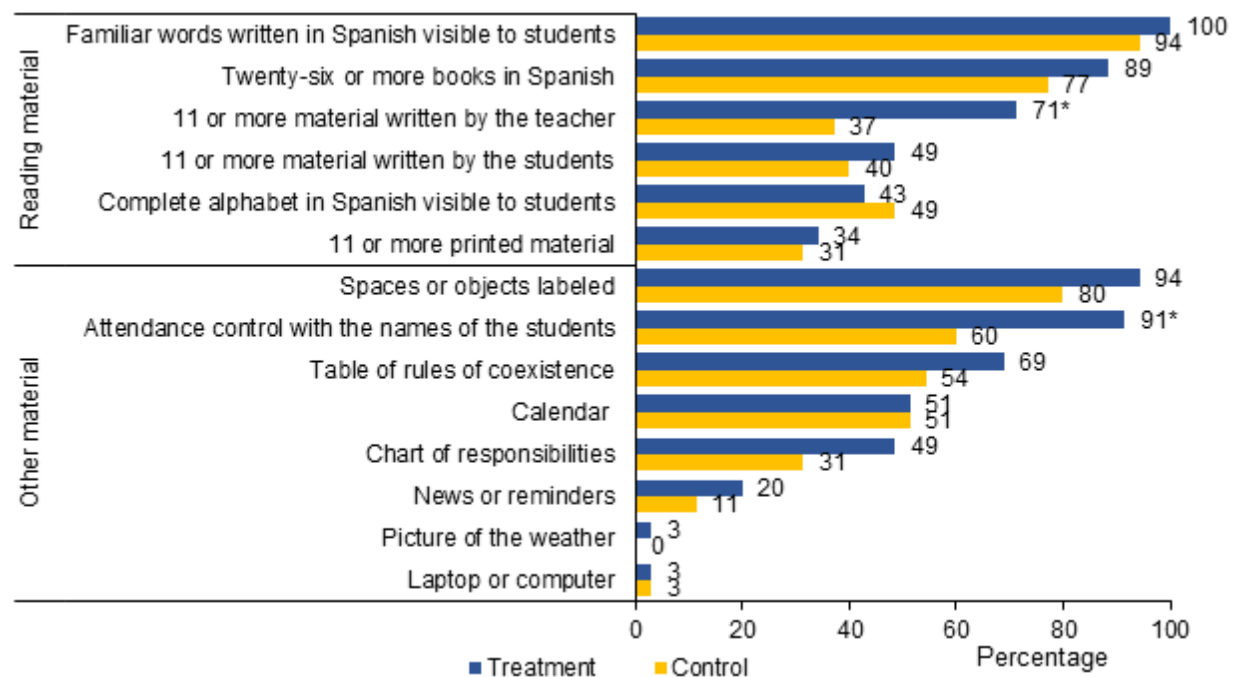
Figure C.1. Treatment group teachers were more likely to report being trained in other pedagogical strategies and classroom management



Source: *Amazonía Lee* Teacher Survey—Final Follow-up 2016.

* Difference in group means is statistically significant at the 0.05 level.

Figure C.2. Literacy resources and materials observed in second grade classrooms



Source: *Amazonía Lee* Classroom observation—Final Follow-up 2016.

* Difference in group means is statistically significant at the 0.05 level.

Table C.4. Treatment group teachers are more likely to report receiving textbook, narrative books, workbooks, guides, and audiovisual resources from *Amazonía Lee*

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Percentage of teachers who reported receiving the following material from MINEDU:				
Textbooks	60.0	57.1	2.9	0.831
Reference books (dictionaries)	-0.0	8.6	-8.6	0.083
Narrative books (story books)	25.7	5.7	20.0*	0.033
Workbooks	80.0	71.4	8.6	0.475
Guide books	45.7	40.0	5.7	0.661
Audiovisual resources	2.9	-0.0	2.9	0.324
Posters, laminates	20.0	20.0	0.0	1.000
School supplies	-0.0	2.9	-2.9	0.324
Other material	2.9	5.7	-2.9	0.571
Percentage of teachers who reported receiving the following material from <i>Amazonía Lee</i>:				
Textbooks	48.6	5.7	42.9*	0.000
Reference books (dictionaries)	0.0	0.0	0.0	
Narrative books (story books)	82.9	0.0	82.9*	0.000
Workbooks	82.9	8.6	74.3*	0.000
Guide books	34.3	8.6	25.7*	0.010
Audiovisual resources	14.3	-0.0	14.3*	0.023
Posters, laminates	28.6	-0.0	28.6*	0.001
School supplies	2.9	-0.0	2.9	0.324
Number of teachers	35	35		

Source: *Amazonía Lee* Teacher survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table C.5. Teachers in the treatment group were more likely to receive materials from *Amazonía Lee* and having a classroom library

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Teachers who reported receiving material from:				
Ministry of Education	94.3	82.9	11.4	0.160
<i>Amazonía Lee</i>	94.3	8.6	85.7*	0.000
Teachers who reported having a library in the classroom	91.4	5.7	85.7*	0.000
Teachers who reported using the classroom library every day	71.4	5.7	65.7*	0.000
Number of teachers	35	35		

Source: *Amazonía Lee* Teacher Survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table C.6. Control teachers were more likely to report as a barrier to teach lack of pedagogical resources and lack of training to teach literacy

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Lack of supports and resources for teaching reading and writing				
Lack of books	37.1	60.0	-22.9	0.073
Lack of pedagogical resources and materials	54.3	80.0	-25.7*	0.037
Lack of sufficient instructional or class time	31.4	34.3	-2.9	0.786
Lack of training in methods to teach literacy	48.6	80.0	-31.4*	0.009
Lack of training in intercultural bilingual education	25.7	22.9	2.9	0.768
Lack of knowledge of native language	14.3	20.0	-5.7	0.535
Lack of classroom equipment or material	54.3	60.0	-5.7	0.661
Lack of support from parents	91.4	94.3	-2.9	0.661
Other	-0.0	8.6	-8.6	0.083
Number of teachers	35	35		

Source: *Amazonía Lee* Teacher survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table C.7. Teachers in both groups reported doing literacy activities at similar rates

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Teachers who reported doing the following literacy activities with second grade students				
Rhymes or games with the sounds of words	97.1	94.3	2.9	0.571
Ask student to make predictions about the readings	94.3	88.6	5.7	0.324
Practice to complete incomplete words	100.0	94.3	5.7	0.160
Practice to complete incomplete sentences	94.3	88.6	5.7	0.422
Quick reading exercises on familiar words or phrases	91.4	94.3	-2.9	0.661
Pre-reading exercises in preparation to read a book when students see the title of the book	97.1	100.0	-2.9	0.324
Demonstrate reading	100.0	97.1	2.9	0.324
Silent reading exercises	100.0	97.1	2.9	0.324
Reading exercises in pairs or small groups	100.0	94.3	5.7	0.160
Group reading exercises in chorus or in a chain	100.0	97.1	2.9	0.324
Practice to develop graphomotor skills (draw letters)	68.6	74.3	-5.7	0.600
Copy words or sentences	94.3	91.4	2.9	0.661
Practice punctuation	88.6	82.9	5.7	0.535
Dictation	82.9	91.4	-8.6	0.324
Practice spelling rules or accents	82.9	91.4	-8.6	0.263
Number of teachers	35	35		

Source: *Amazonía Lee* Teacher survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

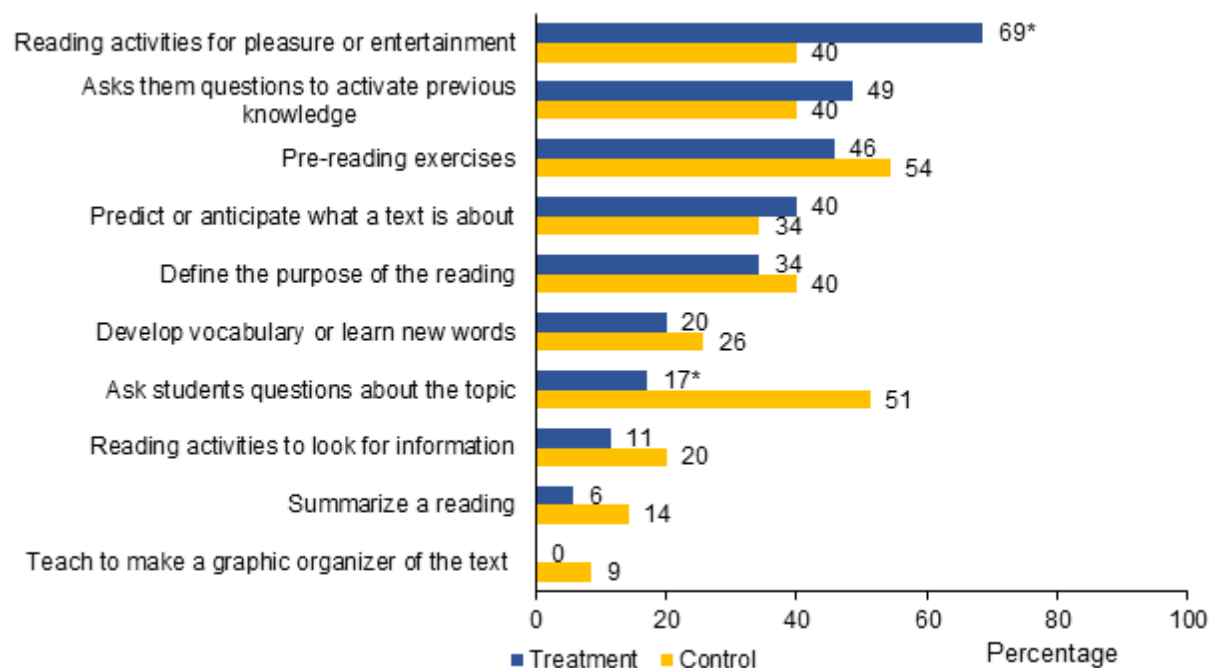
Figure C.3. Both groups performed initial literacy activities daily at similar rates



Source: *Amazonía Lee* Teacher Survey—Final Follow-up 2016.

