ROCKWOOL Foundation Berlin
Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES 085/26

Effective Families or Effective Schools? Experimental Evidence on Fostering Children's Numeracy

Samuel Berlinski, Michele Giannola, Alessandro Toppeta

www.rfberlin.com MARCH 2026

Effective Families or Effective Schools? Experimental Evidence on Fostering Children's Numeracy

Authors

Samuel Berlinski, Michele Giannola, Alessandro Toppeta

Reference

JEL Codes: A2, H52, I25

Keywords: Families, Schools, Human Capital, Numeracy

Recommended Citation: Samuel Berlinski, Michele Giannola, Alessandro Toppeta (2026): Effective Families or Effective Schools? Experimental Evidence on Fostering Children's Numeracy. RFBerlin Discussion Paper No. 085/26

Access

Papers can be downloaded free of charge from the RFBerlin website: https://www.rfberlin.com/discussion-papers

Discussion Papers of RFBerlin are indexed on RePEc: https://ideas.repec.org/s/crm/wpaper.html

Disclaimer

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com

Effective Families or Effective Schools?
Experimental Evidence on Fostering Children's Numeracy

Samuel Berlinski Michele Giannola Alessandro Toppeta*

March 2026

Abstract

We study the relative effectiveness, cost-effectiveness, and interaction of family- and school-based learning interventions using a randomized controlled trial in Colombia that assigns children to a parental engagement program, a teacher professional development program, both, or a control group. Both interventions are grounded in a child-centered learning approach that emphasizes active engagement and the progression from informal to formal mathematical understanding. Each intervention independently generates sizable and statistically similar gains in early numeracy (0.17σ and 0.20σ). Combining them produces no additional learning gains, suggesting that the two interventions act as substitutes over the time horizon and skill domain we study. When benefits accruing to future cohorts are taken into account, the teacher development program becomes at least as cost-effective as, and potentially more cost-effective than, the parental engagement intervention. Our results suggest that, in this setting, strategically concentrating resources on a single binding constraint – either at home or in school – maximizes the short-run learning gains per dollar spent.

*Berlinski: Inter-American Development Bank. Email: samuelb@iadb.org. Giannola: University of Naples Federico II, CSEF and the Institute for Fiscal Studies. Email: michele.giannola@unina.it. Toppeta: SOFI, Stockholm University. Email: alessandro.toppeta@sofi.su.se. We would like to thank Fundación Luker, PREST and STEM Academia for their contribution to implementing the interventions. In particular, Stephan Baillargeon, Monica Bedoya, Mauricio Duque, Annie Fontaine and Maria Valentina Moreno. We also thank Susanna Loeb, Erich Battistin, Henning Hermes, Genia Rachkovski, Miguel Urquiola, Marianne Page and seminar participants at NBER spring meeting (Children), CESifo Economics of Education Conference, HEC Lausanne, UNU-WIDER, Williams College, IFAU (Uppsala), SOFI (Stockholm University), IDB, CESifo Venice Summer Institute on Funding Education, UCL, LSE, Workshop on Early Investments, Family Well-Being, and Child Development (UCopenhagen/VIVE), Advances in Field Experiments (AFE), TUM Workshop on Field Experiments in Economics and Business, IZA/ECONtribute Workshop on the Economics of Education, SITE/SSE, Aarhus University, Helsinki GSE, Research Institute of Industrial Economics (IFN) for their useful comments. Giannola gratefully acknowledges support from the Jacobs-CIFAR Research Fellowship 2026-2028. Isabel Avila provided outstanding research assistantship. The study was registered with the AEA RCT Registry (AEARCTR-0015545). The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.

