“Improving Reading at Scale: Evidence for Structured Literacy Instruction in Argentina’s Primary Schools”

NCT Number: NOT YET ASSIGNED

Date: 01/28/2026

Improving Reading at Scale: Evidence for Structured Literacy Instruction in Argentina's Primary Schools

Florencia Salvarezza. Universidad de la Ciudad de Buenos Aires and the UConn/Haskins Global Literacy Hub. Florencia.salvarezza@udelaciudad.edu.ar

Dan Steinhof, Department of Psychology and Department of Cognitive & Brain Sciences, The Hebrew University of Jerusalem, dan.steinhof@mail.huji.ac.il

Elizabeth Zagata, Center for Education Systems, WestEd, ezagata@wested.org ORCiD: 0000-0003-1208-2691

Luiza de Melo Carvalho, University of Connecticut, Department of Psychological Sciences, Storrs; UConn/Haskins Global Literacy Hub. luiza.carvalho@uconn.edu ORCiD: 0000-0002-2571-9762

Augusto Buchweitz, University of Connecticut, Department of Psychological Sciences, Storrs; UConn/Haskins Global Literacy Hub. augusto.buchweitz@uconn.edu 0000-0003-3791-7472

Noam Siegelman, Department of Psychology and Department of Cognitive & Brain Sciences, The Hebrew University of Jerusalem, noam.siegelman@mail.huji.ac.il, 0000-0002-5005-4838

Kenneth R. Pugh, University of Connecticut, Yale University School of Medicine, Director of Research, Haskins Laboratories, kenneth.pugh@yale.edu

Maureen W. Lovett, The Hospital for Sick Children, the University of Toronto, the Yale Child Study Center and the UConn/Haskins Global Literacy Hub. mwl@sickkids.ca.

ORCiD: 0000-0002-7556-8309

Abstract

Purpose:

Throughout Argentina, schools have struggled with the teaching and evaluation of foundational literacy skills. We report results from a longitudinal controlled evaluation of a teacher training and structured literacy program on foundational literacy outcomes in multiple Buenos Aires schools.

Method:

From 8 intervention and 7 control schools, 537 children participated: 258 intervention and 279 control participants. All were native Spanish speakers entering Grade 1 and tracked for two years. Teachers from intervention schools were provided over 20 hours of professional development yearly in *Aprendo Leyendo* (AL), a Spanish-language structured reading program. AL was taught in Grades 1 and 2 intervention classes, and control schools used their usual reading instruction. Performance data were collected four times: the beginning (T1) and end of Grade 1 (T2), and the beginning (T3) and end of Grade 2 (T4).

Results:

Intervention students demonstrated significant posttest advantages on several reading and reading-related skills, with better T4 performance than controls on passage comprehension and letter-sound naming. Differential growth on two word reading fluency and one letter naming fluency measure were also revealed, and on two component skills, phoneme segmentation and set for variability. Individual differences analyses showed that, generally, children with better component reading skills at T1 demonstrated larger intervention gains.

Conclusion:

Results demonstrate the positive impact of structured literacy Professional Development, PD, and instruction in the early grades and demonstrate that better literacy outcomes may be achieved. Findings support continued use and scaling of structured literacy initiatives to improve reading outcomes for children across Argentina.

Word count = 250

Introduction

Evidence-based foundational literacy practices in Latin America and the Caribbean

In recent years, Latin American and Caribbean countries (LAC) have demonstrated significant improvements in early education access and school enrollment. Out-of-school rates (3%) and school completion rates (94%) for primary education in the region have approached near-universal levels (UNESCO, 2024). However, most LAC countries do not assess early literacy outcomes; in primary education, for example, about a quarter of the region has no reliable national assessment (UNESCO, 2024). Even with use of these assessments, there remain many obstacles to effectively teaching early literacy skills. In Argentina, only 33% of sixth graders have achieved the minimum proficiency level in reading (APRENDER, 2023).

Part of the instructional challenge in LAC countries is related to a historical failure to adopt systematic structured literacy programs and phonics-based early literacy approaches (Stone et al., 2020). There has been longstanding resistance to the Science of Reading, from within the Schools of Education, and many practical challenges to implementing and sustaining evidence-based policies.

There is evidence, however, of some potential solutions through an investment in teacher training in evidence-based practices. Teacher training is a cost-effective strategy for promoting evidence-based foundational literacy practices. In Honduras, one of the poorest countries in LAC, a USAID report revealed a cost of only US $1.00 (one US dollar per year per student) to achieve the same improvement in student performance that other strategies, such as class size reduction (US$23.00 per student) and implementation of end-of-grade assessments (US$52.00 per student) accomplished at far greater cost (Rhodes, 2021).

In the present study, we investigated the effects of teacher training and of the implementation of systematic, phonics-based instruction in the province of Buenos Aires, Argentina. We compared the early literacy performance of children in schools that adopted the program with that of children in control schools that used business as usual, non-phonics based instruction. The primary goals were, first, to carry out a longitudinal evaluation of an evidence-based teacher training program in multiple schools in Buenos Aires; second, to assess the feasibility of the program and its acceptance by teachers; and, third, to examine potential effects of the program on foundational literacy outcomes among early primary education students. The longitudinal nature of the study design also permitted the investigation of additional research questions regarding individual differences in children's response to instruction and an examination of early predictors of reading outcomes.

What works: Cumulative, systematic, and explicit structured literacy instruction

Systematic and explicit structured literacy instruction is the most effective way for teaching children to read (see, e.g., Castles et al., 2018; Ehri et al, 2001). This begins with a phonics-based curriculum that teaches mappings between letters and sounds, letter clusters and sounds, variant letter-sound symbol mappings, and decoding strategies.

Ideally, this approach treats decoding and spelling as reciprocal processes to be learned (Kim & Zagata, 2024). Such instruction serves as a foundation upon which a multidimensional reading system is built. It should be noted that phonics-based instruction is a necessary, but not sufficient, approach to literacy instruction. It is essential to establish the foundational layers of a reading system but other aspects of written language structure also need explicit attention.e,g., orthography, morphology, semantics, and syntax.

Phonics-based structured literacy instruction is the most effective choice for beginning reading instruction independent of the regularity of the mapping of speech sounds to print. Among world languages and their writing systems, there are more "opaque" orthographies, like English and French, with more variable mappings between speech sounds and letter combinations, and more transparent ones, like Spanish and Finnish, which have basically univocal letter-sound mapping (Seymour, 2003). Regardless of these differences, the path to effective early literacy instruction is the same (Caravolas et al., 2012, 2013), and the predictors of achievement are remarkably similar.

Brain imaging research has also provided evidence that the integration of effective mapping of oral and written language is the basis of fluent reading. It has been shown that the brain's network of areas that support reading is constrained by previously-developed speech-processing circuits common to all languages. The more the brain circuitry for speech and for reading converge, the better the actual reading fluency and comprehension of an individual, across multiple languages Rueckl et al., 2015). These findings are compatible with the findings that systematic and explicit teaching of phoneme-grapheme correspondences, syllabic structure, and morphological patterns effectively support the learning of foundational literacy in Spanish.

The struggle with teaching children to read in Latin America

Schools in LAC countries have struggled with the teaching and evaluation of foundational literacy skills. There is a scarcity of both successful evidence-based efforts in the region and rigorous scientific evaluation of early literacy programs (evidence-based, or not; see Stone et al., 2020 for a review). Randomized controlled trials and large cohort observational studies in early literacy and in education are a rarity in Latin America (Stone et al., 2020). The challenge for improving foundational literacy in Latin America includes the need both to provide teachers with effective, evidence-based professional development (PD) and to establish rigorous scientific scrutiny of reading instruction and its efficacy.

One outlier in the region has been the program (and associated study) *Aprendamos todos a leer* (Márquez de Arboleda et al., 2020). This project provided evidence-based teacher training and instructional materials to schools in Colombia, and its efficacy was evaluated rigorously, in 70 schools and with 2,100 students. The study of *Aprendamos todos a leer* showed that children whose teachers were trained to use the program achieved significantly greater gains on early literacy outcomes than their peers, and these differential gains persisted over the years. In other words, students of teachers from the program outperformed their peers in first grade, and continued to outperform these peers through second and third grades (Alvarez Marinelli et al., 2023). These findings are encouraging

and raise the question of whether such innovations can be adapted and transported to other countries and cultures.

The need for evidence-based early literacy instruction in Buenos Aires and Argentina

The Argentine educational system, of which *Ciudad Autonoma de Buenos Aires* (CABA) is part, faces numerous problems and produces generally poor results, both on international tests like *Programme for International Student Assessment* (PISA) and on regional assessments like *Estudio Regional Comparativo y Explicativo* (ERCE). On PISA 2022, the country scored 400.7 points and ranked below Chile (448), Uruguay (430), Costa Rica (415), Mexico (415), and Brazil (410). PISA data shows that 54.4% of students in Argentina do not reach the minimum performance level established by the Organization for Economic Co-operation and Development (OECD). Rather, only 5 out of 10 students achieve the most basic reading level; and only 1% of students in Argentina achieve the highest levels of reading performance for their age (5 and 6). The regional ERCE 2019 test, which classifies students into four levels, showed that 46% of third-grade students in Argentina are at the lowest level (Level I). This means they are unable to locate information in text, or make simple inferences from age-appropriate texts. Only 14% of Buenos Aires students reached Level IV, which indicates the highest performance level in reading.

Literacy results are poor throughout the system, whether in public or private schools. Many families tend to move children into the private system in an attempt to ensure more class days. In private schools, there is more predictability, to some extent, as private schools do not grapple with closures due to labor strikes or infrastructural issues. In the public system, the level of absenteeism for both students and teachers is very high, adding additional strain to a struggling system and further decreasing opportunities for literacy learning in the early primary grades. The public education system at national and municipal levels is faced with structural and contextual challenges that limit the effectiveness and sustainability of interventions. Added to these challenges is a certain ‘normalization’ of absenteeism; this attitude shapes a school culture where attendance is not always considered a priority value.

Until 2025, the early literacy instruction method adopted by CABA was a whole language, balanced approach to early literacy, with no letter-sound instruction nor reading aloud in class. A key element of this approach is the belief that literacy development occurs fostered by exposure to reading and print, by participating and being immersed in environments with books and written words, and by teachers reading to their students. In such an environment, children are supposed to develop the ability to read without direct instruction, as they "construct" this knowledge on their own.

Although Spanish has a transparent orthography, structured literacy instruction remains essential, as transparency does not eliminate all reading acquisition challenges. The development of select reading skills (e.g., fluency measures) in opaque orthographies, like English and French, have been shown to take somewhat longer when compared with more transparent orthographies, like Spanish (Caravolas et al., 2012, 2013). Similar skills and knowledge, however, underpin reading development in opaque and transparent writing systems, as reviewed above.

Pilot work preceding the current research

Before the present investigation, a pilot project was conducted in Buenos Aires schools (Braze, Salvarezza, & Pugh, 2020). Six first grade classrooms and six control classrooms participated, enlisting a total of 328 children. Teachers received PD training and taught the structured literacy program, *Aprendo Leyendo* (AL)*,* in their classrooms. AL is designed for early Spanish-speaking readers and provides systematic phonics-based instruction in decoding and spelling, as well as instruction in reading comprehension strategies, syntax, morphology, vocabulary, and handwriting (see details in Methods). Although intended for full year instruction, the amount of instruction received by participants fell far short of expectations due to high rates of student absenteeism, frequent school closings, and lack of uptake from classroom teachers. As a result, while children in the structured literacy classrooms demonstrated a significant advantage in sentence reading comprehension at the end of the year, other outcomes were equivalent for the two conditions. This work also revealed that almost all children entered first grade insufficiently prepared for early reading instruction. Findings also indicated the need for more comprehensive PD and close attention to teacher PD response and the fidelity with which teachers implemented the *AL* program. Finally, the researchers noted the systemic barriers that resulted in the high rates of absenteeism across all 12 classrooms.