* Difference in group means is statistically significant at the 0.05 level.

Figure C.4. Treatment group more likely to read for pleasure every day and control group more likely to ask about the topic of the reading every day



Source: *Amazonía Lee* Teacher Survey—Final Follow-up 2016.

* Difference in group means is statistically significant at the 0.05 level.

Table C.8. Teachers in both groups reported spending similar time in all activities, except for participating in reading activities

	Treatment adjusted mean	Control adjusted mean	Difference	p-value
Hours per week teachers reported spending in the following activities:				
Administrative tasks	0.6	0.7	-0.0	0.929
Preparing the class	13.5	12.2	1.2	0.380
Meeting with principal or other teachers	1.2	0.9	0.3	0.417
Meeting with students' parents	0.4	0.6	-0.2	0.398
Monitoring students during recess and free time	1.5	1.5	0.1	0.783
Teaching reading	8.2	8.5	-0.3	0.753
Teaching math	7.1	8.0	-1.0	0.224
Teaching social science	2.4	2.5	-0.1	0.837
Teaching science	4.8	3.0	1.7	0.535
Teaching art	1.4	1.7	-0.3	0.111
Planning or managing or correcting exams or homework	4.2	3.4	0.8	0.321
Providing support to students with low academic achievement	3.9	3.1	0.9	0.224
Preparing students for national examinations	3.5	3.3	0.3	0.703
Participating in activities promoting reading after school	0.7	0.1	0.6*	0.022
Number of teachers	35	35		

Source: *Amazonía Lee* Teacher Survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table C.9. Teachers in both groups spent similar time on teaching practices not focused on reading

	Treatment adjusted mean	Control adjusted mean	Difference	p-value
Number of times the following teaching practices were observed during class (average)				
Informal conversations or socialize with students	4.5	4.7	-0.2	0.703
Verify that students understood instructions	0.4	0.3	0.1	0.624
Have informal conversations with students	0.2	0.2	-0.0	0.744
Verify students have the material needed for the activities	0.5	0.4	0.1	0.600
Provide feedback to students' participation in class activities	3.2	3.0	0.2	0.722
Praise students sincerely	4.7	3.8	0.8	0.194
Speak respectfully to students	3.6	3.0	0.6	0.275
Provide opportunities for cooperation with classmates	0.9	0.6	0.3	0.498
Do activities differentiated by students' learning	-0.0	0.1	-0.1	0.324
Offer individualized help to students	1.5	1.7	-0.2	0.658
Number of classrooms	35	35		

Source: *Amazonia Lee* Classroom observation—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table C.10. Teachers in both groups reported using most activities related to classroom climate at similar rates

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Teachers who reported to use the following activities (related with classroom climate) always:				
Praise students specifically and sincerely	54.3	62.9	-8.6	0.324
Show affection or appreciation to students	80.0	74.3	5.7	0.571
Teach students to speak positively about themselves	77.1	65.7	11.4	0.292
Give students responsibilities in the classroom	48.6	71.4	-22.9*	0.030
Display in the classroom the jobs that students have done	71.4	68.6	2.9	0.800
Give students opportunities to cooperate with their peers	71.4	71.4	0.0	1.000
Remind students of rules of coexistence in the classroom	85.7	82.9	2.9	0.744
Number of teachers	35	35		

Source: *Amazonia Lee* Teacher Survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table C.11. Teachers in both groups reported using differentiated instruction practices at similar rates, except for planning to teach for different levels of learning

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Teachers who reported using differentiated instruction	91.4	94.3	-2.9	0.661
Use some type of assessment to identify underperforming students	80.0	91.4	-11.4	0.211
Organize students to teach to customize instruction	91.4	82.9	8.6	0.324
Plan teaching for the different levels of learning	74.3	91.4	-17.1*	0.032
Select differentiated contents for the different levels of learning	62.9	74.3	-11.4	0.324
Use differentiated activities for the different levels of learning	77.1	77.1	0.0	1.000
Monitor individual progress	68.6	68.6	0.0	1.000
Number of teachers	35	35		

Source: *Amazonía Lee* Teacher survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table C.12. Teachers in the treatment report using differentiated instruction less often than control teachers

	Treatment adjusted mean	Control adjusted mean	Difference	p-value
Number of times teachers reported to use the following strategies for differentiated instruction with underperforming students in reading in the last month (average)				
Use some type of assessment to identify underperforming students	5.9	8.9	-2.9*	0.050
Organize students to teach based on students' learning needs	8.4	9.6	-1.1	0.551
Plan teaching for the different levels of learning	6.5	11.0	-4.6*	0.003
Select differentiated contents for the different levels of learning	4.4	8.5	-4.1*	0.024
Use differentiated activities for the different levels of learning	6.3	8.0	-1.7	0.312
Prepare differentiated material	4.9	7.1	-2.2	0.207
Monitor individual progress	11.6	14.5	-2.9	0.119
Number of teachers	35	35		

Source: *Amazonía Lee* Teacher survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table C.13. Treatment group teachers reported working extra hours with underperforming students in reading

	Treatment adjusted mean	Control adjusted mean	Difference	p-value
Number of underperforming students in reading reported by teachers (average)	5.6	5.5	0.1	0.890
Teachers who reported to work additional hours with underperforming students in reading (percentage)	85.7	57.1	28.6*	0.023
Extra hours per week working with underperforming students (average)	3.9	2.7	1.3	0.089
Teachers who refer underperforming students to receive academic reinforcement (percentage)	31.4	11.4	20.0	0.051
Number of teachers	35	35		

Source: *Amazonía Lee* Teacher survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table C.14. Teachers in both groups report using formative assessments at similar rates except for assessed student achievement at the end of the school year

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Teachers who reported making an initial or diagnostic evaluation to make plans for the communication class	88.6	97.1	-8.6	0.183
Teachers who reported using the initial or diagnostic evaluation to organize learning groups among students	82.9	77.1	5.7	0.571
Teachers who reported using the following assessment activities:				
Formative assessments to determine progress or difficulties in reading and writing	97.1	97.1	0.0	1.000
Provides feedback to students about their progress in reading and writing	100.0	97.1	2.9	0.324
Performs an end-of-term assessment to assess students' achievement in reading and writing	77.1	100.0	-22.9*	0.003
Number of teachers	35	35		

Source: *Amazonía Lee* Teacher survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Figure C.5. Skills demonstrated by at least 50 percent of students are similar between groups, except for two skills (reported by teachers)



Source: *Amazonía Lee* Teacher Survey—Final Follow-up 2016.

* Difference in group means is statistically significant at the 0.05 level.

Table C.15. Students reading outcomes were similar between groups

	Impact for girls	p-value	Impact for boys	p-value	Difference in impacts by gender (p-value)
Percentage of students who got a score of zero in:					
Reading familiar words	-16.9*	0.011	-18.2*	0.008	0.784
Decoding task	-9.8	0.102	-15.0*	0.024	0.345
Fluency task	-14.3*	0.007	-18.7*	0.001	0.391
Number of students	386		354		

Source: *Amazonía Lee* Child assessment—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables). We also include baseline score of students and some student's characteristics (2015) and baseline infrastructure characteristics of schools (2014). Effect sizes were calculated before rounding the adjusted difference.

* Difference in group means is statistically significant.

Table C.16. *Amazonía Lee* impacts on reading familiar words, decoding, fluency, and reading comprehension, sample with non-missing baseline data

	Treatment adjusted mean	Control adjusted mean	Difference	p-value	Effect sizes
Number of correct oral comprehension responses	2.8	2.5	0.3*	0.004	0.21
Number of familiar words read correctly	7.3	6.2	1.1*	0.009	0.26
Number of pseudo-words read correctly in one minute (decoding)	18.3	13.0	5.3*	0.000	0.38
Number of words read correctly in one minute (fluency)	37.3	29.8	7.5*	0.000	0.27
Number of correct answers in reading comprehension	2.7	1.9	0.8*	0.000	0.40
Number of students	315	314			

Source: *Amazonía Lee* Child Assessment—Final Follow-Up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables). We also include baseline scores of students and some students' characteristics (2015) and baseline infrastructure characteristics of schools (2014). Effect sizes were calculated before rounding the adjusted difference.