1 Introduction

Children's early numeracy skills are an important predictor of later academic achievement (Duncan et al. 2007) and economic success in adulthood (Hanushek et al. 2015). Building early mathematical competence during pre-primary and primary education is therefore a key milestone to ensure that *all* children reach their full potential. This challenge is especially severe in low- and middle-income countries (LMICs), where most children complete primary school but acquire very limited skills, leaving learning outcomes far below curricular expectations (Angrist et al. 2021, Glewwe & Muralidharan 2016). In Colombia, the context of this study, 76% of 5th-grade students performed at or below the minimum level in mathematics on the national SABER test in 2016 (ICFES 2017), and only 16.6% achieved minimum math proficiency by the end of primary in 2019 (UNESCO 2025).[1]

This *learning crisis* is widely attributed to two interrelated constraints on early skill formation. First, children from disadvantaged backgrounds often enter school with significant learning deficits that emerge well before formal schooling begins (Carneiro & Heckman 2002, 2003). Limited cognitive stimulation at home and low levels of parental support leave many children ill-prepared to meet the demands of classroom instruction (Currie & Almond 2011, Almond et al. 2018, Duncan et al. 2023). Without strong foundational skills, students struggle to keep pace once they enter school, and early gaps tend to persist and worsen over time (Muralidharan et al. 2019).

Second, the quality of teaching in many LMICs is alarmingly low, further compounding the problem once children enter school. While high-quality teaching can significantly improve student learning in both the short and long term, affecting not just academic achievement but also adult outcomes (Araujo et al. 2016, Chetty et al. 2014, Jackson 2018), many teachers lack the pedagogical skills and motivation needed to deliver instruction that is aligned with students' learning needs. Classroom observations from six countries in Sub-Saharan Africa revealed that few public primary school teachers use the active, structured pedagogical practices associated with effective teaching (Bold et al. 2017).

[1] PISA results for 15-year-old students confirm these low levels of achievement (Schleicher 2018). Similar patterns are observed in other LMICs: for example, in India fewer than 28 percent of grade 3 students could master double-digit subtraction, and only 34 percent could recognize two-digit numbers (World Bank 2018).

In response to this learning crisis, a wide range of policies have sought to strengthen early skill formation either by improving the home learning environment or by enhancing classroom instruction.[2] Parental engagement interventions support caregivers in fostering learning through structured activities at home (Duncan et al. 2023), while teacher professional development programs aim to improve pedagogy inside the classroom (Popova et al. 2022). While both approaches have shown promise in isolation (e.g., Angrist & Lavy 2001; Cilliers et al. 2020; York et al. 2019), there is limited causal evidence on their *relative* effectiveness and cost-effectiveness in improving early numeracy, particularly in LMICs.

Even less is known about how these different policy approaches interact over the course of early childhood. Models of skill formation emphasize dynamic complementarities, in which early investments increase the productivity of subsequent ones (Cunha & Heckman 2007, Cunha et al. 2010, Agostinelli & Wiswall 2025). From this perspective, interventions targeting families and teachers may reinforce one another, generating gains that exceed the sum of their parts. At the same time, when interventions target a narrow set of foundational skills over a relatively short developmental window, improvements along one margin – either at home or in school – may substantially relax binding constraints on students' learning, leaving limited scope for additional gains from further inputs along the other margin. Whether policies targeting families and teachers operate as complements or substitutes remains an open empirical question, particularly because complementarities in the skill production function need not directly translate into complementarities in reduced-form policy effects (Todd & Wolpin 2003).

This paper provides causal evidence on these questions using a randomized controlled trial conducted in Colombia between 2021 and 2023, following children as they transition from pre-primary into primary school, a critical developmental window for foundational learning and skill development (Reynolds & Temple 2019). Children were randomly assigned to a parental engagement program, a teacher professional development program, both interventions, or a control group that received neither. This design allows us to estimate the impact of each program in isolation, assess their relative cost-effectiveness, and examine whether combining them generates additional gains beyond those produced by each intervention alone.

[2] See Glewwe & Muralidharan (2016) and Muralidharan (2017) for recent reviews.