What predicts individual differences among children in reading development?

Because there are large differences among children in any classroom, there have been many attempts to identify predictors of their varying trajectories and of their different responses to instruction and intervention. Different sets of cognitive, behavioral, and reading-related predictors have been examined for their ability to predict individual differences in reading outcomes. Phoneme awareness, letter–sound knowledge, and rapid automatized naming have been shown to predict reading development across languages. Recently, Siegelman and colleagues (2021) characterizes additional predictors of individual reading gains from children's differential reliance on psycholinguistic properties, reflecting the functional organization of their reading system (i.e., integrity of print-speech and print-meaning pathways). .

Another promising predictor of individual differences has been revealed with use of an experimental task rather than a standardized test. *Set for Variability* (SfV) (Gibson & Levin, 1975; Venezky, 1999) is typically defined as a learner's ability to correct a mispronunciation of an unfamiliar word to its actual spoken word form. In English and other "opaque" orthographies, SfV is typically tested in a spoken language task where participants are presented in each trial with a decoded phonological form of a word (e.g., /brɪkfast/) and are asked to guess the correct intended word (e.g., /brɛkfast/). Recently, studies have shown that SfV performance correlates highly with reading proficiency among early readers of English (e.g., Steacy et al., 2019) even dominating other more canonical predictors such as phonological awareness (e.g., Steacy et al., 2023). These high correlations have been interpreted as suggesting that the ability to arrive at a correct (i.e., lexical) form of a word from a mispronounced decoded spoken form taps into multiple important skills, including

the integrity of the phonological system, vocabulary knowledge, and print-speech and speech-print relations (see Edwards et al., 2022).

Spanish has a transparent orthography and thus does not allow the creation of decoded spoken word forms as stimuli. In the current study, we therefore designed a new SfV Spanish task, which is meant to tap into the integrity of the phonological system without needing decoded stimuli. In particular, we used as stimuli spoken pseudowords that are close in phonological space to actual words (e.g., /amimal/). Then, similarly to the English SfV task, participants were instructed to guess the intended correct spoken form of the word (e.g., /animal/). With this task at hand, we could then examine whether SfV predicts reading proficiency outcomes even in a transparent orthography. If so, this would suggest that one's ability to correct mispronounced spoken words to arrive at lexical forms is important regardless of the transparency of their writing system (i.e., even when a writing system does not evoke decoded forms).

Research Questions

The overarching goal of the present study was to carry out a longitudinal observation of the effects of an evidence-based teacher training program (see below for details) and corresponding teacher and student instructional materials on foundational literacy outcomes. We conducted four assessments of literacy and literacy-related skills over the span of two years in schools in the province of Buenos Aires, Argentina. We evaluated the early reading skills of children whose teachers were trained in the program and had *Aprendo Leyendo*'s instructional materials, and compared their performance and evolution with the skills of children whose teachers were not trained in the program and did not have the instructional materials. In addition, individual differences in the children's reading growth were assessed with a limited set of predictors.

There are three questions on which the current research is based.

1. Is a two-year evidence-based teacher training program in Spanish-language structured literacy instruction feasible to implement in Buenos Aires schools and acceptable to participating teachers? This question was assessed using descriptive/qualitative survey data collected from participating teachers.
2. Does a two-year evidence-based teacher training program in Spanish-language structured literacy instruction produce gains in reading development for Grade 1 and 2 children relative to their peers in control schools? This issue represents the primary question motivating the research, and it was evaluated using linear mixed-effect models examining performance differences in the teacher training vs. control groups on multiple reading and reading-related outcomes over time.
3. Do individual differences in reading and reading-related skills at entry predict differential reading gains among the children receiving Spanish-language structured literacy instruction? Separate linear mixed-effects models for each potential predictor variable and each outcome were developed to determine whether a variable was associated with reading growth and with better outcomes in the teacher training or the control groups.

Methods

Context for the research

To undertake the present research, an agreement was entered into with the CABA Ministry of Education and the University (Universidad de la Ciudad de Buenos Aires). The Ministry alone selected the participating schools and allocated them to intervention or control groups. Their school selection was guided by the following considerations:

1. That various districts were represented, which in turn represent different demographic groups.
2. That in each district, there was one intervention school and one control school.
3. That the district supervisors did not oppose the research.

The research team was not permitted to participate in school selection or allocation of schools to groups. This constraint prevented the investigators from undertaking randomized assignment of schools to intervention or control conditions. During the research project, high levels of student absenteeism influenced the continuity of classroom work and the evaluations conducted at four different times. As an external research group, direct access to school attendance records was not permitted by the Ministry, thus accurate data on absenteeism cannot be reported here.

At the beginning of the project, demographic data for both intervention and control schools was requested from the Ministry of Education's Statistics and Census unit. The data provided were limited and incomplete, with information only for some families and only at certain data points. The schools belonged to Districts 10, 16, 18, and 21 in the City of Buenos Aires. In each district, the intervention school had its control counterpart. The different districts of Buenos Aires are located in the north, center and south of Buenos Aires city, and the levels of household income vary between them. Families can send their children to any school regardless of where they live, so they tend to choose a school in better neighbourhoods on the belief that those are better.

In an effort to obtain more demographic data, a survey was distributed to the families of participating students to complement the demographic data provided by the Ministry. Argentina has a system of conditional cash transfers, Universal Child Allowance (AUH), aimed at low-income families. This benefit, which reaches approximately 20% of the country's minors, requires schooling and compliance with children's vaccination schedules.

There were 431 parents (in control and intervention group) who responded to the survey (out of 1050 people receiving the survey). Overall, 44% responded to the survey. Only 67 families stated they received the AUH, a state financial aid, and 16 of them also received some public transport benefit. Only 19 parents in total stated they had completed tertiary education, and 4 parents had incomplete primary education. The vast majority of parents stated they had Wi-Fi or broadband internet at home, as well as cell phones; furthermore, a high percentage responded that they had a computer and/or tablet. Although the country has implemented device distribution programs (such as the Conectar Igualdad plan), it has had a discontinuous coverage and an unequal impact in different regions of the country.

This project received an ethics review from Universidad de la Ciudad de Buenos Aires. The selected schools collected parental consent.

Participant sample

The Buenos Aires literacy intervention project began in the 2022 school year (school year starts in March and ends the first week in December). A total of 15 schools were identified to participate, 7 in the control group and 8 in the intervention group. The total sample of students was 537: 258 children participated in the intervention group (*Aprendo Leyendo*), and 279 in the control group. All children were native Spanish speakers and were entering Grades 1 in 2022 at the beginning of data collection (T1). Students were tracked for two years, through the end of their second grade. *Aprendo Leyendo* (AL) was taught in Grade 1 (2022) and Grade 2 (2023) in the intervention schools, and control schools used their traditional methods of reading instruction

In the two years of the project, 22 teachers (15 teachers in 2022 and 7 new teachers in 2023) from the intervention schools were each provided over 20 hours of PD in *Aprendo Leyendo*, a Spanish-language structured literacy program. All the teachers who participated in the training sessions, as they were teachers of the intervention group, completed tertiary level education.

The Spanish-language structured literacy program—*Aprendo Leyendo*

Aprendo Leyendo (AL) is a Spanish-language multisensory structured reading program, designed for children in the early stages of reading development. Using an explicit, sequential, and multisensory approach, AL is designed to help beginning readers achieve automatic accurate decoding and fluent reading of connected text. AL also includes systematic instruction in reading comprehension strategies, syntax, morphology, vocabulary, spelling and handwriting.

Throughout an instructional sequence of 96 Pasos (levels), AL teaches a range of comprehensive literacy content, including: phonemic awareness; letters/graphemes and their grapheme–phoneme correspondences; blending phonemes to read syllables, words, phrases, and the reading of connected text. Further, there are lessons in handwriting, spelling, grammar (including verb conjugations, pronouns, and noun/adjective correspondences, synonyms and antonyms), syntax, morphology, and instruction in reading comprehension strategies.

All these vital components of reading are taught simultaneously in cohesive structured lessons, each subject reinforcing and strengthening the others. An *Aprendo Leyendo* lesson should be delivered daily and last 45-60 minutes. Each lesson has 5 components:

1. **Review** (3 minutes) Review previously taught sounds using the Review Pack.
2. **Introduction of New Material** (15 minutes): New material could be a phonogram, morpheme, or a grammatical rule. While the alphabetic code is being taught, this part of the lesson includes phonemic awareness and handwriting instruction.
3. **Spelling Dictation** (10 minutes) Give a dictation with the new phonogram, using

words and a sentence provided in the Teacher Handbook.

4. **Reading** (20/25 minutes). Read a list of words, phrases, and sentences with children from the Skills Books to develop decoding skills, word recognition, and reading fluency. Reading a decodable and controlled chapter book with children to develop fluency and model comprehension strategies while reading connected text. This portion of the lesson is also an opportunity to develop oral language as the conversation should not be controlled.
5. **Reinforcement**. Provide reinforcement activities including reading comprehension, morphology, syntax, and grammar.

For more detail, please refer to the Curriculum Overview, Instructional Sequence, and digital samples of the materials provided in the supplementary materials.

Aprendo Leyendo's integrated PD for teachers is a crucial design feature of the program as it provides teachers with the tools and knowledge they need to help children develop into independent and competent readers. The PD curriculum included 20 hours of in-person training in year one, 20 hours of continued ongoing support throughout both school years, along with the provision of print and online teacher and student materials.

Teachers in the control schools received no special training, nor the *Aprendo Leyendo* instructional materials, and their students therefore received business-as-usual reading instruction. These classes served as a comparison group against which to evaluate the effectiveness of the AL program implemented in the intervention schools. A key element of the business-as-usual approach at the time is the belief that literacy development occurs fostered by exposure to reading and print, by participating and being immersed in environments with books and written words, and by teachers reading to their students. In such an environment, children are supposed to develop the ability to read without direct instruction, as they "construct" this knowledge on their own.

Teacher Professional Development

PD sessions for the teachers were held in 2022 (April, June, September, December in-person; May by Zoom) and in 2023 (in-person reinforcement in February, virtual in March). Due to a 50% teacher turnover in the second year of implementation, an additional in-person training was held in April 2023 for new teachers; these teachers therefore received half the training of those who participated for both years.

An open communication channel was established with teachers in the intervention group through a series of *WhatsApp* groups, maintained throughout the two years of the program. Additionally, several visits were made to model *Aprendo Leyendo* classes, with teachers observing a trainer and taking notes to apply it in their own classes and improve their work with the program. These visits were carried out at least three times during each year of the two-year project: in 2022 (June, August-September, late October approximately) and in 2023 (April, June-August, October).

In each visit to the intervention schools, a descriptive record was kept of how the students worked, how the teachers presented themselves in relation to the program, and what questions they had for the trainer. The instructional level the students were on was also recorded, and any other detail that caught attention of the trainer or needed to be highlighted.