* Difference in group means is statistically significant at the 0.05 level.

Table C.17. Students characteristics at endline and early reading scores at baseline, sample with non-missing baseline data

	Treatment adjusted mean	Control adjusted mean	Difference	p-value
Student is a boy	48.3	48.8	-0.5	0.806
Age in years at endline (average)	7.5	7.4	0.1	0.112
Current school is same as school in baseline	91.1	91.6	-0.5	0.758
Students not attending school at endline	1.2	0.9	0.3	0.597
Student attends second grade in endline school year	96.2	95.2	1.0	0.553
Endline assessment took place in school	91.8	94.7	-2.8	0.135
Overage for first grade (percentage of students age 7 or older at baseline)	8.1	4.0	4.1*	0.005
Student repeated first grade	1.2	0.6	0.6	0.424
Letter identification score	18.1	18.5	-0.4	0.351
Emergent writing score	10.5	10.8	-0.2	0.656
Fluency, number of correctly read words in one minute	2.1	2.1	-0.0	0.968
Phonemic awareness score	3.1	2.7	0.5*	0.002
Decoding, number of correctly read pseudo-words in one minute	2.0	2.0	-0.0	0.940
Oral comprehension score	2.2	2.1	0.1	0.221
Number of students	315	314		

Source: *Amazonía Lee* Child assessment—Baseline 2015 and Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level

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

APPENDIX D

ADDITIONAL TABLES AND FIGURES FOR SAN MARTÍN

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

Table D.1. Schools at baseline had similar infrastructure characteristics with some exceptions (percentage, unless otherwise indicated)

	Treatment adjusted mean	Control adjusted mean	Difference	p-value	Effect sizes
Programs implemented at baseline					
PELA	22.0	21.2	0.8	0.892	0.02
Redes	41.4	38.2	3.2	0.652	0.07
Education materials provided by MINEDU	99.0	99.0	0.0	0.994	0.00
One Laptop per Child	77.8	72.5	5.2	0.386	0.12
Computer lab/Technology room	54.5	49.5	5.0	0.417	0.10
National children's program Qali Warma	90.0	86.9	3.1	0.424	0.10
Number of risks observed in school (average)					
Physical hazards	1.3	1.4	-0.1	0.599	-0.08
Health risks	1.2	1.3	-0.1	0.681	-0.06
Percentage of schools with multigrade classrooms	25.0	23.2	1.8	0.608	0.04
Utilities and facilities					
Library or resource room	36.0	39.4	-3.4	0.601	-0.07
Piped water supply	79.0	78.4	0.6	0.915	0.02
Unpiped water supply	7.0	11.2	-4.2	0.325	-0.15
Potable drinking water supply	65.0	55.6	9.4	0.120	0.19
Electricity	89.0	89.6	-0.6	0.890	-0.02
Permanent building	92.0	93.9	-1.9	0.580	-0.08
Painted interior walls	93.0	92.9	0.1	0.970	0.01
Painted exterior walls	70.7	68.7	2.0	0.741	0.04
Classrooms have ceilings	100.0	98.0	2.0	0.154	0.20
Perimeter fence	80.0	72.8	7.2	0.181	0.17
Working restroom facilities for children	100.0	98.0	2.0	0.156	0.20
Latrines or urinals	72.0	74.2	-2.2	0.719	-0.05
Functional toilets	87.0	81.8	5.2	0.180	0.14
Separate facilities for boys and girls	87.0	82.8	4.2	0.356	0.12
Hand-washing facilities	86.0	81.8	4.2	0.374	0.11
Working waste disposal	92.0	86.9	5.1	0.267	0.17
Plumbing waste disposal	37.4	43.4	-6.0	0.244	-0.12
Septic or similar) disposal	60.0	56.2	3.8	0.466	0.08
Other waste disposal absorption bed or trench, river)	5.0	4.0	1.0	0.741	0.05
Kitchen	71.0	66.7	4.3	0.463	0.09
Vegetable garden	47.0	49.4	-2.4	0.731	-0.05
Teachers' lounge	27.0	26.5	0.5	0.936	0.01
Outdoors recreational space	90.0	86.8	3.2	0.461	0.10
Gymnasium or Sport Facilities	5.0	1.0	4.0	0.105	0.23
Infirmary	2.0	6.0	-4.0	0.154	-0.21
Computers for students	92.0	91.8	0.2	0.966	0.01
Computers for teachers	69.7	62.7	7.0	0.266	0.15
Internet connectivity	42.0	38.4	3.6	0.528	0.07
Phone service	15.0	17.2	-2.2	0.632	-0.06
School is in urban area	62.0	63.0	-1.0	0.881	-0.02
School offers both shifts	39.0	40.0	-1.0	0.885	-0.02
Number of schools	100	100			

Source: *Amazonia Lee* infrastructure observation—Baseline 2014.

Note: Regression-adjusted means. Standard deviations in brackets. P-values in parentheses.

* Difference in group means is statistically significant at the 0.05 level.

Table D.2. Students in both groups have similar characteristics at endline but significant differences in overage and repetition of first grade at baseline (percentage, unless otherwise indicated)

	Treatment adjusted mean	Control adjusted mean	Difference	p-value
Student is a boy	50.0	51.0	-0.9	0.276
Age in years at endline (average)	7.4	7.4	0.0	0.062
Current school is same as school in baseline	91.7	93.3	-1.6	0.118
Students not attending school at endline	0.5	0.3	0.2	0.508
Student attends second grade in endline school year	97.6	97.3	0.2	0.709
Endline assessment took place in school	95.3	96.8	-1.5*	0.040
Overage for first grade at baseline (percentage of students age 7 or older at baseline)	3.9	1.7	2.3*	0.001
Student repeated first grade	0.4	0.0	0.4*	0.019
Number of students	824	822		

Source: *Amazonía Lee* Child assessment—Baseline 2015 and Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table D.3. Children in both groups had similar early reading scores at baseline

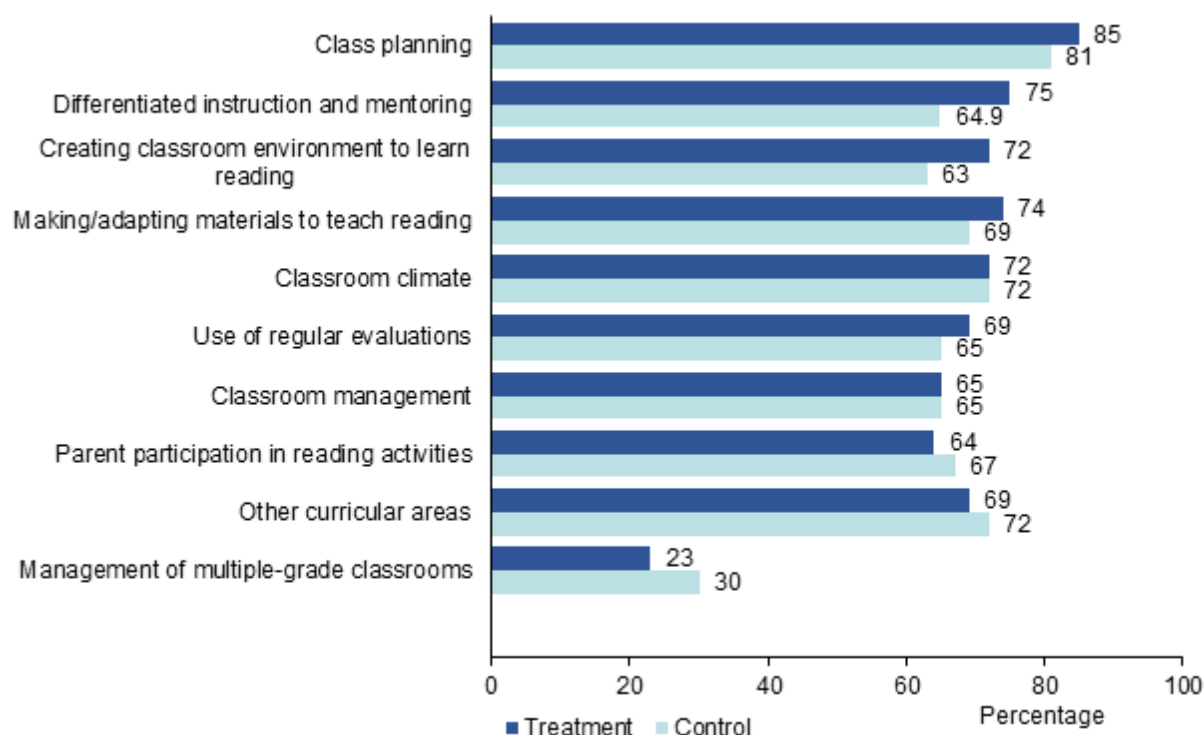
	Treatment adjusted mean	Control adjusted mean	Difference	p-value
Test scores of 1st-grade emergent literacy assessment at baseline				
Letter identification score	18.5	18.5	0.0	0.946
Emergent writing score	11.9	12.0	-0.2	0.497
Fluency, number of correctly read words in one minute	2.2	2.3	-0.1	0.470
Phonemic awareness score	3.4	3.3	0.1	0.090
Decoding, number of correctly read pseudo-words in one minute	1.7	1.9	-0.2	0.289
Oral comprehension score	2.8	2.7	0.1	0.355
Number of students	824	822		

Source: *Amazonía Lee* Child assessment—Baseline 2015 and Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Figure D.1. Teachers in the two groups report being trained in other pedagogical strategies at similar rates



Source: *Amazonía Lee* Teacher Survey—Final Follow-up 2016.