The parental engagement program was implemented prior to school entry, during the summer preceding first grade (December 2021–February 2022). Building on evidence that text-message interventions can effectively support learning at home (Mayer et al. 2019), the program encouraged caregivers to engage children in simple, developmentally appropriate activities embedded in daily routines and using common household objects. While the focus was on numeracy, the activities also supported broader early cognitive skills, helping children develop foundational abilities that prepare them for school. The teacher professional development program, implemented after school entry, targeted classroom instruction during first and second grade. First-grade teachers were trained in 2022, and second-grade teachers in 2023, when the study cohort transitioned to second grade. Drawing on prior evidence that structured teacher support can improve learning outcomes (Popova et al. 2022), the program provided multiple in-person workshops, weekly in-class coaching, and ongoing online support to strengthen pedagogy and enhance children's mathematics learning.

Both interventions build on recent evidence showing that child-centered environments that foster curiosity and active exploration can effectively support learning (Alan & Mumcu 2024, Ashraf et al. 2021, Granger et al. 2012). They operationalize this approach through structured, play-based learning activities and the use of concrete manipulatives that engage children's natural reasoning about quantities, patterns, and relationships. Building on children's intuitive, non-symbolic numerical understanding (Dehaene 2011) and explicitly connecting it to formal mathematical symbols and procedures, these interventions are designed to address a key pedagogical challenge documented in previous work: the difficulty of translating intuitive mathematical competence into school-relevant skills (Dillon et al. 2017, Banerjee et al. 2025).

We measure learning using standardized achievement tests administered at the end of first grade and at the beginning and end of second grade. To shed light on the mechanisms, we complement test-score data with detailed measures of home and school inputs, including phone surveys with caregivers, in-person interviews with teachers and in-classroom observations.

Both interventions independently improve children's numeracy skills. Relative to the control group, children assigned only to the parental engagement program scored 0.174 standard deviations (SD) higher on a summary numeracy index (p-value$<$0.01), while those assigned only to the teacher development program improved by 0.203 SD (p-value<0.05). These effects are similar in magnitude and statistically indistinguishable (the p-value for the difference is 0.741). Children who received both interventions improved by 0.175 SD (p-value<0.05), an effect that is not statistically different from either intervention alone. Thus, combining the two programs does not generate additional learning gains beyond those produced by each individually, suggesting that the two interventions function as substitutes in this context.

Examining impacts by skill type reveals that effects are smallest for basic numeracy skills, such as number writing and reading, where the performance of control group students was already high, and largest for more advanced skills such as solving sequences, equations, and word problems, where control-group children correctly answered only 26% of items. Consistent with Jervis et al. (2024), the gains from the parental engagement program are sustained, persisting even after the intervention ends and children transition into primary school. Over time, the effects of the teacher development grow, increasing from 0.12 SD at the end of first grade to 0.26 SD by the end of second grade, potentially reflecting the cumulative exposure to trained teachers, or improvements in program quality over time.

Consistent with the educational philosophy of the programs, we find that both caregivers and educators engaged more actively with children. The parental engagement program increased the frequency of learning-oriented interactions at home, such as counting, reading, and storytelling, with engagement rising during the period in which caregivers received the prompts and declining after the intervention ended. The teacher professional development program led to substantial changes in instructional practice: treated teachers adopted more active and structured teaching methods, increased student participation, reduced classroom distractions, and demonstrated stronger alignment with the national curriculum and child-centered pedagogy. These improvements in teachers' practices were large across survey waves and increased modestly over time.

We further examine the policy implications of these findings by comparing the relative cost-effectiveness of the interventions. Over the study period, the parental engagement program emerges as the more cost-effective approach, reflecting its low per-child cost and meaningful impacts on early numeracy, consistent with the view that earlier investments generate higher returns than later ones (Heckman 2006). At the same time, this comparison depends on how investments in teachers are conceptualized. While the parental engagement intervention is more cost-effective for the initial cohort, accounting for benefits to future cohorts implies that teacher development becomes at least as cost-effective, and potentially more cost-effective, under different amortization assumptions. The absence of additional learning gains from the combination of parental engagement and teacher professional development interventions has an important implication. In this setting, providing both programs jointly yields lower improvements in numeracy per dollar spent than implementing either intervention alone. Across all scenarios, therefore, the combined intervention is dominated.