'Business as Usual' Teaching in CABA

The general approach to reading instruction in CABA could be characterized as a "global/ whole language" or "balanced" approach. A key element of the global approach is that literacy development occurs fostered by exposure to reading and print, by participating and being immersed in environments with books and written words, and by teachers reading to their students. In such an environment, children are supposed to develop the ability to read without direct instruction, as they "construct" this knowledge on their own. This approach is based on the idea that learning to read and write is as natural as early language acquisition. Explicit teaching of decoding and reading is considered unnecessary as it is boring and interferes with a process the child must undertake. When the global approach introduces phonological awareness work, it is called a balanced or integral approach. It is important to note that the addition of phonological awareness does not imply that instruction becomes explicit and structured, as reading acquisition is still considered a process of construction by the student, rather than a skill taught by the teacher.

Control teachers and classroom observations

Teachers in the control schools did not receive AL training, and their students received the reading instruction typically conducted in CABA. To document these practices, the research team visited the control schools to observe a language class and to record the instructional practices using a class observation rubric. The observation rubric for control classes contained items related to teaching letters, reading and writing practices in class, vocabulary work, and practices related to phonological awareness. Less than 50% of Language Arts classes were found to be working on these skills.

The team observed that, in most of the classes, the teacher read aloud, and in very few cases, students read on their own. Only some teachers asked questions about the text that they read. 90% of control group classrooms had alphabet and word charts as part of the classroom decoration, and a large proportion of teachers worked with upper-case print on the board, in some cases alternating between upper-case and lower-case.

The observers noted that most teachers corrected writing assignments. An important aspect noted in control group schools was the lack of homogeneity in routines and work methods related to literacy. There appeared to be no uniform method or way of working among classrooms.

***Teaching Materials*:**

Aprendo Leyendo distributed 3260 individual materials to work with the AL class groups. These included 1620 student skillsbooks, 1620 decodable and controlled chapter books, and 20 teacher sets (teacher manuals, teacher editions to the chapter books, and alphabet and review card packs).

***Teacher Surveys*:**

At the end of each school year in the study, the research team conducted a survey with teachers about the program and their experiences with its implementation. Out of a total of 22 participating teachers during both years, 15 out of 15 completed the questionnaire in 2022 and 11 out of 14, in 2023. The survey consisted of 38 items.

The survey was administered at the end of the first year, in December 2022, and then again at the end of the second year, December 2023. Information was collected on teachers' professional trajectories, the use of reading science practices in teaching, opinions regarding the *Aprendo Leyendo* training program, the implementation of the AL program in their classrooms, and students' reading results and motivation.

Measures of Reading and Reading-related Skills

Student performance data were collected at four time points: T1 (pre-intervention), T2 (post-year-1 intervention), T3 (pre-year-2 intervention), and T4 (post-year-2 intervention). Due to the Argentinian school schedule, no school classes were held between T2 and T3 thus any change during that period reflected either continued growth or loss of skills over the holiday period.

To evaluate early literacy development in Spanish, two primary assessments were administered: *Indicadores Dinámicos del Éxito en la Lectura* (IDEL; Baker et al., 2007) and the *Tejas LEE* (Texas Education Agency, 2010). These assessments were administered across four time points (T1–T4), depending on the subtest. Student assessment was carried out by two teams of graduate student examiners from the Universidad de la Ciudad de Buenos Aires. Examiners were trained and supervised by university faculty under the direction of Professor F. Salvarezza. Assessments were conducted in the students' own schools.

IDEL Subtests. IDEL is a reliable and valid (Nelson, 2001; Watson; 2004) formative assessment designed to measure foundational literacy skills among Spanish-speaking emergent readers. The following IDEL subtests were used at multiple time points:

- **Letter Naming Fluency (idel_fnl):** Students named as many randomly ordered upper- and lower-case letters as possible within one minute (this measure was collected in all time points: T1–T4). The test's score is the number of letter names that can be produced correctly in 1 minute (appearing below as idel_letter_naming_fluency). Note that standard practice for this test is to use mixed case materials. We deviated from that practice and used all uppercase materials. In Argentina, most printed material for beginning readers is written using uppercase letters only.
- **Phoneme Segmentation Fluency (idel_fsf_tlp):** Students orally segmented individual phonemes from orally presented words of one to three syllables. Two scoring rubrics provide separate scores for the task, one based on phonemes (Phoneme Segmentation Fluency) and another based on syllables (idel_fsf_sil). Scores from this task are available in T1–T4.

- **Nonsense Word Correct Fluency (idel_Nonsense Word Correct Fluency):** measures fluency in nonsense word reading. Two scoring rubrics provide scores for number of letter-sounds correct per minute (idel_fps_tsl) and number of (non)words pronounced correctly (idel_Nonsense Word Correct Fluency). Students read aloud a list of pseudowords (e.g., ro, lali, sepi), measuring their ability to apply letter-sound correspondences and blend them into whole words (available in T1–T4).
- **Oral Reading Fluency (idel_flo):** This subtest assessed accurate and fluent oral reading of a connected passage within one minute (T2–T4).

***Tejas LEE Subtests*.** The *Tejas LEE* is a researcher-developed assessment of Spanish reading skills for students in grades K–3. Based on project access, these subtests were administered primarily at T4 and serve as an independent outcome evaluation. The *Tejas Lee* included the following components:

- **Words Read Correctly Per Minute:** A composite score combining accuracy and fluency in reading grade-level word lists (conducted in T3–T4).
- **Passage Fluency (lectura_fluidez):** Students read connected text aloud for one minute, measuring word reading fluency (conducted in T4 only).
- **Passage Comprehension (lectura_compr):** After reading a text aloud, students answered eight comprehension questions targeting both explicit and implicit information (T4 only).
- **Dictation (l_dictado):** Students wrote ten dictated words of varying complexity, presented in isolation and within a sentence context to avoid homophone confusion (T4).
- **Letter Sound (letter_sound):** Students named the corresponding sound for each of 30 randomly ordered upper- and lower-case letters (T2–T4).

Rapid Automatic Naming (RAN). A *Rapid Automatic Naming (RAN)* test assesses the speed and accuracy with which an individual can orally identify a series of familiar stimuli, such as common colors and everyday items, presented in a randomized, repeated sequence. This timed measure evaluates automaticity in lexical retrieval, a skill closely linked to reading fluency and efficiency. This test was administered at T1 only.

Nonword repetition. A Nonword Repetition task assesses phonological working memory by requiring individuals to repeat unfamiliar, phonotactically plausible sequences of sounds that do not form real words. Performance on this task reflects the capacity to temporarily store and reproduce novel phonological information, a skill linked to language development and reading acquisition.This test was administered at T1 and T4.

Peabody Picture Vocabulary Test. The *Peabody Picture Vocabulary Test* measures an individual's ability to comprehend spoken words by having them select, from a set of four pictures, the image that best represents the meaning of each word presented orally by the examiner. This test was administered at T1 only.

Set for Variability. The *Set for Variability* task measures an individual's ability to correctly identify a mispronounced or phonologically altered version of a word and match it to the correct lexical form (e.g, *amimal* x animal; *torduga* x tortuga) This task assesses flexible word recognition processes that support decoding irregular or less familiar words. This test was administered at T1 and T4.

Results

Feasibility and acceptability results

The first research question addressed whether a two-year evidence-based teacher training program in Spanish-language structured literacy instruction was feasible to implement in Buenos Aires schools and acceptable to participating teachers. This question was assessed using descriptive/qualitative survey data collected from participating teachers. No formal analyses of these descriptive data were undertaken but an overview of teacher responses is summarized below.

Overall, participating teachers reported very favorable opinions about the simplicity and feasibility of implementing the program in the classroom and its efficacy with students. The survey results revealed that:

- 100% of teachers agreed that they would recommend the training course to their colleagues.
- 94% of them believe that their students are learning to read better than in previous years.
- 100% believe that their students improved their reading comprehension skills compared to previous years.
- 94% of respondents believe that their students were more motivated to learn than in previous years.
- 100% of respondents believe that the methodology helps them improve other learning areas.

These descriptive data support the conclusion that the PD Program and the associated AL instructional materials were not only feasible to implement in a variety of classrooms, but teachers felt very positive about their students' response to the program and their own ability to foster reading development in these children.

Main Analyses of Results

The main analyses focused primarily on comparing gains in reading measures in the intervention versus the control group. First, to make sure any differences in gains were not related to pre-existing differences between the two groups, we examined whether there are any differences at Time 1 (T1) between the control and intervention groups. We established the groups had comparable T1 reading skills, with the exception of some marginal

differences, which, if anything, favor the control group (see analyses of T1, below). Second, We examined group-level differences in reading measures between groups over time. We did so by running statistical models that estimate whether there is a main effect of group (across timepoints, i.e., evidence for a general advantage for one group over the other across timepoints tested) and an interaction effect between group and time (i.e., evidence for differential growth between the groups). Third, we examined whether individual differences in component skills of reading collected at T1 predicted reading gains, and whether there are component skills that are differentially predictive of skill in one group versus the other. To this aim, we ran statistical models examining whether T1 individual differences measures interact with group and time to predict reading outcomes.

We note that in the Results section, we review in the text (and in accompanying Figures) all significant ($p < .05$) and marginally significant ($p < .1$) effects. The rationale in reporting marginal effects as well was both to provide a fuller picture of the findings, and, given the large number of comparisons involved (i.e., multiple dependent variables), to get estimates of the proportion of findings that show both significant and marginal significant effects (recall that under the null hypothesis, p-values are uniformly distributed, hence the expected proportion of significant findings is 1/20, and marginal findings is 2/20).

Baseline Differences in Reading Skill Dependent Variables at T1

Two assessments of reading skills were measured in T1: Letter Naming Fluency and Nonword Fluency. As can be seen in Table 1, there was a marginally significant difference between groups in the Letter Naming Fluency task (t(367)=1.83, p=0.069, d=0.19). However, this difference had to do with the control group numerically outperforming the intervention group (with a smaller numeric difference in the same direction also in nonword reading, see Table 1). Such baseline differences means that the reported advantages of the intervention group reported below are observed despite a slight numeric disadvantage at T1.

Table 1 further presents baseline differences in component skills of reading collected at T1 (i.e., set for variability, vocabulary, RAN, phonological awareness (segmentation), and nonword repetition). As can be seen, there were no significant differences between the control and intervention group in these tests. There was again a marginal difference in phonological awareness (t(366)=1.69, p=0.092, d=0.18), of the same trend, with the control group showing numerically better performance. The control group showed similar numeric (statistically insignificant) trends in several other measures (Table 1).

Table 1. Comparison of component skills at T1 for the control and intervention groups

Variable	*t*	*p*	*d*	Mean (SD)	
				Control	Experimental
Letter Naming	**1.83**	**0.069**	**0.19**	**19 (14.7)**	**16.1 (15.2)**
Nonword Reading	1.11	0.270	0.12	6.14 (11.1)	4.88 (10.7)
Set for Variability	1.58	0.115	0.17	22.34 (10.03)	20.8 (8.26)
Vocabulary	0.14	0.891	0.01	66.11 (15.97)	65.84 (21.47)
Rapid Naming (Colors)	-0.37	0.711	-0.04	94.74 (14.64	95.29 (13.37)
Rapid Naming (Object)	0.62	0.538	0.06	94.19 (13.71)	93.23 (15.76)
Phoneme Segmentation	**1.69**	**0.092**	**0.18**	**13.9 (13.82)**	**11.49 (13.56)**
Nonword Repetition	1.36	0.175	0.14	13.07 (5.22)	12.33 (5.25)

Notes: Mean reading skill performance and component skills performance at T1 for the control and experimental groups, along with results of *t* tests (assuming unequal variances); "*d*" represents Cohen's d.