* Difference in group means is statistically significant at the 0.05 level.

Table D.4. More teachers in the treatment group report having a library in the classroom and receiving material from *Amazonía Lee*

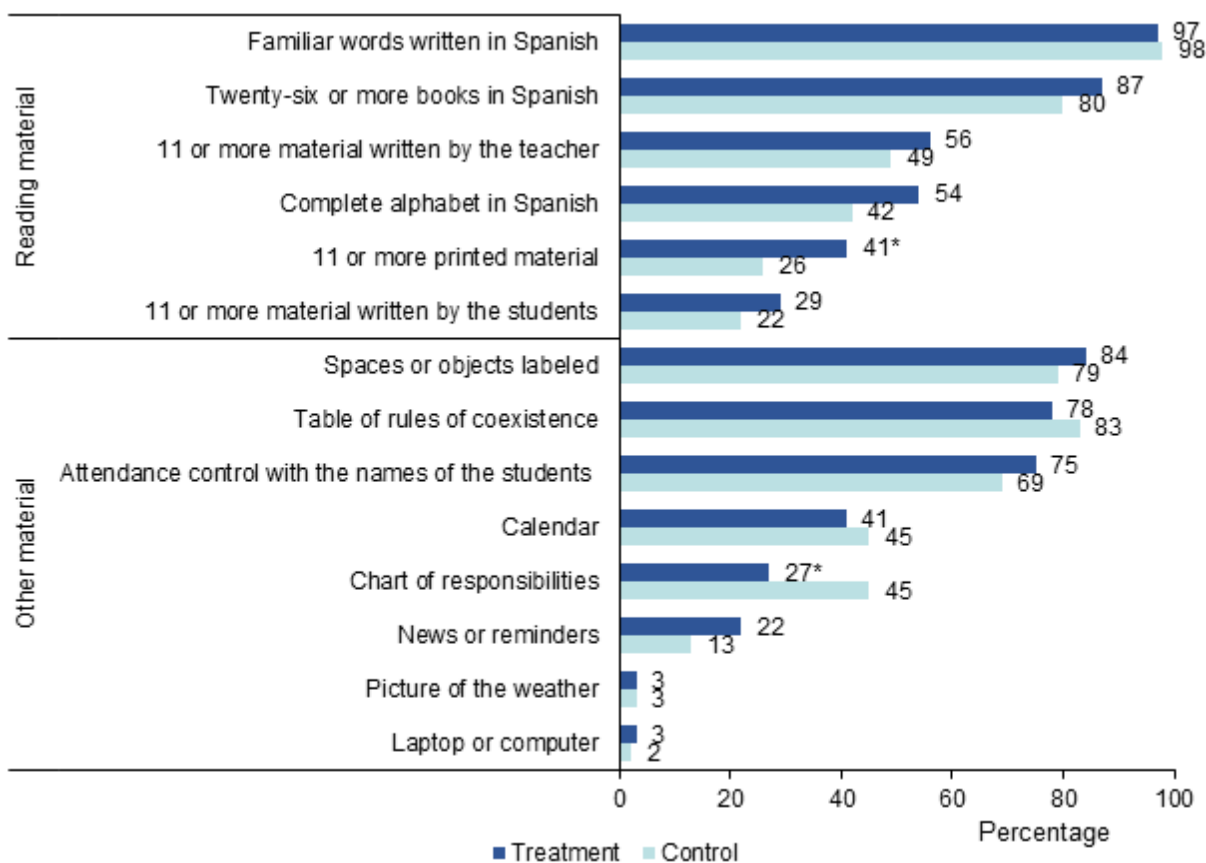
	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Teachers who reported receiving material from:				
Ministry of Education	82.0	84.0	-2.0	0.707
<i>Amazonía Lee</i>	76.0	2.0	74.0*	0.000
Teachers who reported having a library in the classroom				
	62.0	2.0	60.0*	0.000
Teachers who reported using the classroom library every day				
	43.0	1.0	42.0*	0.000
Number of teachers	100	100		

Source: *Amazonía Lee* Teacher Survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Figure D.2. Observed literacy resources and materials were similar in both groups



Source: *Amazonía Lee* Teacher Survey—Final Follow-up 2016.

* Difference in group means is statistically significant at the 0.05 level.

Table D.5. Control teachers are more likely to report lack of training to teach literacy as a factor limiting their ability to teach students to read

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Lack of supports and resources for teaching reading and writing				
Lack of books	53.0	58.0	-5.0	0.438
Lack of pedagogical resources and materials	69.0	62.0	7.0	0.310
Lack of sufficient instructional or class time	27.0	21.0	6.0	0.275
Lack of training in methods to teach literacy	45.0	60.0	-15.0*	0.035
Lack of training in intercultural bilingual education	21.0	22.0	-1.0	0.863
Lack of knowledge of native language	19.0	19.0	0.0	1.000
Lack of classroom equipment or material	58.0	67.0	-9.0	0.235
Lack of support from parents	88.0	82.0	6.0	0.241
Other	2.0	4.0	-2.0	0.417
Number of teachers	100	100		

Source: *Amazonía Lee* Teacher survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table D.6. Teachers reported receiving materials from MINEDU at similar rates

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Percentage of teachers who reported receiving the following material from MINEDU:				
Textbooks	52.0	57.0	-5.0	0.495
Reference books (dictionaries)	2.0	2.0	0.0	1.000
Narrative books (story books)	19.0	15.0	4.0	0.417
Workbooks	72.0	77.0	-5.0	0.426
Guide books	40.0	35.0	5.0	0.438
Audiovisual resources	1.0	4.0	-3.0	0.181
Posters, laminates	24.0	20.0	4.0	0.468
School supplies	33.0	42.0	-9.0	0.181
Other material	1.0	7.0	-6.0*	0.033
Percentage of teachers who reported receiving the following material from <i>Amazonía Lee</i>:				
Textbooks	34.0	2.0	32.0*	0.000
Reference books (dictionaries)	1.0	0.0	1.0	0.320
Narrative books (story books)	5.0	-0.0	5.0*	0.025
Workbooks	62.0	1.0	61.0*	0.000
Guide books	31.0	1.0	30.0*	0.000
Audiovisual resources	1.0	0.0	1.0	0.320
Posters, laminates	25.0	0.0	25.0*	0.000
School supplies	4.0	0.0	4.0*	0.045
Number of teachers	100	100		

Source: *Amazonía Lee* Teacher survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table D.7. Teachers in both groups reported doing literacy activities with similar rates

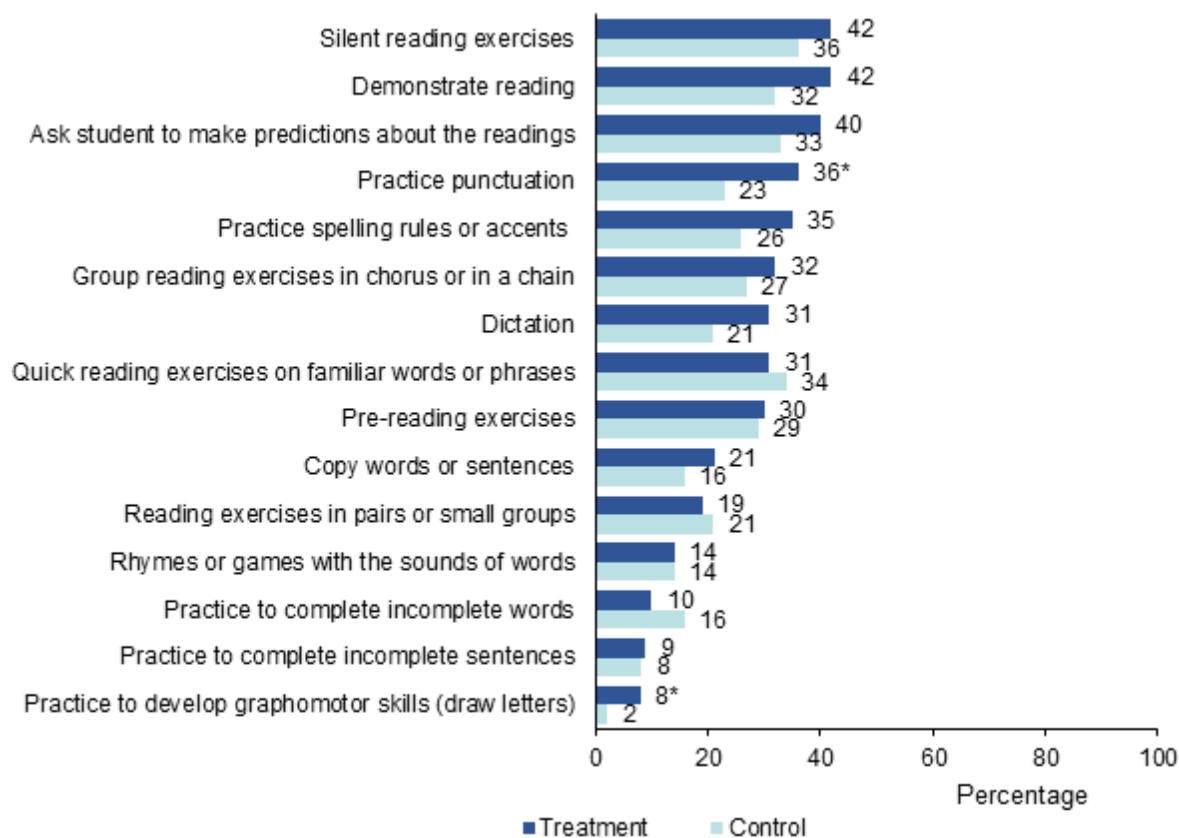
	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Teachers who reported doing the following literacy activities with second grade students				
Rhymes or games with the sounds of words	97.0	98.0	-1.0	0.657
Ask student to make predictions about the readings	96.0	89.0	7.0	0.052
Practice to complete incomplete words	90.0	93.0	-3.0	0.408
Practice to complete incomplete sentences	92.0	95.0	-3.0	0.408
Quick reading exercises on familiar words or phrases	93.0	98.0	-5.0	0.096
Pre-reading exercises in preparation to read a book when students see the title of the book	100.0	98.0	2.0	0.158
Demonstrate reading	100.0	98.0	2.0	0.158
Silent reading exercises	99.0	99.0	0.0	1.000
Reading exercises in pairs or small groups	96.0	98.0	-2.0	0.417
Group reading exercises in chorus or in a chain	99.0	96.0	3.0	0.181
Practice to develop graphomotor skills (draw letters)	72.0	76.0	-4.0	0.468
Copy words or sentences	95.0	89.0	6.0	0.134
Practice punctuation	95.0	97.0	-2.0	0.482
Dictation	96.0	91.0	5.0	0.167
Practice spelling rules or accents	90.0	91.0	-1.0	0.810
Number of teachers	100	100		