Our main contribution is threefold. First, we provide causal evidence that child-centered pedagogical approaches can meaningfully enhance early numeracy by leveraging children's intuitive understanding of quantities and guiding them to formal mathematical concepts (Alan & Mumcu 2024, Ashraf et al. 2021, Granger et al. 2012). In particular, we show that the same pedagogical principles are effective whether applied at home by caregivers or in the classroom by teachers, demonstrating the versatility of child-centered, exploration-driven learning across multiple contexts.

Second, we provide a rigorous comparison of the relative effectiveness and cost-effectiveness of family- and school-based early learning interventions in a low-income setting. Although both programs improve numeracy, consistent with previous evidence (Banerjee et al. 2007, Muralidharan et al. 2019, Rege et al. 2024, Grönqvist et al. 2025), their joint provision does not generate additional gains. Over the time horizon and skill set we study, a single intervention is more cost-effective than implementing both, highlighting that strategically concentrating resources on a single binding constraint – either at home or in school – can maximize short-run learning gains per dollar spent. This result is particularly relevant for policymakers seeking to efficiently allocate scarce resources across competing programs.

Third, our findings speak to the literature on skills formation by providing evidence on how family- and school-based interventions interact, an important question in models emphasizing dynamic complementarities in early investments (Cunha & Heckman 2007, Cunha et al. 2010, Agostinelli & Wiswall 2025).[3] Policy effects reflect not only the underlying technology of skill formation but also behavioral responses by parents, teachers, and students (Das et al. 2013, Todd & Wolpin 2003). We find that parental engagement and teacher professional development each improve early numeracy, but that combining the two yields no additional gains. One possible explanation for the lack of additive effects is that the two interventions relax the same binding instructional constraint, so that once children receive effective numeracy support along this dimension, additional inputs over the time horizon we study generate no further gains. Importantly, our results do not rule out complementarities in the underlying skill formation process, which may operate across broader skill domains or over longer developmental horizons.

The paper is organized as follows. Section 2 presents the setting and interventions. Section 3 describes the study design and Section 4 presents the data. Section 5 presents the internal validity of the experiment, the main results with an exploration of the mechanism, and the costs and benefits of the different interventions. Section 6 concludes.

2 Setting and Interventions

2.1 Study setting

In Colombia, compulsory education is universal and free of charge. Formal schooling begins with a compulsory pre-primary year, known as Transición, which enrolls children aged 5 to 6 and marks the formal entry point into the school system (OECD 2016). Transición is followed by five years of primary education, covering Grades 1 through 5. The school year runs from February to early December, and primary school students are typically taught by a single classroom teacher who is responsible for instruction across all subjects.

Our study is conducted in Manizales, the capital of the Department of Caldas. Manizales is a medium-sized city with a public school system serving both urban and rural populations. Approximately 14% of residents live below the poverty line, and 7% live in rural areas. Although primary school students in Manizales performed slightly above the 2016 national average on

[3] Gilraine (2016) and Johnson & Jackson (2019) find evidence of complementarity between early and later interventions in the US, while Adhvaryu et al. (2024), Carneiro et al. (2022), Goff et al. (2025), Gunnsteinsson et al. (2022), Rossin-Slater & Wüst (2020) find evidence of substitutability.

the national standardized mathematics assessment (Pruebas Saber Matemáticas), learning outcomes remain a significant concern. Official statistics indicate that 47% of students scored at or below the minimal-knowledge threshold in standardized tests (Alcaldía de Manizales 2017).

Against this backdrop, the local Secretary of Education expressed a strong interest in identifying and testing effective strategies to address these persistent learning challenges. The two interventions evaluated in this study were developed and implemented through a collaboration between the Secretary of Education of Manizales and the local NGO Fundación Luker. Fundación Luker was responsible for implementing the program on the ground, with technical guidance and support from both local and international pedagogical experts.

Both interventions share a common philosophy that places children at the center of the learning process, rather than treating them as passive recipients of facts or procedures. They operationalize this approach through structured, play-based activities and concrete manipulatives that engage children's natural reasoning about quantities, patterns, and relationships. Building on children's intuitive, non-symbolic understanding of numeracy and explicitly connecting it to formal mathematical symbols and procedures, the programs target foundational skills that support later learning.

