



EVALUATION REPORT

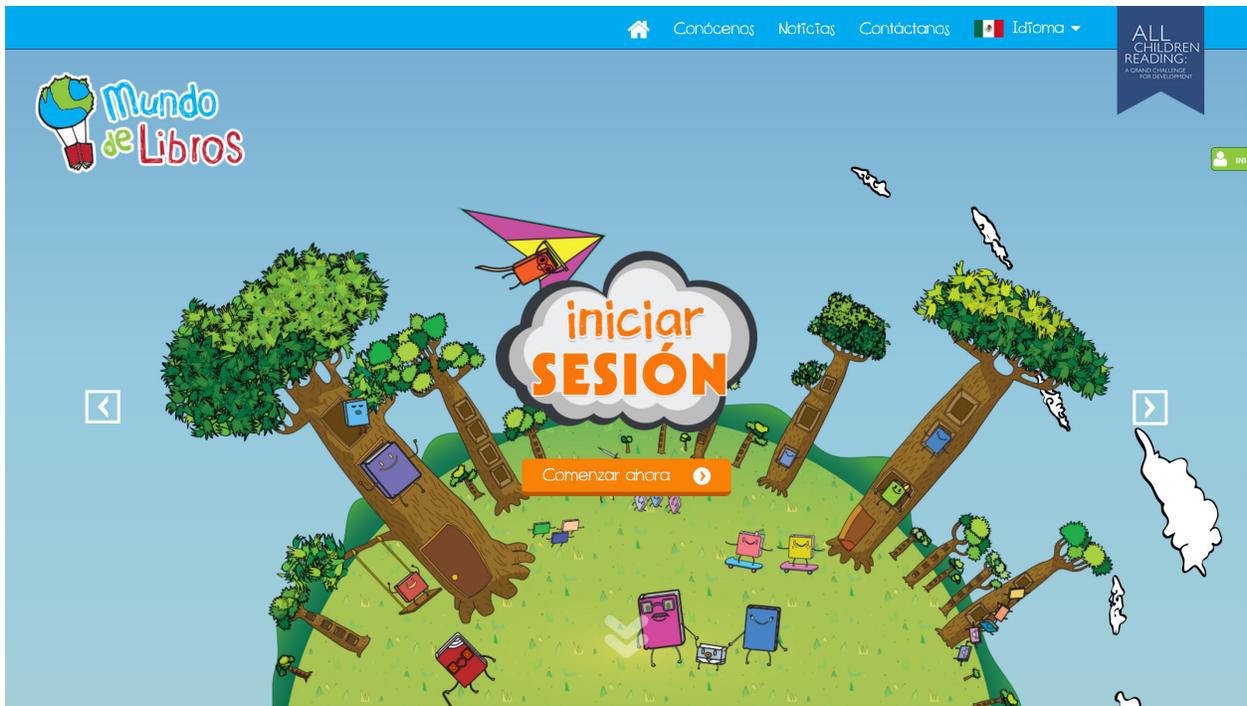
Mundo de Libros: Matching Children with Level-Appropriate Books and Engaging Families

Implemented by Qué Funciona para el Desarrollo in Mexico

OCTOBER 2017

Prepared by
School-to-School International (STS)
For All Children Reading: A Grand Challenge for Development





Mundo de Libros: Matching Children with Level-Appropriate Books and Engaging Families

Implemented by Qué Funciona para el Desarrollo in Mexico

Table of Contents

	List of Acronyms	4
I.	Executive Summary	5
	Key Findings	6
II.	Project Description	7
III.	Research Purpose and Design	9
	Sample	10
IV.	Fieldwork Preparation and Data Collection	12
	EGRA Instrument	12
	Additional Instruments	12
	Institutional Review Boards	13
	Baseline EGRA	13
	Endline EGRA	14
	End-of-Project Interviews	14
V.	Project Implementation	15
	Development	15
	Implementation	16
	Management	17
	Fidelity of Implementation	18
	Project Dosage	19
VI.	EGRA Data Analysis	22
	Considerations	24
VII.	EGRA Results	25
	EGRA Results by Subtask	26
	EGRA Results by Subgroup	34
	Key Factors for Success	35
VIII.	Adaptive Subtask Results	38
IX.	Scalability	41
	Relevance	41
	Credibility	42
X.	Cost Analysis	43
XI.	Conclusions	47
XII.	Annexes	50
	Annex A: Baseline and Endline EGRA Instrument	51
	Annex B: Reading Habits and Attitudes Survey	62
	Annex C: Reading Habits and Attitudes Survey Results and Composites	66
	Annex D: EGRA Descriptive Statistics and Additional Tables	71
	Annex E: EGRA Reliability Results	77

List of Acronyms

ACR GCD	All Children Reading: A Grand Challenge for Development
AORF	Adaptive Oral Reading Fluency
CFWPM	Correct Familiar Words per Minute
CLSPM	Correct Letter Sounds per Minute
CNWPM	Correct Nonwords per Minute
CWPM	Correct Words per Minute
EGRA	Early Grade Reading Assessment
EOP	End-of-Project
FOI	Fidelity of Implementation
GoM	Government of Mexico
IBSA	Información, Bibliotecas y Sistemas Avanzados
IRB	Institutional Review Board
M&E	Monitoring and Evaluation
MdL	Mundo de Libros
NICRA	Negotiated Indirect Cost Recovery Agreement
OECD	Organization for Economic Co-operation and Development
ORF	Oral Reading Fluency
PISA	Program for International Student Assessment
PNL	National Basic Education Reading Program
QFD	Qué Funciona para el Desarrollo
SD	Standard Deviation
SES	Socioeconomic Status
STS	School-to-School International
TVIP	Peabody Picture Vocabulary Test
USAID	United States Agency for International Development

I. Executive Summary

All Children Reading: A Grand Challenge for Development (ACR GCD)—a partnership between the United States Agency for International Development (USAID), World Vision, and the Australian Government—is an ongoing series of grant and prize competitions that leverage science and technology to source, test, and disseminate scalable solutions to improve literacy skills of early grade learners in developing countries. Round 2 of ACR GCD, which started in 2014 and continues through 2017, supports technology-based innovations to improve early grade reading outcomes in developing countries.¹ These technology-based innovations concentrate on three focus areas:

1. Mother tongue instruction and reading materials
2. Family and community engagement
3. Children with disabilities

ACR GCD Round 2 increased its focus on the assessment of early grade reading skills to understand the ability of technology-based innovations to improve the literacy skills of early grade learners. To measure this, ACR GCD uses the Early Grade Reading Assessment (EGRA) to systematically assess reading skills across all Round 2 grantees. The EGRA is an oral assessment that measures students' most basic foundational literacy skills in the early grades—specifically, recognizing letters of the alphabet, reading simple words, understanding sentences and paragraphs, and listening with comprehension. The EGRA methodology was developed under EdData II and has been applied in more than 30 countries and 60 languages.² The EGRA instruments used by ACR GCD grantees were adapted to reflect the specific context of each grantee's project, including adaptations for students who have low vision or are blind and students who are deaf or hard of hearing.

Qué Funciona para el Desarrollo (QFD)—an ACR GCD Round 2 grantee—implemented the Mundo de Libros (MdL) project in collaboration with Fundación Proacceso. The MdL project aimed to improve reading skills—specifically, foundational skills—and reading habits of students enrolled in Grades 1 through 3 in Estado de México (the State of Mexico). QFD developed an innovation, the MATCH algorithm specifically for this project, which assesses students' literacy level and the difficulty level of books to make personalized reading recommendations. They also built a web-based platform to share these recommendations via individual student profiles. In addition, the project granted students access to libraries that offered a high-quality selection of children's books; developed a system for leveling books; and provided workshops for parents and caretakers to support children's reading.

The MdL project began in February 2015, and implementation under the initial contract concluded in February 2017.³ To understand how the project impacted participating students' reading skills and habits, School-to-School International (STS) and QFD conducted EGRAs twice during the project. Baseline data were collected from December 2015 to April 2016,⁴ and endline data were collected from January to February 2017. EGRAs were supplemented by a reading habits and attitudes survey that explored changes in students' behaviors.

During the endline data collection, STS also conducted end-of-project (EOP) interviews with the MdL project staff, librarians, parents, and students. The interviews were designed to explore any lessons learned from the project's implementation, elicit data that would lead to better understandings about how the project impacted students, and allow funders and researchers to assess the potential scalability of the MdL project.

¹ All Children Reading. (2017, June). About us. Retrieved from <http://allchildrenreading.org/about-us/>

² EdData II was a contract mechanism funded by USAID from January 1, 2004 to December 31, 2013. Implemented by RTI International, the purpose of EdData II was to improve the accuracy, timeliness, accessibility, and use of data for education policy and program planning. See <http://www.rti.org/sites/default/files/brochures/eddataii.pdf> for additional details

³ The project contract ended on April 15, 2017.

⁴ Baseline data were collected in seven libraries in November and December 2015. Baseline data were collected in three additional libraries from February to April 2016.

The following report presents a summary of lessons learned from project implementation, EGRA results, survey results, and scalability assessments.⁵

Key Findings

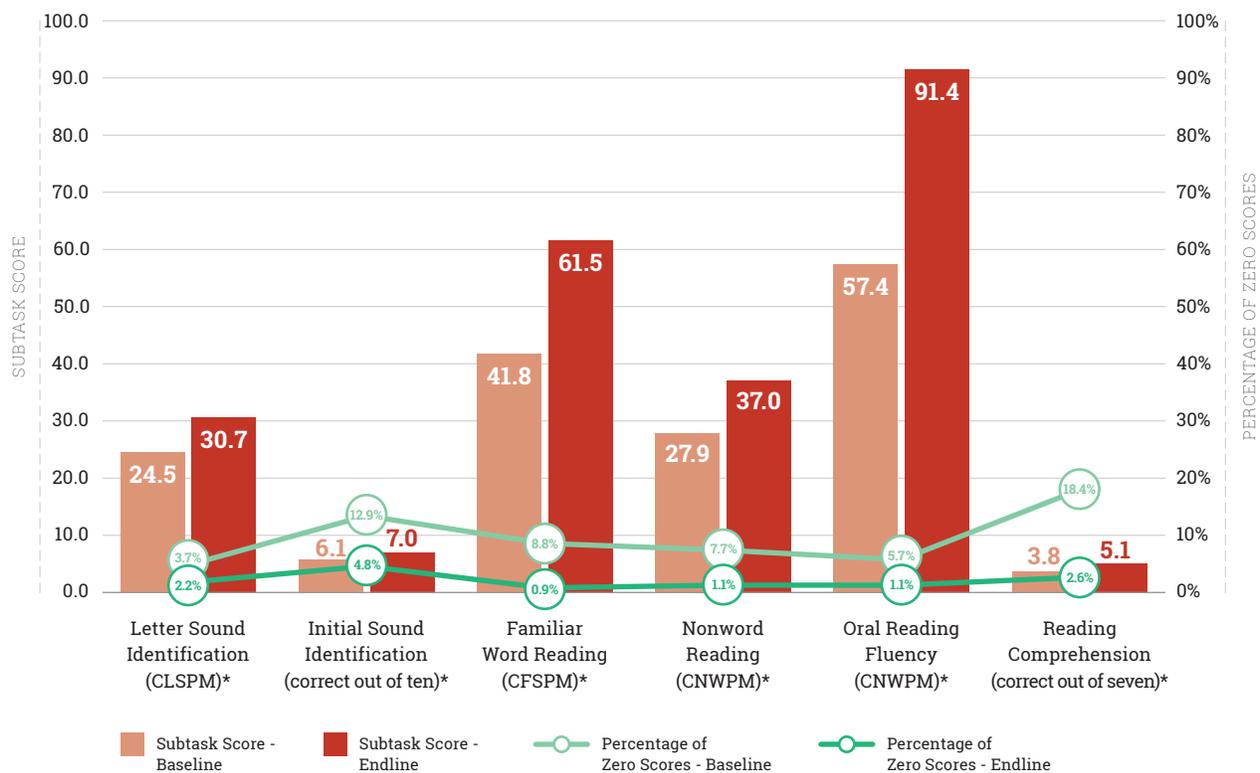
- **Students who participated in the MdL project showed significant reading gains across all EGRA subtasks; however, it is not possible to assess whether the reading gains experienced by MdL participants were associated with the intervention.** The greatest gains from baseline to endline assessment were observed on the oral reading fluency (ORF) and familiar word reading subtasks; students gained an average of 34.0 correct words per minute (CWPM) and 19.6 correct familiar words per minute (CFWPM). Also, the proportion of students receiving zero scores dropped significantly across subtasks. The greatest decrease was seen on the reading comprehension subtask, where the percentage of students receiving zero scores dropped 15.8 percentage points from 18.4 percent at baseline to only 2.6 percent at endline. Grade 1 students showed greater reading gains than their peers in Grades 2 or 3, but there was no difference in reading gains between boys and girls who participated in the project. Because the project's research design did not include a comparison group of students who received no exposure to the project,⁶ and because the project had low registrant uptake and fidelity of implementation (FOI), it is not possible to know whether gains from baseline to endline were due to the intervention or an additional year of schooling.
- **The MdL project's participation rate and FOI were low.** Of the students who enrolled in the MdL project, 39.4 percent did not check out a single library book, and 66.5 percent did not log on to the MdL web-based platform. Only 19.7 percent of students who attended libraries that offered workshops had a parent or caretaker attend a workshop. On average, students checked out 12.7 books and logged on to the MdL web-based platform 1.2 times during the project's implementation period.
- **Student participation in the MdL project varied widely across the ten digital libraries.**⁷ As part of the MdL project, QFD supplied digital libraries with more than 720 copies of 295 unique children's books titles. The average number of books checked out per student by library ranged from 2.8 books to 44.0 books over the project's implementation period. The average number of platform logins per student by library ranged from 5.9 logins to 0.3 logins during the implementation period. For both book checkouts and platform logins, the library with the highest student average was at least two times greater than the next most active library.
- **QFD piloted two innovative technology components in the MdL project: an algorithm that provides personalized book recommendations to students and a web-based platform that allows students to access their recommendations.** Unfortunately, development and implementation challenges did not allow for a conclusive understanding of the potential reading gains associated with these technologies. QFD coped with challenges throughout the development of these technologies and, as a result, learned what may be needed to improve these components as well as the quality of implementation in the future.

5 The MdL project continued to enroll students into the project after the initial baseline data collection. In total, the project reached 856 students, although results presented in this report are only for 457 students who were assessed during the original baseline and at endline.

6 The QFD research design included a group of students (intervention D) who did not receive personalized book recommendations nor were their parents and caretakers offered workshops. Although these students did not receive two of the key components of the project, they still had access to the libraries, books, and web-based platform. Therefore, they do not represent a comparison group that demonstrates the absence of the project's interventions. See Research Purpose and Design.

7 Digital libraries are community spaces to consult and create digital content, access information, read, learn, and meet academic, personal, professional, and social needs. They are equipped with computers and tablets with internet access; however, before the MdL project, they had no physical stock of books nor library furniture.

Figure 1: Mean Results⁸ by EGRA Subtask⁹



II. Project Description

In Mexico, many primary school students are not reading at grade level.¹⁰ School curricula¹¹ and teaching practices often follow a “one size fits all” approach to teaching literacy, and schools lack a systematic method to grant students access to appropriate materials at appropriately precise (within grade) reading levels.¹² Additionally, many parents are often unsure how best to help their children learn to read.¹³ Evidence suggests that children are encouraged in learning when parents play an active role in reading activities and when books are tailored to their reading level and interests.^{14, 15, 16, 17, 18, 19}

⁸ Results in the figures are based on the combined results of students in Grade 1, Grade 2, and Grade 3 unless otherwise noted.

⁹ An asterisk (*) indicates that the endline subtask mean was significantly higher than the baseline subtask mean and that the percentage of zero scores at endline was significantly lower than at baseline at $p < 0.05$. $N = 457$.

¹⁰ Díaz, M. A., & Flores, G. (Organizadores). (2010). México en PISA 2009. *México: Instituto Nacional para la Evaluación de la Educación. Explorador Excale—3o de Primaria* (Excale database for third grade students in Mexico). <http://www.inee.edu.mx/explorador/>

¹¹ The national curricula focus on reading decoding and fluency in Grades 1 and 2, and on reading comprehension starting in Grade 3.

¹² For instance, the National Reading Program (PNL in Spanish) roughly classifies some books by grade or education level.

¹³ Ortega Hesles, M. (2012). Learning from the Pilot Study of a Cluster Randomized Trial: Summer Reading Interventions Targeting Third Grade Students in Mexico (unpublished qualifying paper, Harvard Graduate School of Education).

¹⁴ Allington, R. L., & McGill-Franzen, A. (1989). Different programs, indifferent instruction in *Beyond Separate Education: Quality Education for All*, edited by Dorothy K. Lipsky and Alan Gartner. Baltimore, MD: Paul H. Brookes.

¹⁵ Worthy, J. (1996) A matter of interest: Literature that hooks reluctant readers and keeps them reading in *The Reading Teacher*, 50(3), 204–212.

¹⁶ Allington, R. L. (2002). You can't learn much from books you can't read in *Educational Leadership*, 60(3), 16–19.

¹⁷ Snow, C. (2002) Reading for understanding: Toward an R&D program in reading comprehension. *Santa Monica, CA: Science and Technology Institute, RAND Education*. Retrieved from https://www.rand.org/pubs/monograph_reports/MR1465.html.

¹⁸ National Research Council. (1998). Preventing reading difficulties in young children. *Washington, DC: The National Academies Press*. Retrieved from <https://doi.org/10.17226/6023>.

¹⁹ Lin, Q. (2003, October). Parent involvement and early literacy. *Global Family Research Project (formerly Harvard Family Research Project)*. Retrieved from <http://www.hfrp.org/publications-resources/browse-our-publications/parent-involvement-and-early-literacy>.

To address these challenges, QFD, a nonprofit organization with a deep history in conducting research studies of social and economic development programs in Mexico, developed the MdL project. Implemented in collaboration with Fundación Proacceso, a local nonprofit organization that provides technological education for social and economic development in underserved communities in Mexico,²⁰ the MdL project aimed to improve the reading skills and habits of students in Grades 1 through 3 in Mexico.

The project had three core components:

- 1. Access to digital libraries that offered a high-quality collection of children's books.** Students received a project passport—similar to a library card—that allowed them to check out books and keep track of due dates.
- 2. A web-based MdL platform that provided book recommendations to students through individual platform profiles.**²¹ Depending on the students' group assignment (see Research Purpose and Design), their profile on the MdL platform gave them either personalized book recommendations appropriate to their vocabulary level and reading skills or a random selection of book recommendations. Personalized recommendations were determined by QFD's MATCH algorithm, which considered both the student's reading level at baseline and each book's level of difficulty.
- 3. Workshops and materials for parents or caretakers.** Depending on the libraries' group assignment (see Research Purpose and Design), parents and caretakers of participating students were invited to attend workshops hosted at the libraries and to receive supplemental materials. The main objectives of these workshops and materials were to promote parent and caretaker engagement, provide information and strategies to improve children's reading practices, and advise on how to create a rich literacy environment at home.

The MdL project incorporated several technological elements, including an algorithm that provides students with personalized book recommendations and the use of a web-based platform accessed through computers and tablets.

The MdL project was implemented in ten digital libraries operated by Fundación Proacceso in Estado de México. Fundación Proacceso operates a total of 50 digital libraries throughout Estado de México; these buildings are located within or adjacent to public education facilities and are equipped with desktop computers and tablets for use by members of the community. Library services, accessed for free, frequently host computer and languages classes or reading clubs for children and adults. Prior to the project the digital libraries did not contain any books. Therefore, the MdL project purchased and provided each participating library with more than 720 copies of 295 unique children's books titles. The MdL book catalog was diverse in terms of difficulty and topics, thus ensuring that every student had a variety of choices. The MdL project also created child-friendly spaces within the digital libraries, including adding bookshelves, floor mats, signage, and chairs. Librarians, who are Fundación Proacceso staff members and not employed by QFD, were responsible for managing the book catalog and monitoring student book checkouts and returns using Koha, a free and open-source integrated library system.²² Students could check out up to two books at a time and were not required to access the MdL platform prior to checking out books.

For the second component of the MdL project, QFD created individual usernames and profiles for every student; this allowed them to access the web-based platform. After logging on to the platform, students could choose an avatar,²³ see reading recommendations, filter book titles according to interests, and search for specific titles,

20 Fundación Proacceso. (2017, June). Quiénes somos. Retrieved from <http://www.proacceso.org.mx/index.php/quienes-somos/>

21 Individual profiles are accessible by logging in with a unique username and password to <http://www.mundodelibros.mx>.

22 See <https://koha-community.org> for more details.

23 Avatars on the MdL platform were icons that represented the students.

authors, or keywords. The platform also allowed students to rate books they had read on a scale of one to five after returning them. The key technological innovation of the MdL project, and a core part of the platform component, was the MATCH algorithm developed specifically for this project. The MATCH algorithm provided students with personalized book recommendations based on two factors: the student's reading level as measured through the baseline EGRA and the books' difficulty score as defined by QFD. Book difficulty scores were defined using quantitative parameters—such as sentence length, word length, and words per sentence—and qualitative parameters—such as graphics, content, and style. A panel of reading experts assessed 30 books to determine a set of criteria associated with a book's difficulty level. QFD then tested and selected a formula that was used to assign a difficulty level to all the MdL books.

For the third component of the project, QFD hosted workshops for parents and caretakers at participating libraries every two months for a total of five workshops over the course of the MdL project. The MdL team members led the workshops which centered around a variety of topics, including good reading habits (e.g. recommended reading session length and activities to do before and after reading), the importance of summer reading, and reading resources available outside of school. QFD distributed informational handouts during each workshop that attendees could distribute to others in the household. Parent or caretaker attendance at workshops was tracked through attendance forms and subsequently entered in Koha.

Project implementation began in seven libraries in January 2016. Three libraries that were part of the original cohort of ten were replaced due to low enrollment in the MdL project. Baseline assessments were conducted at those locations starting in March 2016; implementation in the replacement libraries started shortly thereafter.

III. Research Purpose and Design

The goal of the MdL project was to improve the reading skills and habits of students in Grades 1 through 3 through parental engagement, access to a digital library with a high-quality collection of children's books, and a web-based platform that matched students with books based on their reading level. The research conducted by STS and QFD sought to answer the following research questions specific to the MdL project:

1. Does access to book recommendations based on the MATCH algorithm improve vocabulary, reading scores, and reading habits of early grade readers?
2. Do workshops improve parental engagement in their children's reading?
3. Does access to workshops for parents improve their children's vocabulary, reading scores, or reading habits?

In addition, EOP research was conducted to answer the following ACR GCD supplemental questions common to all ACR GCD grantees:

1. How successful was the rollout of the project?
2. How did the project influence or impact adults' (teachers, parents, community members) knowledge, skills, or attitude regarding their role in helping children read?
3. How did the project influence certain subsets of the student population more than others based on identifiable contextual factors?
4. How much did the development, implementation, and management aspects of the project cost?
5. Are the project and technology suitable for scaling?

To examine the initial research questions, STS and QFD collected EGRA data twice during the project. During the baseline and endline data collection, assessors also administered a reading habits and attitudes survey to students. Qualitative and cost data were collected to answer ACR GCD's supplemental questions.