Group-Level Intervention Effects

Next, we tested group-level differences between the two groups over time. For reading outcomes that were collected at more than one time point, we used linear mixed-effect models, with fixed effects for group (effect-coded, control coded as (-1)), time (inserted as an effect-coded categorical variable, T1 coded as (-1)), and their interaction. Models also had by-participant random intercepts, which was the maximal random-effect structure justified by the design that converged (Barr et al., 2013; a model with additional random slopes for time was not identifiable given the number of parameters to be estimated and the number of data points). Models here and below were implemented in R, using the *lme4* package (Bates et al., 2015), with significance of fixed effects estimated using the *lmerTest* package (Kuznetsova et al., 2017). Because we were interested in the overall effects of time, group, and their interaction, rather than specific time point comparisons, we report F tests for the full terms (using the anova() function in R). Effect sizes were estimated using the *effectsize* package (Ben-Shachar et al., 2020). Estimated effects over time are then visually depicted for significant and marginally significant effects (plots generated using the interactions package, Long, 2024). F-values and significance levels for the effects of group, time, and their interaction are presented in Table 2. For the three measures that were assessed only at T4 (Text Reading Fluency, Text Reading Comprehension, and Dictation), we conducted t-tests to examine the difference between the two groups. The results of these comparisons are presented in Table 3.

As expected, in all dependent variables with more than one assessment, there was a significant main effect of time (see Table 2), meaning that, on average, participants

improved their reading with time. Importantly, several models also revealed greater improvement in the intervention than the control group (a significant main effect of group). Thus, students in the intervention group had significantly better performance in text reading comprehension, measured at T4 (t(386)=2.27, p=.024, d=.23; Fig. 1A), and in letter-sound naming, measured at T2-T4 (F(1, 425.4)=18.53, p<.001, η^2p=.04; Fig.1B), alongside marginally significant better performance in text reading fluency (i.e., number of correctly read words per minute in a text), measured at T4 (t(302)=1.69, p=.092, d=0.19; Fig. 1C), and in sentence comprehension, measured at T2-T4 (F(1, 422.8)=3.61, p=.058, η^2p=.008; Fig. 1D).

In addition, other reading measures showed greater improvement in the intervention group through time, as evidenced by significant group-by-time interactions. Thus, students in the intervention group also showed significantly greater gains in letter naming fluency, measured at T1-T4 (F(3, 1132.1)=3.8, p=.01, η^2p=.01; Fig. 2A), in oral word reading fluency, measured at T2-T4 (F(2, 748.1)=3.15, p=.043, η^2p=.008; Fig. 2B), and in word list reading fluency (i.e., number of correctly read words per minute in a word list), measured at T3-T4 (F(1, 276.6)=6.68, p=.009, η^2p=.02; Fig. 2C). Overall, then, in 7 out of 10 dependent variables, there was significant or marginally significant statistical evidence for either an overall advantage for the intervention group over the control group, for a better growth in the intervention group through the timepoints measured, or both.

Table 2. Group-level Analyses: F-values of terms from linear mixed-effect models predicting reading skills from Group (experimental vs. control), and Time.

Variable	Time-points	Group	Time	Group X Time
Letter Naming	T1, T2, T3, T4	F(1, 428.9)=0.04	F(3, 1132.1)=709.64***	F(3, 1132.1)=3.80**
Nonword Naming	T1, T2, T3, T4	F(1, 428.9)=0.97	F(3, 1134.4)=617.46***	F(3, 1134.4)=1.08
Oral Word Fluency	T2, T3, T4	F(1, 428.8)=0.15	F(2, 748.1)=748.12***	F(2, 748.1)=3.15*
Words List Fluency	T3, T4	F(1, 374.6)=0.71	F(1, 276.6)=755.18***	F(1, 276.6)=6.86**
Word Comprehension	T2, T3, T4	F(1, 422.1)=1.68	F(2, 765)=46.50***	F(2, 765)=0.47
Sentence Comprehension	T2, T3, T4	F(1, 422.8)=3.61†	F(2, 764.3)=80.66***	F(2, 764.3)=1.73
Letter-Sound Naming	T2, T3, T4	F(1, 425.4)18.52***	F(2, 764.8)=123.13***	F(2, 764.8)=0.39

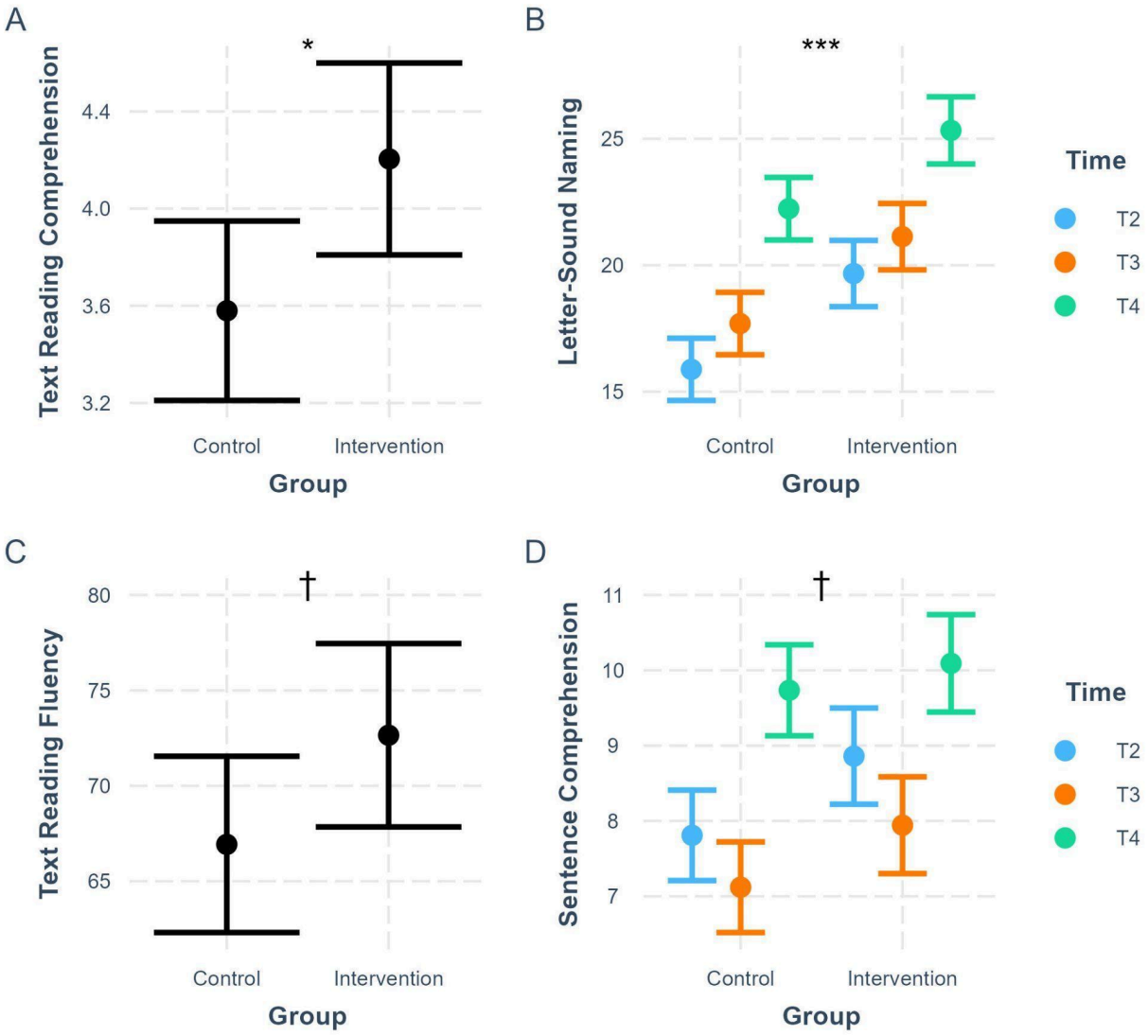
Notes: † p < .1; * p < .05; ** p < .01; *** p < .001

Table 3. Results of t-tests comparing reading skills measured only at T4 between the control and intervention groups

Variable measured at T4	*t*	*p*	*d*	Mean (SD)	
				Control	Experimental
Text Reading Fluency	1.69	0.092	0.19	66.93 (29.72)	72.65 (29.27)
Text Reading Comprehension	2.27	0.024	0.23	3.58 (2.65)	4.2 (2.76)
Dictation	1.03	0.305	0.11	5.39 (3.03)	5.7 (2.91)

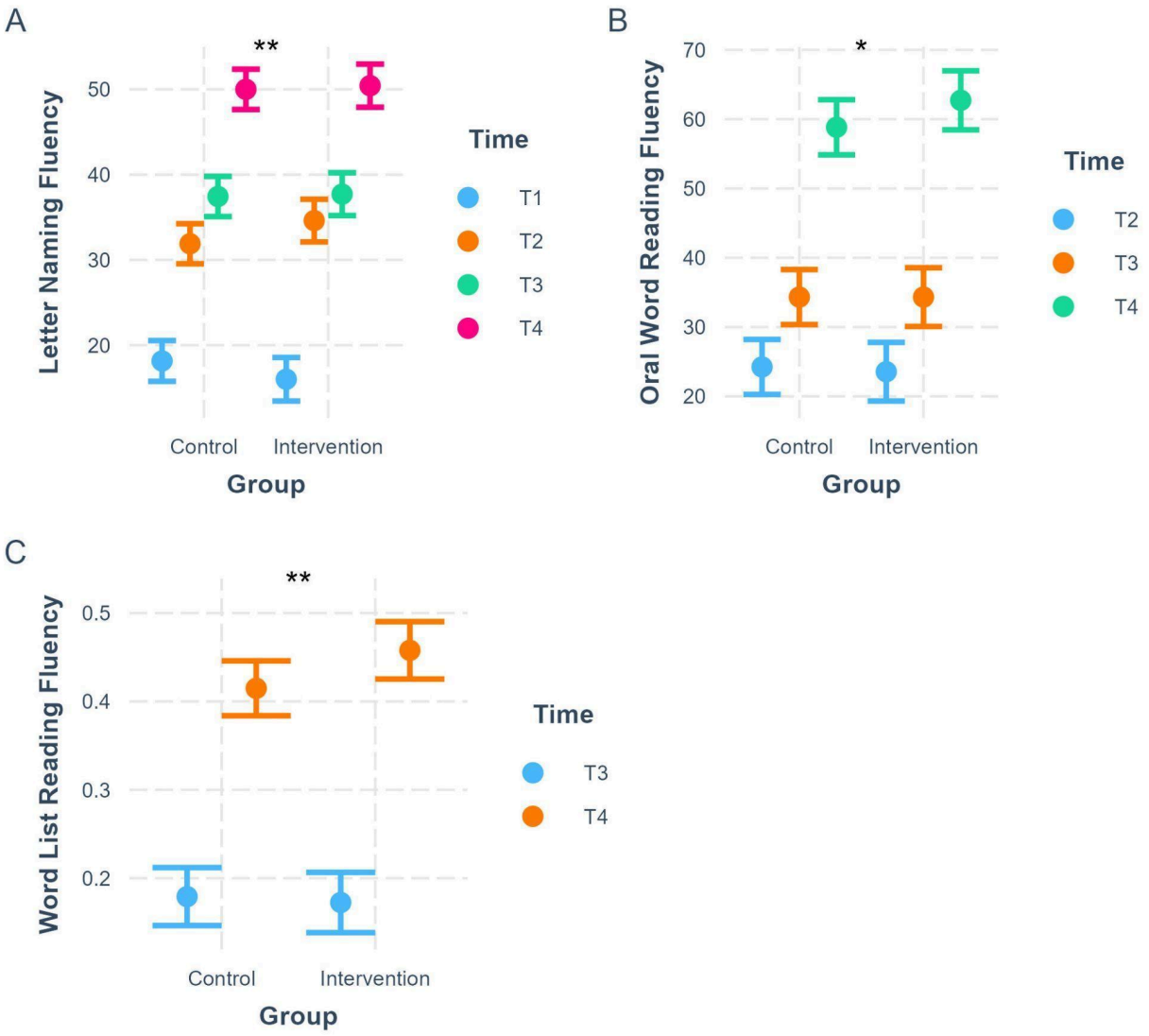
Notes: Mean reading skill performance at T4 for the control and experimental groups, along with results of *t* tests (assuming unequal variances); *"d"* represents Cohen's d.