2.2 The Parental Engagement Intervention

In the summer preceding children's entry into first grade (between December 2021 and February 2022), we implemented *Aprender en Casa* (Learning at Home), a three-month parental engagement intervention designed to improve engaged caregiving and strengthen children's school readiness, with a particular focus on foundational mathematics skills. The program combined culturally appropriate learning materials and structured activities with ongoing behavioral support for primary caregivers delivered via weekly text messages.

Families received a workbook that included a set of structured play-based activities that caregivers could engage in with their child. Designed by a team of international and local experts, the activities were intended to enrich the quality and quantity of care and stimulation that caregivers provided to their child at home. The pedagogical approach emphasized playful interaction embedded in daily routines, using everyday household items – such as cups, plates, paper, scissors, and colored pencils – so that activities were low-cost and easy to implement (we describe program costs in Section 5.7).

An important goal of the workbook was the development of early numeracy skills through play. The activities leveraged children's intuitive, non-symbolic understanding of quantities, such as comparing magnitudes, estimating, and recognizing patterns, and were designed to help connect these abilities to formal symbolic representations, including numerals and basic arithmetic. For example, children practiced categorization and counting by sorting plastic bottle caps, engaged in one-to-one correspondence while setting the table (e.g., "How many plates do we need?"), and developed spatial reasoning through constructing towers with cups. These numeracy-focused activities were complemented by broader school readiness tasks, including early literacy skills, time management (e.g., building routines for the child), and fine and gross motor skills, offering an integrated approach to early learning at home. For example, several activities targeted fine motor skills through tracing or coloring within boundaries, supporting handwriting readiness. Although our evaluation did not directly measure these skills, they likely contributed to overall school readiness.

The workbook was distributed to all families in the study areas, including those assigned to the control group. In addition, families assigned to the treatment group received online support to facilitate engagement with the workbook activities through a series of weekly text messages. This approach draws on recent evidence demonstrating that text messaging can be an effective, low-cost tool to support parents in their caregiving role (Doss et al. 2019, Mayer et al. 2019, Kalil et al. 2025). The messages were designed to enhance parent–child interactions and complement the workbook activities. The messaging strategy closely follows the framework developed by York et al. (2019). Caregivers received three messages per week:

- A *FACT* message on Mondays, offering information on a specific aspect of child learning and development. These messages aimed to highlight the importance of building particular skills that contribute to numeracy and school readiness.

- A *TIP* message on Wednesdays, providing practical guidance on how to implement games or activities that reinforce the fact. These tips were designed to fit naturally into household routines and used everyday items or the workbook itself.

- A *GROWTH* message on Fridays, intended to encourage and motivate parents by reinforcing the week's content and recognizing their efforts.

Message content was adapted to align with the workbook's activities and to reflect the local culture and context. Appendix Figure A1 displays an example of three messages sent to caregivers in a particular week.

The rationale for this component of the program was to support parental engagement by addressing common barriers that often limit caregiver involvement in their child's learning. First, caregivers may lack clear information about the skills children are expected to develop before entering first grade. The *FACT* messages were designed to close this information gap by highlighting the importance of foundational skills. Second, even motivated parents may face behavioral obstacles, such as uncertainty about how to engage effectively, limited time, or low confidence, which the *TIP* messages addressed by offering simple, actionable guidance. Finally, the *GROWTH* messages aimed to sustain engagement and build parental self-efficacy through consistent encouragement and reinforcement. Because the workbook was distributed to all families, the intervention does not relax any material constraints parents may face. Instead, it targets informational and behavioral barriers to engaged parenting.

To maintain contact with families in both experimental arms, we also sent a small number of neutral, non-instructional text messages to families in both the treatment and control groups. These messages conveyed general information unrelated to parenting practices or learning activities (e.g., reminders about the start of the school year or holiday greetings). In total we scheduled 47 texts to families in the treatment group and 7 texts to families in the control group.