Because of challenges during the development of the MATCH algorithm, technical difficulties with the MdL platform, and low project participation and implementation fidelity, the research questions cannot be answered as stated. Therefore, instead of assessing gains associated with specific components or the entire MdL project, this report focuses on average changes in students' reading skills as a result of exposure to the MdL project and an additional year of schooling. See the Sample, Project Implementation, and Data Analysis sections for more details.

Sample

QFD used a multistep selection process to choose libraries and students for participation in the MdL project. To begin, QFD conducted an online survey of all 50 libraries run by Fundación Proceso in Estado de México and used those results to establish two selection criteria: (a) libraries must have at least two librarians working full time, and (b) libraries must have reliable internet connectivity. The selection criteria identified 14 libraries as eligible for selection. QFD randomly selected ten of those 14 to participate in the project.

QFD promoted the MdL project at the libraries and in nearby primary schools to recruit students. Promotion efforts in schools included five-minute presentations to school principals and teachers to explain the project and distribution of flyers to school staff. The QFD team also made flyers available for parents, caretakers, and students at participating libraries. Interested students visited a participating library to receive registration materials. These materials included a description of the program, a copy of QFD's privacy policy (required by law), a consent form, and a registration form that asked for contact and sociodemographic information for the student and his or her parents or caretakers. Students self-selected to participate and were registered on a first come, first served basis.

QFD conducted baseline data collection on the initial sample of 459 students from December 2015 to January 2016 in seven libraries. In the three remaining libraries, fewer than ten children enrolled. As a result, these libraries, and the children who registered with them, were removed from the sample and replaced with three new libraries. QFD repeated the recruitment process and collected baseline data from the registered students at the newly added libraries from February to April 2016.

The MdL project's original research design consisted of four different groups:²⁴

- **Intervention A:** Students receive personalized book recommendations on the MdL platform through the MATCH algorithm, and libraries offer workshops for parents and caretakers.
- **Intervention B:** Students receive personalized book recommendations on the MdL platform through the MATCH algorithm, and libraries do not offer workshops for parents and caretakers.
- **Intervention C:** Students receive a random selection of book recommendations on the MdL platform (i.e. no MATCH algorithm), and libraries offer workshops for parents and caretakers.
- **Intervention D:** Students receive a random selection of book recommendations on the MdL platform (i.e. no MATCH algorithm), and libraries do not offer workshops for parents and caretakers.

Table 1 provides a summary of component assignment by intervention group.

²⁴ The project intended to evaluate the effects of book recommendations through the MATCH algorithm and parent workshops, not access to books. For this reason, and because of the difficulty in constructing a comparison group given the project design, STS and QFD determined that a comparison group without access to libraries and books should not be part of the research design.

Book recommendations were assigned at the student level. Students across all ten libraries were randomly assigned to receive either personalized book recommendations based on the MATCH algorithm (intervention A and B) or random book recommendations (intervention C and D). All students, regardless of their intervention group assignment, could access the web-based platform and check out books from the library.

Workshops for parents and caretakers were assigned at the library level. Five of the ten libraries selected for participation were randomly assigned to offer workshops for parents and caretakers; all students who registered with those libraries were assigned to either intervention A or C.

A total of 575 students were assessed at baseline, and 457 students were assessed at endline. Table 2 provides characteristics of the final student sample used for reporting.²⁵

STS, with support from World Vision, conducted EOP interviews from January 30 to February 9, 2017. Interviews explored contextual factors that may have impacted the project's implementation and student reading gains. Responses also identified considerations for the future scalability of the project. EOP interview details are provided in Table 3.

Table 1: Research Design of Project Groups

Project Group	Students received MATCH algorithm recommendations	Libraries offered workshops to parents and caretakers
Intervention A	Yes	Yes
Intervention B	Yes	No
Intervention C	No	Yes
Intervention D	No	No

Table 2: EGRA Sample Characteristics²⁶

Characteristic	Number of Students	
Grade at baseline	Grade 1	145
	Grade 2	174
	Grade 3	138
Gender	Girls	242
	Boys	215
Group	Intervention A	144
	Intervention B	81
	Intervention C	146
	Intervention D	86

Table 3: EOP Interview Sample

Type of Interview	N	Description
Project management	5	Five QFD staff members
Stakeholder	1	One representative of Fundación Proacceso
Librarian	7	Seven librarians from five different libraries
Parent and caretaker	38	Mothers, fathers, and grandmothers with children registered at the same five libraries
Student	14	Participating students from the same five libraries
School principal	6	Six principals from primary schools that have students registered at one of the same five libraries
Total	71	

²⁵ In total, the project reached 856 students. Results presented are for those students assessed in the original baseline data collection.

²⁶ Due to low project uptake and FOI during the pilot year, the analysis in this report does not compare results between intervention groups as originally intended in the research design. See Project Implementation, Data Analysis, and EGRA Results for more details.

QFD purposively selected five libraries to conduct EOP interviews with an eye toward representing a wide range of characteristics, including the location's degree of urbanization (urban and rural), student engagement levels in the library (high and low), and parent workshop assignment (offered and not offered). All librarians who were present at the five selected libraries on the day of EOP interviews were interviewed, and any parent, caretaker, or student who visited the selected libraries on the day of EOP interviews were asked to participate. Assessors also administered endline EGRAs to students during those visits.

Additionally, STS and World Vision visited primary schools near the five libraries to conduct interviews with the schools' principals. At one school, teachers and a reading coach were invited by the school's principal to participate in a focus group discussion. Project management interviews were also conducted with QFD staff members and a representative of Fundación Proacceso.

IV. Fieldwork Preparation and Data Collection

EGRA Instrument

The Spanish EGRA instrument used in assessing the MdL project was adapted from an existing EGRA. In June 2015, the psychometrics consulting firm MetCuantus conducted an adaptation workshop to update the existing EGRA for the MdL project. MetCuantus and QFD pilot tested the MdL's Spanish EGRA instrument with 225 students in one preschool and three public primary schools in Estado de México. The final MdL EGRA instrument was administered at both baseline and endline.²⁷

The EGRA used in the MdL project consists of six standard subtasks: letter sound identification, initial sound identification, familiar word reading, nonword reading, ORF, and reading comprehension. QFD added two additional subtasks to the EGRA instrument: adaptive oral reading fluency (AORF) and adaptive reading comprehension. The intent of these additional adaptive subtasks was to better differentiate the reading comprehension level of students by providing two different reading passages in addition to the ORF passage; the first passage was shorter and easier while the second passage was longer and more difficult. Depending on a student's performance on the ORF and reading comprehension subtasks, he or she was routed to one of the AORF passages and asked corresponding questions in the adaptive reading comprehension subtask. QFD compared performance on the timed ORF and reading comprehension subtasks with performance on the untimed AORF and adaptive reading comprehension subtasks in hopes of better understanding students' reading fluency and comprehension levels (see Table 8 and 9 for additional details).

Additional Instruments

QFD and MetCuantus designed a reading habits and attitudes survey that was administered at baseline and endline. The survey includes three yes-or-no questions and 14 questions that use a four-point Likert scale; both sets of questions focus on personal and family literacy behaviors.

The Peabody Picture Vocabulary Test (known as the Test de Vocabulario en Imágenes Peabody, or TVIP, in Spanish) was also administered at baseline and endline. The TVIP measures receptive (listening) vocabulary acquisition.²⁸ However, because the TVIP must be normed with national test results, and because it is generally used as a diagnostic monitoring tool, the results are not presented in this report.

²⁷ For more details on the EGRA adaptation process, see QFD's baseline report.

²⁸ Dunn, L. M., Lugo, D. E., Padilla, E.R., and Dunn, L.M. (1986). Test de vocabulario en imágenes peabody (TVIP). Retrieved from <http://www.pearsonclinical.com/language/products/100000487/test-de-vocabulario-en-imagenes-peabody-tvip.html>.

Institutional Review Boards

Institutional review boards (IRBs) are responsible for ascertaining the acceptability of proposed research in terms of institutional commitments and regulations, applicable laws, standards of professional conduct and practice, and ethical and societal norms. IRBs examine subject recruitment procedures, proposed remuneration, and the informed consent process. IRBs also evaluate the potential risks and benefits to participants outlined in each protocol.

In consultation with World Vision, QFD staff completed the required research ethics training and submitted the EGRA instrument and research design to Solutions IRB, a private IRB based in the U.S. QFD received approval to conduct their research prior to baseline data collection and received an extension covering endline data collection.

Baseline EGRA

Baseline assessor trainings were held in November 2015 to prepare for operational baseline data collection (see Table 4). The training took place over eight days. Four days were dedicated to classroom training, and four days focused on practical training in schools. Assessors participated in assessor accuracy testing during the practical training.²⁹ Assessor accuracy testing is conducted to ensure consistency in scoring between assessors and to measure the degree to which assessors agree in their assessment decisions. At least 90 percent consistency is the minimum requirement; this means that at least 90 percent of assessors' ratings must be consistent with the list of acceptable responses. By the end of the eight-day training, all assessors met the 90 percent threshold.

Table 4: Fieldwork Preparation and Data Collection Timeline

Task	Dates
EGRA adaptation and pilot test	June 2015
Assessor training	November 2015
Baseline EGRA operational data collection	December 2015–April 2016
Endline EGRA operational data collection including refresher training	January–March 2017
EOP interviews	January–February 2017

Following the assessor training and the student enrollment period, assessors conducted the baseline EGRA, TVIP, and reading habits and attitudes survey data collections. Data were collected at seven libraries in December 2015 and January 2016 and at three replacement libraries in February, March, and April 2016. Baseline assessments were primarily administered at libraries. In select cases, baseline assessments were administered at schools where a significant number of students participating in the MdL project attended; this was only done if QFD received authorization from the school's principal. When the baseline data collections were administered at a library, registrants were called in advance to schedule individual appointments for assessments.

²⁹ Assessor accuracy testing is similar to interrater reliability testing. According to the EGRA Toolkit (2nd Edition), assessor accuracy refers to the testing conducted during training, while interrater reliability is conducted during operational data collection.

Endline EGRA

Endline EGRA data were collected from January to March 2017. As with the baseline data collection, assessments were conducted at libraries and, in select cases, at schools. Prior to the endline operational data collection, QFD conducted an assessor refresher training that included assessor accuracy testing and review of the EGRA instrument and administration.

End-of-Project Interviews

STS and World Vision conducted EOP interviews from January 30 to February 9, 2017. The purpose of the interviews was to explore the contextual factors that may have impacted the variations in implementation and differing results between schools and between students. EOP interviews were conducted with six groups of project participants: project management, stakeholders, librarians, parents and caretakers, students, and school principals.

Project management interviews consisted of 25 open-ended questions related to general information about the project and the intervention timeline, characteristics of the implementing organizations, perceptions of project design and implementation quality, and considerations for scalability. Librarians were asked 21 open-ended questions related to the MdL project and its components, challenges they faced in implementing the project with fidelity, student and parent engagement in the project, and the project's potential for scalability. Parents and caretakers responded to 18 open-ended questions about their own and their children's engagement in the MdL project, how they used the library, how they support their students' reading at home, and, when applicable, what they learned in the parent workshops. Students were asked 18 open-ended questions related to their engagement in the MdL project, their disposition toward reading, what they did when they visited the library, and how their reading is supported by their family. Stakeholders and school principals were asked questions about the quality of reading materials in Mexican primary schools, priorities for addressing literacy challenges for early grade readers, and the MdL project's potential for scaling.



V. Project Implementation

The MdL project began on February 16, 2015, and ended on April 15, 2017. This section presents implementation challenges, solutions, and successes that help answer the ACR GCD research question: *How successful was the rollout of the intervention?*

Development

The MdL project was conceptualized based on lessons learned from a research study conducted by the project's principal investigator and program manager.³⁰ The study revealed that there is a potential mismatch between students' actual reading levels and the level of the books provided through the school curriculum. To respond to this, the MdL project proffered a methodology for determining reading levels and a book-matching algorithm to counteract the one-size-fits-all teaching strategy within the Mexican school system. The project's technological components, which are complex and were time-consuming to develop, were created specifically for the project. Their development resulted in significant challenges throughout the life of the project.

The MATCH algorithm, which used data on a student's individual reading level as well as ratings of book difficulty to recommend books to students in interventions A and C, was identified by all interviewed staff as the central component of the MdL project. QFD worked with MetCuantus to develop the MATCH algorithm; they began their work in August 2015, and QFD finalized the algorithm in April 2016 after project implementation in libraries had already begun. The QFD team noted in quarterly reports and during EOP interviews that establishing the quantitative and qualitative parameters that defined a book's difficulty required significant research and took more time than expected.

QFD team members, MetCuantus, and a quantitative research assistant worked to develop and define the algorithm's quantitative parameters. A literacy specialist was contracted to create a rubric to evaluate the qualitative parameters of books. The MdL team then determined the relative weight of each parameter. After assessing 30 books and finalizing the book-difficulty formula, the task of assessing the books purchased by the MdL project remained. Each of these steps necessitated significant labor resources and technical expertise and required feedback and fine-tuning between all members of the MdL team. Despite the time and resources invested in the development of the MATCH algorithm during the development phase of the project, project management expressed the need for continual improvement and refinement of the algorithm throughout the project's implementation. The QFD team noted challenges with MetCuantus's capacity and quality of work during the development of the algorithm and during the creation of assessment tools.

Another large technical component of the MdL project was the development and rollout of the web-based platform for students. Development of the platform began in February 2015, and the completed platform was released in February 2016. QFD worked with Zenit, a local IT consulting firm, to develop the platform. While Zenit initially indicated that they had the internal knowledge necessary to design all aspects of the platform, QFD ultimately had to engage trained librarians to identify the different modules and sections that the platform should contain, which delayed timelines. Further short-term technical assistance was contracted through the library consulting firm Información, Bibliotecas y Sistemas Avanzados (IBSA) to help the QFD team produce reports effectively through Koha.

In addition to developing the MATCH algorithm and web-based platform, QFD also had to procure the physical components of the libraries, namely books and furniture. QFD collaborated with literacy experts, including an elementary school librarian, to create a preliminary list of titles for the MdL book catalog. However, before the

³⁰ Ortega Hesles, M. (2012). Learning from the pilot study of a cluster randomized trial: Summer reading interventions targeting third grade students in Mexico (unpublished qualifying paper, Harvard Graduate School of Education).

list could be finalized and books purchased, QFD had to finalize the parameters of the MATCH algorithm to ensure that the catalog could be categorized through the algorithm. QFD began to procure books between July and September 2015, with the intent of building a catalog with a range of levels and topics. Despite some challenges in procuring enough copies of specific books due to limited stock, QFD purchased 295 different titles for the MdL libraries by the start of the project's implementation. Once the books were procured, a library curator was responsible for manually digitizing the book catalog into Koha and generating physical barcodes. A QFD research associate was responsible for manually entering student data into Koha. The QFD team noted in EOP interviews that the branding and bar-coding of books was a more time- and labor-consuming task than anticipated.

Overall, QFD's project management team shared that the development phase of the MdL project was challenging; it faced unanticipated delays and required more labor and investment than expected.

Implementation

In addition to challenges during the development of the MdL project, QFD also cited significant challenges during implementation. These began during implementation start-up and the uneven enrollment of interested students by participating libraries. Seven of the ten libraries originally selected successfully enrolled a sufficient number of interested students in the MdL project to allow for baseline assessments and project rollout to begin as scheduled in January 2016. However, in three of the ten selected libraries, enrollment was so low that the QFD team had to substitute in new libraries. This resulted in delayed implementation in those replacement libraries, which meant that students in different libraries had different lengths of exposure to the MdL project.

Another significant challenge was the irregular use of the MdL platform by students. One research question asks how the MATCH algorithm, and its tailored book recommendations, impacted students' reading; to measure this, students would have had to use the MdL platform to access their personalized recommendations. However, the project did not require students to use the platform before checking out books; therefore, many students did not log in and instead chose books on their own. Additionally, because implementation began in several libraries before finalizing the MdL platform, students at those libraries were not initially able to use the web-based platform to view and select books. In those cases, project management observed that students continued to enter the library and select books without the platform even after the platform was launched. This challenge was not just restricted to early implementation libraries; the MdL project management observed that student use of the platform was inconsistent across all libraries throughout the project.

In qualitative EOP interviews conducted with students, parents, and caretakers, many expressed that they or their children preferred to choose books on their own rather than rely on the platform to recommend books. Respondents cited a lack of computer literacy, technical challenges with the platform, or a child's central interest in books' illustrations as the main reasons that they did not log on. Specifically, some parents stated that they were embarrassed to use the platform with their children because they did not know how to use the computer; others cited lack of time. Additionally, very few students had access to a computer at home to view the platform, and, in some libraries, the computers and tablets that students could have used to log on to the platform were often already in use by attendees of the libraries' computer classes. QFD did observe that some librarians were more proactive in encouraging students to use the platform than others; this may explain some of the variation in platform use across project sites.

To respond to infrequent platform usage, QFD began implementing reading games with prizes that necessitated students log in to the platform to participate. One such game, launched in the fall, asked students to complete reading and writing literacy tasks related to the Día de los Muertos holiday.³¹ Another game, known as a *lectometro* challenge, offered virtual medals to students who logged in to the platform to register books they

31 Día de los Muertos, or "Day of the Dead," is a cultural holiday celebrated throughout Mexico, particularly in the central and southern regions.

had read. Project data indicate that these games increased platform use, but it is unclear if the impact was consistent across libraries or sustainable.

Additionally, the web-based platform experienced technical issues throughout the implementation phase of the MdL project which may have negatively impacted student use. Most notably, in October and November 2016, the server for the MdL platform server was damaged, leaving both the MdL platform and the Koha system inaccessible for several weeks. Members of QFD's project management team also noted other technical challenges and glitches with the platform throughout the project. These included accessibility issues due to library firewalls and redundant prompts that asked students to rate the same books every time they logged in.

In addition to low web-based platform use rates, project monitoring data also indicated that students' library attendance was inconsistent. This was a significant challenge as the MdL project depended on students in all intervention groups regularly visiting the libraries. The MdL team noted that student attendance and book loans peaked immediately after the start of implementation, but that activity tended to decrease over time. The MdL team identified summer vacations and changes in weather between the rainy season or the hot season as possible reasons for variations in student attendance. In other instances, the libraries themselves seemed to be the catalyst for high or low levels of engagement. For example, librarians who were more active in the local community were better able to encourage students to come to the library; other librarians hosted highly attended digital classes during which parents would bring their children to the library with them. Conversely, two libraries were in tourist zones, and during the high tourism season, students registered at those libraries worked in commercial activities rather than visit the library. During EOP interviews, most parents stated that a lack of time was the greatest impediment in bringing their children to the library.

The final component of the MdL project—parent workshops and materials that promote parental engagement, provide information and strategies to structure children's reading practices, and advise on how to create a rich literacy environment at home—also faced implementation challenges. The first parent workshop was held in January 2016, and the fifth and final workshop was held in December 2016. The MdL team noted very low attendance throughout the workshops. In some cases, low attendance was attributable to overlap with a specific event or mitigating factors such as summer vacation from school, poor weather, or lack of workshop promotion due to librarian turnover. QFD also noted that the workshops were promoted as *talleres*, which connotes a formal training; qualitative data indicated that many parents were intimidated or discouraged from attending because of this terminology. To remedy this challenge, the QFD team tested different ways to market the workshops to motivate parents and caretakers to attend, including calling the workshops *convivios*, or get-togethers. Regardless, across workshops and libraries, the average attendance rate was less than ten percent of eligible parents and caretakers.

Management

The MdL project was one of the first intervention projects to be implemented by QFD, which historically has focused on conducting research and evaluating social and economic development projects. This endeavor into project implementation necessitated a steep management learning curve for QFD staff, who recognized some challenges in their abilities to deliver this aspect of the project.

The MdL project, particularly in the development phase, was strongly supported by consultants or consulting firms that were not always able to effectively deliver products as per their scopes of work. One example of this was Zenit; their delay in producing the final version of the MdL platform was partially due to insufficient internal capacity. Since the QFD team relied on external sources to produce key project deliverables, they had to invest resources in the oversight of consultants—a significant challenge for an organization that was relatively new to project implementation.

To implement the MdL project, QFD partnered with Fundación Proacceso, which provided both use of their librarians as well physical space at their preexisting digital libraries for the MdL book corner. This significantly reduced the operational cost of the project; however, because librarians had other responsibilities and were not direct employees of the MdL project, there were challenges in the management and engagement of personnel which was key to the project's success. QFD staff expressed that some librarians were more enthusiastic about the project than others. Specifically, some librarians were reported as more willing to engage with the community about the MdL project and encourage the use of the web-based platform, while others just fulfilled their minimum duties as required. Addressing librarians' performance on the MdL project required QFD to liaise with Fundación Proacceso's management rather than make autonomous decisions. Additionally, by partnering with Fundación Proacceso to access physical space and librarian personnel, the QFD team had to adhere to administrative requirements put in place by Fundación Proacceso and their partners, most notably the Government of Estado de México. At times, this delayed decision-making on or the promotion of the MdL project since approval was required from multiple partners. These administrative challenges were mirrored in the project's attempts to engage with local primary schools: some school principals and teachers were more willing to participate in promoting the MdL project than others. Management of these relationships and the ability to encourage their enthusiasm about the MdL project was critical to its success.

Finally, because the QFD had to hire a team with project management skills to effectively implement the MdL project, they experienced common human resources challenges, including staff turnover, difficulty managing different personalities, and evolving job descriptions. Despite their significant experience in project research and evaluation, QFD did not have the equivalent experience in project implementation in the education sector. Feedback during EOP interviews suggests that the initial project team was not structured or staffed with individuals who had the right type of experience; this may have impacted the effectiveness of the project roll-out and implementation. Over time, however, QFD hired highly capable team members who effectively oversaw project activities, rapidly addressed challenges faced by the project, and supported the engagement of librarians, parents, caretakers, students, and primary schools. Due to QFD's exceptional capacity in research and evaluation, they constantly monitored the MdL project implementation to recognize challenges and develop potential solutions. This dedication to using data to improve the life of the project is evident in QFD quarterly reports and progress updates, which articulated well the project's challenges and solutions. In instances where implementation challenges were insurmountable, the QFD team recognized these potential design problems and proactively thought of ways to change future iterations of the project so that it could continue to improve.

Fidelity of Implementation

By definition, FOI is the accurate and consistent application of an agreed upon procedure. FOI research is used to assess the degree to which a project is implemented as intended. Measuring FOI helps implementers and researchers understand and differentiate between what was supposed to happen and what actually happened during the life of a project. When FOI is high and an intervention group experiences gains, then it is possible to associate gains with the intervention; this, in turn, makes it possible to recommend scaling the intervention. FOI also makes it possible to identify which components of an intervention are most strongly associated with outcomes. When FOI is low and gains are low, it is impossible to know whether the reason for low gains is a poor design or poor implementation. FOI can also be coupled with monitoring and evaluation (M&E) to provide feedback to implementers during the project cycle to improve adherence to project design in the case of low FOI.³²

32 Creative Associates International, Inc. (2015). Fidelity of implementation (FOI) how-to guide (unpublished). Washington, D.C.: USAID.