Source: *Amazonía Lee* Teacher survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Figure D.3. Initial literacy activities performed every day were similar for both groups with two exceptions



Source: *Amazonía Lee* Teacher Survey—Final Follow-up 2016.

* Difference in group means is statistically significant at the 0.05 level.

Figure D.4. Treatment group more likely to develop vocabulary every day

Source: *Amazonía Lee* Teacher Survey—Final Follow-up 2016.

* Difference in group means is statistically significant at the 0.05 level.

Table D.8. Teachers in both groups were observed performing general teaching practices with similar frequency

	Treatment adjusted mean	Control adjusted mean	Difference	p-value
Number of times the following teaching practices were observed during class (average)				
Informal conversations or socialize with students	5.3	5.3	-0.1	0.716
Verify that students understood instructions	0.7	0.9	-0.2	0.502
Have informal conversations with students	0.4	0.2	0.1	0.196
Verify students have the material needed for the activities	0.6	0.5	0.1	0.453
Provide feedback to students' participation in class activities	2.8	2.8	-0.1	0.858
Praise students sincerely	4.4	4.4	-0.0	0.977
Speak respectfully to students	5.0	5.0	-0.0	0.987
Provide opportunities for cooperation with classmates	0.4	0.6	-0.2	0.372
Do activities differentiated by students' learning	0.1	0.1	-0.0	0.872
Offer individualized help to students	2.6	2.5	0.1	0.751
Number of classrooms	100	100		

Source: *Amazonía Lee* Classroom observation—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table D.9. Teachers reported using most activities related to classroom climate at similar rates

	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Teachers who reported to use the following activities related with classroom climate always:				
Praise students specifically and sincerely	67.0	70.0	-3.0	0.650
Show affection or appreciation to students	74.0	73.0	1.0	0.867
Teach students to speak positively about themselves	69.0	81.0	-12.0	0.051
Give students responsibilities in the classroom	56.0	70.0	-14.0*	0.034
Display in the classroom the jobs that students have done	64.0	65.0	-1.0	0.887
Give students opportunities to cooperate with their peers	62.0	76.0	-14.0*	0.022
Remind students of rules of coexistence in the classroom	84.0	83.0	1.0	0.854
Number of teachers	100	100		

Source: *Amazonía Lee* Teacher Survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table D.10. Teachers in both groups report spending similar hours in teaching related activities, except for monitoring students during recess class

	Treatment adjusted mean	Control adjusted mean	Difference	p-value
Hours per week teachers reported spending in the following activities:				
Administrative tasks	1.7	1.5	0.2	0.666
Preparing the class	11.0	10.8	0.2	0.845
Meeting with principal or other teachers	1.4	1.3	0.1	0.628
Meeting with students' parents	1.2	1.3	-0.1	0.548
Monitoring students during recess and free time	1.4	1.8	-0.4*	0.009
Teaching reading techniques	9.0	7.8	1.2	0.061
Teaching math	7.7	7.6	0.1	0.860
Teaching social science	2.3	2.4	-0.1	0.758
Teaching science	4.4	4.7	-0.3	0.851
Teaching art	1.7	1.7	0.0	0.875
Planning or managing or correcting exams or homework	3.0	3.0	0.1	0.840
Providing support to students with low academic achievement	4.1	3.7	0.4	0.360
Preparing students for national examinations	3.1	4.0	-0.9	0.060
Participating in reading activities	0.4	0.6	-0.2	0.131
Number of teachers	100	100		

Source: *Amazonía Lee* Teacher Survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table D.11. Teachers reported using differentiated instruction practices at similar rates, except for planning to teach for different levels of learning more times per month

	Treatment adjusted mean	Control adjusted mean	Difference	p-value
Teachers reported using differentiated instruction with underperforming students in reading (percentage)	90.0	93.0	-3.0	0.441
Number of times teachers reported to use the following strategies for differentiated instruction with underperforming students in reading in the last month (average)				
Use some type of assessment to identify underperforming students	4.8	5.4	-0.7	0.419
Organize students to teach based on students' learning needs	6.9	6.7	0.2	0.774
Plan teaching for the different levels of learning	8.5	6.0	2.5*	0.007
Select differentiated contents for the different levels of learning	6.1	5.7	0.4	0.603
Use differentiated activities for the different levels of learning	6.1	5.4	0.7	0.376
Prepare differentiated material	6.4	6.1	0.3	0.762
Monitor individual progress	11.9	12.4	-0.5	0.683
Number of teachers	100	100		

Source: *Amazonía Lee* Teacher survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table D.12. Teachers in both groups work extra hours with underperforming students in reading but control teachers are more likely to refer them to receive academic reinforcement

	Treatment adjusted mean	Control adjusted mean	Difference	p-value
Number of underperforming students in reading reported by teachers (average)	4.3	4.2	0.1	0.695
Teachers who reported to work additional hours with underperforming students in reading (percentage)	80.0	71.0	9.0	0.129
Extra hours per week working with underperforming students (average)	4.0	4.0	0.1	0.876
Teachers who refer underperforming students to receive academic reinforcement outside class (percentage)	38.0	59.0	-21.0*	0.001
Number of teachers	100	100		

Source: *Amazonía Lee* Teacher survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Table D.13. Teachers in both groups report using formative assessments at similar rates