2.3 The Teacher Development Intervention

To support children's learning in school, we implemented a teacher professional development program aimed at improving pedagogical practice and fostering early mathematical competencies. The program equipped educators with practical tools and structured strategies to promote the development of foundational numeracy skills.

A defining feature of the program is its pedagogical approach. Unlike traditional methods that emphasize rote memorization and passive learning, the curriculum promotes an active pedagogy in which children co-construct mathematical understanding through hands-on exploration, guided interaction, and structured discussion with teachers and peers. Active teaching approaches of this kind have been shown to generate substantial learning gains across diverse contexts, including both high-income and low- and middle-income countries (Alan & Mumcu 2024, Ashraf et al. 2021, Granger et al. 2012).[4] Consistent with this philosophy, teachers are trained to use carefully sequenced lessons built around manipulative activities that encourage students' active engagement with mathematical concepts.

The pedagogical model builds on evidence that young children possess and develop intuitive, non-symbolic mathematical competencies – such as comparing quantities, recognizing patterns, and estimating magnitudes – long before formal instruction begins (Dehaene 2011). Each learning unit is constructed around a problem-based situation and incorporates a progression of concrete and semi-concrete manipulative activities designed to connect these informal, non-symbolic intuitions to formal symbolic representations. This approach allows children to develop meaningful links between intuitive reasoning and formal, symbolic mathematical representations. For example, teachers introduce number sequences using concrete, familiar objects and color patterns, enabling children to reason about quantities and patterns in an intuitive, hands-on way before transitioning to symbolic expressions such as numerals (e.g., 2, 4, 6, . . .).

To support effective implementation, teachers received a detailed instructional guide, and schools were equipped with a set of classroom materials designed specifically for early numeracy development. These resources were carefully adapted to the Colombian context through collaboration with local universities, the Colombian Academy of Sciences, the Ministry of Education, and Fundación Luker. Although these materials had been distributed widely across Colombia in previous years, their actual classroom use remained extremely limited due to the lack of adequate training and pedagogical support for teachers. Without clear guidance on implementation, many teachers were unable to integrate the materials into daily instruction, resulting in minimal adoption.

Our experimental design reflects this institutional reality. Both treatment and control schools received the same materials, but only teachers in treatment schools received training on how to

[4] Some implementations of active learning strategies have also found null (De Barros et al. 2024) or even negative results (Berlinski & Busso 2017) in low- and middle-income countries.

use them. This allows us to isolate the effect of the professional development program from the effect of simply providing additional physical resources, an important distinction given that materials alone have consistently been shown to have limited impact on learning without accompanying pedagogical change (Glewwe et al. 2009, 2004, Muralidharan 2017).

In line with best practices in the teacher professional development literature (Popova et al. 2022), the program combined several complementary components: six full-day, in-person training workshops (each lasting approximately six hours and spaced throughout the academic year), weekly in-person classroom coaching, and continuous online support. During the workshops, tutors introduced the instructional sequence, modeled classroom activities, and guided teachers through hands-on practice with the materials. Sessions included plenary discussions to review the thematic content, followed by small-group work where teachers deepened their mastery of the activities, including demonstrations using concrete mathematical tools such as base-ten blocks and the abacus. Across all components, training emphasized a carefully sequenced progression from concrete, manipulable experiences to abstract symbolic representations.

To further support teachers and ensure fidelity in the implementation of the program, tutors visited classrooms weekly to provide individualized, real-time feedback and to help teachers integrate the new practices into their everyday routines. Online on-demand support complemented these visits by offering continuous pedagogical guidance throughout the school year.

The teacher development intervention was offered to all first-grade teachers in treatment schools in the 2022 school year, beginning with the first in-person workshop in January. Teachers received the full sequence of six workshops and weekly coaching over the academic year. When students advanced to second grade in 2023, their second-grade teachers received training under similar conditions. As a result, children in schools randomized to the teacher-development program were assigned to the intervention over two consecutive school years.