As part of their projects, all ACR GCD Round 2 grantees conduct FOI research during the implementation period. The primary objectives of FOI for grantees were to

1. Understand what FOI is and why it is important throughout the life of the project
2. Identify essential components, activities, and questions for each phase of project implementation
3. Create relevant, project-specific FOI tools to monitor registrants' adherence to the intervention plan

STS held a series of FOI meetings with each ACR GCD grantee to develop project-specific FOI tools and an implementation plan for FOI research. After finishing the FOI sessions, ACR GCD grantees were expected to pilot test their FOI tools and collect data. Grantees were advised to collect a minimum of one round of FOI data; two or more rounds of data collection were considered ideal.

The collected data serves several purposes:

1. To indicate where revisions in data collection tools were necessary
2. To highlight where improvements in implementation were needed
3. When combined with assessment results, to provide evidence that gains were associated with the intervention (if possible)

QFD used several means to collect FOI data during the project. Project dosage data, which was tracked through Koha, provides details on whether students and parents were participating in activities and engaging in project components as intended. The MdL team also captured FOI data from parents, caretakers, and students through interviews conducted by EQUIDE, an applied research center at the Universidad Iberoamericana in Mexico. Finally, QFD collected data at the library-level to articulate any differences in implementation across sites. In the following section, project dosage from Koha is analyzed in detail to assess uptake in and FOI of project components.

Project Dosage³³

The MdL project consisted of three components. The first component of the project, access to libraries and children's books, was assessed by examining the proportion of students who checked out books from the library. Library activity details, including the number of students who checked out at least one book and those who did not check out any books, are detailed in Table 5. **Across all intervention groups, 60.6 percent of students were active in the MdL project, meaning they checked out at least one book during the project; 39.4 percent of students were inactive, meaning they did not check out a single book during the project.** The proportion of active students was highest in intervention B and lowest in intervention A.

Table 5: Student Library Activity by Group³⁴

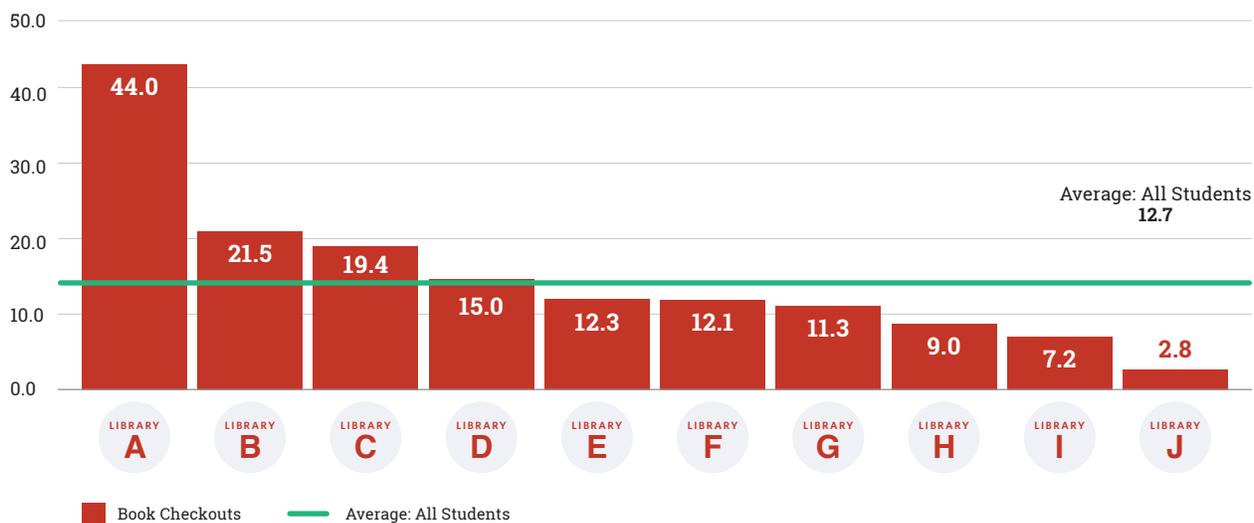
Activity Level/ Book Checkouts	Intervention A		Intervention B		Intervention C		Intervention D		All Students	
	n	Percentage (%)	n	Percentage (%)						
Inactive (did not check out any books)	64	44.4	25	30.9	60	41.1	31	36.0	180	39.4
Active (checked out at least one book)	80	55.6	56	69.1	86	58.9	55	64.0	277	60.6
Total	144	100.0	81	100.0	146	100.0	86	100.0	457	100.0

³³ Dosage was calculated for the final sample of 457 students that were assessed at both baseline and endline.

³⁴ In intervention A, students received personalized book recommendations through the MATCH algorithm, and libraries offered workshops for parents and caretakers. In intervention B, students received personalized book recommendations, but libraries did not offer workshops. In intervention C, students received a random selection of book recommendations, and libraries offered workshops. In intervention D, students received a random selection of book recommendations, and libraries did not offer workshops. See Research Purpose and Design.

Figure 2 illustrates the variation in the number of books checked out per student during the project. The average number of books checked out across libraries was 12.7 books per student. However, students in library A checked out, on average, twice as many books as students in library B; students in library J only checked out an average of 2.8 books during the project.

Figure 2: Average Book Checkouts per Student by Library



The second component of the MdL project, a web-based platform that provided book recommendations to students, necessitated that students were randomly assigned to receive MATCH algorithm-based book recommendations. Students in the MATCH algorithm group and the non-MATCH algorithm group took two steps: (1) logged into their profiles; and (2) used the platform’s recommendations to decide which books to check out. The Koha data summarized in Table 6 presents student login data based on their assignment to a MATCH algorithm group or a non-MATCH algorithm group.

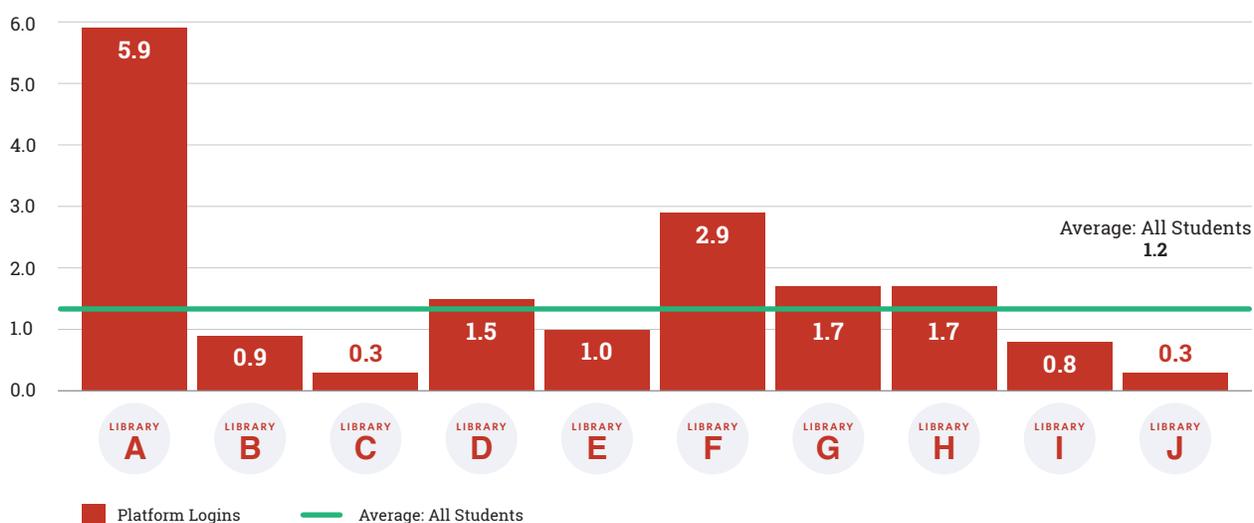
Table 6: Student Platform Use by MATCH Algorithm Assignment

Platform Logins	No MATCH		MATCH		All Students	
	<i>n</i>	Percentage (%)	<i>n</i>	Percentage (%)	<i>n</i>	Percentage (%)
No logins	157	67.7	147	65.3	304	66.5
At least one login	75	32.3	78	34.7	153	33.5
Total	232	100.0	225	100.0	457	100.0

Regardless of MATCH assignment, only 33.5 percent of students logged on to the platform at least one time during the project. For those students assigned to receive personalized book recommendations through the MATCH algorithm, 65.3 percent of students did not log in to the platform even once during the project, meaning that only 34.7 percent of students assigned to receive personalized book recommendations had the opportunity to see those recommendations. The proportion of students who logged into the platform from non-MATCH group was similar to the MATCH group: only 32.3 percent of students logging on at least once during the project.

Figure 3 shows the average number of platform logins per student during the project. The average number of logins per student across libraries was 1.2. As with book checkouts, library A had the highest average of platform logins per student; this was twice as many as in library F which had the second highest average.

Figure 3: Average Platform Logins per Student by Library



QFD did not have an easy way to access data on whether students who logged into the platform ultimately checked out books that had been recommended to them. As a result, it is difficult to determine if students who received personalized book recommendations took these books home and benefited from the MATCH algorithm (see Considerations). Qualitative data collected by QFD and STS indicates that most students did not use the platform recommendations when selecting which books to check out. Instead, students chose books based on the book’s illustrations or topic and peers’ recommendations.

Uptake and FOI of the third component of the project—workshops and materials for parents or caretakers—was measured through parent and caretaker attendance at the workshops. Five of the MdL libraries—representing 63 percent of students—were randomly assigned to offer parent workshops. The number of workshops attended by parents and caretakers is presented in Table 7. In total, 290 students enrolled in libraries that held workshops for parents and caretakers; of the 290, **only 19.7 percent of the eligible students had a parent or caretaker attend at least one workshop**, while 80.3 percent of students’ parents or caretakers did not attend a single workshop. No students’ parents or caretakers attended all five workshops held by the MdL project.

Table 7: Parent or Caretaker Workshop Attendance

Number of Workshops Attended	<i>n</i>	Percentage (%)
Did not attend any workshops	233	80.3
One	35	12.1
Two	15	5.2
Three	5	1.7
Four	2	0.7
Five	0	0.0
Total	290	100.0

Program dosage indicates that there were very low project uptake and FOI for the MdL project in its pilot stage, as well as wide variation in project engagement across libraries. To associate observed gains with the MdL intervention, students, parents, and caretakers must have participated in the project and received the project dosage according to their intervention group assignment. Although a majority of students did check out at least one book during the project, only a third of students logged on to the MdL platform even once, and over 80 percent of students' parents and caretakers who were assigned to participate in workshops did not attend a single one. Because of this low uptake and because project dosage was not received as intended, it was not possible to assess if the reading gains of students registered in the MdL project were associated with the intervention (see Considerations).

VI. EGRA Data Analysis

EGRA data were analyzed using Microsoft Excel and IBM SPSS Statistics. Only students who had data at both baseline and endline were included. EGRA subtask results were matched by student and compared by time period to calculate reading gains over the life of the project.³⁵ Subtasks' mean fluencies and scores are reported, as are standard deviations (SD) relevant to those mean values.³⁶ Gain scores were computed as the difference between endline and baseline for each subtask, and student reading performance was evaluated across subgroups of students, including grade and gender. Zero scores³⁷ were also calculated for all subtasks. Differences between student scores at baseline and endline were tested for significance using paired t-test analysis. Differences in gain scores between grade levels and gender were tested for significance using analysis of variance and independent samples t-test analysis,³⁸ and differences in the proportion of zero scores were tested for significance using chi-square test.³⁹ Results with statistically significant differences are reported throughout with an asterisk. Where results are not statistically significant, it is not possible to assume that there is any difference between the baseline and endline results.

For each subtask, decision rules were applied to assess whether outliers would need to be removed. For example, if the time remaining for a timed subtask resulted in a fluency rate that was outside a reasonable range, then that student's fluency rate was not included in the analyses. Reasonable ranges for the time remaining were based on multiple factors, including the rate at which letters or words in the language tested are typically read, the distribution—or relative performance—of students in the sample, and the mean fluency rates with and without the outlier data point(s). After consideration of the reasonable ranges in the data, one outlier was removed.⁴⁰

For timed subtasks, rates were calculated per second and multiplied by sixty seconds to compute the rate per minute. This assumes that, if there were additional items included on the timed subtask, the child would have continued responding at the same rate. As a result, for some subtasks, average rates were higher than the number of items on the subtask.

Table 8 and 9 provide details on the EGRA and adaptive subtasks, including how results were calculated.

35 Because of rounding, mean changes reported may not always equal endline value minus baseline value.

36 SD describes how much observed values vary from the mean. A smaller SD indicates that most of values are close to the mean; a larger SD indicates that values are further from the mean. This report provides mean fluencies and scores of the entire sample of students participating in the MdL project. SDs are listed to understand the variability of the scores within the sample.

37 Students receive a zero score if they are unable to correctly identify a single item on a subtask. In this report, zero scores are shown as the number of students and/or as the percent of the total students unable to correctly identify a single item on a subtask.

38 Analysis of variance is a statistical model that is used to analyze the differences between group means, which helps identify differences in the sample that can be generalized to the population. The independent-sample t-tests compares the difference between the means of two independent groups on the same dependent variable.

39 The chi-square test is a statistical test comparing proportion of students with zero scores that were observed in the data against what was expected.

40 On endline, one outlier student had subtask rates and scores more than three SDs above the mean.

Table 8: EGRA Subtask and Data Analysis Method

Subtask	Type	Analysis
Letter sound identification	Timed	Letter sound identification is measured as the number of correct letter sounds read in one minute (CLSPM). Letter sound identification is a measure of alphabet knowledge. Each student had the opportunity to read up to 100 upper- and lower-case letters.
Initial sound identification	Untimed	Initial sound identification is measured as the number of correct initial sounds identified out of ten. Initial sound identification is a measure of phonological awareness. Each student had the opportunity to identify ten beginning phonemes that are different from two others in a series of words.
Familiar word reading	Timed	Familiar word reading is measured as the number of correct familiar words read in one minute (CFWPM). Familiar word reading measures word recognition and decoding. Each student had the opportunity to read up to 50 high-frequency words.
Nonword reading	Timed	Nonword reading is measured as the number of correct “nonwords” read in one minute (CNWPM). Nonword reading measures decoding. Each student had the opportunity to read up to 50 one or two syllable nonwords.
Oral reading fluency (ORF)	Timed	ORF is measured as correct words read in one minute (CWPM). ORF is a decoding and reading fluency measure. Each student had the opportunity to read 59 words. The ORF passage formed the textual basis for the reading comprehension subtask.
Reading comprehension	Untimed	Reading comprehension is measured as the number of correct answers verbally delivered to the assessor based on questions asked about the passage read as part of the ORF subtask. Each student had the opportunity to answer up to five factual and two inferential questions.

Table 9: Adaptive Subtask and Data Analysis Method

Subtask	Type	Analysis
Adaptive oral reading fluency (AORF)	Untimed	AORF is measured as the number of correct words read within a passage. Students were presented one of two different stories according to their performance in the reading comprehension subtask; three correct answers is the threshold. Students routed to the short passage (Outcome B) had the opportunity to read 97 words, and students routed to the longer passage (Outcome C) had the opportunity to read 164 words. The passages varied in difficulty in terms of word, sentence, and paragraph length. The AORF passage formed the textual basis for the adaptive reading comprehension subtask.
Adaptive reading comprehension	Untimed	Adaptive reading comprehension is measured as the number of correct answers verbally delivered to the assessor based on questions asked about the corresponding passage from the AORF subtask. Students had the opportunity to answer up to four factual questions and two inferential questions.

Considerations

Intervention Groups

The MdL research design included four intervention groups, all of which were expected to receive different project inputs. Because of low FOI during the pilot year of the project, it was not possible to compare the results of different intervention groups. Uptake by registrants in the project's components was so low that it is not possible to assume that students received different dosage according to their intervention group assignment, making comparison across groups misleading. The analysis in this report will not compare results between intervention groups as originally intended in the research design. Instead, the analysis combines all students into one group to compare baseline and endline EGRA results.

Further, because all students are combined into one group and their performance compared at baseline and endline, there is no comparison group available for analysis. The purpose of comparison groups is to provide a measure of the changes that occurred in the absence of a project or intervention. Because no comparison group is presented, the findings of this report should be understood as the changes that occurred from a combination of the MdL project plus an additional year of schooling. It is not possible to isolate fully how much of the measured change from baseline to endline is due to the project and how much is due to an additional year of schooling.

Research Questions

As previously noted, the three MdL research questions could not be answered by this analysis because project uptake and FOI were not high enough to compare the reading outcomes of students who received different interventions.

In addition to the implementation challenges in the MdL project's pilot year, the QFD team did not capture all the data necessary to be able to answer the project's research questions. For the first research question—*Does access to book recommendations based on the MATCH algorithm improve vocabulary, reading scores, and reading habits of early grade readers*—there was no straightforward way to examine if students who logged into the platform and received personalized book recommendations subsequently checked out books recommended to them. As a result, this report does not assess if students benefited from the personalized recommendations.

Additionally, QFD did not capture data to measure the second research question—*Do workshops improve parental engagement in children's reading*. The MdL team attempted several different ways to collect data directly from parents on their engagement in their children's reading, including sending paper surveys home to parents and administering questionnaires over the phone (See Implementation). However, they received very low response rates. Ultimately, students were asked to respond to questions on the reading habits and attitudes survey about their parents' behaviors and interactions with their reading; data was also collected through qualitative EOP interviews with parents and caretakers. Because of the limitations of these data sources, coupled with low attendance at parent workshops, this research question was not answerable.

The third research question—*Does access to workshops for parents improve their children's vocabulary, reading scores, or reading habits*—was dependent on parents' attendance at the MdL workshops. Because over 80 percent of students with eligible parents did not have a parent or caretaker attend a single workshop, data were insufficient to answer this question.

VII. EGRA Results⁴¹

This section presents EGRA results to understand whether the reading skills of students participating in the MdL project increased from baseline to endline. The section contains findings across EGRA subtasks as well as detailed results for each subtask by grade. EGRA results are also explored by gender and by key factors for success. Students are combined into one intervention group ($N=457$) because of low project uptake and FOI during the MdL project’s pilot year.

Overall, the results presented in Figure 4 indicate that **students enrolled in the MdL project showed improved reading skills at endline compared to baseline**; this is attributed to a combination of an additional year of schooling and participation in the project. The difference in the gains was statistically significant across grade levels, with students in Grade 1 consistently having greater gains on all subtasks than students in Grade 2 and Grade 3.

Figure 4: Average Gain Scores from Baseline to Endline by EGRA Subtask and Grade⁴²

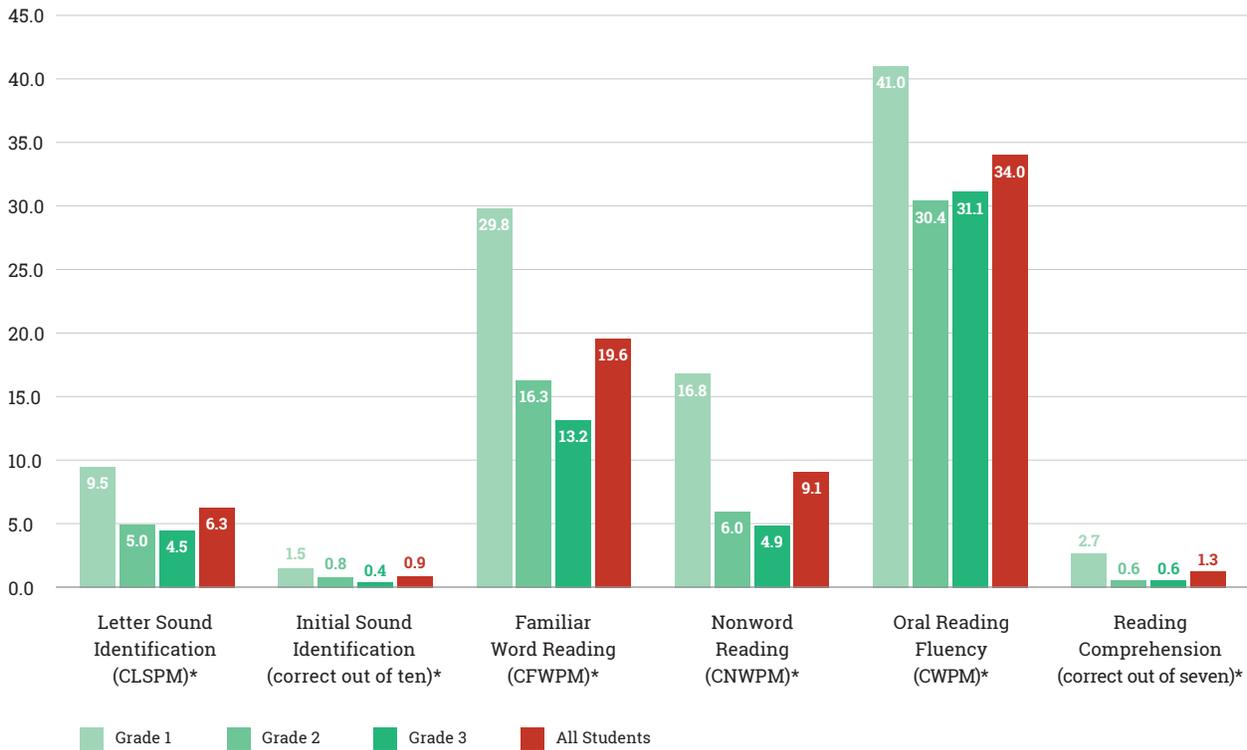
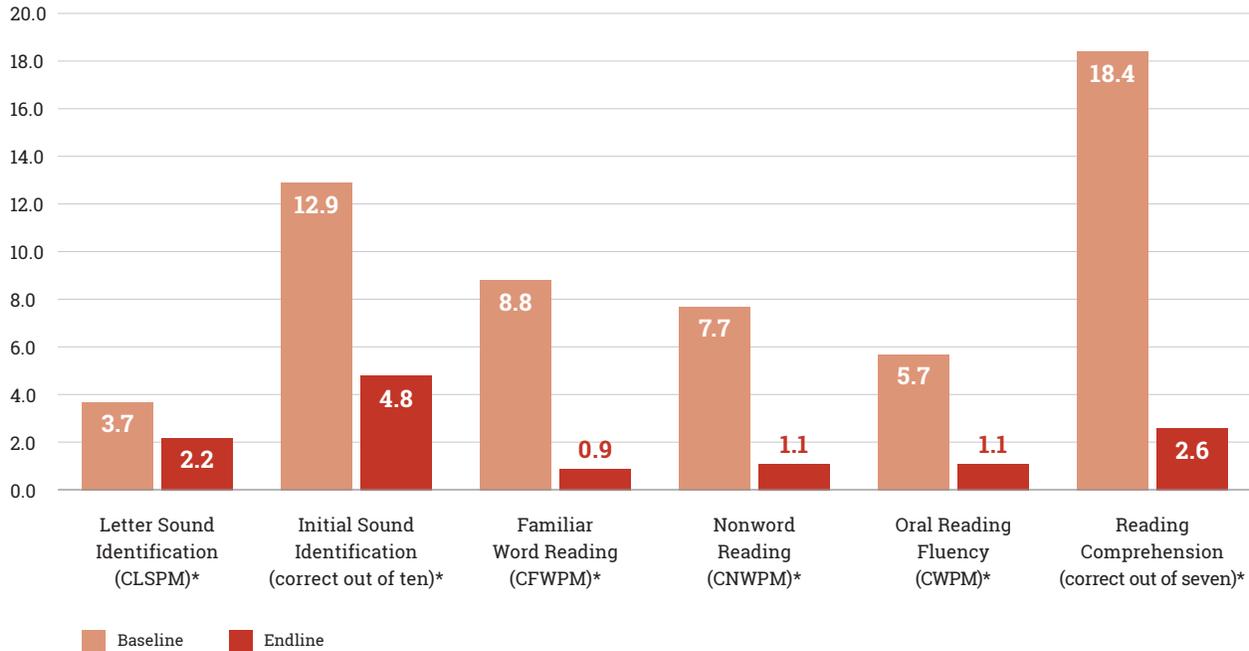


Figure 5 shows the percentage of students who received zero scores at baseline and endline. The proportion of students who received zero scores was lower at endline than at baseline on all subtasks, meaning that **more students could identify or answer at least one item correctly at endline than at baseline**.