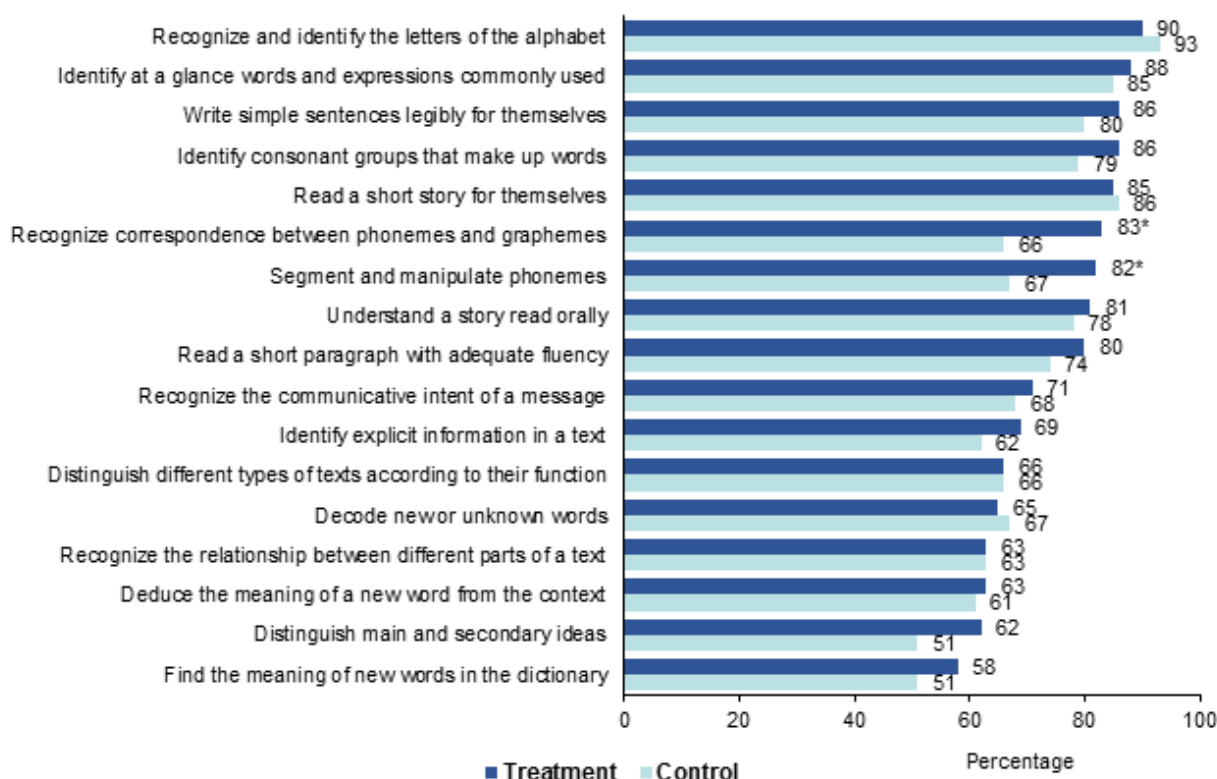
	Treatment adjusted mean (percentage)	Control adjusted mean (percentage)	Difference	p-value
Teachers who reported making an initial or diagnostic evaluation to make plans for the communication class	93.0	94.0	-1.0	0.765
Teachers who reported using the initial or diagnostic evaluation to organize learning groups among students	89.0	93.0	-4.0	0.287
Teachers who reported using the following assessment activities:				
Formative assessments to determine progress or difficulties in reading and writing	99.0	100.0	-1.0	0.320
Provides feedback to students about their progress in reading and writing	100.0	100.0	0.0	.
Performs an end-of-term assessment to assess students' achievement in reading and writing	93.0	98.0	-5.0	0.058
Number of teachers	100	100		

Source: Amazonía Lee Teacher survey—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level.

Figure D.5. Skills demonstrated by at least 50 percent of students are similar between groups (reported by teachers)



Source: *Amazonia Lee* Teacher Survey—Final Follow-up 2016.

* Difference in group means is statistically significant at the 0.05 level.

Table D.14. Percentage of boys and girls that got a score of zero in reading assessment task were similar in both groups

	Impact for girls	p-value	Impact for boys	p-value	Difference in impacts by gender (p-value)
Zero score in the following tests:					
Reading familiar words	-1.0	0.550	-0.3	0.845	0.809
Decoding task	-0.2	0.913	1.2	0.529	0.640
Fluency task	-2.8	0.104	-1.5	0.393	0.612
Number of students	815		831		

Source: *Amazonia Lee* Child assessment—Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables). We also include baseline score of students and some student's characteristics (2015) and baseline infrastructure characteristics of schools (2014). Effect sizes were calculated before rounding the adjusted difference.

* Difference in group means is statistically significant.

Table D.15. *Amazonía Lee* impacts on reading familiar words, decoding, fluency, and reading comprehension, sample with non-missing baseline data

	Treatment adjusted mean	Control adjusted mean	Difference	p-value	Effect sizes
Number of correct oral comprehension responses	3.0	2.9	0.1	0.155	0.08
Number of familiar words read correctly	8.3	8.2	0.1	0.441	0.03
Number of pseudo-words read correctly in one minute (decoding)	20.3	20.6	-0.3	0.538	-0.02
Number of words read correctly in one minute (fluency)	44.5	44.9	-0.4	0.643	-0.02
Number of correct answers in reading comprehension	3.1	3.0	0.1	0.481	0.03
Number of students	706	640			

Source: *Amazonía Lee* Child Assessment—Final Follow-Up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables). We also include baseline scores of students and some students' characteristics (2015) and baseline infrastructure characteristics of schools (2014). Effect sizes were calculated before rounding the adjusted difference.

* Difference in group means is statistically significant at the 0.05 level.

Table D.16. Students characteristics at endline and early reading scores at baseline, sample with non-missing baseline data

	Treatment adjusted mean	Control adjusted mean	Difference	p-value
Student is a boy	51.0	50.4	0.6	0.505
Age in years at endline (average)	7.4	7.4	0.0	0.051
Current school is same as school in baseline	91.3	93.6	-2.3*	0.047
Students not attending school at endline	0.4	0.2	0.2	0.511
Student attends second grade in endline school year	98.1	98.8	-0.7	0.170
Endline assessment took place in school	95.5	97.7	-2.2*	0.014
Overage at for first grade(percentage of students age 7 or older at baseline)	3.3	1.5	1.8*	0.022
Student repeated first grade	0.5	-0.1	0.5*	0.019
Letter identification score	18.6	18.8	-0.1	0.554
Emergent writing score	12.0	12.4	-0.4	0.135
Fluency, number of correctly read words in one minute	2.2	2.4	-0.2	0.212
Phonemic awareness score	3.5	3.4	0.1	0.364
Decoding, number of correctly read pseudo-words in one minute	1.8	2.2	-0.4	0.127
Oral comprehension score	2.8	2.8	0.0	0.667
Number of students	706	640		

Source: *Amazonía Lee* Child assessment—Baseline 2015 and Final Follow-up 2016.

Note: Regression-adjusted means ordinary least squares and include covariates to account for the design (strata variables).

* Difference in group means is statistically significant at the 0.05 level

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

APPENDIX E

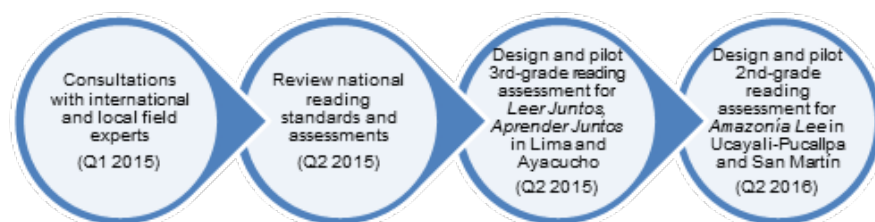
DEVELOPMENT OF READING ASSESSMENT INSTRUMENTS AND CLASSROOM OBSERVATION PROTOCOL

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

Design and pilot of the endline Early Grade Reading Assessment (EGRA)

Mathematica, in partnership with the Peru-based Grupo de Análisis para el Desarrollo (GRADE), designed and piloted the endline reading assessments for the *Amazonía Lee* impact evaluation as well as the impact evaluation of *Leer Juntos, Aprender Juntos*, a related study being conducted around the same time by the same research team in Peru. The assessment was developed in five phases as shown in Figure E1. During the first quarter of 2015, the study team consulted with Sylvia Linan-Thompson, international expert in early grade reading instruction and measurement, to draft the guidelines for the measurement of student reading outcomes at endline. We aimed to select EGRA tasks that were appropriate for the students' grade level and the Peruvian context and which would discriminate students' reading achievement levels at endline. We excluded tasks that were likely to result in floor effects among 2nd- and 3rd-graders (for example, phonemic awareness, letter identification); focused on decoding, fluency, and reading comprehension tasks; and developed reading passages that were culturally relevant for students in rural or indigenous communities. During the second quarter of 2015, in collaboration with GRADE's measurement expert in Peru, we reviewed national reading standards for the target grades, identified reading assessments previously used in Peru, and selected EGRA tasks and grade-appropriate reading passages. We piloted the 3rd-grade endline assessment in Lima and Andahuaylas (for *Leer Juntos, Aprender Juntos*) in 2015 and subsequently piloted the 2nd-grade endline assessment for *Amazonía Lee* in Ucayali and Pucallpa. The design of *Amazonía Lee*'s reading assessment was largely based on the version previously developed for third graders in the *Leer Juntos, Aprender Juntos* study, but calibrated to the skill level of second graders in the Amazon region.

Figure E.1. Process for the development of the endline reading assessment



Early grade reading standards in Peru

It was important for the students in the current study's assessment be aligned with prevailing national standards. According to the Peruvian National Curriculum for early primary grades, students should develop oral language skills (oral expression and comprehension, argumentation and debate), reading comprehension, and writing skills. With regard to reading comprehension, at the end of 2nd and 3rd grade, students are expected to be able to:

Second grade	Third grade
• Formulate reading hypotheses based on the title, abstract, and images, and verify them reading the text • Infer the meaning and function of words within sentences • Read texts orally or silently, infer the meaning, and identify sequences of events and informational content	• Understand narrative, descriptive, informative, and instructional texts • Identify the purpose of a text, formulate and verify hypotheses about it • Infer the meaning of new words • Recognize and use reading comprehension strategies • Read narrative, descriptive, informative, and instructional texts fluently • Recognize the type and content of a text • Establishes connections between a text and everyday situations

The National Assessment for reading comprehension (known as *Evaluación Censal de Estudiantes* [ECE]) is a low-stakes, group-administered, diagnostic assessment of reading comprehension. It is administered in Spanish to 2nd-graders whose mother tongue is Spanish, and to 4th-graders attending intercultural bilingual (EIB) schools who are learning Spanish as a second language. Specifically, 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).