Figure 1. Estimated Effect of Group on Text Reading Comprehension (Top Left Panel, A); Estimated Interaction Effect of Group and Time on Letter-Sound Naming (Top Right Panel, B); Estimated Effect of Group on Text Reading Fluency (Bottom Left Panel, C); and Estimated Interaction Effect of Group and Time on Sentence Comprehension (Bottom Right Panel, D).



Notes: †<.1; * <.05; **<.01; ***<.001

Figure 2. Estimated Interaction Effect of Group and Time on Letter Naming Fluency (Top Left Panel, A); Estimated Interaction Effect of Group and Time on Oral Word Reading Fluency (Top Right Panel, B); and Estimated Interaction Effect of Group and Time on Word List Reading Fluency (Bottom Panel, C).



Notes: †<.1; * <.05; **<.01; ***<.001

Individual Differences in Component Skills as Predictors of Reading Gains

Lastly, we tested whether individual differences in component skills of reading measured at T1 differentially predicted performance and gains over time in the two groups. Again, for reading outcomes that were collected at more than one time point, we used linear mixed-effect models, with fixed effects for group (effect-coded as above), time (effect-coded

as above), the component skills (scaled and centered), all their two-way interactions, and their three-way interaction. We ran separate models with each component skill of reading as a predictor, for each dependent variable. The only exception was the two sub-scales of Rapid Automatized Naming (colors and objects) which were highly correlated (r = .56) and thus replaced by a composite score (i.e., average of Z-scores of the two sub-scales). Models also again had by-participant random intercepts. For simplicity and given the large number of models tested, we report the F-values for the two interactions of interest (i.e., those that involve Group): Component Skill by Group, and Component Skill by Group by Time (see Table 4), which reflect differential performance or growth in the two groups depending on the baseline component skill level. For the three measures that were assessed only at T4 (i.e., Text reading fluency, Text Reading Comprehension, and Dictation), we conducted a linear model with fixed effects for group (effect-coded, same as above), the component skill (scaled and centered), and their interaction. For brevity we report for these models the estimated effect of the interaction (Table 5).

While not all component skills and/or outcomes showed significant interaction effects involving Group (see Tables 4 and 5), some specific component skills and reading outcomes did yield significant or marginally significant interactions. In particular, we found that students with higher scores in the phoneme segmentation task, measured at T1, had significantly higher scores in the intervention than the control group in text reading comprehension, measured at T4 (F(1, 329)=3.98, p=.047, η^2p=.01; Fig. 3A). Higher T1 scores in phoneme segmentation were marginally associated with better performance in the intervention than the control group (i.e., Group by Phoneme Segmentation interaction) in letter naming, measured in T1-T4 (F(1, 364.4)=3.21, p=.074, η²p=.009; Fig. 3B), nonword naming, measured in T1-T4 (F(1, 367)=3.13, p=.078, η²p=.008; Fig. 3C), and oral word reading fluency, measured in T2-T4 (F(1, 362.8)=3.69, p=.056, η²p=.01; Fig. 3D).

Set for Variability scores at T1 interacted significantly with Group to predict oral word reading fluency, measured in T2-T4 (F(1, 357.4)=4.03, p=.045, η^2p=.01), word list reading fluency, measured in T3-T4 (F(1, 318.2)=4.65, p=.032, η^2p=.01), and text reading comprehension, measured in T4 (F(1, 324)=7.25, p=.007, η^2p=.02). In all these cases, students with higher Set for Variability T1 scores had better reading skills in the intervention vs. control group (see Figures 4A-4C). Lastly, T1 vocabulary scores, T1 scores in the Set for Variability task, and T1 nonword repetition scores, showed significant *three*-way interactions with Group and Time, such that students with higher baseline scores in these component skills had better gains in the intervention group than the control group in word list reading fluency, measured in T3-T4 (Vocabulary: F(1, 250.8)=4.04, p=.046, η^2p=.02; Fig. 5A; Set for Variability: F(1, 238.7)=6.63, p=.011, η^2p=.03; Fig. 5B; Nonword Repetition: F(1, 248.5)=5.42, p=.021, η^2p=.002; Fig. 5C).

These results suggest that, together with the general advantage in reading outcomes associated with the intervention at the group-level, the intervention was more effective for

some children, in particular those with higher component reading skills at baseline. We return to these findings in the General Discussion.

Table 4. Individual-differences analysis for T1 variables that predict reading outcomes

Variable	Component skills (CS)	CS X Group	CS X Group X Time
Letter Naming	Set for variability	$F(1, 359.1)=1.9$	$F(3, 993.9)=0.93$
	Vocabulary	$F(1, 366)=0$	$F(3,1014.6)=0.5$
	RAN-composite	$F(1, 358.5)=2.5$	$F(3, 1003.3)=0.39$
	Phoneme Segmentation	**$F(1, 364.4)=3.21$†**	$F(3, 1009.6)=0.85$
	Nonword Repetition	$F(1, 363.8)=0.004$	$F(3, 1006.7)=1.16$
Nonword Naming	Set for variability	$F(1, 361.5)=2.6$	$F(3, 997.8)=0.53$
	Vocabulary	$F(1, 371.6)=0.28$	$F(3, 1019.3)=0.63$
	RAN-composite	$F(1, 360.1)=1.37$	$F(3, 1006.2)=1.21$
	Phoneme Segmentation	**$F(1, 367)=3.13$†**	$F(3, 1014.3)=0.68$
	Nonword Repetition	$F(1, 366.6)=0.098$	$F(3, 1011.1)=1.16$
Oral Word Reading Fluency	Set for variability	**$F(1, 357.4)=4.03$***	$F(2, 637.3)=2.28$
	Vocabulary	$F(1, 367)=0.51$	$F(2, 651.6)=0.18$
	RAN-composite	$F(1, 356)=0.07$	$F(2, 641.9)=1.46$
	Phoneme Segmentation	**$F(1, 362.8)=3.69$†**	$F(2, 648.6)=0.19$
	Nonword Repetition	$F(1, 362.3)=0.08$	$F(2, 646.1)=1.36$
Words List Reading Fluency	Set for variability	**$F(1, 318.2)=4.65$***	**$F(1, 238.7)=6.63$***
	Vocabulary	$F(1, 332.7)=1.56$	**$F(1, 250.8)=4.04$***
	Ran-composite	$F(1, 321.2)=0.23$	$F(1, 239.8)=1.42$
	Phoneme Segmentation	$F(1, 325.4)=2.49$	$F(1, 243.9)=1.86$
	Nonword Repetition	$F(1, 325.3)=0.03$	**$F(1, 248.5)=5.42$***
Word-Level Processing	Set for variability	$F(1, 353.5)=0.39$	$F(2, 639.6)=0.31$
	Vocabulary	$F(1, 360.9)=0.41$	$F(2, 653.5)=0.98$
	RAN-composite	$F(1, 351.8)=2.15$	$F(2, 644.7)=0.34$
	Phoneme Segmentation	$F(1, 359.1)=0.47$	$F(2, 650.2)=0.15$
	Nonword Repetition	$F(1, 358.7)=0.58$	$F(2, 650.6)=0.98$
Sentence Comprehension	Set for variability	$F(1, 351.4)=0.07$	$F(2, 636.7)=1.63$
	Vocabulary	$F(1, 360.2)=0.1$	$F(2, 651.4)=0.31$
	RAN-composite	$F(1, 352.6)=0.51$	$F(2, 643.8)=0.68$
	Phoneme Segmentation	$F(1, 358)=0.21$	$F(2, 647.8)=1.16$
	Nonword Repetition	$F(1, 356.3)=0.78$	$F(2, 646.7)=2.24$
Letter-Sound Naming	Set for variability	$F(1, 352.7)=0.42$	$F(2, 636.7)=1.28$
	Vocabulary	$F(1, 363.5)=0.45$	$F(2, 352.3)=0.52$
	RAN-composite	$F(1, 353.1)=2.26$	$F(2, 641.3)=0.44$
	Phoneme Segmentation	$F(1, 359.8)=1.92$	$F(2, 648.2)=1.07$
	Nonword Repetition	$F(1, 358.3)=0.3$	$F(1, 646.8)=0.91$

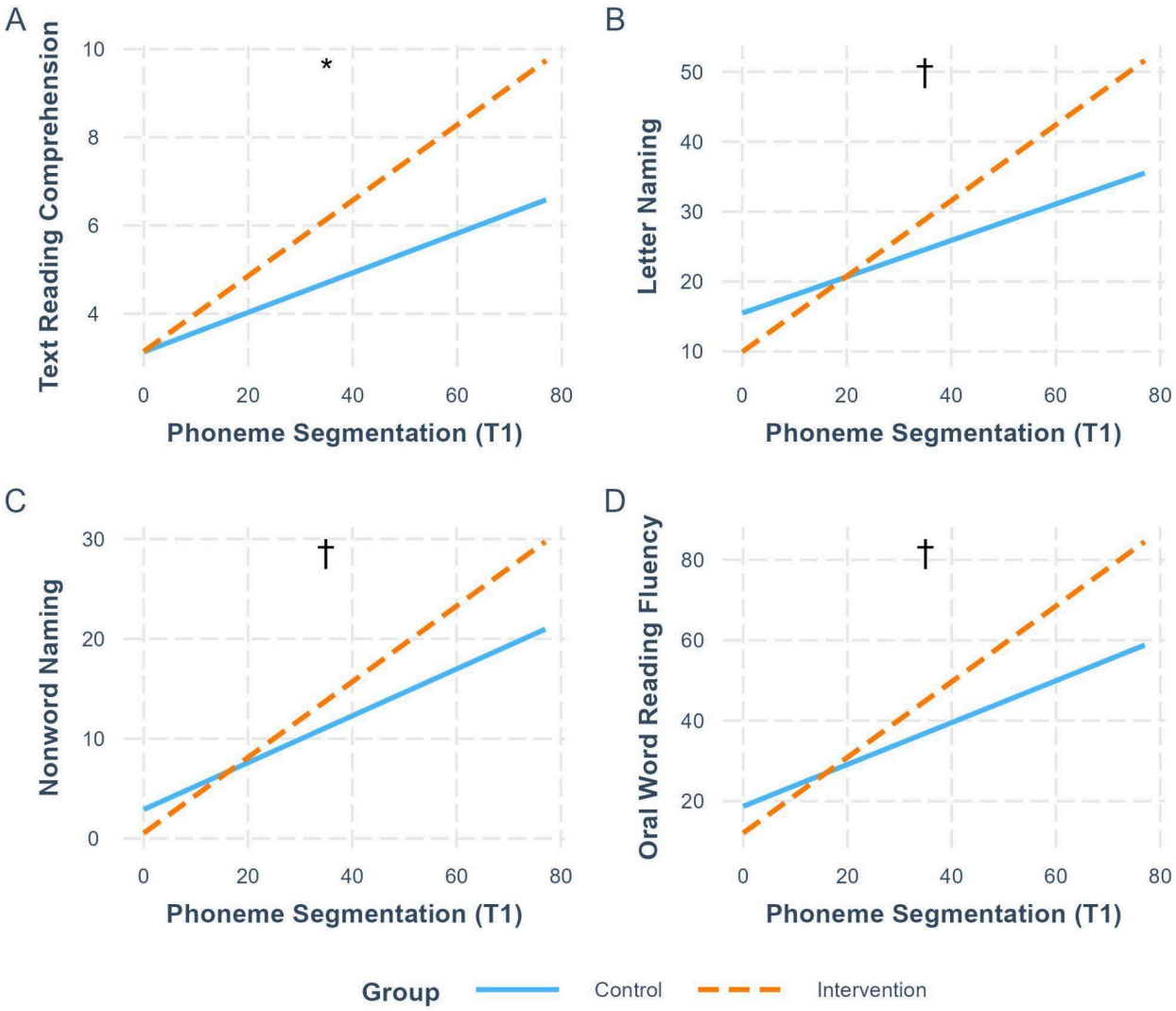
Legend: RAN = rapid automatized naming; CS = component skills; F-values of terms from linear mixed-effect models predicting reading skills from Group (intervention vs. control), Time, Component Skills, and their two- and three-way interactions. For brevity, F-values are presented for the component skill by Group and the three-way interaction, only; † $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$