41 Fluency rates were calculated for all timed subtasks (letter sound identification, familiar word reading, nonword reading, and ORF) per second and multiplied by 60 seconds to compute the rate per minute. This calculation considers the amount of time remaining and assumes that, if there were additional items included on the timed subtask, the child would have continued responding at the same rate. As a result, for some subtasks, average rates were higher than the number of items on the subtask.

42 An asterisk (*) indicates that the gain scores are significantly different across grade levels at $p < 0.05$. $N=457$; Grade 1 $n=145$; Grade 2 $n=174$; Grade 3 $n=138$.

Figure 5: Percentage of Students Receiving Zero Scores by EGRA Subtask at Baseline and Endline⁴³



EGRA results by subtask are detailed in the following section, as well as the percentage of zero scores by subtask.

EGRA Results by Subtask

Letter Sound Identification

The letter sound identification subtask measures students' understanding of the alphabetic principle, which states that each letter of the alphabet corresponds to a specific sound. To demonstrate letter sound identification, students must identify the appropriate sounds for each letter. The ability to match letters with correct sounds is critical to reading fluency and comprehension. For this subtask, each student was presented with a stimulus of 100 letters and asked to read aloud as many of the sounds as they could in one minute.⁴⁴ Results for this subtask are reported as a rate of CLSPM.

Mean results for the letter sound identification subtask are presented in Figure 6. On average, **letter sound identification rates increased from baseline to endline for all students**. Specifically, at baseline, students identified 24.5 letter sounds within a minute on average, while at endline, students identified 30.7 letter sounds within a minute on average. The mean identification rates that students achieved at endline were significantly higher than those at baseline (Figure 2). Across grades, students had significantly higher results at endline than at baseline. On average, students in Grade 1 identified 9.5 additional letter sounds within a minute, students in Grade 2 identified 5.0 additional letter sounds within a minute, and students in Grade 3 identified 4.5 additional letter sounds within a minute.

⁴³ An asterisk (*) indicates the percentage of students receiving zero scores was significantly different baseline and endline at $p < 0.05$. $N = 457$.

⁴⁴ There is an auto stop rule in all the timed EGRA subtasks. In this case, the test was discontinued if a student was unable to correctly name any of the first ten letters on the stimulus.

Figure 6: Mean Results at Baseline and Endline by Grade—Letter Sound Identification (CLSPM)⁴⁵

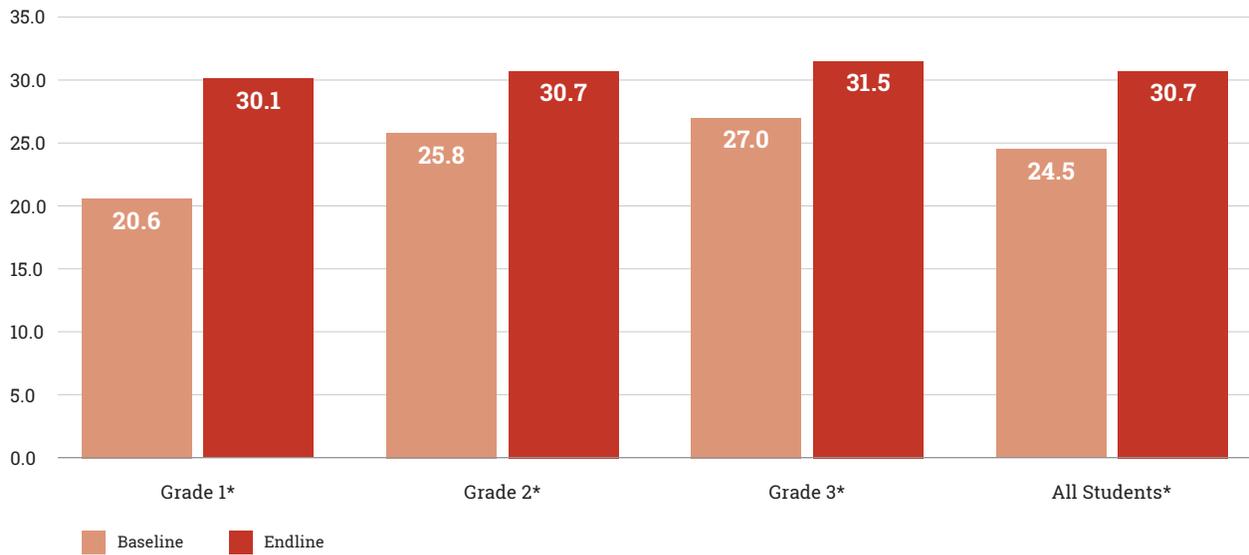
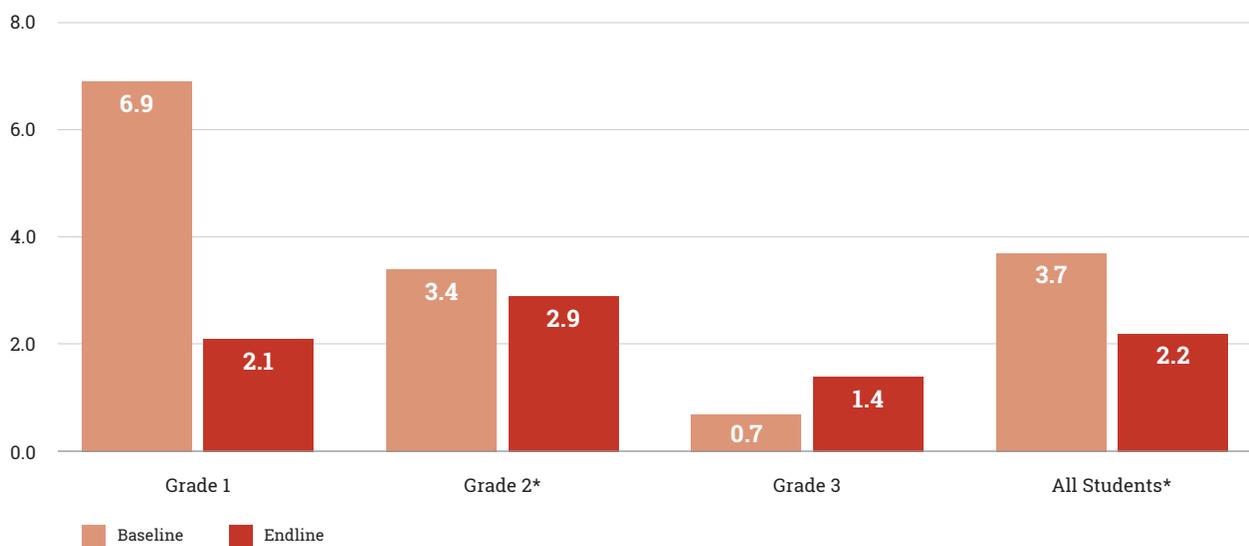


Figure 7 presents the proportions of students receiving zero scores at baseline and endline. Overall, 3.7 percent of students received zero scores at the baseline while 2.2 percent of students received zero scores at endline; the proportion of students receiving zero scores at endline was significantly smaller. The results show that the proportion of students in Grade 1 and Grade 2 receiving zero scores at endline was lower than the proportion of students receiving zero scores at baseline; conversely, a greater proportion—equal to a 0.7 percentage-point increase—of Grade 3 students received zero scores at endline than at baseline. The difference in the proportion of zero scores at baseline was significantly smaller only for students in Grade 2.

Figure 7: Percentage of Students Receiving Zero Scores by Grade at Baseline and Endline—Letter Sound Identification (%)⁴⁶



45 An asterisk (*) indicates the average gain score were significantly different across grade levels at $p < 0.05$. *N* sizes: All students $N = 457$; Grade 1 $n = 145$; Grade 2 $n = 174$; Grade 3 $n = 138$.

46 An asterisk (*) indicates the proportion of students receiving zero scores at endline was significantly smaller than the proportion of students receiving zero scores at baseline at $p < 0.05$. *N* sizes: All students $N = 457$; Grade 1 $n = 145$; Grade 2 $n = 174$; Grade 3 $n = 138$.

Initial Sound Identification

The initial sound identification subtask measures students' ability to identify the initial sounds of words. The ability to identify isolated sounds within a word is a test of phonemic awareness and indicates that a student understands that words are made up of sounds—an understanding he or she can then use to associate sounds with letters, which is a building block of decoding. In this subtask, the assessor read ten words, and students were asked to indicate the initial sound—or phoneme—of each word verbally. This was an untimed subtask.

Mean results for the initial sound identification subtask are presented in Figure 8. On average, **at endline, students identified 0.9 additional initial sounds than at baseline.** Specifically, at baseline, students identified 6.1 initial sounds out of ten, while at endline, students identified 7.0 initial sounds. The mean scores students achieved at endline were significantly higher than those at baseline across all grades.

Figure 8: Mean Results at Baseline and Endline by Grade—Initial Sound Identification (Correct out of Ten)⁴⁷

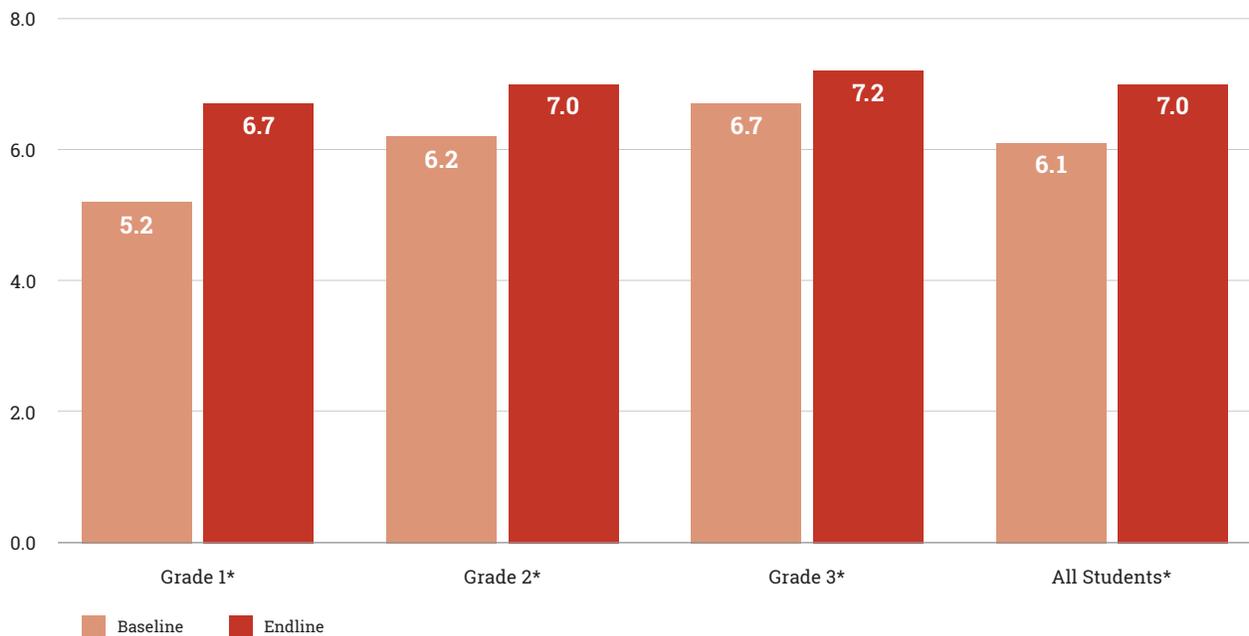


Figure 9 presents the proportions of students receiving zero scores at baseline and endline. **Overall, 12.9 percent of students received zero scores at baseline, and 4.8 percent of students received zero scores at endline;** this difference was statistically significant. The results show that the proportion of students in all grades receiving zero scores at endline was lower than the proportion of students receiving zero scores at baseline. The greatest decrease was observed among Grade 2 students, where the proportion of students who received zero scores on this subtask decreased by 8.6 percentage points.

⁴⁷ An asterisk (*) indicates the average gain score were significantly different across grade levels at $p < 0.05$. N sizes: All students $N = 457$; Grade 1 $n = 145$; Grade 2 $n = 174$; Grade 3 $n = 138$.

Figure 9: Percentage of Students Receiving Zero Scores by Grade at Baseline and Endline—Initial Sound Identification (%)⁴⁸

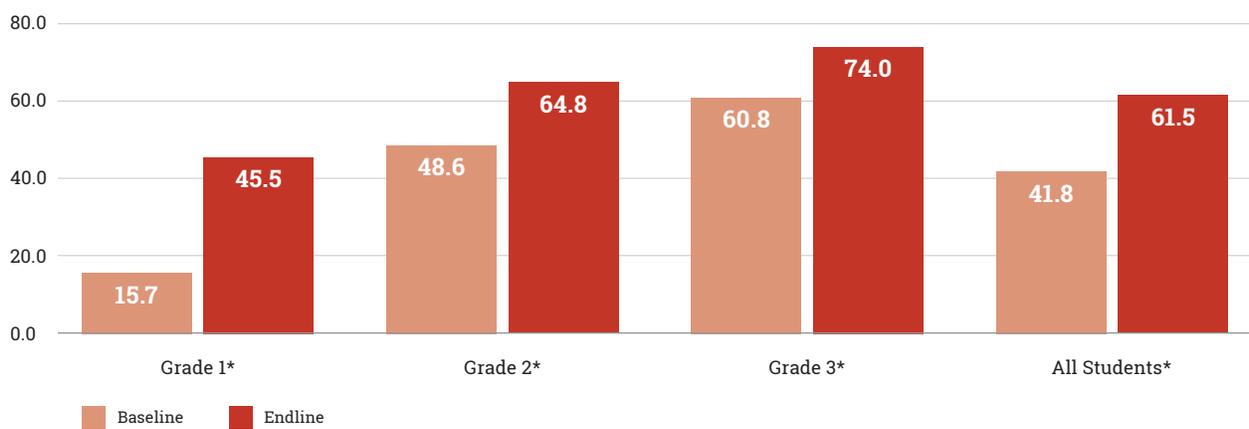


Familiar Word Reading

Knowledge of familiar words and the ability to read them quickly enables a student to read with automaticity—a skill critical to learning to read with fluency and comprehension. In the familiar word reading subtask, students were presented with 50 familiar words⁴⁹ and asked to read as many as they could within one minute. The subtask was discontinued if a student was unable to name correctly any of the first five familiar words. Results for this subtask are reported as a rate of CFWPM.

Mean results for the familiar words reading subtask are presented in Figure 10. On average, **familiar word reading rates increased from baseline to endline for all students**. At baseline, students read 41.8 familiar words within a minute, while students read 61.5 familiar words within a minute at endline; the mean rates that students achieved at endline were significantly higher than those at baseline. Students across grade levels had significantly higher mean results at endline than at baseline; as with other subtasks, students in Grade 1 made greater gains than students in Grade 2 and Grade 3. Grade 1 students read 29.8 additional familiar words within one minute at endline.

Figure 10: Mean Results at Baseline and Endline by Grade—Familiar Word Reading (CFWPM)⁵⁰



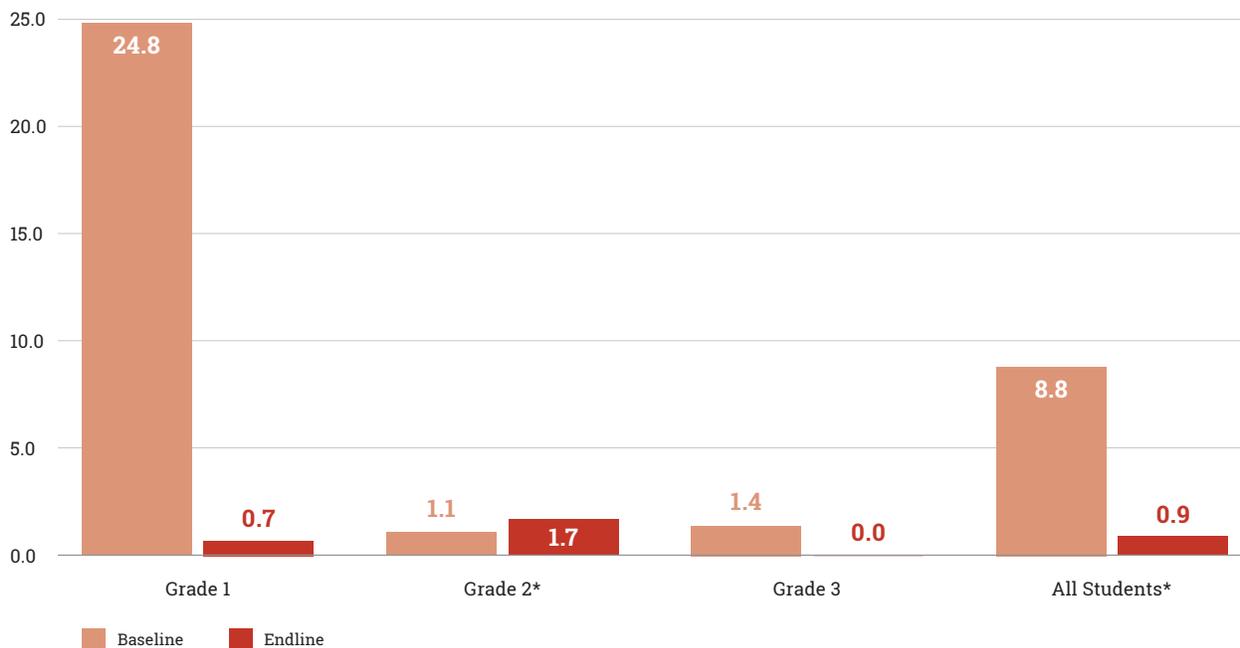
⁴⁸ An asterisk (*) indicates the proportion of students receiving zero scores at endline was significantly smaller than the proportion of students receiving zero scores at baseline at $p < 0.05$. $N = 457$.

⁴⁹ The words in this subtask were derived from frequently used words for the age group.

⁵⁰ An asterisk (*) indicates the average gain score were significantly different across grade levels at $p < 0.05$. N sizes: All students $N = 457$; Grade 1 $n = 145$; Grade 2 $n = 174$; Grade 3 $n = 138$.

Figure 11 presents the proportions of students receiving zero scores at baseline and endline. **Overall, 8.8 percent of students received zero scores at the baseline, while only 0.9 percent of students received zero scores at endline;** this difference was statistically significant. A greater proportion of Grade 2 students received zero scores at endline than at baseline, representing a 0.6 percentage-point increase. The proportion of students receiving zero scores was not significantly different at endline than at baseline for students in Grade 1 or 3.

Figure 11: Percentage of Students Receiving Zero Scores by Grade at Baseline and Endline—Familiar Word Reading (%)⁵¹



Nonword Reading

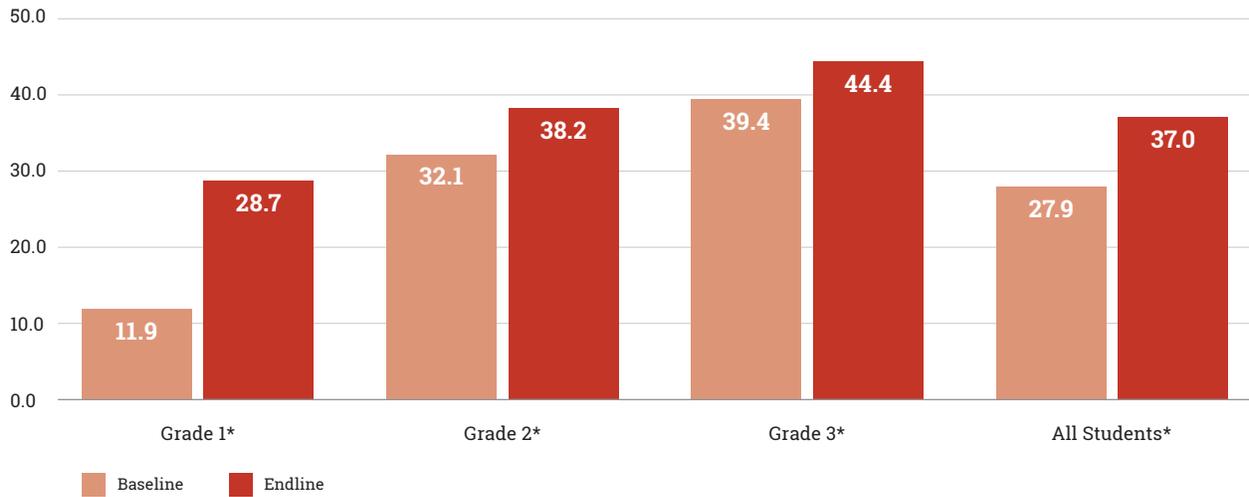
The nonword reading subtask measures students’ decoding ability by presenting them with words that they would not be able to recognize due to familiarity. Many students in the early grades learn to memorize or recognize a range of familiar words. Thus, to assess their decoding skills, students are presented with invented nonsense words, which require them to sound out each letter and syllable to decode a word. During this timed subtask, the assessor presented each student with 50 nonwords and asked them to read as many as possible in one minute.⁵² Results for this subtask are reported as a rate of CNWPM.

Mean results for the nonword reading subtask are presented in Figure 12. On average, **nonword reading rates increased from baseline to endline for all students.** At baseline, students read 27.9 nonwords per minute, while at endline, students read 37.0 nonwords per minute. The mean rates students achieved at endline were significantly higher than those at baseline. Students across grade levels had significantly higher scores at endline than at baseline, and students in Grade 1 made greater improvements than students in Grade 2 and Grade 3 on average. Specifically, Grade 1 students read 16.8 additional nonwords per minute at endline compared to baseline.

⁵¹ An asterisk (*) indicates the proportion of students receiving zero scores at endline was significantly different than the proportion of students receiving zero scores at baseline at $p < 0.05$. *N* sizes: All students $N = 457$; Grade 1 $n = 145$; Grade 2 $n = 174$; Grade 3 $n = 138$. Statistical significance was determined using the chi-square test which compares the proportion of students with zero scores that were observed in the data against what was expected.