Selection of EGRA tasks for the endline reading assessment and pilot

We reviewed aggregate national and regional data on 2013 ECE scores, Mathematica's baseline emergent reading assessment, and scores for EGRA tasks from the Young Lives study (third round 2009) to inform the design of the endline assessment and target the level of difficulty appropriate for the study regions. Young Lives is a long-term international study that seeks to improve understanding of the causes and consequences of childhood poverty and to inform the development and implementation of future policies and practices that will reduce it. Young Lives has been following two cohorts of children in Peru: one cohort of 2,000 children who were born in 2001–02 (younger cohort) and a second cohort of 1,000 in each country who were born in 1994–95 (older cohort). Children from the younger cohort of the Young Lives study were assessed with three tasks adapted from EGRA (at ages 7 to 8).³⁰ The words and texts were adapted for Peru, in many cases making them more demanding than the original EGRA texts because the Young Lives pilot test showed that more difficult items were needed to assess the full range of children's reading skills. The Young Lives EGRA included these three tasks:

- Familiar-word identification: children were presented with 60 words (randomly ordered) and asked to read them in order. Scores reflect the number of words read correctly in 60 seconds.
- Passage reading fluency and comprehension: children were presented with a short text (130 words) and asked to read it in 60 seconds. Fluency scores reflect the number of correct

³⁰ Cueto, S. and Leon, J. (2012). Psychometric characteristics of cognitive development and achievement instruments in Round Three of Young Lives. Technical note 25, 1–113. https://www.younglives.org.uk/sites/www.younglives.org.uk/files/YL-TN25_Cueto.pdf

words read correctly per minute. Subsequently, children were allowed to read the text silently and were asked eight reading comprehension questions.

- Listening comprehension: children heard a story read aloud and responded to six oral comprehension questions.

We used scores from the Young Lives sample in Andahuaylas³¹ on the EGRA tasks described above as a benchmark to select EGRA tasks for the 2nd- and 3rd-grade reading assessment.

We also reviewed reading passages from national reading comprehension exams, the Young Lives study, and reading materials produced by the Ministry of Education of Peru, and selected 26 readings to consider for use in the reading comprehension task of the endline assessment. We analyzed the level of difficulty of the reading passages, calculating the length of the sentences (average number of words per sentence), Type Token Ratio³² (degree of lexical variation), and length of the text (total number of words). Based on these three measures, we constructed a text difficulty index that ranged between 4 and 11 and allowed us to classify the readings in four categories: very easy, easy, difficult and very difficult.

We selected four subtasks for the endline reading assessment (oral comprehension, decoding pseudo-words, fluency, and reading comprehension), and designed three different assessment forms (Form A, Form B and Form C). Each form included a different set of reading passages and reading comprehension questions of three levels of difficulty (low, moderate and high). To calibrate the level of difficulty, we tested several reading fluency and reading comprehension passages that differed in topic, length, vocabulary, and sentence structure. We conducted a two-phase pilot study to test which assessment form was best targeted to the study's population. Students who participated in the pilot received one of the three assessment forms (selected randomly). The three forms were piloted with roughly the same number of students. Table E.1. shows the structure of the assessment and examples of the items used in the first phase of the pilot. The first pilot phase took place in 24 schools in Ucayali-Pucallpa with 137 2nd-grade students (randomly selected and as evenly divided by gender and across assessment forms as possible). We conducted psychometric item analysis of the reading comprehension subtask to determine which items worked best: (1) item Discrimination Index to learn the extent to which success on an item correlated with success on the whole subtask and how well the item distinguished high and low performers, (2) Cronbach's alpha to determine internal consistency reliability, and (3) Difficulty Index to measure how hard the subtest was to complete³³. We discarded items that showed a very low discrimination index, and selected reading comprehension questions of varying levels of difficulty.

³¹ 96 students in rural areas of Andahuaylas; most were attending 2nd grade at the time of the assessment.

³² Type Token Ratio (TTR) is calculated dividing the total number of different words in the text by the total number of words.

³³ Difficulty defined as the probability of a test-taker answering the question incorrectly.

Table E.1. Endline EGRA for pilot 1 (Form B)

Task	Description	Example	Scoring
Oral comprehension	Assessors read children a short story on a familiar topic and then asked children six comprehension questions about the story they heard.	<i>Ana llevaba muchos libros. De pronto, se tropezó con una silla y todos los libros se cayeron al piso. Su amigo Carlos estaba cerca y la ayudó a recoger los libros.</i>	Correct responses to oral comprehension questions. Range 0–5
Decoding (pseudo words)	Children read out loud as many pseudo-words as they could in 60 seconds. Pseudo-words are vowel-consonant combinations that follow Spanish phonological and spelling rules but are not actual meaningful words.	lete lise bofa mise garo ...	Correct pseudo-words read in 60 seconds. Range 0–50
Reading fluency	Children were asked to read a short story out loud, quickly but carefully, in 60 seconds.	<i>Un día, llegó a la comunidad un zorro que venía de tierras lejanas. El animal empezó a hacer muchas maldades. Se comía a las gallinas, robaba ovejas y hería a los perros. Viendo esta peligrosa amenaza, las hormigas reunieron a las arañas, abejas y otros insectos. Juntos hicieron un plan para expulsar al zorro. Una vez que vieron dormido al zorro, las arañas se metieron por su nariz, las hormigas por sus ojos, las abejas por sus orejas y otros insectos invadieron sus patas causándole picazón. El zorro muy fastidiado huyó al cerro y habitó allí para siempre.</i>	Correct words read in 60 seconds. Range 0–98
Reading comprehension	Children read a short story out loud, then were asked six comprehension questions (literal and inferential) that could be answered only by having read the passage. Lookbacks were not allowed.	<i>Children read a short story out loud, then were asked six comprehension questions (literal and inferential) that could be answered only by having read the passage. Lookbacks were not allowed.</i>	Correct responses to oral comprehension questions.
Passage 1	<i>Después de comer, Pedro se dio cuenta de que su perrito “Dino” no estaba en la casa. Preocupado salió a buscarlo a la calle. Le preguntó a sus vecinos pero nadie lo había visto. Finalmente, lo encontró jugando en el parque. Pedro lo abrazó y se puso muy feliz al ver que su perrito estaba bien.</i>	<i>Después de comer, Pedro se dio cuenta de que su perrito “Dino” no estaba en la casa. Preocupado salió a buscarlo a la calle. Le preguntó a sus vecinos pero nadie lo había visto. Finalmente, lo encontró jugando en el parque. Pedro lo abrazó y se puso muy feliz al ver que su perrito estaba bien.</i>	Range 0–6

Task	Description	Example	Scoring
Passage 2	<i>Un día, una hormiga se cayó a un río. Una paloma, que estaba pasando por allí, ayudó a la hormiga a salir del agua. Después, un hombre vio a la paloma y la quiso matar. La hormiga, al ver lo que el hombre quería hacer, le picó en el pie. El hombre tiró sus armas por el dolor que sentía y la paloma pudo escapar y salvar su vida.</i>	<i>Un día, una hormiga se cayó a un río. Una paloma, que estaba pasando por allí, ayudó a la hormiga a salir del agua. Después, un hombre vio a la paloma y la quiso matar. La hormiga, al ver lo que el hombre quería hacer, le picó en el pie. El hombre tiró sus armas por el dolor que sentía y la paloma pudo escapar y salvar su vida.</i>	Range 0–6
Passage 3	<i>Un gallo vivía en un árbol de mango, desde donde cantaba todas las madrugadas. Un día, llegó un lorito que no tenía casa y se acostó en el árbol del gallo. Aquella noche, antes de dormir, el lorito advirtió seriamente al gallo:</i> - <i>Amigo gallo, estoy cansado y quiero dormir. No vayas a cantar.</i> - <i>Lo siento, lorito. Yo debo dar la hora a mi dueño, por eso canto todas las madrugadas.</i> - <i>Entonces, voy a decirle a tu dueño que te cocine.</i> - <i>Lo siento, lorito. Yo ya soy un gallo viejo, nadie querrá comerme.</i> - <i>Si haces bulla, te coseré el pico -dijo el lorito.</i> - <i>Lo siento, lorito. Yo igual debo cantar en la madrugada.</i> <i>El lorito estaba tan ofuscado por las respuestas del gallo que toda esa noche se la pasó renegando y no pudo dormir. Poco antes del amanecer, el lorito escuchó un ruido y vio un zorro hambriento a punto de devorarlo. De pronto, el gallo despertó y empezó a cantar: ¡Kikirikiii!</i> <i>El zorro se espantó y se escapó de inmediato. El lorito entonces le dijo al gallo que su canto le había salvado la vida. Así, decidió irse a vivir a otro árbol, para no molestar más al gallo que lo había salvado.</i>	<i>Un gallo vivía en un árbol de mango, desde donde cantaba todas las madrugadas. Un día, llegó un lorito que no tenía casa y se acostó en el árbol del gallo. Aquella noche, antes de dormir, el lorito advirtió seriamente al gallo:</i> - <i>Amigo gallo, estoy cansado y quiero dormir. No vayas a cantar.</i> - <i>Lo siento, lorito. Yo debo dar la hora a mi dueño, por eso canto todas las madrugadas.</i> - <i>Entonces, voy a decirle a tu dueño que te cocine.</i> - <i>Lo siento, lorito. Yo ya soy un gallo viejo, nadie querrá comerme.</i> - <i>Si haces bulla, te coseré el pico -dijo el lorito.</i> - <i>Lo siento, lorito. Yo igual debo cantar en la madrugada.</i> <i>El lorito estaba tan ofuscado por las respuestas del gallo que toda esa noche se la pasó renegando y no pudo dormir. Poco antes del amanecer, el lorito escuchó un ruido y vio un zorro hambriento a punto de devorarlo. De pronto, el gallo despertó y empezó a cantar: ¡Kikirikiii!</i> <i>El zorro se espantó y se escapó de inmediato. El lorito entonces le dijo al gallo que su canto le había salvado la vida. Así, decidió irse a vivir a otro árbol, para no molestar más al gallo que lo había salvado.</i>	Range 0–7