Table 5. Individual-differences analysis for measures evaluated at T4

Variable at T4	Component skills (CS)	CS X Group
Text Reading Fluency	Set for variability	$F(1, 256)=1.43$
	Vocabulary	$F(1, 260)=0.7$
	RAN-composite	$F(1, 260)=0.76$
	Phoneme Segmentation	$F(1, 259)=0.68$
	Nonword Repetition	$F(1, 259)=0$
Text Comprehension	Set for variability	**$F(1, 324)=7.25$****
	Vocabulary	$F(1, 331)=0.14$
	RAN-composite	$F(1, 329)=1.07$
	Phoneme Segmentation	**$F(1, 329)=3.98$***
	Nonword Repetition	$F(1, 328)=0.71$
Dictation	Set for variability	$F(1, 324)=1.81$
	Vocabulary	$F(1, 331)=0.33$
	RAN-composite	$F(1, 329)=0.65$
	Phoneme Segmentation	$F(1, 329)=0.12$
	Nonword Repetition	$F(1, 328)=0.15$

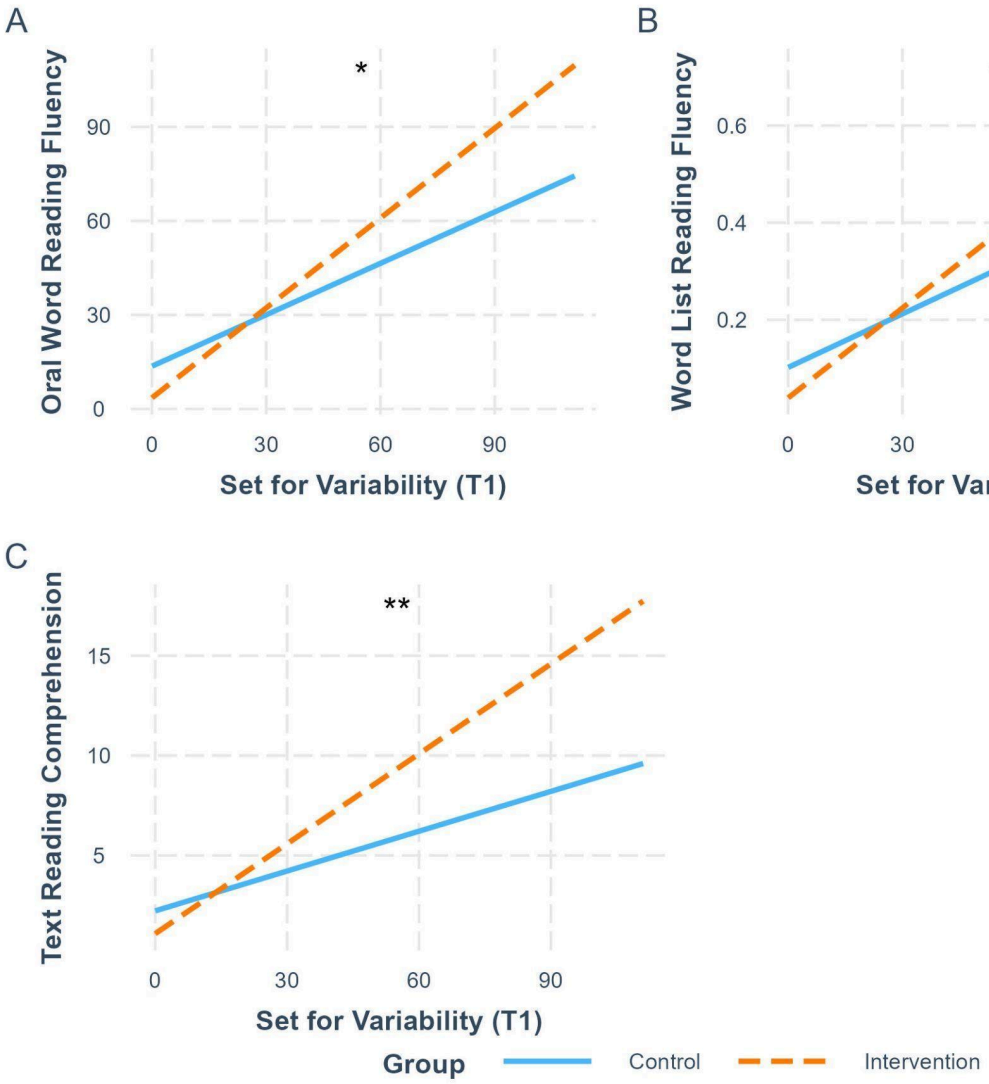
Legend: RAN = rapid automatized naming; CS = component skills; F-values of terms from linear mixed-effect models predicting reading skills from Group (intervention vs. control), Time, Component Skills, and their two- and three-way interactions. For brevity, F-values are presented for the component skill by Group and the three-way interaction, only; † $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$

Figure 3. Estimated Interaction Effect of Group and Phoneme Segmentation at T1 on Text Reading Comprehension (Top Left Panel, A); Estimated Interaction Effect of Group and Phoneme Segmentation at T1 on Letter Naming (Top Right Panel, B); Estimated Interaction Effect of Group and Phoneme Segmentation at T1 on Nonword Naming (Bottom Left Panel, C); and Estimated Interaction Effect of Group and Phoneme Segmentation at T1 on Oral Word Reading Fluency (Bottom Right Panel, D).



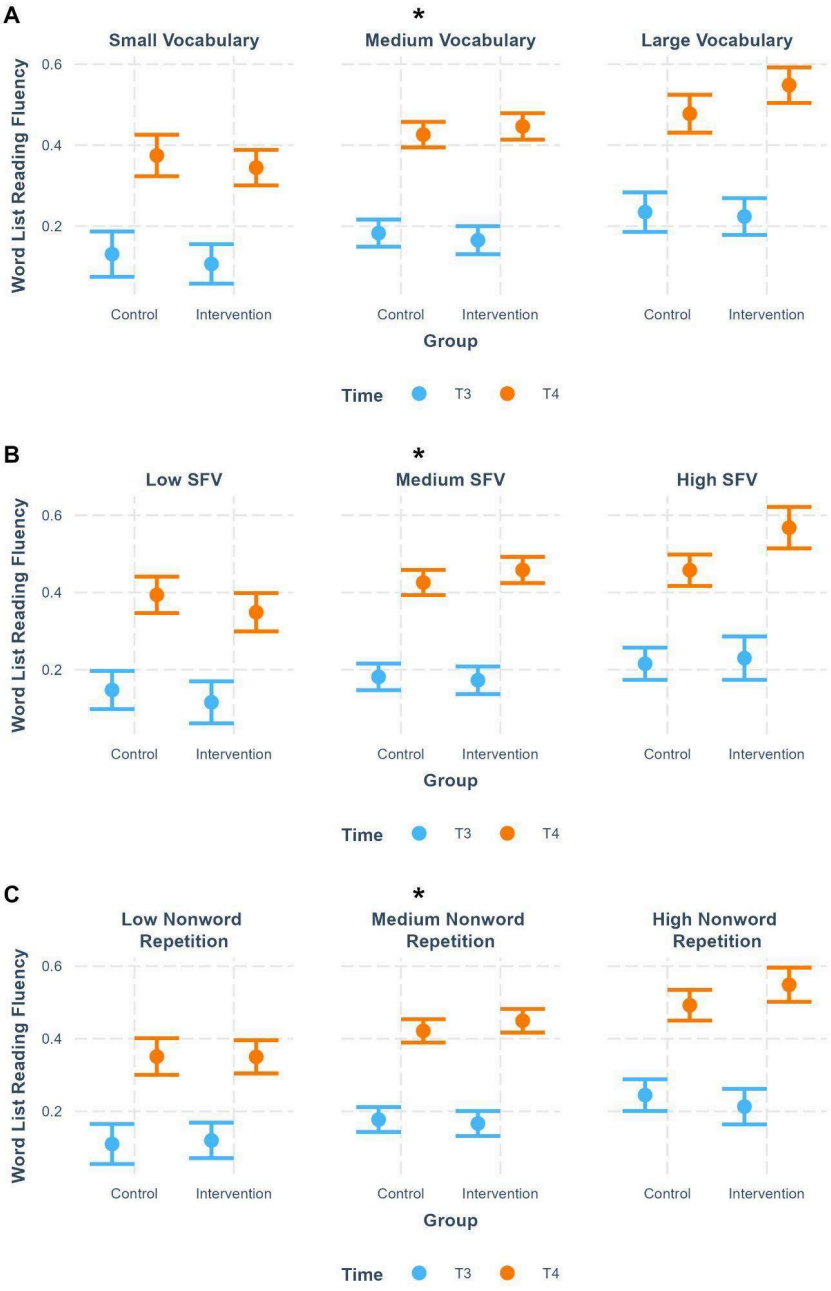
Notes: †<.1; * <.05; **<.01; ***<.001

Figure 4. Estimated Interaction Effect of Group and Set for Variability at T1 on Oral Word Reading Fluency (Top Left Panel, A); Estimated Interaction Effect of Group and Set for Variability at T1 on Word List Reading Fluency (Top Right Panel, B); Estimated Interaction Effect of Group and Set for Variability in T1 on Text Reading Comprehension (Bottom Panel, C).



Notes: †<.1; * <.05; **<.01; ***<.001

Figure 5. Estimated Interaction Effect of Group and Vocabulary size in T1 on Word List Reading Fluency over time (Top Panel, A); Estimated Interaction Effect of Group and Set for Variability in T1 on Word List Reading Fluency over time (Middle Panel, B); Estimated Interaction Effect of Group and Nonword Repetition in T1 on Word List Reading Fluency over time (Bottom Panel, C).