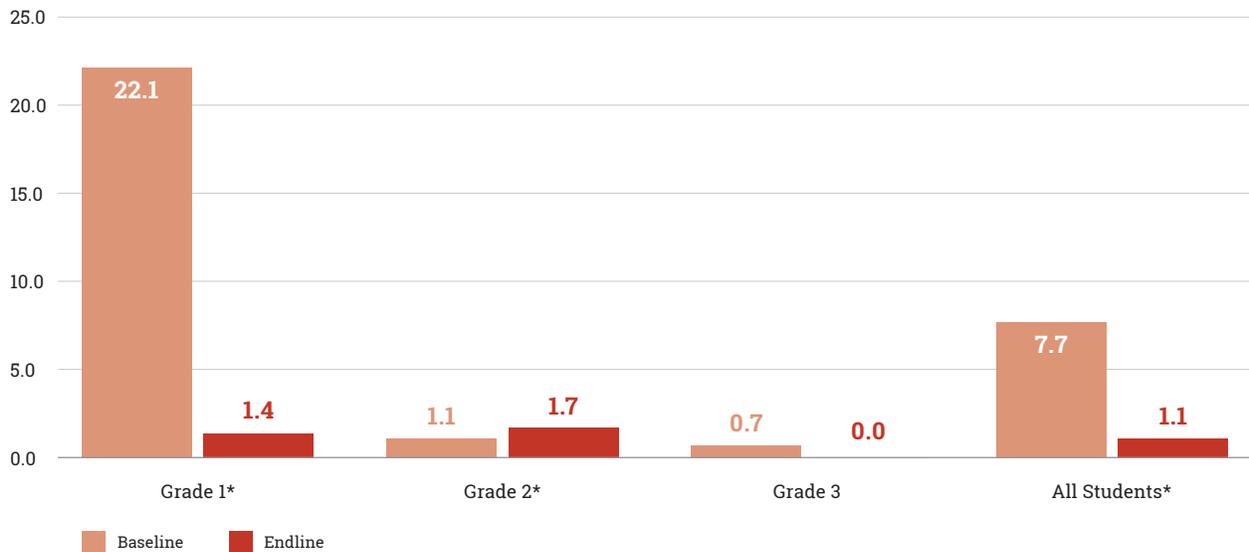
⁵² After one minute, the student was asked to stop. The subtask was discontinued if a student was unable to correctly read any the first ten nonwords.

Figure 12: Mean Results at Baseline and Endline by Grade—Nonword Reading (CNWPM)⁵³



The proportions of students receiving zero scores at baseline and endline are presented in Figure 13. **Overall, 7.7 percent of students received zero scores at baseline while 1.1 percent of students received zero scores at endline;** this difference was statistically significant. The results show that the proportion of students in Grade 1 and Grade 3 who received zero scores at endline is lower than at baseline. The proportion of students receiving zero scores among Grade 1 students decreased by 20.7 percentage points, compared with a decrease of 0.7 percentage points among Grade 3 students. Like the familiar word reading subtask, a greater proportion of Grade 2 students received zero scores at endline than at baseline—a 0.6 percentage-point increase. The proportion of Grade 3 students receiving zero scores at endline was not significantly different than at baseline.

Figure 13: Percentage of Students Receiving Zero Scores by Grade at Baseline and Endline—Nonword Reading (%)⁵⁴



53 An asterisk (*) indicates the average gain score were significantly different across grade levels at $p < 0.05$. *N* sizes: All students $N = 457$; Grade 1 $n = 145$; Grade 2 $n = 174$; Grade 3 $n = 138$.

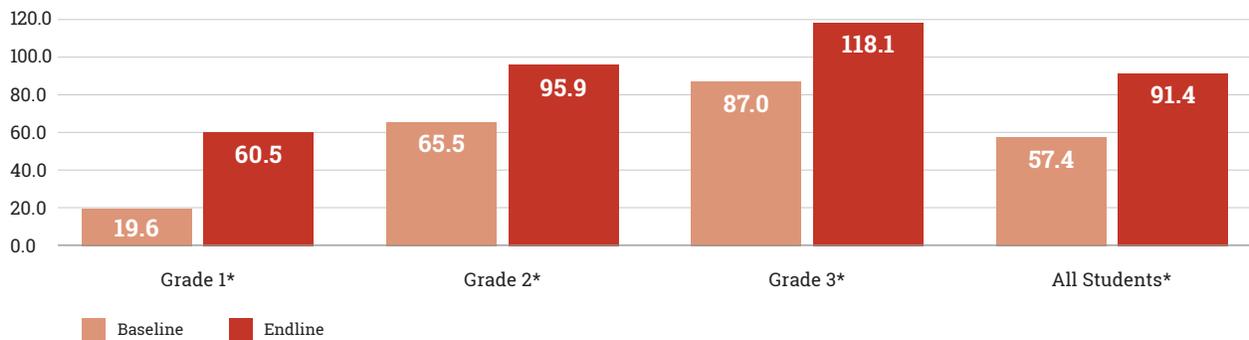
54 An asterisk (*) indicates the proportion of students receiving zero scores at endline was significantly different than the proportion of students receiving zero scores at baseline at $p < 0.05$. *N* sizes: All students $N = 457$; Grade 1 $n = 145$; Grade 2 $n = 174$; Grade 3 $n = 138$.

Oral Reading Fluency

The ORF subtask measures students' overall reading competence. It is the culmination of translating letters into sounds, merging sounds to become words, linking words to become sentences, relating the text to meaning, and making inferences to fill in missing information. A student's ORF score is dependent on the skills assessed in previous subtasks since students need to have some mastery of letter sounds and decoding to read fluently. Students had the opportunity to read up to 59 words in the ORF passage. Results for this subtask are measured as a rate of CWPM.

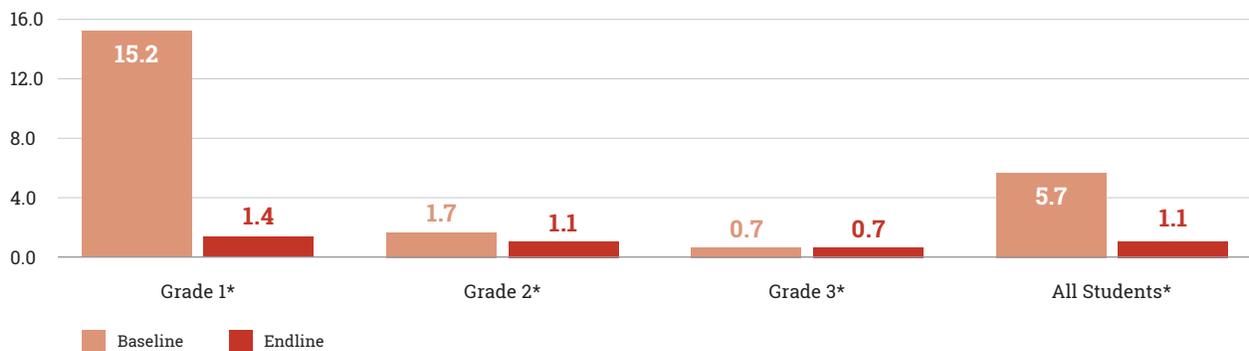
Mean results for ORF are presented in Figure 14. On average, **ORF rates increased from baseline to endline for all students**. Students read at a rate of 57.4 CWPM at baseline and 91.4 CWPM at endline; this difference was statistically significant. Students across grade levels had significantly higher fluency rates at endline than at baseline, with students in Grade 1 making the greatest improvements over time.

Figure 14: Mean Results at Baseline and Endline by Grade—Oral Reading Fluency (CWPM)⁵⁵



The proportions of students receiving zero scores at baseline and endline are presented in Figure 15. **Overall, 5.7 percent of students received zero scores at the baseline while 1.1 percent of students received zero scores at endline**; this difference is statistically significant. The proportion of students receiving zero scores decreased among Grades 1 and 2 students—a 13.8 and 0.6 percentage-point decrease, respectively. The proportion of Grade 3 students who received zero scores on the ORF subtask stayed the same at baseline and endline.

Figure 15: Percentage of Students Receiving Zero Scores by Grade at Baseline and Endline—Oral Reading Fluency (%)⁵⁶



⁵⁵ An asterisk (*) indicates the average gain score were significantly different across grade levels at $p < 0.05$. *N* sizes: All students $N = 457$; Grade 1 $n = 145$; Grade 2 $n = 174$; Grade 3 $n = 138$.

⁵⁶ An asterisk (*) indicates the proportion of students receiving zero scores at endline was significantly smaller than the proportion of students receiving zero scores at baseline at $p < 0.05$. *N* sizes: All students $N = 457$; Grade 1 $n = 145$; Grade 2 $n = 174$; Grade 3 $n = 138$. Statistical significance was determined using the chi-square test which compares the proportion of students with zero scores that were observed in the data against what was expected.

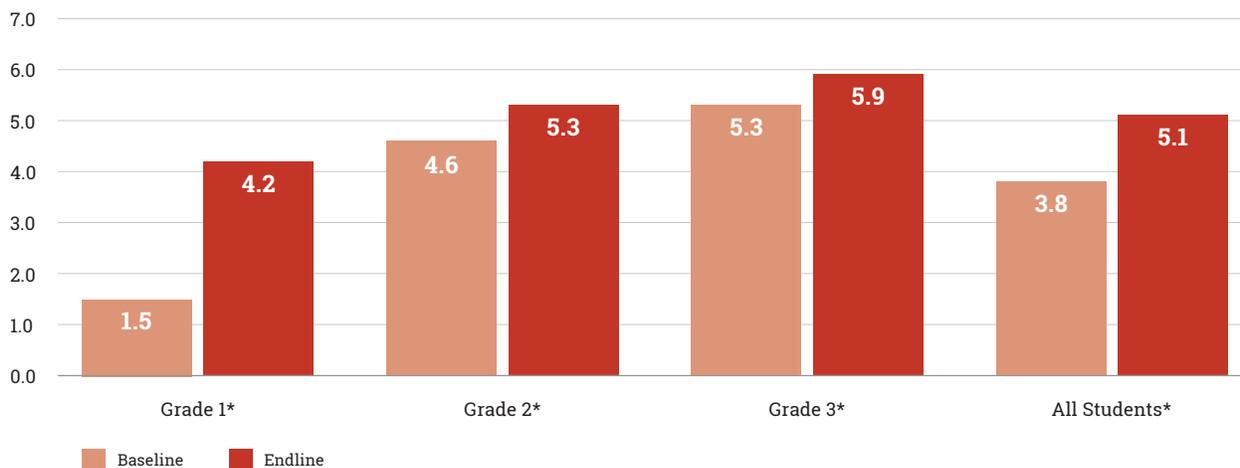
Reading Comprehension

Comprehension is the purpose of reading. Once students learn the sound-letter relationship (alphabetic principle) and can decode and read with automaticity, they become increasingly able to understand the meaning of a text. This subtask assesses that ability.

For the reading comprehension subtask, the assessor removed the story used in the ORF subtask and then asked each student up to seven comprehension questions based on what he or she read. The number of questions asked depended on how many words each student read on the ORF subtask. For instance, if a student read just the first five to ten words, he or she would be asked only the first comprehension question. Similarly, if a student read all 59 words, he or she would be asked all seven questions. Students who received zero scores on the ORF subtask also received a zero score on the reading comprehension subtask because no questions were presented to them. Additionally, any student who could not correctly answer a single reading comprehension question received a zero score on this subtask.

Mean results for reading comprehension are presented in Figure 16. **On average, students correctly answered 1.3 additional questions at endline than at baseline.** Specifically, at baseline, students correctly answered 3.8 questions, while at endline, students correctly answered 5.1 questions; the mean scores students achieved at endline were significantly higher than at baseline. Students across grade levels had significantly higher scores at endline than at baseline; students in Grade 1 made greater improvements than students in Grade 2 and Grade 3.

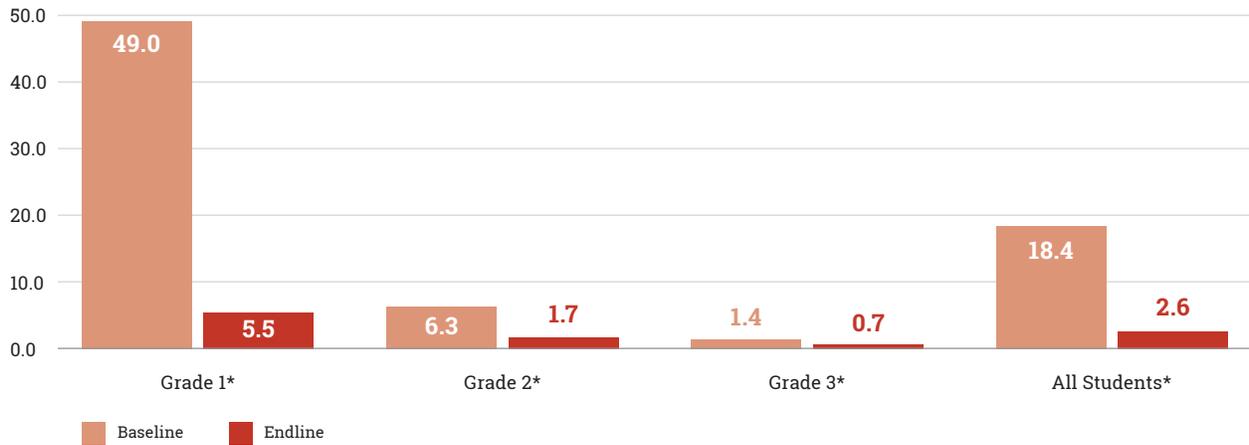
Figure 16: Mean Results at Baseline and Endline by Grade—Reading Comprehension (Correct out of Seven)⁵⁷



The proportion of students receiving zero scores at baseline and endline is presented in Figure 17. **Overall, 18.4 percent of students received zero scores at the baseline while 2.6 percent of students received zero scores at endline;** this difference is statistically significant. The results show that across all grade levels the proportion of students receiving zero scores at endline was lower than the proportion of students receiving zero scores at baseline. The largest decrease was observed among Grade 1 students: the proportion of Grade 1 students receiving zero scores decreased by 43.5 percentage points, compared with decreased of 4.6 and 0.7 percentage points among Grade 2 and Grade 3 students, respectively.

⁵⁷ An asterisk (*) indicates the average gain score were significantly different across grade levels at $p < 0.05$. N sizes: All students $N = 457$; Grade 1 $n = 145$; Grade 2 $n = 174$; Grade 3 $n = 138$.

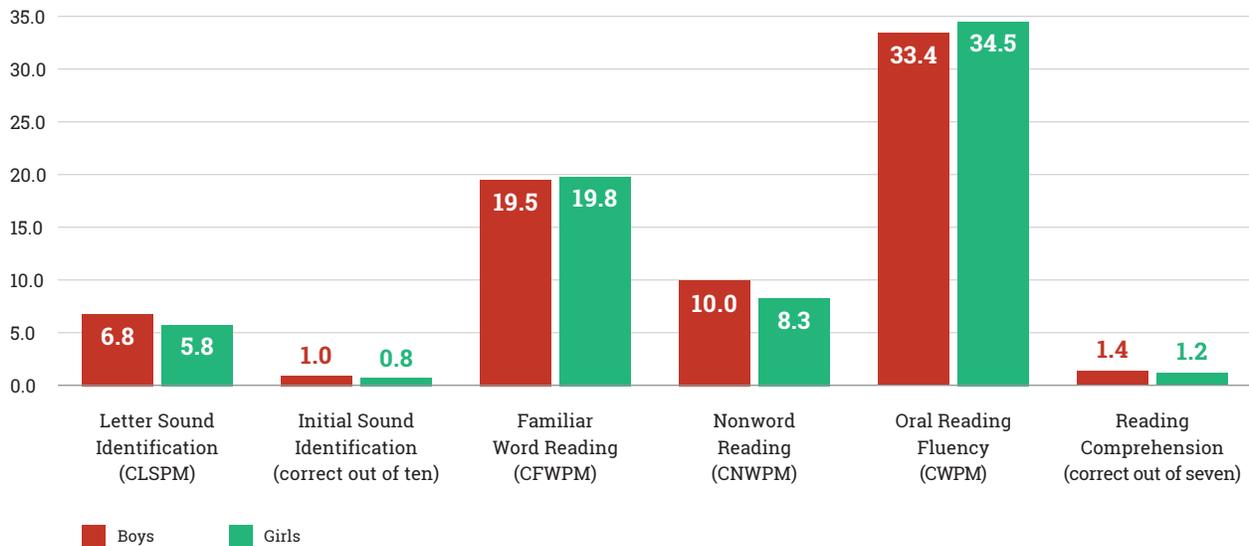
Figure 17: Percentage of Students Receiving Zero Scores by Grade at Baseline and Endline—
Reading Comprehension (%)⁵⁸



EGRA Results by Subgroup

In addition to calculating student gains from baseline to endline across all students and by grade, results were analyzed by students' gender. There were 242 girls and 215 boys in the sample, and the gain scores of girls and boys are presented in Figure 18.

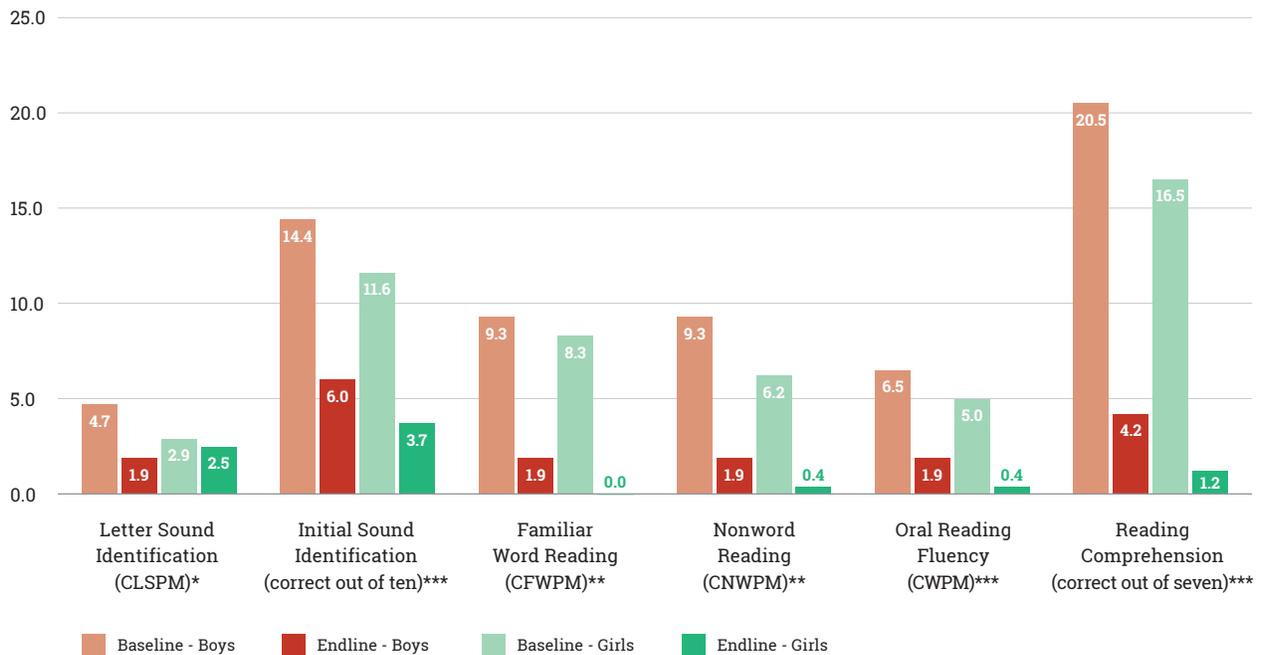
Figure 18: Average Gain Scores by EGRA Subtask and Gender



Across subtasks, girls and boys in the MdL program made comparable gains. On the familiar word reading and ORF subtasks, girls had slightly higher rate gains than boys; on the letter sound identification, initial sound identification, nonword reading, and reading comprehension subtasks, boys had slightly higher gain scores than girls. However, **there were no significant differences between girls' and boys' gain scores across subtasks.**

58 An asterisk (*) indicates the proportion of students receiving zero scores at endline was significantly smaller than the proportion of students receiving zero scores at baseline at $p < 0.05$. *N* sizes: All students $N = 457$; Grade 1 $n = 145$; Grade 2 $n = 174$; Grade 3 $n = 138$.

Figure 19: Percentage of Students Receiving Zero Scores by EGRA Subtask and Gender at Baseline and Endline (%)⁵⁹



The proportions of girls and boys who received zero scores are presented in Figure 19. Across subtasks, both girls and boys had lower proportions of zero scores at endline than at baseline. On three subtasks—initial sound identification, ORF, and reading comprehension—the proportion of both girls and boys receiving zero scores at endline were significantly smaller than the proportion who did so at baseline. Additionally, the proportion of boys receiving zero scores at endline was significantly smaller than the proportion of boys receiving zero scores at baseline on familiar word reading and nonword reading. Finally, the proportion of girls receiving zero scores at endline was significantly smaller than the proportion of girls receiving zero scores at baseline on the letter sound knowledge subtask. Across subtasks, the decrease in the proportion of zero scores from baseline to endline was comparable for girls and boys. For girls and boys, the largest percentage-point decrease was observed on the reading comprehension subtask, on which the proportion of zero scores decreased by 15.3 percentage points for girls and 16.3 percentage points for boys.

Key Factors for Success

Several factors were explored to develop an understanding of students’ experiences in the MdL project and to examine variation in reading gains based on these experiences. Feedback from QFD and from observations and interviews with participants indicates there was variation in student engagement that was potentially based on characteristics of the libraries. Table 10 presents characteristics of the different libraries and the proportion of students who attended libraries with those characteristics.

⁵⁹ An asterisk (*) indicates the proportion of girls receiving zero scores at endline was significantly smaller than the proportion of girls receiving zero scores at baseline at $p < 0.05$. $n = 242$. Two asterisks (**) indicate the proportion of boys receiving zero scores at endline was significantly smaller than the proportion of boys receiving zero scores at baseline at $p < 0.05$. $n = 215$. Three asterisks (***) indicate the proportion of girls and boys receiving zero scores at endline was significantly smaller than the proportion of girls and boys receiving zero scores at baseline at $p < 0.05$. Sample size: Girls $n = 242$; Boys $n = 215$.

Many libraries in the MdL project are in marginalized urban locations; combined they served 77.0 percent of students who participated in the project. About one-third of students attended libraries that are inside or adjacent to a primary school, while two-thirds of students attended libraries that are inside or adjacent to high schools or universities. Overall, 45.3 percent of students attended libraries with a dedicated space that is in an enclosed area and restricted to children; comparatively, 54.7 percent of students attended libraries with an open floor plan.

In total, four libraries serving 35.9 percent of students are located in secure or safe areas, while the remaining 64.1 percent of students were served by six libraries located in insecure or unsafe areas.⁶⁰ Most libraries had three librarians available to support the MdL project and participating students; although, in nine of ten libraries, there was turnover in at least one librarian during the project. The MdL project started implementation in seven libraries in January 2016, at the early start of the program, meaning that 79.4 percent of students started attending libraries as part of the MdL program in that month.