Overall, findings from the pilot suggested that easier items or tasks were needed to capture the skills of lower-performing students who had not yet mastered foundational reading skills. For example, 60 of 137 students were unable to complete the fluency task and were not given the

reading comprehension task to attempt. Another 3 students attempted and were unable to complete the reading comprehension task. Duration of the assessment was slightly higher than expected. Average administration time was approximately 30 minutes for Form A, 19 minutes for Form B, and 18 minutes for Form C. Based on these findings, we adjusted the endline assessment forms in three ways: we (1) included simple word reading tasks, (2) chose simpler and shorter reading vignettes, and (3) reduced the length of the reading comprehension section to one passage. We revised and piloted two new assessment forms with the best-performing reading passages and items. We then conducted the second phase of the pilot to test modifications to the assessment. The second pilot phase took place in 12 schools in San Martin with 152 2nd-grade students (randomly selected and as evenly divided by gender and assessment forms as possible). We conducted a second iteration of psychometric analysis and chose the items from the final version of the endline assessment. The reading passage and reading comprehension questions selected for the final version of the assessment showed good discrimination indices (ranging from .34 to .70) and good internal consistency reliability (Cronbach's $\alpha = .72$). Table E.2. shows the structure and content of the reading assessment we used to measure student reading outcomes at endline.

Table E.2. Amazonía Lee Endline EGRA

Task	Description	Example	Scoring
Oral comprehension	Assessors read children a short story on a familiar topic and then asked children six comprehension questions about the story they heard.	<i>María quiere hacerle un lindo regalo a su mamá. Su cumpleaños se acerca y quiere darle una sorpresa. Por la tarde, luego de salir de la escuela, María fue a buscar semillas rojas y negras y las metió en una bolsa. Cuando llegó a su casa, ella hizo un bonito collar de semillas para su mamá.</i>	Correct responses to oral comprehension questions. Range 0 – 6
Familiar word reading	Children read a list of words they likely had seen before and were expected to read at their grade level. <i>Stop rule:</i> The task was discontinued if the child failed to read the first five words correctly.	sol, casa, toma, niño, pelota, rojo, puerta, escribe, comida, triste	Correct words read. Range 0 -10
Decoding (pseudo-words)	Children read out loud as many pseudo-words as they could in 60 seconds. Pseudo-words are vowel-consonant combinations that follow Spanish phonological and spelling rules but are not actual meaningful words. <i>Stop rule:</i> The task was discontinued if the child failed to read the first five pseudo-words.	Lete lise bofa mise garo...	Correct pseudo-words read in 60 seconds. Range 0–50

Task	Description	Example	Scoring
Reading fluency	Children were asked to read a short story out loud, quickly but carefully, in 60 seconds.	<i>Pepe es un león que vive en la selva junto a los demás leones de su manada.</i> <i>Pepe es el jefe de su manada, por eso defiende a los demás leones y marca su territorio. Cuando las leonas salen a cazar, Pepe y los otros leones protegen a los cachorros mientras ellos juegan y aprenden. Los leones son dormilones. Pueden dormir hasta veinte horas al día. A Pepe le gusta dormir sobre la rama de un árbol.</i>	Correct words read in 60 seconds. Range 0–77
Reading comprehension	Children read a short story out loud, then were asked six comprehension questions (literal and inferential) that could be answered only by having read the passage. Lookbacks were not allowed.	Children read a short story out loud, then were asked six comprehension questions (literal and inferential) that could be answered only by having read the passage. Lookbacks were not allowed.	Correct responses to oral comprehension questions.
Passage 1	<i>Un pollito amarillo caminaba por un puente. De repente se resbaló y cayó al río.</i> - <i>¡Ayúdenme, no sé nadar! -gritó.</i> - <i>Pasaba por ahí un ratón que vio al pollito y fue a ayudarlo pero él también se cayó al agua.</i> - <i>¿Y ahora qué hacemos? Yo tampoco sé nadar -dijo el ratón.</i> - <i>Veo algo que flota hacia nosotros -respondió el pollito.</i> - <i>Los dos se subieron al tronco y remaron juntos hacia la orilla. Cuando por fin llegaron a tierra, ambos gritaron emocionados:</i> <i>¡Estamos salvados!</i>	<i>Un pollito amarillo caminaba por un puente. De repente se resbaló y cayó al río.</i> - <i>¡Ayúdenme, no sé nadar! -gritó.</i> - <i>Pasaba por ahí un ratón que vio al pollito y fue a ayudarlo pero él también se cayó al agua.</i> - <i>¿Y ahora qué hacemos? Yo tampoco sé nadar -dijo el ratón.</i> - <i>Veo algo que flota hacia nosotros -respondió el pollito.</i> - <i>Los dos se subieron al tronco y remaron juntos hacia la orilla. Cuando por fin llegaron a tierra, ambos gritaron emocionados:</i> <i>¡Estamos salvados!</i>	Range 0–6

Classroom observation protocol

The classroom observation instrument was designed to measure of the literacy classroom environment and literacy instruction practices teachers used at baseline and endline. The observation protocol included aspects of teachers' instructional practices, proportion of time spent in literacy-related activities, and classroom characteristics. To capture teachers' instructional practices, we used a checklist and rating format that could be completed in a one-hour classroom observation. To capture class time spent on literacy-related activities, we used a time-sampling approach in which observers recorded presence or absence of specific literacy-related activities during 5 minute intervals throughout the communication class. The classroom observation protocol was not intended as a resource-intensive measure of instructional quality with multiple observation periods to assess teacher performance. Table E.3. describes the sections and characteristics of the classroom observation protocol and school infrastructure.

Table E.3. Description of classroom and school observation instrument

<i>Leer Juntos, Aprender Juntos</i> observation instrument	Description	Instrument sources
Classroom overview	Teacher's gender and contract type, content subject area observed, number of students, classroom layout, students' seating arrangement, and intercultural bilingual education approach	OLLI
Time use	Teacher's use of class time in different types of literacy-related activities: alphabet knowledge (e.g., identifying letters of the alphabet), phonologic awareness and knowledge (e.g., identifying initial word sounds), nonprint oral language use (e.g. narrating traditional oral stories or personal experience), print-related language and reading (e.g., silent reading, oral reading); vocabulary development (e.g., find synonyms); reading comprehension (e.g., identify main characters and story events); writing (e.g., text production, dictation) and other opportunities for reading practice. For each activity type, observers recorded the duration and structure (whole-class, group, or pair) of the activity.	LBTO LISn ELLE ^a
General instructional practices	Extent to which teachers used general instructional practices like checking for understanding, giving clear instructions, providing feedback, and facilitating cooperative learning activities.	CETT UPCH
Literacy –related instructional practices	Extent to which teachers used literacy instruction practices: book-reading instruction, fluency, comprehension and vocabulary development, phonemic awareness and phonics, oral language development, and general teaching practice.	LBTO CETT UPCH
Classroom literacy environment	Availability of books and print materials, variety and overall quality of books and materials, types of print materials (commercial, or student- or teacher-made), variety of books (informational text versus narrative), and availability of writing resources.	ELLE
Classroom management supports	Availability of student rosters and attendance displays, weekly student role assignments and responsibilities, student behavior rules, calendar and announcements.	UPCH

Note: OLLI: Observation of Language and Literacy Instruction (Mathematica Policy Research 2012); LBTO: Literacy Boost Teacher Lesson Observation (Save the Children Malawi, SUPER Fellow 2011); LISn: Language Interaction Snapshot (Atkins-Burnett et al. 2010a); ELLE: Early Language and Literacy Environment (Atkins-Burnett et al. 2010b); CETT: Centers of Excellence for Teachers Best Practices Rating Form (Culver et al. 2006). UPCH, Centro Andino de Excelencia para Capacitación de Maestros training guides.

^a The ELLE instrument was adapted with permission from (1) the Early Language and Literacy Classroom Observation Tool (ELLCO)— Research Edition (Smith et al. 2002); (2) the Child/Home Early Language and Literacy Observation Tool (CHELLO) (Neuman et al. 2007); and (3) the ELLCO Addendum (Castro 2005).

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

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

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