Notes: †<.1; * <.05; **<.01; ***<.001

Discussion

The present research was designed to assess the efficacy of early structured literacy intervention delivered by teachers who received 40 hours of PD and additional support in structured literacy methods, throughout two school years. Eight Buenos Aires schools served as intervention schools, and seven were control schools. A Spanish structured literacy curriculum, *Aprendo Leyendo* (AL), was taught in Grades 1 and 2 in the intervention schools by teachers receiving PD and ongoing support. Control schools used their traditional methods of reading instruction, and their teachers did not receive the PD. AL is a Spanish-language reading program, designed for children in the early stages of reading development.

Intervention students demonstrated significant post-program advantages on several reading and reading-related skills, with better T4 performance than controls on passage comprehension and letter-sound naming. Differential growth on two word reading fluency and one letter naming fluency measure were revealed, and significant gains on two component skills, phoneme segmentation and set for variability. Individual differences analyses showed that, generally, children with better component reading skills at Grade 1 entry demonstrated the greatest intervention gains. Overall, the results of the study provide evidence that teacher training can have a significant impact on early literacy outcomes in complex educational environments; moreover, the impact is driven by an intervention that can be implemented with relative ease and timeliness, and, especially, cost-effectiveness.

The benefits for students of well-trained, effective teachers have been well established. But these benefits may be especially pronounced in poorer countries with substantial social and economic challenges. In lower-income and lower middle-income countries, for example, there is significant instructional variability and the gaps in teacher preparation are pronounced (Dickinson, Connor, & Hadley, 2023). As regional economic challenges vary greatly in LAC countries, so does teacher training and access to effective and updated materials. Initiatives that emphasize teacher professional development have shown encouraging results, especially when they are aligned closely with the Science of Reading and include continued, systematic implementation and evaluation of outcomes.

In a similar PD and intervention program, in Chile, Pallante and Kim (2013) evaluated the Collaborative Language and Literacy Instruction Project (CLLIP), a multicomponent literacy model that provided teachers with targeted training on foundational literacy skills (phonological awareness, phonics, vocabulary) and in comprehension and writing. Teachers received instructional materials that integrated theoretical background, practical

demonstrations, and collaborative lesson planning. An evaluation revealed significant improvement in student outcomes. Kindergartners in the CLLIP classrooms were better than their non-CLLIP peers at letter naming, word reading, vocabulary, and phonemic segmentation; and first-grade students, in turn, were better at vocabulary, reading fluency, and comprehension. These effects, ranging from small (d = .18 in kindergarteners' word reading) to large (d =.70 in kindergarteners' letter-naming fluency), demonstrated the potential of structured, coaching-based teacher training to accelerate early literacy growth. As discussed before, reliable effects of teacher training were found in interventions that combine evidence-based professional development with structured materials, as seen in Aprendamos todos a leer in Colombia (Márquez de Arboleda et al., 2020) and CLLIP in Chile (Pallante & Kim, 2013).

In the present study, the teachers' students were evaluated on a range of reading and reading-related skills using a Spanish reading assessment administered at four time points, before and after Grade 1, and before and at the end of Grade 2. The project was a controlled research evaluation because students' learning trajectories on the performance measures were compared with those of same-grade students in control schools. The control students were assessed at the same time intervals on the same measures, but their teachers did not receive the PD training and students received reading instruction according to the business-as-usual model of their schools.

Findings of the data analyses provide positive evidence of the value of the PD and structured literacy instruction provided to teachers and their students in the intervention schools. Teachers who received the PD and taught the Aprendo Leyendo program had students who demonstrated greater growth than their control peers on important dimensions of reading and reading-related skills. On two of the IDEL measures, children in the Intervention schools showed differential growth in letter naming fluency and in phoneme segmentation fluency relative to controls. The Group x Time interaction for letter naming fluency revealed that while the intervention students were lower performing at T1, they 'caught up' and were equivalent to the control students at T4. In fact, most of this growth occurred during the first year of instruction (T1 to T2). For phoneme segmentation fluency, a similar pattern was observed. These findings are documented in the Supplementary Materials, in Supplementary Table S1 and Supplementary Figure S1.

Another aspect of the present study is that participating teachers reported very favorable opinions about the simplicity and feasibility of implementing the program in the classroom and its efficacy with students. We do not provide a formal analysis of teacher surveys; however, 100% of teachers in the training program (i) agreed that they would recommend the training course to their colleagues, (ii) believed that their students improved their reading comprehension skills compared to previous years, and (iii) believed that the methodology helps them improve other learning areas. Also, 94% of the participating teachers believed their students were learning to read better than in previous years and

that their students were more motivated to learn than in previous years. The present study, thus, suggests that there were significant differences in the reading achievement of students whose teachers were trained with the program and, additionally, that the program was met with high teacher acceptance and satisfaction. These findings differ from those of Braze et al. (2020) who reported a lack of uptake and buy-in from classroom teachers in their study.

Growth on another foundational reading skill was evaluated using the Tejas LEE Letter-Sound measure. The Tejas LEE is considered a better-designed Spanish-language reading assessment tool and only became available to the team after the project was underway. Because of its strengths, however, we decided to incorporate it in the assessments as we collected the data. The Tejas LEE Letter-Sound measure was one of the first included and thus was available for testing at T2, T3, and T4. At T2, the Intervention group was already superior to the control group and maintained this superiority at T3. At T4, the two groups were not significantly different. Results indicated strong significant Group and Time main effects, but a nonsignificant interaction. One could speculate that, had the measure been available at T1, we might have observed such an interaction. Given the scores of the intervention group at T1 on some of these measures, it is quite possible that their superiority at T2 and T3 may be attributable to the structured literacy instruction they received in Year 1.

Two more advanced reading skills were also assessed on the Tejas LEE, although at later time points. Word reading fluency (words read correctly per minute) was available at T3 and T4, and a significant Group x Time interaction was revealed. The intervention group performed lower than controls at T3, but was superior to them at T4. This measure of word reading efficiency is an important outcome, and a building block of higher-order reading skills. The ultimate goal of reading instruction is to help children become independent and competent readers, able to enjoy reading for pleasure and learning, and capable of reading with fluency and comprehension. Thus a key outcome included the passage comprehension measure on the Tejas LEE. This subtest was only available at T4 unfortunately, but it served as an independent assessment of where the two groups placed in general terms at the end of the two years. The intervention group was superior to the control group on Tejas LEE Passage Comprehension at T4 ($p < .02$), with a small but significant advantage (Cohen's $d = .23$) on this important reading outcome.

While the conditions under which this research was undertaken were not ideal, a common lament for real-world evaluation studies, the results are positive and demonstrate the efficacy of the combined PD and structured literacy intervention provided to the intervention schools. Although the children in intervention schools often began at somewhat lower levels of skill than their control school peers, they made significant gains and demonstrated differential positive growth on several important outcomes of reading development. The

positive response of the teachers involved is testimony to the success of the project from the front lines.

The present findings add to the sparse empirical literature on the efficacy of early education initiatives in the LAC. Rebecca Stone and her colleagues conducted a systematic evaluation of evidence (Campbell Systematic Reviews, 2020) on the effectiveness and fidelity of implementation of various programs implemented in the LAC region that aim to improve early grade learning (EGL) outcomes. A review of available quantitative evidence revealed that teacher training, nutrition, and technology programs were not reliably associated with improved early grade learning outcomes in the LAC. A narrative synthesis of the programs, however, did identify factors that could facilitate a positive impact from educational programs on EGL outcomes. One of these determinants was a combination of teacher training with ongoing coaching and communication from mentors, a design feature of the teacher training program implemented in the present study. Stone et al. (2020) caution that the evidence base in LAC is too small to permit strong conclusions. Nevertheless, they do note some evidence supporting the importance of phonological awareness to EGL outcomes, and they speculate that improved preschool programs could facilitate better outcomes in elementary grades.

The individual differences analyses summarized in the Results section revealed that the AL intervention, while generally beneficial, was more effective for some children, in particular those children with higher component reading skills at baseline. This is an example of Matthew effects in reading (Stanovich, 1986, 2000), a phenomenon by which “the rich get richer and the poor get poorer”. Stanovich argues for greater appreciation of “reciprocally facilitating relationships” (Walberg et al., 1983, 1984) that may be a major cause of large individual differences in reading outcomes and other educational achievement data. It is also noteworthy that one of the predictors of reading gain was set for variability performance (i.e., mispronunciation correction task): A task we adapted for the purpose of the current study to Spanish, by creating stimuli that are close phonological neighbors to a target spoken word. Notably, we found that set for variability, a measure presumably reflecting the integrity of the phonological system (see Edwards et al., 2022; Steacy et al., 2023), is relevant not only in opaque writing systems, where it has been mostly studied to date, but also in Spanish, a transparent writing system.

The interaction between where children start and their eventual reading outcomes suggests an urgent need to do more on pre-reading phonological readiness and other reading readiness skills in preschool to facilitate maximum benefits from reading instruction in the early elementary grades. Recognition of the important role of early phonological awareness and the inherent advantage enjoyed by more skilled children at school entry leads to the realization that preschool programs could exert a potentially large positive influence on EGL outcomes. In Argentina, education is compulsory from the age of four years. From the present data and those reviewed by Stone et al. (2020), a comprehensive framework for

education in the preschool years could be advocated. Preschool curricula rich in oral language development and including instructional activities to promote phonological awareness, vocabulary, print awareness, and pre-reading skills could enhance the academic readiness of thousands of young children currently ill-prepared to begin reading instruction.

Overall, these findings can inform educational administration and policy development. The success of the present project demonstrates the impact of structured literacy PD and instruction in the early grades and demonstrates that better literacy outcomes may be achieved with research-based structured literacy programming for all beginning readers. These findings support continued use and ultimate scaling of implementation for structured literacy initiatives to improve literacy outcomes for children across Argentina.

Many positive program evaluation findings remain siloed in academic publications and may never exert an impact on actual classroom practices. In the present instance, the exact opposite scenario occurred. In 2024, co-author FS presented the findings to the Ministry of Education and other authorities in Buenos Aires. CABA had a whole language, balanced approach to literacy, with no letter-sound instruction nor reading aloud in class. After the results of this research were presented to the Ministry and other authorities, however, the Ministry decided that all public schools in CABA would shift from balanced to explicit literacy instruction, and that the curriculum and books and materials would be rewritten according to the evidence found in the research. From 2025 on, CABA has a structured and explicit literacy curriculum and materials across schools and districts. This is a remarkable timeline for the translation of research into meaningful classroom practice.