Table 10: Library Characteristics Descriptive Statistics

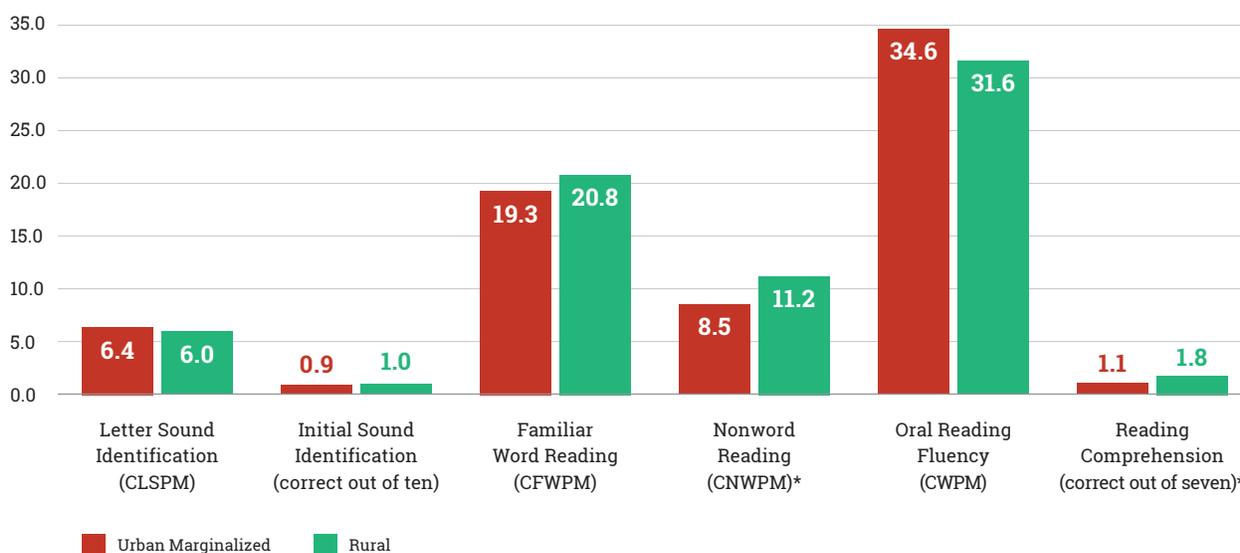
Library Characteristic		Library	Total Students	
		<i>n</i>	<i>n</i>	Percentage of Students (%)
Urban or rural classification	Urban marginalized	8	352	77.0
	Rural	2	105	23.0
Proximity to schools	Attached to primary school	3	153	33.5
	Attached to high school	4	155	33.9
	Attached to university	3	149	32.6
Library structure	Open floor plan	5	207	45.3
	Enclosed MdL library space	5	250	54.7
Safety level	Secure or safe	4	164	35.9
	Not secure or not safe	6	293	64.1
Number of librarians	Two librarians	2	77	16.8
	Three librarians	8	380	83.2
Librarians turnover during project	Yes	9	421	92.1
	No	1	36	7.9
Library start date	January 2016	7	363	79.4
	March 2016	1	48	10.5
	April 2016	1	35	7.7
	May 2016	1	11	2.4

To assess whether the characteristics of the library where students attended were related to the differences in student reading outcomes, a correlational analysis was conducted. This analysis found that the location of the library—urban marginalized versus rural—was the factor that contributed to significant differences in student reading gains. The gains of students that attended urban marginalized and rural libraries are presented in Figure 20 and indicate that students who attended libraries in rural settings had significantly larger gains on the nonword

⁶⁰ Safety levels were determined by the QFD team.

reading and reading comprehension subtasks. There were no significant differences in student reading outcomes based on other characteristics of the libraries.

Figure 20: Average Gain Scores by EGRA Subtask and Urban or Rural Classification⁶¹



To better understand other factors that may have influenced changes in students’ reading outcomes from baseline to endline, questions from the reading habits and attitudes survey were compiled into four composites or groups of questions related to each other. These include disposition towards reading, engagement in the program, family reading support, and socioeconomic status (SES). Each composite consists of a series of items related to a specific theme that may have affected students’ early grade reading skill acquisition; composites were then assigned a maximum score equal to the total number of items in the composite. Unlike other ACR GCD grantees, there is no composite with the MdL project specifically asking students about their comfort with and use of technology.⁶² Table 11 provides descriptive statistics for the four composites. See Annex C for full composite questions, response options, and frequencies.

Table 11: Descriptive Statistics for Composite Scores⁶³

Composite	N	Mean	Std. Deviation	Minimum	Maximum
Disposition to reading	453	3.1	0.7	0.3	4.0
Engagement in program	348	5.1	0.7	1.7	6.0
Family reading support	457	4.6	1.7	0.0	8.0
Socioeconomic status	439	6.6	1.6	1.8	9.0

⁶¹ An asterisk (*) indicates the average gain score were significantly different between students visiting rural and urban libraries at $p < 0.05$. N sizes: All students $N = 457$; Urban marginalized $n = 352$; Rural $n = 105$.

⁶² In the reading habits and attitudes survey, students were asked if they knew how to use a computer. For additional details, see Annex B for response options and frequencies.

⁶³ For each composite, students who were not asked an item or who skipped an item were not included in calculations. For this reason, the N values vary across composites.

For each composite, students' scores were categorized as high or low using the composite's mean score as a cutoff. The average change in EGRA gain scores was then calculated by high or low category to determine if students' composite scores influenced their reading gains. There was no difference in reading gains based on students' disposition to reading composite scores, but there was a difference in reading gains based on the three other composites. Specifically, students categorized as having low engagement in the MdL project had significantly higher gains in the familiar word reading, nonword reading, and ORF subtasks than their peers categorized as having high engagement in the MdL project. Students categorized as having low family reading support made significantly higher gains in the familiar word reading and nonword reading subtasks than their peers categorized as having high family engagement. Finally, students categorized as low SES made significantly higher gains on the nonword reading and reading comprehension subtasks compared to their peers categorized as high SES (see Annex Tables D.12 to D.15).

Finally, the number of student book checkouts and platform logins were examined to determine if there was a relationship between these variables and student reading outcomes. The analysis revealed a weak but significant relationship between the total number of student book checkouts and ORF, indicating that students with more book checkouts tended to have higher gains on the ORF subtask (see Annex Table D.16).⁶⁴

VIII. Adaptive Subtask Results

Based on students' performance on the ORF and reading comprehension subtasks on the baseline EGRA, students completed two additional untimed subtasks—AORF and adaptive reading comprehension. Three outcomes were possible for their adaptive subtasks at baseline:

1. *Outcome A:* For students who did not read enough of the ORF story to go on to the AORF (i.e., read less than 11 words in one minute), the assessment ended after the EGRA reading comprehension subtask on the assumption they would not be able to read other passages of comparable difficulty.
2. *Outcome B:* For students who answered less than three reading comprehension questions correctly, the assessor presented them with an easier AORF story. The easier AORF passage—based on words sentences, and sentence and paragraph length—was called "Toto" and had a total of 97 words.
3. *Outcome C:* For students who answered three or more of the seven comprehension questions correctly, the assessor presented them with a more difficult story. The difficult AORF reading passage was called "Rufo" and had a total of 164 words.

Each AORF reading passage had a corresponding adaptive reading comprehension subtask with six questions.

To enable comparisons over time, students read the same AORF story at endline that they read at baseline even if they read enough of the ORF subtask story to be routed into the more difficult AORF story. The number of students who qualified for a harder reading passage at endline than at baseline was tracked to report the proportion of students who advanced over time. At endline, 101 students read "Toto" (Outcome B), 252 students read "Rufo" (Outcome C), and 94 students did not advance to the adaptive subtasks during the baseline assessment (Option A). Ten students erroneously read a different AORF passage at endline than at baseline and were excluded from the analysis.

64 $r=.140, p=.003, N=457$.

As Table 12 indicates, at endline, students who read “Toto” read 7.8 more words correctly than at baseline. In comparison, students who read “Rufo” read 0.3 fewer words correctly at endline than at baseline.⁶⁵ Students in Grade 1 made larger improvements on the AORF subtask than students in Grade 2 and 3 (see Annex Tables D.7 to D.8).

Table 12: Average Gain Scores by Story—Adaptive Oral Reading Fluency

Adaptive Story	<i>n</i>	Average Gain	SD
“Toto” (correct words read out of 97)	101	7.8	19.6
“Rufo” (correct words read out of 164)	252	-0.3	11.8

Additionally, as indicated in Table 13, students who read “Toto” answered 1.4 more questions correctly at endline than at baseline, while students who read “Rufo” answered 0.4 more questions correctly at endline than at baseline. Students in Grade 1 made larger improvements than students in Grade 2 and Grade 3 (see Annex Tables D.9 to D.10).

Table 13: Average Gain Scores by Story—Adaptive Reading Comprehension

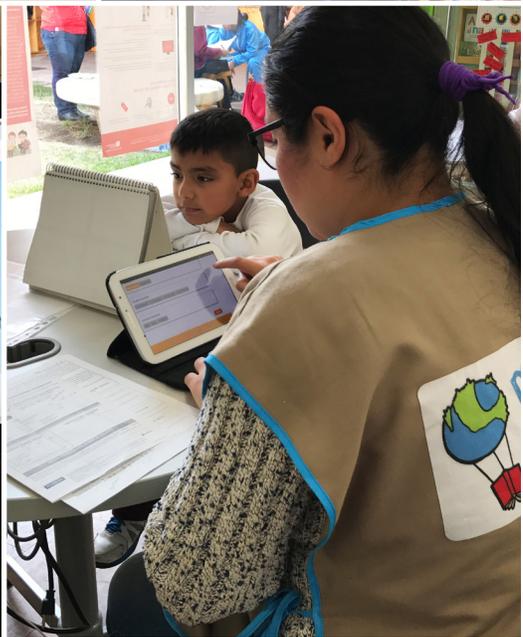
Adaptive Story	<i>n</i>	Average Gain	SD
“Toto” (correct out of seven)	101	1.4	1.5
“Rufo” (correct out of seven)	252	0.4	1.4

The analysis also showed that some students qualified for a different adaptive story at endline than they did at baseline when applying the outcome rules detailed above to their results on the endline ORF and reading comprehension subtasks. As indicated in Table 14, 39.2 percent of all students qualified for a harder passage at endline than what they took at baseline, while 2.6 percent qualified for an easier passage at endline. The majority of students—58.2 percent—qualified for the same level passage at endline as at baseline.

Table 14: Number of Students Who Qualified for the Easier, Same, or Harder Story from Baseline to Endline

Change from Baseline	<i>n</i>	Percentage of total (%)
Qualified for the easier story	12	2.6
Qualified for the same story	266	58.2
Qualified for the harder story	179	39.2

⁶⁵ Gains from baseline to endline on the adaptive oral reading fluency and adaptive reading comprehension subtasks were not tested for statistical significance.



IX. Scalability

Stakeholders are increasingly interested in assessing the scalability of interventions in addition to their results or impacts. To scale up a project means to expand, replicate, adapt, and sustain a successful project in a new geographic area and to reach more beneficiaries over time.⁶⁶ ACR GCD grantees have implemented small-scale pilot projects, and an important consideration after each project is the feasibility of replicating or expanding the technology-based innovation and project models to a different or larger population or area.

To inform this decision, STS conducted scalability assessments for ACR GCD grantees guided by the following research question: *Are the project and technology suitable for scaling?* STS used an indirect approach that relies on qualitative descriptions of project performance around seven parameters of sustainability:

1. Credibility
2. Observability
3. Relevance
4. Relative Advantage
5. Ease of Transfer and Adoption
6. Testability
7. Sustainability of Funding

The seven parameters were adapted from the USAID-funded Scalability Assessment Tool developed by Management Systems International.⁶⁷ The tool includes seven sections and 28 questions. STS used data from EOP interviews, EGRA results, literature reviews, and project M&E to assess scalability parameters. **In its pilot year, the MdL project faced significant development challenges and low FOI. Additionally, because it was not possible to analyze the reading outcomes as intended in the research design or to attribute any observed reading gains to the project, the MdL project should not yet be considered for scale-up.**

The MdL project design is not stable enough to be assessed under the seven parameters. Instead, the following section will present a discussion on the relevance and credibility parameters, which address the question of whether the MdL components were attempting to solve pressing problems or gaps in policy for early grade readers in Mexico.

Relevance

The relevance parameter is based on the knowledge that an intervention must be relevant to the context in which it is being implemented to be scalable. It should effectively address a problem that is recognizable and considered important by stakeholders.

There is evidence that the components of the MdL project are relevant. There is a need for additional early grade reading support in Mexico. Though the country is not the worst-performing in the Latin American region, there does not appear to have been significant improvements in reading skills over the past decade. Results from the 2015 Program for International Student Assessment (PISA) indicate that Mexico performs below the Organization for

⁶⁶ Cooley, L., & Linn, J.F. (2014, September). Taking innovations to scale: Methods, applications and lessons. *Washington, D.C.: Results for Development Institute*. Retrieved from https://www.usaid.gov/sites/default/files/documents/1865/v5web_R4D_MSI-BrookingsSynthPaper0914-3.pdf

⁶⁷ Ibid.

Economic Co-operation and Development (OECD) average in reading with 42 percent of students performed below Level 2 in reading; this has remained unchanged since 2009.⁶⁸ Additionally, parental engagement in their children's literacy practices varies across sociocultural environments, which indicates a need for better parent and caretaker education on their role in their children's education.⁶⁹

It is unclear if the MdL project addressed a priority on the policy agenda for stakeholders in Mexico. In 2012, Government of Mexico (GoM) expenditure on education was 5.2 percent of the total gross domestic product, which was slightly higher than the average for other countries in Latin America.⁷⁰ However, expenditure on primary schools as a percentage of government expenditure on education has steadily decreased from 40.4 percent in 2002 to 34.7 percent in 2012.⁷¹ Additionally, there was no specific policy indicating that access to leveled books for primary school students is a top priority for the GoM. The GoM initiated its National Basic Education Reading Program (PNL) in the late 2000s; however, over time and with changes in administrations, PNL's funding has not remained consistent. Instead, the GoM has diverted education funds into different initiatives, such as the Quality Schools Program, do are not specifically target reading.

Primary school principals who participated in EOP interviews reported that they do not feel there is strong guidance on how to support students with different reading levels within the same grade. Principals also acknowledged that they do not believe that their schools have adequate or appropriate reading materials for students. Based on feedback received during EOP interviews, only a few primary schools had physical libraries on-site; of those, some were used for other purposes, such as computer storage, and others were locked and not accessible by students. Instead of a school library, classrooms had small book corners where books received through the PNL were kept. No principal said that they felt that the books they received from the GoM were appropriate for student reading levels, sufficient, or engaging for students.

These issues in schools are exacerbated by the lack of books in homes. Based on results from the 2015 National Reading Survey, only 18.1 percent of Mexicans reported having children's books at home.⁷² Parents and caretakers corroborated this in EOP interviews and expressed that the costs of good children's books in Mexico were prohibitive; they further noted that there were no public libraries with books that were appropriate and interesting for their children.

Credibility

The credibility parameter is based on the knowledge that an intervention or innovation must be credible to be supported and taken to scale through either replication or expansion. This aspect of scalability assesses if various stakeholders—including potential adopters, funders, implementers, and beneficiaries—believe that the model has a strong evidence base that may include existing empirical research or anecdotal information.

Due to low implementation fidelity, it is not possible to assess adequately the credibility of the MdL project as piloted. Further empirical research could better illuminate the evidence base for the project model. The MdL project was developed to link students with level-appropriate books. The key component of the MdL project, the MATCH algorithm that provided personalized book recommendations according to students' assessed level of vocabulary and reading skills, was developed in the model of reading level frameworks—such as The Lexile®

68 Country note: Programme for international student assessment (PISA) results from PISA 2015. (2016) OECD. Retrieved from <https://www.oecd.org/pisa/PISA-2015-Mexico.pdf>

69 Reese, L., Aarau, R.M., & Bazán, A.R. Mexican parents' and teachers' literacy perspectives and practices: Construction of cultural capital. *International Journal of Qualitative Studies in Education*, 25.8, 983-1003. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531546/>

70 UNESCO Institute for Statistics. Accessed via World Bank Open Data (<http://data.worldbank.org/>).

71 Ibid.

72 Cornejo Nacional para la Cultura y las Artes (Conaculta) [National Council for Culture and Arts of Mexico]. (2015). Encuesta Nacional de Lectura 2015. https://observatorio.librosmexico.mx/files/encuesta_nacional_2015.pdf.

Framework for Reading⁷³ and the Developmental Reading Assessment reading levels.⁷⁴ There was not, to the knowledge of QFD, any comparable leveling method within Spanish-speaking countries; the MATCH algorithm was designed to fill this gap.

Nevertheless, given poor access to books in schools and at home, stagnant levels of reading skills among early grade students, and uneven parental engagement, it is not clear if a web-based platform with personalized book recommendations is the best solution to address the current challenges facing Mexican students in learning to read. The majority of students and parents who participated in EOP interviews did not see logging on to the computer to access their recommendations as fundamental or necessary; instead, nearly all students selected the books they wanted to read based on recommendations from their friends or by looking at the book's pictures. Additionally, many parents and caretakers noted that they were not comfortable using computers, and this dissuaded them from helping their children—or, at a minimum, encouraging them—to log on to the MdL platform. It is possible that improvements in design and implementation could mitigate these challenges in the future.

X. Cost Analysis

A cost analysis is often a component of scalability assessments as it helps decision makers and stakeholders understand the feasibility of replication with given budgetary constraints. Because ACR GCD grantees implement new approaches, they often allot significant financial resources to developing new materials that could be used on a recurring basis. To better understand the funding requirements of the MdL project, a cost analysis was conducted to present the total cost of the intervention and to understand the investments that would be needed for project replication or scale-up.

USAID guidance on conducting cost analyses on early grade reading projects suggests that the “ingredients method”⁷⁵ be used to calculate costs in the following categories:

- Management and associated technical costs
- Development costs
- Implementation costs

Project staff completed a costing template with guidance from World Vision and STS. Costs were outlined based on the activities from the project work plan, and each expenditure was classified based on the three categories above. Invoiced costs were used for analysis from the beginning of the project through March 2017.⁷⁶ Though costs specific to the close-out of the project are not included in this analysis, these would be categorized as implementation and management costs. The absences of these costs should be considered when comparing the proportion of project budget spent on the three categories. Table 15 provides a detailed breakdown of costs by category based on MdL's project activities.

73 <https://lexile.com/>

74 <http://www.scholastic.com/parents/resources/article/book-selection-tips/assess-dra-reading-levels>

75 RTI International. (2015). Measurement and research support to education strategy goal 1: Early grade reading costing template and guidance. Washington, D.C.: USAID. http://pdf.usaid.gov/pdf_docs/PBAAF458.pdf

76 The total budget of the project was \$514,385. This includes \$338,971 from ACR GCD and \$175,414 in QFD cost share.

Table 15: Cost Analysis

Activity	Management	Development	Implementation
Objective 1 – Project startup			
<i>Activity 1.1</i> - Attend start-up workshop	\$ -	\$ 3,476	\$ -
<i>Activity 1.2</i> - Recruit main team members	\$ -	\$ 2,100	\$ -
<i>Activity 1.3</i> - Select and purchase equipment and supplies	\$ -	\$ 480	\$ -
<i>Activity 1.4</i> - Meet with potential collaborators	\$ -	\$ 9,705	\$ -
<i>Activity 1.6</i> - Communication	\$ -	\$ -	\$ 7,923
Objective 2 – Design tests to assess children's reading abilities			
<i>Activity 2.1</i> - Adapt EGRA to the Mexican/regional context	\$ -	\$ 5,004	\$ -
<i>Activity 2.2</i> - Design complementary assessments (reading comprehension) and surveys (socio-emotional)	\$ -	\$ 3,233	\$ -
<i>Activity 2.3</i> - Pilot tests (EGRA and complementary tests) and surveys (socio-emotional)	\$ -	\$ 4,952	\$ -
<i>Activity 2.4</i> - Analyze the results from the pilot and adjust the assessment and survey	\$ -	\$ 1,335	\$ -
<i>Activity 2.5</i> - Define abilities classification (groups) based on the students' levels defined with the assessments	\$ -	\$ 4,299	\$ -
<i>Activity 2.6</i> - Program the assessments and survey onto tablets	\$ -	\$ 1,416	\$ -
<i>Activity 2.7</i> - Purchase Peabody Picture Vocabulary Test licenses (Spanish version)	\$ -	\$ -	\$ 175
Objective 3 – Classify, select and buy books for the project			
<i>Activity 3.1</i> - Meet with publishers and other relevant organizations	\$ -	\$ 3,352	\$ -
<i>Activity 3.2</i> - Define the text's parameters and topics of interest to select the books	\$ -	\$ 6,888	\$ -
<i>Activity 3.3</i> - Define the books' levels, based on the texts' parameters	\$ -	\$ 5,707	\$ -
<i>Activity 3.4</i> - Select the books for the project's catalog	\$ -	\$ 3,280	\$ -
<i>Activity 3.6</i> - Electronically record the books' characteristics for classification	\$ -	\$ 9,455	\$ -
<i>Activity 3.7</i> - Order books	\$ -	\$ -	\$ 13,261
<i>Activity 3.8</i> - Request drawings or comments related to the books	\$ -	\$ -	\$ 6,774
Objective 4 – Create the technology-based platform (website)			
<i>Activity 4.1</i> - Design brand image, logo and name	\$ -	\$ 2,988	\$ -
<i>Activity 4.2</i> - Design the platform, including educational content	\$ -	\$ 11,711	\$ -
<i>Activity 4.3</i> - Define algorithm to link reader's levels with the appropriate book levels	\$ -	\$ 7,499	\$ -
<i>Activity 4.4</i> - Create a database with all the participant's information, book catalog, quizzes, loans' history, etc.	\$ -	\$ 8,644	\$ -
<i>Activity 4.5</i> - Pilot the web-based platform and make adjustments	\$ -	\$ 2,886	\$ -

Table 15: Cost Analysis (continued)

Activity	Management	Development	Implementation
Objective 5 – Design tests to assess children’s reading abilities			
Activity 5.1 - Visit and selection of library sites for the project	\$ -	\$ -	\$ 3,543
Activity 5.2 - Label books (USAID identity) and implement bar code system	\$ -	\$ -	\$ 6,379
Activity 5.3 - Buy, label and set-up furniture (i.e. bookshelves and bean bag chairs)	\$ -	\$ -	\$ 6,212
Activity 5.4 - Train the librarians	\$ -	\$ -	\$ 1,796
Activity 5.5 - Design and distribution of promotion materials	\$ -	\$ -	\$ 21,987
Activity 5.6 - Design, print and distribute registration forms	\$ -	\$ -	\$ 3,050
Activity 5.7 - Register children in the system and assign them a specific profile	\$ -	\$ -	\$ 1,662
Activity 5.8 - Assign libraries to experimental groups	\$ -	\$ -	\$ 3,997
Objective 6 – Implementation of the program			
Activity 6.1 - Train evaluators	\$ -	\$ -	\$ 10,065
Activity 6.2 - Administer baseline assessment and survey	\$ -	\$ -	\$ 12,926
Activity 6.3 - Design bi-monthly workshops with parents	\$ -	\$ 5,567	\$ -
Activity 6.4 - Implement the workshops with parents	\$ -	\$ -	\$ 4,342
Activity 6.5 - Monitor program’s implementation	\$ -	\$ -	\$ 12,040
Activity 6.7 - Analyze reading and book loans’ patterns	\$ -	\$ -	\$ 4,638
Activity 6.8 - Administer exit assessments and surveys	\$ -	\$ -	\$ 10,699
Activity 6.9 - Library operation	\$ -	\$ -	\$ 70,124
Activity 6.10 - Other activities	\$ -	\$ -	\$ 13,335
Objective 7 – Data analysis and reporting			
Activity 7.1 - Design surveys for parents (baseline and endline)	\$ -	\$ 6,085	\$ -
Activity 7.2 - Plan and design qualitative instruments	\$ -	\$ 6,408	\$ -
Activity 7.3 - Analyze qualitative and quantitative data	\$ -	\$ -	\$ 5,504
Activity 7.4 - Prepare final report	\$ -	\$ -	\$ 3,913
Total⁷⁷	\$ 104,703	\$ 116,471	\$ 224,344
Percentage of Total (%)	23.5	26.1	50.4

⁷⁷ Total may not equal sum of individual lines due to rounding.

The management category includes costs that are not directly related to the implementation of activities; these are likely to vary widely based on who is overseeing the implementation of the intervention. **Management costs for the MdL project represent 23.5 percent of the costs expended** and include: the cost of maintaining the project office in Mexico City; personnel salaries; travel, lodging, and per-diem costs for technical consultants; and, other indirect rates and fees.⁷⁸

Development includes the costs related to the development of materials, survey instruments, programs, and other content that would not need to be redeveloped in the scale-up of a project. **The development costs for the MdL project represent 26.1 percent of the costs expended.** The major expenses within this category were related to the creation of the web-based platform—including the algorithm—and the selection and classification of books for the MdL catalog. The categorization of development costs for the MdL project assumes that the platform and the book catalog would remain the same for future iterations of the project.

The implementation cost category is arguably the most relevant for stakeholders who are considering scaling up a project or intervention. This category includes all the recurrent activities and costs that would need to be expensed should the project be replicated, including materials printing and distribution, training, M&E, events and presentations, workshops, and human resources activities. **For the MdL project, implementation costs represented 50.4 percent of the total project cost.** Within this cost category, the largest expenses were related to the operation and monitoring of the libraries and the purchase of the books for the MdL book catalog. The categorization of implementation costs for the MdL project assumes that the project would be implemented in different libraries from those in the pilot year and that additional books would need to be purchased for these new libraries.

Projects sometimes benefit from in-kind services, institutional support, or preexisting relationships with stakeholders or governments that may provide the project with tangible benefits, although it may be difficult or not possible to monetize the costs. Examples of this include local volunteers, strong capacity or support from a large non-governmental organization, or relationships with local governments that could ease logistics and procedures. The MdL project was able to install their book catalogs within preexisting library spaces that already had information and communications technology equipment, including tablets and computers; as a result, the project did not incur costs for these. Further, because the MdL project could use librarians paid through Fundación Proceso's digital libraries, the project did not have to hire and pay for full-time librarians to staff the MdL library; instead, QFD only had to pay a small bonus to each librarian for their work on the MdL project.

⁷⁸ Management costs are inclusive of a 17% flat fee charged for Negotiated Indirect Cost Recovery Agreement (NICRA), which captures indirect costs including regional management and technical support, the local RBI country representative, Overseas Operations management (RBI headquarters), Program Quality and Support (RBI headquarters), and shipping and procurement costs. This also captures miscellaneous headquarters-based services that were provided to the project including finance, internal auditing, human resources, executive management, board, and global knowledge and information management. This analysis assumes that no NICRA expenses were also billed as independent line items, although it should be recognized that some double-counting may have occurred.

XI. Conclusions

QFD, an organization with expertise in education and extensive experience in research and evaluation, conceptualized the MdL project as a way to provide students in early primary grades access to appropriate and engaging reading materials that fit their level and interests to improve their reading skills. QFD paired standard components—a diverse children’s book catalog and workshops for parents and caretakers—with innovative technologies—a web-based platform and an algorithm that provided students with personalized book recommendations, both accessible through computers and tablets. The project’s goal hinged on parents and caretakers attending workshops and students logging on to the platform and receiving personalized book recommendations from the MATCH algorithm.

In its pilot year, the MdL project faced significant development and implementation challenges that impeded the team’s ability to deliver the intervention to students, parents, and caretakers as intended. Data indicate that although 60.6 percent of students enrolled in the MdL project checked out at least one book during the project, only 33.5 percent logged on to the platform at least once. The average number of books checked out per student and the average number of logins per student across the nine to eleven months of implementation were 12.7 books and 1.2 logins, respectively. Further, out of the 290 students who attended libraries that held parent and caretaker workshops, 80.3 percent did not have a parent or caretaker attend a single workshop. Because of findings that indicate very low project uptake and FOI, it was not possible to assess if the reading gains of the MdL students were associated with the intervention. The EGRA results presented in this report should be interpreted as the reading gains from a combination of an additional year of schooling plus participation in the project.

The following are lessons that should be considered for any future interventions incorporating components of the MdL project.

Lessons Learned



Initiatives like the MdL project that require complex, custom technology components need to have sufficient time and funds allocated for development, pilot testing, and finalizing before the start of the intervention.

The MdL project necessitated the comprehensive development of two major technological components: the web-based platform and the MATCH algorithm. Neither of these technologies had advanced beyond the theoretical stage at the start of QFD’s grant from ACR GCD, and as a result, it took significant time and financial investment to fully conceptualize and operationalize these two components. The MdL project began implementation in several libraries before the platform was finalized, and students affiliated these libraries may have grown accustomed to accessing books without using the platform to receive book recommendations. Funders and implementers should consider personnel capacities and accurately predict the financial and time investments necessary to develop complex custom technologies before implementing projects. They should also consider leveraging the expertise of implementers who have experience managing projects and coupling them with researchers to maximize financial investments in innovative pilot projects.



Implementers should not assume that technology in and of itself will be valuable or necessary to project participants.

The MdL project's theory of change hinged on students logging on to the web-based platform to receive personalized book recommendations. Although there were computers and tablets available in the digital libraries, usage of the web-based platform was very low. Only 33.5 percent of students logged on to the platform at least once during the project, and students only logged in to the platform an average of 1.2 times throughout the entire project. More specifically, 65.3 percent of students who were randomly assigned to receive book recommendations based on the MATCH algorithm did not log on to the platform even once during the project; these students were not exposed to personalized book recommendations. When interviewed, students said they preferred to choose books because of the drawings, specific characters, or recommendations from their peers or librarians—indicating that students did not feel the platform was essential for their library experience. For future iterations of the project, QFD should find a way to make the MdL platform a more integrated part of the library intervention, either through better engaging students in the platform or by making use of the platform compulsory. QFD could also explore making the MATCH algorithm directly available to teachers or librarians so that they can provide recommendations to students.



Students who participated in the MdL project had significant reading gains from baseline to endline after an additional year of schooling and exposure to the MdL project.

Across grades, students gained 34.0 CWPM from baseline to endline on the ORF subtask and 19.6 CFWPM on the familiar word reading subtask. Although the proportions of students receiving zero scores were relatively low at baseline, there were decreases in the proportion of zero scores across all subtasks from baseline to endline. The largest reduction in the proportion of students receiving zero scores was observed on the reading comprehension subtask, on which there was a 15.8 percentage-point decrease from baseline to endline. Across subtasks, Grade 1 students showed the largest improvements in reading skills.⁷⁹ Though these results are promising, because of low project uptake and FOI it is not possible to know how much of students' reading gains were attributable to the MdL project—case in point, students categorized as having low engagement in MdL had higher gains on the familiar word reading, nonword reading, and ORF subtasks on average than their peers who were categorized as having higher engagement.



It is important that the research design of pilot projects include a comparison group, which does not receive any intervention components.

The research design for the MdL project aimed to capture differences in reading outcomes among students who received personalized book recommendations through the MATCH algorithm and students who received random book recommendations. However, the research design did not include a group of students who did not have access to a library or its books. Without a comparison group, it is impossible to know whether changes from baseline to endline were due to the intervention or simply to maturation. In future iterations of the MdL project, implementers should strongly consider including a comparison group that does not have exposure to any of the project's components.

⁷⁹ Given it is their first year in school, Grade 1 students have the greatest opportunity for growth across all subtasks, while students in Grade 2 and 3 already show a ceiling effect with high fluency scores at baseline.



When considering research design, project management should ensure reliable sources of data to answer all research questions.

The MdL project had three project-specific research questions that aimed to understand the reading outcomes associated with different components of the project. However, limited or no data were available to answer two of those questions. No data were readily available to link whether students who viewed book recommendations on the MdL platform ultimately checked out or read the recommended books, and parental engagement was measured through student questionnaires due to insufficient response rates to parental questionnaires. Even if the project been implemented with fidelity, the lack of reliable data would have limited the project's ability to answer its established research questions. Improvements should be made to capture the link between book recommendations and books read, and more reliable ways to capture changes in parental engagement should be introduced.



Initiatives like the MdL project that implement components in disparate physical sites—such as ten separate libraries—should constantly monitor variations in implementation to understand project outcomes better.

The MdL project was implemented in ten different digital libraries each with slightly different characteristics across Estado de México. This resulted in significant variability of student engagement in project components between libraries. For example, the average number of book checkouts per student by library ranged from 44.0 books to 2.8 books, and the average number of platform logins per student by library ranged from 5.9 logins to 0.3 logins. Although this analysis did not determine any specific library characteristics that correlated with reading gains, this could be due to the minimal amounts of project dosage received by students. In the future, the MdL project should continue to monitor implementation at each library site closely and attempt to minimize variation across sites.



To adequately engage parents and caretakers, initiatives like the MdL project must find creative ways to incentivize or encourage parents and caretakers to participate.

One of the biggest challenges of the MdL project was the extremely low attendance rate at parent and caretaker workshops—on average, less than ten percent across workshops. Only 19.7 percent of students had a parent or caretaker attend at least one workshop. Parents and caretakers expressed excitement about the MdL project in interviews, but most all cited lack of time as the reason for not attending project workshops. The MdL project attempted to engage parents better by changing the way workshops were marketed and providing informational handouts for those unable to attend, but engagement remained low throughout the project. If project management deems parental engagement a key component in future iterations of the MdL project, they should explore different implementation models, such as home visits with reading coaches.

Annexes

Annex A: Baseline and Endline EGRA Instrument

Vamos a dar inicio a la prueba de EGRA. Lea al estudiante solamente las instrucciones que están en **negritas** (esto es su “guión”).

[Las instrucciones para el aplicador aparecen entre corchetes y en letras cursivas. NO deben leerse al niño]

[Es importante establecer un ambiente relajado con el niño a través de la conversación inicial. El niño debe percibir la situación como un juego o ejercicio y no una prueba. En la presentación inicial no debe tomarse más de dos minutos.]

Observaciones:

Sección 1: Conocimiento de los sonidos de las letras

Aquí tienes una serie de letras para que me digas su sonido. Por favor dime solamente los SONIDOS de estas letras, no sus nombres.

Empiezo con un ejemplo: el sonido de esta letra *[señala la F]* es "/ffff/". Ahora inténtalo tú con esta otra letra.

Dime el sonido de esta letra *[señala la letra "T"]* :

[Si el niño/a responde correctamente, diga] : bien, el sonido de esta letra es "/tttt/".

[Si el niño/a no responde correctamente, diga] : el sonido de esta letra es /tttt/".

Ahora dime el sonido de esta letra *[señala la letra "a"]* :

[Si el niño/a responde correctamente, diga] : bien, el sonido de esta letra es "/aaa/".

[Si el niño/a no responde correctamente, diga] : el sonido de esta letra es "/aaa/".

¿Comprendes lo que debes hacer? Cuando te diga "comienza", dime los SONIDOS lo más rápido y lo mejor que puedas.

Comienza aquí y continúa en esta dirección *[indicar con su dedo dónde iniciar y en la dirección de izquierda a derecha]*. Si hay una letra cuyo sonido no conoces, sáltatelo y continua con la siguiente letra. Ahora me voy a quedar en silencio y te voy a escuchar, a menos que necesites ayuda. ¿Listo(a)? Comienza por favor.

Ejemplo: f T a

E	s	n	r	D	G	c	o	q	F	10
O	A	d	S	R	n	t	a	i	N	20
r	u	e	T	r	s	a	o	n	d	30
A	e	k	U	O	t	c	P	m	E	40
x	r	C	y	L	a	i	D	l	E	50
W	K	J	o	r	c	ñ	Z	O	X	60
o	l	S	i	n	a	s	u	e	M	70
d	f	b	r	Ñ	z	i	S	g	R	80
l	u	e	A	m	L	s	t	E	i	90
A	e	N	Q	p	a	Y	j	B	w	100

Ejercicio Detenido

Tiempo (segundos) si leyó en menos de 1 minuto:

Sección 2: Identificación del sonido inicial

Sabemos que cada letra tiene un sonido, por ejemplo, la letra M suena así: /mmm/. Ahora, voy a leerte algunas palabras para que me digas su primer sonido.

Por favor escucha bien y dime el sonido con el que comienza cada palabra. Empiezo con “mamá”; el primer sonido de “mamá” es /mmm/. Practiquemos juntos ¿Cuál es el primer sonido de mamá? /mmm/

*[Si lo hace incorrecto, diga] : Practiquemos de nuevo la palabra “mamá”, cuyo primer sonido es /mmm/.
[Marcar con énfasis el sonido /mmm/]. Dime cuál el primer sonido de “mamá”.*

[Si dice, /mmm/ diga] : ¡Muy bien! El primer sonido es /mmm/.

Practiquemos con otra palabra, ¿Cuál es el primer sonido de la palabra “la”?

[Si dice, /lll/ diga] : ¡Muy bien!, El primer sonido de “la” es /lll/.

[Si lo dice incorrectamente diga] : El primer sonido de “la” es /lll/.

¿Entendiste lo que vamos a hacer? ¿Listo? Te voy a dar la primera palabra. ¿Cuáles el primer sonido de la palabra? *[Repita cada vez estas instrucciones, repitiendo la palabra una segunda vez.]*

¿Cuáles el primer sonido de la palabra _____ ?

¿Cuál es el primer sonido de la palabra?	Sonido	Marque correcto o incorrecto		
sol	/s/	<input type="checkbox"/> Correcto	<input type="checkbox"/> Incorrecto	<input type="checkbox"/> No responde
ratón	/r/	<input type="checkbox"/> Correcto	<input type="checkbox"/> Incorrecto	<input type="checkbox"/> No responde
nieto	/n/	<input type="checkbox"/> Correcto	<input type="checkbox"/> Incorrecto	<input type="checkbox"/> No responde
dedo	/d/	<input type="checkbox"/> Correcto	<input type="checkbox"/> Incorrecto	<input type="checkbox"/> No responde
robo	/r/	<input type="checkbox"/> Correcto	<input type="checkbox"/> Incorrecto	<input type="checkbox"/> No responde
casa	/c/	<input type="checkbox"/> Correcto	<input type="checkbox"/> Incorrecto	<input type="checkbox"/> No responde
taco	/t/	<input type="checkbox"/> Correcto	<input type="checkbox"/> Incorrecto	<input type="checkbox"/> No responde
mar	/m/	<input type="checkbox"/> Correcto	<input type="checkbox"/> Incorrecto	<input type="checkbox"/> No responde
pato	/p/	<input type="checkbox"/> Correcto	<input type="checkbox"/> Incorrecto	<input type="checkbox"/> No responde
barco	/b/	<input type="checkbox"/> Correcto	<input type="checkbox"/> Incorrecto	<input type="checkbox"/> No responde

Ejercicio Detenido (estudiante dijo los primeros cinco sonidos incorrectamente)

Sección 3: Lectura de palabras simples

Aquí tienes una serie de palabras para que las leas, una por una. Te voy a dar un ejemplo: esta palabra es “el”. Ahora inténtalo tú con esta otra palabra. *[señale la siguiente palabra: mi]* Léela en voz alta.

[si el estudiante responde correctamente, diga] : Muy bien: “mi”.

[si el estudiante no responde correctamente, diga] : Esta palabra es “mi”.

Si hay una palabra que no conozcas, no te preocupes, continua con la siguiente palabra. ¿Entendiste lo que vamos a hacer? Cuando te diga “comienza”, lee las palabras lo más rápido y lo mejor que puedas. Cuando pase un tiempo voy a decir “alto” para que te detengas. Pon tu dedo debajo de la primera palabra.

¿Listo? Comienza por favor.

con	peso	jefe	rana	come	5
reino	la	poco	eso	solo	10
lado	cerca	De	que	piña	15
casa	el	tela	luna	hada	20
no	niña	cara	más	así	25
gata	mano	vela	kilo	paz	30
fiel	como	ala	nada	el	35
por	feliz	cena	buscar	río	40
cola	linda	uno	hijo	vida	45
dolor	alegre	queso	otro	dulce	50

Ejercicio Detenido

Tiempo (segundos) si leyó en menos de 1 minuto:

Sección 4: Lectura de palabras sin sentido

Aquí tenemos una serie de palabras inventadas, las vas a leer por favor. Te voy a dar un ejemplo: esta palabra inventada es “ut”. Ahora inténtalo tú con esta otra. Léela en voz alta [señale: dif].

[Si el estudiante dice “dif”, diga] : “Muy bien: dif”.

[Si el estudiante no dice “dif” correctamente, diga] : Esta palabra inventada es “dif”.

¿Entendiste lo que vamos a hacer? Cuando te diga “comienza”, lee las palabras lo más rápido y lo mejor que puedas por favor. Cuando pase un tiempo voy a decir “alto” para que te detengas. Pon tu dedo debajo de la primera palabra.

¿Listo? Comienza por favor.

pamo	vede	doso	repa	peno	5
sadi	helo	mase	pepu	quele	10
rapu	gaba	ferre	cupa	cavi	15
side	colo	dipu	nide	pacu	20
letu	ficu	lono	depe	rabu	25
invi	eslo	arti	epta	osla	30
ibos	abto	edno	actu	optu	35
epcu	olsi	undo	endo	ursi	40
carte	gravu	invlo	lecda	abuto	45
papre	protu	ultla	cuema	imate	50

Ejercicio Detenido

Tiempo (segundos) si leyó en menos de 1 minuto:

Sección 5a: Lectura y comprensión de un párrafo

Aquí tienes un cuento, quiero que lo leas en voz alta. Si hay una palabra que no conozcas, no te preocupes, continua con la siguiente palabra. Cuando termines, te haré algunas preguntas sobre el cuento.

¿Entendiste lo que vamos a hacer? Cuando te diga "comienza", lee el cuento lo más rápido y mejor que puedas.

Después de un rato voy a decir "alto" para que te detengas. Pon tu dedo debajo de la primera palabra y continúa en esta dirección (*indicar con el dedo donde iniciar y en la dirección de izquierda a derecha*). ¿Listo (a)?

Ahora, por favor, comienza.

Quando termine de leer el niño/a dígame: Ahora te voy a hacer unas preguntas sobre el cuento que leíste, ¿listo/a? Después de leer cada pregunta, dé al niño/a cuando mucho 15 segundos para responder. Marque la casilla adecuada a la respuesta.

Anita se tomó rápido el atole y se quemó la lengua.

La sentía como si fuera de madera.

No es la primera vez que se quema.

Una vez metió la mano en el fuego y se quemó dos dedos.

Le ardieron por tres días; luego se le hizo una ampolla que su mamá le reventó con una aguja caliente.

[Si se detuvo el ejercicio, acabar EGRA → No pasar a Toto ni Rufo]

Ejercicio Detenido

Tiempo (segundos):

Sección 5b: Lectura y comprensión de un párrafo

[Quite el texto al estudiante después de que lo haya leído y diga]. [Las preguntas en itálicas indican que las repuestas pueden ser posibles o imposibles, en lugar de correctas o incorrectas].

¡Muy bien! Ahora te voy a hacer algunas preguntas sobre el cuento.

1. ¿Quién se quemó la lengua? (Respuesta esperada: Anita, Ana, la niña)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

2. ¿Cómo se tomó Anita el atole? (Respuesta esperada: Rápido, muy rápido, de prisa)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

3. ¿Cómo estaba el atole que se tomó Anita? (Respuesta esperada: Caliente, hirviendo)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

4. ¿Cómo que sentía Anita su lengua cuando se quemó con el atole? (Respuesta esperada: Como de madera, dura)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

5. ¿Cuántos dedos se quemó Anita cuando metió las manos al fuego? (Respuesta esperada: Dos)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

6. ¿Con qué le reventaron las ampollas cuando se quemó los dedos? (Respuesta esperada: Con una aguja, aguja caliente)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

7. ¿Qué debe hacer Anita para no volverse a quemar? (Respuesta esperada: No tocar cosas calientes, tener cuidado)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

Si no contestó ninguna pregunta → Detener el ejercicio

Si contestó entre 1 y 3 de manera correcta → Pasar a Toto

Si contestó 4 o más de manera correcta → Pasar a Rufo

Sección 6T: TOTO Lectura y comprensión

Aquí tienes otro cuento para que lo leas. Quiero que lo leas en voz alta. Cuando termines, te haré algunas preguntas sobre el cuento.

¿Entendiste lo que vamos a hacer? Cuando te diga “comienza”, lee el cuento lo más rápido y mejor que puedas. Pon tu dedo debajo de la primera palabra.

¿Listo? Comienza ahora por favor.

Tengo un perrito que se llama Toto. Mi mamá me lo regaló; lo encontró en la calle con hambre y enfermo. Le di comida y agua; lo llevé al veterinario para que lo curara.

Con el tiempo se puso sano e inquieto. Es muy travieso. Con sus dientes muerde lo que está a su alcance: el bote del agua, un calcetín que dejé en el piso, un carrito que olvidé en mi cuarto. Todos los días lo llevo al parque para que corra y juegue con otros perritos.

Ahora está grande, ladra mucho y cuida la casa.

Ejercicio Detenido

Tiempo:

Sección 6T: TOTO Lectura y comprensión de un párrafo

[Quite el texto al estudiante después de que lo haya leído y diga] [Las preguntas en itálicas indican que las respuestas pueden ser posibles o imposibles, en lugar de correctas o incorrectas]. [No hay tiempo pero se registra cuánto tiempo le tomó]

Ahora te voy a hacer algunas preguntas sobre el cuento.

1. ¿Quién es el personaje de este cuento? (Respuesta esperada: Toto, un perro, un perrito)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

2. ¿Con quién llevaron a Toto para que lo curaran? (Respuesta esperada: Veterinario, doctor, doctor de animales)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

3. ¿A dónde llevaban a Toto todos los días? (Respuesta esperada: Al parque)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

4. ¿Estaba Toto bien cuidado y alimentado antes de que se lo encontrarán? (Respuesta esperada: No, hambre, enfermo)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

5. ¿Por qué Toto, mordía todo lo que estaba cerca de él? (Respuesta esperada: Porque le gustaba, porque sentía comezón, por juguetón, por inquieto, por hambre)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

6. ¿Por qué Toto cuida la casa? (Respuesta esperada: Porque ya es grande, porque le dan comida, porque es su deber/trabajo)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

[Una vez concluido todo el ejercicio, agradecer al niño su participación excepto en los casos que se siga con otro ejercicio.] ¡Muchas gracias por ayudarme!

Sección 6R: RUFO Lectura y comprensión

Aquí tienes otro cuento para que lo leas. Quiero que lo leas en voz alta. Cuando termines, te haré algunas preguntas sobre el cuento.

¿Entendiste lo que vamos a hacer? Cuando te diga “comienza”, lee el cuento lo más rápido y mejor que puedas. Pon tu dedo debajo de la primera palabra.

¿Listo? Comienza ahora por favor.

Rufo era un perrito que se estaba transformando en un perro adulto que no tenía nada que hacer a diferencia de otros animales. Pero él quería ayudar en la granja.

Una tarde tomaba una siesta cuando vio a la gorda y alegre Celia cargando una canasta bajo el brazo. “Ahí va Celia a recoger huevos. La voy a ayudar.”

Salió brincando ligero y feliz. En un segundo llegó al nido más cercano y agarró con los dientes los huevos rojos y tibios; pero el huevo fue hecho para ser tratado con delicadeza. Y los dientes de rufo eran duros.