Supplementary material

Group level effects on component skills

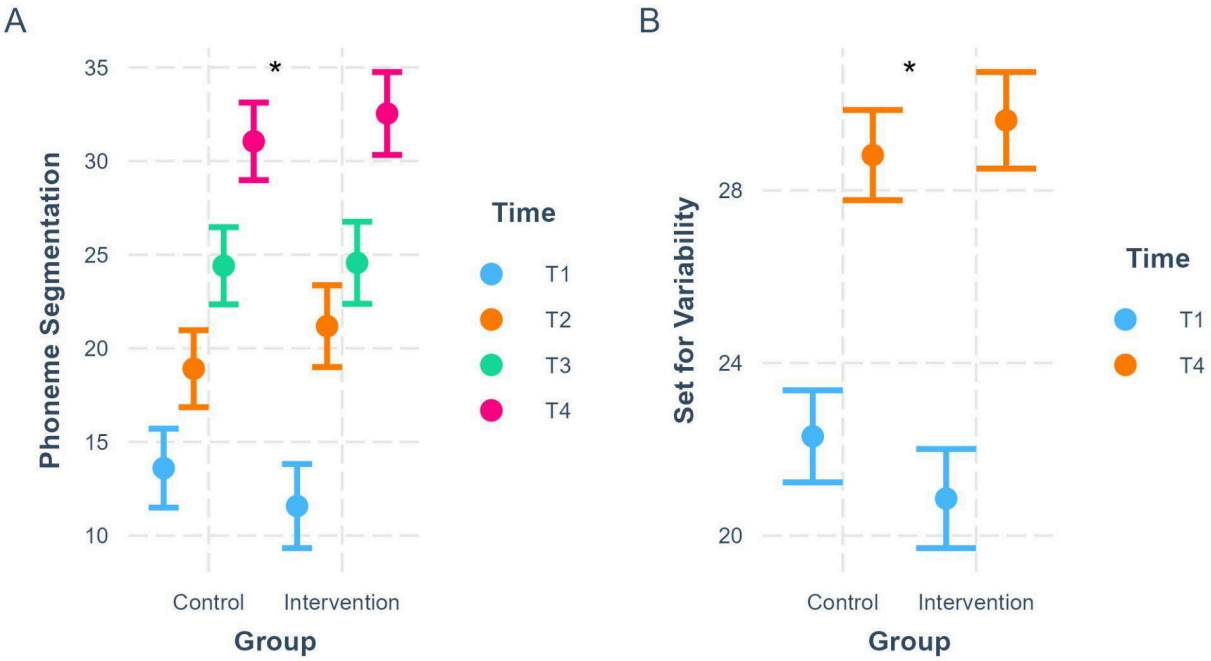
Similar to the group-level analysis of the group and time effects on reading skills, we analyzed the group and time effects on component skills that were collected at more than one time point. In this analysis, we used linear mixed-effect models, with fixed effects for group (effect-coded, control coded as (-1)), time (inserted as an effect-coded categorical variable, T1 coded as (-1)), and their interaction. Models also had by-participant random intercepts, which was the maximal random-effect structure justified by the design. F-values and significance levels for the effects of group, time, and their interaction are presented in Table S1. The three measures that were assessed in more than one time point were Phoneme Segmentation, Set for variability, and Nonword Repetition. For all the component skills with more than one assessment, there was a significant main effect of time (see Table S1), meaning that, on average, participants improved in their component skills with time. Two models also revealed greater improvement in the intervention through time compared to the control group, as evidenced by significant group-by-time interactions. Thus, students in the intervention group showed significantly greater gains in phoneme segmentation, measured at T1-T4 (F(3, 1140.94)=2.97, p=.031, η^2p<.01; Fig. S1A), and in Set for Variability, measured at T1 and T4 (F(1, 385.59)= 5.71, p=.017, η^2p=.01; Fig. S1B)

Table S1. Group-level Analyses: F-values of terms from linear mixed-effect models predicting component skills from Group (intervention vs. control), and Time.

Variable	Time points	Group	Time	Group X Time
Phon.	T1, T2, T3, T4	F(1, 430.85)=0.04	F(3, 1140.94)=217.71***	F(3, 1140.94)=2.97*
SfV	T1, T4	F(1, 419.09)=0.97	F(1, 385.59)=264.19***	F(1, 385.59)= 5.71*
NW	T1, T4	F(1, 405.35)=0.17	F(1, 353.70)=195.31***	F(2, 353.70)=1.8

Notes: †<.1; * <.05; **<.01; ***<.00; Phon.. = Phoneme Segmentation; SfV = Set for Variability; NW = Nonword Repetition.

Figure S1. Estimated Interaction Effect of Group and Time on Phoneme Segmentation (Left Panel, A); and Estimated Interaction Effect of Group and Time on Set For Variability (Right Panel, B).



Notes: †<.1; * <.05; **<.01; ***<.001

References

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. Remedial and Special Education, 23(5), 300–316. https://doi.org/10.1177/07419325020230050501

Alvarez Marinelli, H., Berlinski, S., Busso, M., & Martinez Correa, J. (2023). Improving early literacy through teacher professional development: Experimental evidence from Colombia. *Journal of Public Economics Plus*, *4*, 100019. https://doi.org/10.1016/j.pubecp.2023.100019

APRENDER (Resultados APRENDER 2023, p. 35). (2023). Ministerio de Capital Humano. Argentina

Baker, D. L., Cummings, K. D., Good, R. H., & Smolkowski, K. (2007). *Indicadores Dinámicos del Éxito in la Lectura (IDEL®): Summary of decision rules for intensive, strategic, and benchmark instructional recommendations in kindergarten through third grade*. Dynamic Measurement Group.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language, 68*(3), 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of statistical software, 67*, 1-48.

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of open source software*, *5*(56), 2815.

Braze, D., Salvarezza, F., & Pugh, K.R. (2020) Multi-sensory structured reading instruction for Spanish-speaking First Grade students in Buenos Aires public schools. Unpublished research report.

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., Salas, N., Schöffelová, M., Defior, S., Mikulajová, M., Seidlová-Málková, G., & Hulme, C. (2012). Common Patterns of Prediction of Literacy Development in Different Alphabetic Orthographies. Psychological Science, 23(6), 678-686. https://doi.org/10.1177/0956797611434536

Caravolas, M., Lervåg, A., Defior, S., Seidlová Málková, G., & Hulme, C. (2013). Different Patterns, but Equivalent Predictors, of Growth in Reading in Consistent and Inconsistent Orthographies. *Psychological Science*, *24*(8), 1398-1407. https://doi.org/10.1177/0956797612473122

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological science in the public interest*, 19(1), 5-51.

Edwards, A. A., Steacy, L. M., Siegelman, N., Rigobon, V. M., Kearns, D. M., Rueckl, J. G., & Compton, D. L. (2022). Unpacking the unique relationship between set for variability and word reading development: Examining word- and child-level predictors of performance. *Journal of Educational Psychology*, 114(6), 1242–1256. https://doi.org/10.1037/edu0000696

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. Review of educational research, 71(3), 393-447.

Gibson, E. J., & Levin, H. (1975). *The psychology of reading.* The MIT Press.

Kim, Y. S. G., & Zagata, E. (2024). Enhancing reading and writing skills through systematically integrated instruction. *The Reading Teacher*, *77*(6), 787-799.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, *82*, 1-26.

Long JA (2024). *interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions*. doi:10.32614/CRAN.package.interactions

Márquez de Arboleda, Á., Cotto Girón, E. I., Niño Rocha, N., Ayala Guio, S. C., Mancilla Rodríguez, L. L., Calderón Jiménez, F. J., & Duarte Amézquita, C. A. (2020). *Aprendamos todos a leer: Guía para el docente: Primer semestre: Unidades 1 y 2*. https://doi.org/10.18235/0002592

Nelson, M. (2001). Assessing the early literacy skills of young English learners: The use of DIBELS in Spanish. Doctoral dissertation, University of Oregon.

Rhodes, R. (2021). *The Future of Literacy in Latin America: How Evidence from LAC Reads Can Help Close Early Reading Gaps*. USAID. file:///Users/aub20002/Downloads/LACReadsEvidence_02252021.pdf

Rueckl, J. G., Paz-Alonso, P. M., Molfese, P. J., Kuo, W. J., Bick, A., Frost, S. J., ... & Frost, R. (2015). Universal brain signature of proficient reading: Evidence from four contrasting languages. *Proceedings of the National Academy of Sciences*, 112(50), 15510-15515.

Seymour, P. H. K. K., Aro, M., Erskine, J. M., Wimmer, H., Leybaert, J., Elbro, C., Lyytinen, H., Gombert, J. E., Le Normand, M. T., Schneider, W., Porpodas, C., Ragnarsdottir, H., Tressoldi, P., Vio, C., De Groot, A., Licht, R., Iønnessen, F. E., Castro, S. L., Cary, L., Olofsson, Å. (2003). Foundation literacy acquisition in European orthographies. *Br.J.Psychol.*, *94*(Pt2), 143–174 https://doi.org/10.1348/000712603321661859

Siegelman, N., Rueckl, J. G., van den Bunt, M., Frijters, J. C., Zevin, J. D., Lovett, M. W., ... & Morris, R. D. (2022). How you read affects what you gain: Individual differences in the functional organization of the reading system predict intervention gains in children with reading disabilities. *Journal of Educational Psychology*, 114(4), 855.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360-407. doi:10.1598/RRQ.21.4.1

Stanovich, K. E. (2000). *Progress in understanding reading. Scientific foundations and new frontiers*. New York, NY: Guilford Press.

Steacy, L. M., Compton, D. L., Petscher, Y., Elliott, J. D., Smith, K., Rueckl, J. G., ... & Pugh, K. R. (2019). *Development and prediction of context-dependent vowel pronunciation in elementary readers*. Scientific Studies of Reading, 23(1), 49-63.

Steacy, L. M., Edwards, A. A., Rigobon, V. M., Gutierrez, N., Marencin, N. C., Siegelman, N., ... & Compton, D. L. (2023). Set for variability as a critical predictor of word reading: Potential implications for early identification and treatment of dyslexia. *Reading Research Quarterly*, 58(2), 254-267.

Stone, R., Hoop, T., Coombes, A., & Nakamura, P. (2020). What works to improve early grade literacy in Latin America and the Caribbean? A systematic review and meta-analysis. *Campbell Systematic Reviews*, *16*(1). https://doi.org/10.1002/cl2.1067

Stuebing, K. K., Barth, A. E., Trahan, L. H., Reddy, R. R., Miciak, J., & Fletcher, J. M. (2015). *Are Child Cognitive Characteristics Strong Predictors of Responses to Intervention?* A Meta-Analysis. Review of Educational Research, 85(3), 395–429. **https://doi.org/10.3102/0034654314555996**

Texas Education Agency & The University of Houston System. (2010). *Tejas LEE (Lectura en Español) Spanish Early Reading Assessment*. Brookes Publishing Company.

UNESCO. (2024). *Global Education Monitoring Report, 2024/5*. GEM Report UNESCO. https://doi.org/10.54676/EFLH5184

Venezky, R. L. (1999). *The American way of spelling: The structure and origins of American English orthography*. Guilford Press.

Videla, B., Nistal, M. & Orlicki, E. (2024). *Ausentismo estudiantil en secundaria: percepción y dimensiones.* Observatorio Argentinos por la Educación.

Walberg, H.J. & Tsai, S.-L. 1983. "Matthew Effects in Education”, *American Educational Research Journal*, 20: 359-373.

Walberg, H.J., Strykowski, B.E, Rovai E., & Hung, S.S.(1984). Exceptional performance. *Review of Educational Research*, 54, 87-112.

Watson, J. M. (2004). *Examining the reliability and validity of the Indicadores Dinámicos del Éxito en la Lectura (IDEL): A research study*. University of Oregon

Buenos Aires Ciudad (2019). Diseño curricular 2019. Prácticas del Lenguaje. Primer ciclo escuela primaria.

Buenos Aires Ciudad (2021). *Contenidos priorizados para el ciclo lectivo 2021.* Nivel Primario.

Buenos Aires Ciudad (2024). *Diseño Curricular 2024. Orientaciones para la enseñanza y la evaluación*. Ministerio de Educación.Diseño curricular - Nivel primario | Buenos Aires Ciudad

Buenos Aires Ciudad (2024). *Diseño curricular 2024. Marco general.* Ministerio de Educación. Nivel primario Ciudad de Buenos Aires