Cuando Celia llegó, encontró a Rufo decepcionado frente a un nido lleno de huevos quebrados.

¡Fuera de aquí perro vagabundo! Siempre digo que el gallinero no es lugar para los perros. Rufo quiso protestar pero se acordó de la mancha amarilla en que se había transformado el nido. De verdad Celia tenía razón.

Cabizbajo e infeliz, se fue a acostar al fondo de la huerta.

Ejercicio Detenido

Tiempo:

Sección 6R: Lectura y comprensión de un párrafo

[Quite el texto al estudiante después de que lo haya leído y diga] [Las preguntas en *itálicas* indican que las respuestas pueden ser posibles o imposibles, en lugar de correctas o incorrectas]. [No hay tiempo pero se registra cuánto tiempo le tomó]

Ahora te voy a hacer algunas preguntas sobre el cuento.

1. ¿Quién es el personaje de este cuento? (Respuesta esperada: Rufo, un perro, un perrito)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

2. ¿Qué quería hacer el perro en la granja? (Respuesta esperada: Ayudar en los trabajos, recoger huevos)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

3. ¿Qué pasó con los huevos cuando Rufo los agarró? (Respuesta esperada: Se quebraron, se rompieron)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

4. ¿Qué era la mancha amarilla que Rufo dejó en el nido? (Respuesta esperada: Las yemas de los huevos, la parte adentro de los huevos)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

5. ¿Por qué Celia dijo que un gallinero no es lugar para perros? (Respuesta esperada: Porque no saben trabajar, porque Rufo quebró todos los huevos, porque asustó a las gallinas)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

6. ¿Cómo se sintió Rufo cuando se fue a dormir? (Respuesta esperada: Triste, con pena, con vergüenza)

Correcto Incorrecto No respondió Respuesta (en caso de duda)

[Una vez concluido todo el ejercicio, agradecer al niño su participación excepto en los casos que se siga con otro ejercicio.] ¡Muchas gracias por ayudarme!

Annex B: Reading Habits and Attitudes Survey

Vamos a pasar a la última actividad.

Sección A

Te voy hacer algunas preguntas para conocerte mejor. *[Marcar con X la casilla que corresponda a su respuesta]*

	Si	No	NS/NC	
1. ¿Te gusta leer?				
2. ¿Sabes usar una computadora o tableta?				
3. ¿Tus padres leen libros contigo?				
4. ¿Recogiste el pasaporte de Mundo de Libros para poder sacar libros de la biblioteca?				¿por qué no?
<i>[Contestó SI en pregunta 4]</i>				
5. ¿Has ido la biblioteca por libros?				¿por qué no?

*[Si en la pregunta 4 contestó: NO → Saltar la sección B, pasar directo a Sección C omitiendo preguntas con * ,
SI → Pasar a Sección B, además contestar preguntas con * en Sección C]*

Sección B

1. ¿Qué es lo que más te gusta de Mundo de Libros?

[Dejar abierta pero marcar las opciones mencionadas. Si dicen más de 1, indicar orden]

- Libros (cantidad) → muchos libros, diversidad de títulos
- Libros (calidad) → bonitos, interesantes, tienen dibujos, tamaño
- Mobiliario → Libreros, caja, muebles, decoración, foamy
- Página de Internet
- Actividades/premios → Retos
- Facilidades (préstamos, no cuesta, etc.)
- Otro: _____

2. ¿Dónde has visto estas imágenes?

[Enseñar estímulo con imagen de la página].

- Página
- No las ha visto
- Otra respuesta

3. ¿Para qué has utilizado a la página?

[Dejar abierta pero marcar las opciones mencionadas. Si dicen más de 1, indicar orden]

- Recomendaciones
- Ver escudo (lectómetro)
- Avatar
- Recibir estrellas
- Búsquedas (título, autor, ubicación)
- Calificar libros
- Otro: _____

[Continuar en la sección C]

4a. Ah, viene de la página de Mundo de Libros. ¿y por qué no la usas?

[en tono amigable]

4b. ¿Sabes para qué sirve?

- No
- Si, para: _____

[Continuar en la sección C]

Sección C

Ahora vamos a usar estos cuadrados para contestar unas preguntas [enseñar hoja con cuadrados]

[Señalando cada cuadrado que se menciona]

- El pequeño significa: muy poco o nunca
- El siguiente significa: poco
- Este mediano significa: algo
- Y el más grande: mucho o siempre

[Si dice la respuesta oral, por ejemplo: “mucho”, márcala como 3] [Respuestas como “no tengo libros” tomarla como 1]

¿Entendiste lo que vamos a hacer?

Por ejemplo, enséñame:

					No contestó
	1	2	3	4	
1. ¿Qué tan seguido ves la televisión?					
2. ¿Qué tan seguido ves a tus papás leyendo en casa?					
3. ¿Qué tanto te gusta leer?					
4. ¿Qué tantos libros para niños hay en tu casa?					
5. ¿Qué tan seguido tus padres te escuchan leer en voz alta?					
6. ¿Qué tan seguido platicas con alguien de tu familia sobre lo que lees?					
7. ¿Qué tan seguido tus padres te leen un libro?					
8. ¿Qué tan divertido es lo que lees en la escuela?					
9.* ¿Qué tan divertidos son los libros de Mundo de Libros?					
10. ¿Qué tan difícil es lo que lees en la escuela?					
11.* ¿Qué tan difíciles son los libros de Mundo de Libros?					
12. ¿Qué tan seguido le preguntas a tus padres el significado de las palabras que no entiendes?					
13. ¿Qué tan seguido te ayudan a leer tus papás?					
14. ¿Qué tan seguido alguien de tu familia te da premios o dice cosas agradables cuando haces bien tu tarea?					
15.* ¿Qué tanto crees que Mundo de Libros te ha ayudado a mejorar tu lectura?					
16.* ¿Qué tanto te gusta Mundo de Libros?					

Sección D

Ya para acabar, te voy a preguntar unas cosas sobre de la casa donde vives.

[La mayor parte del tiempo en caso de que viva en diferentes casas]

1. Contándote a ti, ¿cuántas personas viven en tu casa? [Contando niños y adultos]

personas

2. En tu casa, ¿tienen _____ ?

	No	Si	NC/NS	Otro
Horno microondas				
Lavadora de ropa				
Computadora o Laptop				
Internet				
Auto o camioneta				

3. ¿Cuántos _____ hay en tu casa?

	0	1	2	3 o más	NC/NS	Otro
Baños (escusado, WC)						
Televisiones que sirvan						
Teléfono celulares						

4. La mayoría de las veces, ¿qué transporte usas para llegar a la escuela?

- | | | |
|------------------------------------|---|---|
| <input type="checkbox"/> Caminando | <input type="checkbox"/> Bicicleta | <input type="checkbox"/> Transporte escolar |
| <input type="checkbox"/> Metro | <input type="checkbox"/> Metrobus | <input type="checkbox"/> Autobús/Combi |
| <input type="checkbox"/> Taxi | <input type="checkbox"/> Auto/Camioneta | <input type="checkbox"/> Otro: _____ |

¡Muchas gracias tu tiempo y ayuda! Hemos acabado.

¿Hay algo más que me quieras decir o preguntar?

Annex C: Reading Habits and Attitudes Survey Results and Composites

Table C.1: Disposition to Reading Composite

Questions and Response Options		Frequency	Percentage (%)
Do you like reading?	No	33	7.2%
	Yes	422	92.3%
	No answer	2	0.4%
How much do you like reading?	Very little / not at all	27	5.9%
	A little	39	8.5%
	Some	100	21.9%
	A lot	291	63.7%
How fun are the things you read at school?	Not fun	27	5.9%
	A little fun	31	6.8%
	Somewhat fun	96	21.0%
	Very fun	301	65.9%
	No answer	2	0.4%
How hard are the things you read at school?	Not at all hard	101	22.1%
	A little hard	97	21.2%
	Somewhat hard	166	36.3%
	Very hard	92	20.1%
	No answer	1	0.2%

Table C.2: Engagement in Program Composite

Questions and Response Options		Frequency	Percentage (%)
Did you pick up your passport to let you check out books from Mundo de Libros?	No	97	21.2%
	Yes	358	78.3%
	No answer	2	0.4%
Have you visited the library to check out a book?	No	75	16.4%
	Yes	283	61.9%
	Other	0	0.0%
	No answer	0	0.0%
	Not applicable	99	21.7%
How fun are Mundo de Libros books?	Not at all fun	6	1.3%
	A little fun	3	0.7%
	Somewhat fun	33	7.2%
	Very fun	308	67.4%
	No answer	107	23.4%
How hard are Mundo de Libros books?	Not at all hard	119	26.0%
	A little hard	54	11.8%
	Somewhat hard	76	16.6%
	Very hard	102	22.3%
	No answer	106	23.2%
How much do you think Mundo de Libros has helped to improve your reading skills?	Not at all	8	1.8%
	A little	11	2.4%
	Somewhat	45	9.8%
	A lot	288	63.0%
	No answer	105	23.0%
How much do you like Mundo de Libros?	Not at all	8	1.8%
	A little	6	1.3%
	Somewhat	33	7.2%
	A lot	306	67.0%
	No answer	104	22.8%

Table C.3: Family Reading Support Composite

Questions and Response Options		Frequency	Percentage (%)
How often do you see your parents reading?	Never	166	36.3%
	Once in awhile	85	18.6%
	Sometimes	91	19.9%
	A lot / always	115	25.2%
Do your parents read books with you?	No	175	38.3%
	Yes	282	61.7%
	No answer	0	0.0%
How often do your parents listen to you read out loud?	Never	74	16.2%
	Once in awhile	76	16.6%
	Sometimes	129	28.2%
	A lot / always	178	38.9%
How often do you talk with a family member about what you read?	Never	83	18.2%
	Once in awhile	79	17.3%
	Sometimes	134	29.3%
	A lot / always	161	35.2%
How often do your parents read books to you?	Never	164	35.9%
	Once in awhile	86	18.8%
	Sometimes	115	25.2%
	A lot / always	92	20.1%
How often do you ask your parents about words you did not understand?	Never	85	18.6%
	Once in awhile	105	23.0%
	Sometimes	105	23.0%
	A lot / always	162	35.4%
How often do your parents help you to read?	Never	104	22.8%
	Once in awhile	76	16.6%
	Sometimes	126	27.6%
	A lot / always	151	33.0%
How often does a family member reward you or say a nice thing to you when you read?	Never	59	12.9%
	Once in awhile	52	11.4%
	Sometimes	90	19.7%
	A lot / always	256	56.0%

Table C.4: Socioeconomic Status Composite

Questions and Response Options		Frequency	Percentage (%)
Does your house have a microwave?	No	151	33.0%
	Yes	302	66.1%
	Other	2	0.4%
	Doesn't know / No answer	1	0.2%
	Missing	1	0.2%
Does your house have a washing machine?	No	57	12.5%
	Yes	396	86.7%
	Other	2	0.4%
	Doesn't know / No answer	1	0.2%
	Missing	1	0.2%
Does your house have a computer or laptop?	No	170	37.2%
	Yes	278	60.8%
	Other	7	1.5%
	Doesn't know / No answer	1	0.2%
	Missing	1	0.2%
Does your house have internet access?	No	120	26.3%
	Yes	333	72.9%
	Other	1	0.2%
	Doesn't know / No answer	2	0.4%
	Missing	1	0.2%
Does your house have a car or truck?	No	179	39.2%
	Yes	272	59.5%
	Other	5	1.1%
	Doesn't know / No answer	0	0.0%
	Missing	1	0.2%
How many bathrooms are in your house?	None	1	0.2%
	One	232	50.8%
	Two	154	33.7%
	Three	68	14.9%
	Doesn't know / No answer	1	0.2%
	Missing	1	0.2%

Table C.4: **Socioeconomic Status Composite** (continued)

Questions and Response Options		Frequency	Percentage (%)
How many TVs are in your house?	None	5	1.1%
	One	120	26.3%
	Two	143	31.3%
	Three	186	40.7%
	Four	1	0.2%
	Doesn't know / No answer	1	0.2%
	Missing	1	0.2%
How many cellphones are in your house?	None	2	0.4%
	One	37	8.1%
	Two	85	18.6%
	Three	321	70.2%
	Four	2	0.4%
	Doesn't know / No answer	9	2.0%
	Missing	1	0.2%

Annex D: EGRA Descriptive Statistics and Additional Tables

Table D.1: Letter Sound Identification Rate (CLSPM)

Grade at Baseline	N/n	Baseline			Endline			Mean Gain
		Mean	SD	Zero Score (%)	Mean	SD	Zero Score (%)	
Grade 1	145	20.6	15.7	6.9%	30.1	14.8	2.1%	9.5
Grade 2	174	25.8	15.4	3.4%	30.7	14.4	2.9%	5.0
Grade 3	138	27.0	13.4	0.7%	31.5	14.7	1.4%	4.5
Total: All Students	457	24.5	15.1	3.7%	30.7	14.6	2.2%	6.3

Table D.2: Initial Sound Identification Score (Correct out of Ten)

Grade at Baseline	N/n	Baseline			Endline			Mean Gain
		Mean	SD	Zero Score (%)	Mean	SD	Zero Score (%)	
Grade 1	145	5.2	3.7	15.2%	6.7	3.1	6.9%	1.5
Grade 2	174	6.2	3.6	14.4%	7.0	2.9	5.7%	0.8
Grade 3	138	6.7	3.2	8.7%	7.2	2.6	1.4%	0.4
Total: All Students	457	6.1	3.5	12.9%	7.0	2.9	4.8%	0.9

Table D.3: Familiar Word Reading Rate (CFWPM)

Grade at Baseline	N/n	Baseline			Endline			Mean Gain
		Mean	SD	Zero Score (%)	Mean	SD	Zero Score (%)	
Grade 1	145	15.7	16.3	24.8%	45.5	21.1	0.7%	29.8
Grade 2	174	48.6	21.5	1.1%	64.8	24.6	1.7%	16.3
Grade 3	138	60.8	22.5	1.4%	74.0	24.6	0.0%	13.2
Total: All Students	457	41.8	27.5	8.8%	61.5	26.2	0.9%	19.6

Table D.4: Nonword Reading Rate (CNWPM)

Grade at Baseline	N/n	Baseline			Endline			Mean Gain
		Mean	SD	Zero Score (%)	Mean	SD	Zero Score (%)	
Grade 1	145	11.9	12.2	22.1%	28.7	11.4	1.4%	16.8
Grade 2	174	32.1	12.3	1.1%	38.2	13.2	1.7%	6.0
Grade 3	138	39.4	11.0	0.7%	44.4	13.0	0.0%	4.9
Total: All Students	457	27.9	16.4	7.7%	37.0	14.0	1.1%	9.1

Table D.5: Oral Reading Fluency Rate (CWPM)

Grade at Baseline	N/n	Baseline			Endline			Mean Gain
		Mean	SD	Zero Score (%)	Mean	SD	Zero Score (%)	
Grade 1	145	19.6	21.5	15.2%	60.5	27.5	1.4%	41.0
Grade 2	174	65.5	31.4	1.7%	95.9	40.6	1.1%	30.4
Grade 3	138	87.0	32.5	0.7%	118.1	41.2	0.7%	31.1
Total: All Students	457	57.4	39.8	5.7%	91.4	43.6	1.1%	34.0

Table D.6: Reading Comprehension Score (Correct out of Seven)

Grade at Baseline	N/n	Baseline			Endline			Mean Gain
		Mean	SD	Zero Score (%)	Mean	SD	Zero Score (%)	
Grade 1	145	1.5	1.9	49.0%	4.2	2.0	5.5%	2.7
Grade 2	174	4.6	2.0	6.3%	5.3	1.5	1.7%	0.6
Grade 3	138	5.3	1.6	1.4%	5.9	1.1	0.7%	0.6
Total: All Students	457	3.8	2.5	18.4%	5.1	1.7	2.6%	1.3

Table D.7: Adaptive Oral Reading Option B “Toto” Score

Grade at Baseline	Adjusted Value N	Baseline			Endline			Gain from Baseline to Endline
		Mean	SD	Zero Score (%)	Mean	SD	Zero Score (%)	
Grade 1	41	74.5	26.6	0.0%	92.7	4.0	0.0%	18.2
Grade 2	35	91.2	10.5	0.0%	91.5	4.1	0.0%	0.3
Grade 3	25	90.3	13.9	0.0%	91.5	12.2	0.0%	1.2
Total: All Students	101	84.2	20.8	0.0%	92.0	6.9	0.0%	7.8

Table D.8: Adaptive Oral Reading Option C “Rufo” Score

Grade at Baseline	Adjusted Value N	Baseline			Endline			Gain from Baseline to Endline
		Mean	SD	Zero Score (%)	Mean	SD	Zero Score (%)	
Grade 1	22	155.7	16.6	0.0%	158.0	6.7	0.0%	2.4
Grade 2	120	157.5	7.7	0.0%	157.4	15.4	0.8%	-0.1
Grade 3	110	159.3	6.5	0.0%	158.3	6.1	0.0%	-1.0
Total: All Students	252	158.1	8.4	0.0%	157.9	11.5	0.4%	-0.3

Table D.9: Adaptive Reading Comprehension “Toto” Score (Correct out of Six)

Grade at Baseline	Adjusted Value N	Baseline			Endline			Gain from Baseline to Endline
		Mean	SD	Zero Score (%)	Mean	SD	Zero Score (%)	
Grade 1	41	2.7	1.4	9.8%	4.7	0.9	0.0%	2.0
Grade 2	35	3.4	1.3	2.9%	4.5	1.2	0.0%	1.1
Grade 3	25	4.0	1.8	4.0%	4.8	1.2	0.0%	0.8
Total: All Students	101	3.2	1.6	5.9%	4.6	1.1	0.0%	1.4

Table D.10: Adaptive Reading Comprehension “Rufo” Score (Correct out of Six)

Grade at Baseline	Adjusted Value N	Baseline			Endline			Gain from Baseline to Endline
		Mean	SD	Zero Score (%)	Mean	SD	Zero Score (%)	
Grade 1	22	3.7	1.4	0.0%	4.5	1.2	0.0%	0.9
Grade 2	120	4.4	1.4	0.8%	4.7	1.2	0.8%	0.2
Grade 3	110	4.6	1.3	0.0%	5.0	1.1	0.0%	0.5
Total: All Students	252	4.4	1.4	0.4%	4.8	1.2	0.4%	0.4

Table D.11: Average Gain Scores by Gender

Subtask	Gender			
	Girls		Boys	
	<i>n</i>	Mean	<i>n</i>	Mean
Letter sound identification (CLSPM)	242	5.8	215	6.8
Initial sound identification (correct out of ten)	242	0.8	215	1.0
Familiar word reading (CFWPM)	242	19.8	215	19.5
Nonword reading (CNWPM)	242	8.3	215	10.0
Oral reading fluency (CWPM)	242	34.5	215	33.4
Reading comprehension (correct out of seven)	242	1.2	215	1.4

Table D.12: Average Gain Scores by Disposition to Reading Composite

Subtask	Low Disposition		High Disposition	
	<i>n</i>	Mean	<i>n</i>	Mean
Letter sound identification (CLSPM)	238	6.0	215	6.7
Initial sound identification (correct out of ten)	238	0.8	215	1.0
Familiar word reading (CFWPM)	238	18.9	215	20.8
Nonword reading (CNWPM)	238	9.0	215	9.3
Oral reading fluency (CWPM)	238	33.3	215	35.0
Reading comprehension (correct out of seven)	238	1.2	215	1.3

Table D.13: Average Gain Scores by Engagement in Program Composite

Subtask	Low Engagement		High Engagement	
	<i>n</i>	Mean	<i>n</i>	Mean
Letter sound identification (CLSPM)	195	6.6	153	5.3
Initial sound identification (correct out of ten)	195	0.9	153	0.9
Familiar word reading (CFWPM)*	195	21.2	153	16.3
Nonword reading (CNWPM)*	195	10.5	153	6.5
Oral reading fluency (CWPM)*	195	35.7	153	30.3
Reading comprehension (correct out of seven)	195	1.4	153	1.1

An asterisk (*) indicates the average gain score on the subtask were significantly different between students' groups at $p < 0.05$.

Table D.14: Average Gain Scores by Family Reading Support Composite

Subtask	Low Family Support		High Family Support	
	<i>n</i>	Mean	<i>n</i>	Mean
Letter sound identification (CLSPM)	209	5.5	248	6.9
Initial sound identification (correct out of ten)	209	0.8	248	1.0
Familiar word reading (CFWPM)*	209	21.9	248	17.8
Nonword reading (CNWPM)*	209	10.2	248	8.2
Oral reading fluency (CWPM)	209	35.4	248	32.7
Reading comprehension (correct out of seven)	209	1.4	248	1.2

An asterisk (*) indicates the average gain score on the subtask were significantly different between students' groups at $p < 0.05$.

Table D.15: Average Gain Scores by Socioeconomic Status Composite

Subtask	Low SES		High SES	
	<i>n</i>	Mean	<i>n</i>	Mean
Letter sound identification (CLSPM)	208	6.3	231	6.6
Initial sound identification (correct out of ten)	208	1.0	231	0.8
Familiar word reading (CFWPM)*	208	21.3	231	18.6
Nonword reading (CNWPM)*	208	10.3	231	8.1
Oral reading fluency (CWPM)*	208	34.7	231	34.2
Reading comprehension (correct out of seven)	208	1.7	231	1.0

An asterisk (*) indicates the average gain score on the subtask were significantly different between students' groups at $p < 0.05$.

Table D.16: Correlational Analysis of Subtask Gains and Book Checkouts or Platform Logins

Subtask	Mean	1	2	3	4	5	6	7	8
Letter sound identification (CLSPM) gain	6.3	-							
Initial sound identification (correct out of ten) gain	0.9	-	-						
Familiar word reading (CFWPM) gain	19.6	-	-	-					
Nonword reading (CNWPM) gain	9.1	-	-	-	-				
Oral reading fluency (CWPM) gain	34.0	-	-	-	-	-			
Reading comprehension (correct out of seven) gain	1.3	-	-	-	-	-	-		
Book loans (number of books loaned)	20.9	-0.018	-0.031	0.031	0.015	.140**	0.050	-	-
Platform logins (number of platform logins)	3.5	-0.037	-0.049	-0.022	-0.060	0.058	0.024	-	-

N=457

** sig. at $p < .001$

Annex E: EGRA Reliability Results

Table E.1: Reliability Results for Endline EGRA

Subtask	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Letter sound identification (percentage correct)	0.535	0.874
Initial sound identification (percentage correct)	0.496	0.896
Familiar word reading (percentage correct)	0.851	0.821
Nonword reading (percentage correct)	0.774	0.833
Oral reading fluency (percentage correct)	0.814	0.829
Reading comprehension (percentage correct)	0.710	0.844
	EGRA Coefficient Alpha	0.872

Table E.2: Reliability Results for Baseline EGRA

Subtask	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Letter sound identification (percentage correct)	0.517	0.920
Initial sound identification (percentage correct)	0.460	0.930
Familiar word reading (percentage correct)	0.914	0.858
Nonword reading (percentage correct)	0.886	0.866
Oral reading fluency (percentage correct)	0.905	0.859
Reading comprehension (percentage correct)	0.851	0.869
	EGRA Coefficient Alpha	0.904