3 Experimental Design

The teacher development intervention was randomized at the school level, while the parental engagement intervention was randomized at the individual level. The cross-randomization generates four mutually exclusive assignments: children who received only the parental engagement intervention (P), children who received only the teacher development intervention (T), children who received both interventions (P+T), and children who received neither intervention. Henceforth, we refer to this latter group as the control group of children (C).

3.1 School-level Randomization

In October 2021, towards the end of the school year, we invited public primary schools to participate in the study. School principals received official invitation letters from the Secretary of Education, informing them about the study and inviting them to take part in the following school year. Out of 60 schools that received an invitation letter, 50 agreed to participate. We refer to these as *study schools*. We randomly assigned half of the study schools to the teacher development program and half to continue with business-as-usual practices. Control schools received the program in later years. Randomization was stratified by school size. Following the recommendations in Athey & Imbens (2017), we formed strata of size four, ensuring that each stratum included at least two treatment and two control units.[5]

All first-grade teachers in treatment schools were offered the program in the 2022 school year (we discuss teacher-level take-up later in Section 5.1). No teacher in control schools was offered the program in that year. In the 2023 school year, when the students advanced to second grade, second-grade teachers in treatment schools were also offered the program (and no second-grade teachers were offered the program in control schools).

3.2 Individual-level Randomization

In November 2021, Fundación Luker staff attended end-of-year parent/teacher meetings in all study schools to recruit families of children attending *Transición* for the parental engagement intervention. During these meetings, staff members used the official administrative roster to distribute the workbooks to caregivers, collected their phone numbers, and their written consent to receive the text messages. A total of 1160 child caregivers signed up. Children were then randomly assigned to the parental engagement intervention, stratifying by school size and child gender, resulting in 601 children being randomized to the treatment group and 559 to the control

[5] One stratum included only two schools.

group.[6]

3.3 Treatment Groups

The cross-randomization of the school-level teacher development and the individual-level parental engagement intervention resulted in the following allocation of children: 311 children received only the parental engagement intervention (P), 290 children received only the teacher development intervention (T), 290 children received both interventions (P+T), and 269 children received neither intervention (C).

Caregivers were not informed about whether their child's school was assigned to the teacher development program. Likewise, neither first-grade nor second-grade teachers were informed about whether individual children in their classrooms were assigned to the parental engagement intervention.

By design, within each school we have both children randomized *in* and *out* of the parental engagement intervention. If the parental engagement program were to generate spillovers on untreated children within the same school, we would only identify a lower bound on the effect of parental engagement on children's outcomes. In Section 5.5.2, we use the fact that the individual-level randomization generated exogenous variation in the proportion of children assigned to the parental engagement intervention within a school to test for such spillovers.

4 Data

We assess the effects of the interventions using data on students' learning collected at schools as well as rich data from parents and teachers to monitor engagement with the interventions and their pedagogical practices at home and in school. We combine these data with administrative data from the Integrated Enrollment System (Sistema Integrado de Matricula, SIMAT), which is the national database for the registration of students in public education, and from the Sistema de Información Nacional de Educación Básica y Media (SINEB), which provides detailed

[6] Some of the caregivers present at the information sessions were not in the administrative roster and were randomized to the parental engagement intervention in a later batch. We include randomization batch fixed-effects in our analysis.

information on teachers and administrative staff in the official education sector.

4.1 Student Achievement

Learning assessments were administered in schools by trained enumerators, who were blind to students' treatment status and tested students individually using a tablet. For each task, enumerators prompted children to respond and recorded whether their answer was correct or incorrect. These assessments, part of routine data collection led by Fundación Luker and the Secretary of Education of Manizales, evaluated numeracy and literacy skills using adapted versions of the Early Grade Mathematics Assessment (EGMA) and the Early Grade Reading Assessment (EGRA) (RTI-International 2009). These instruments are widely used to monitor educational progress and were adapted here to include age-appropriate tasks that also captured higher-order reasoning skills. Similar assessments have been used to evaluate educational interventions in this setting (e.g., Alvarez Marinelli et al. 2024, Barrera-Osorio et al. 2020, Berlinski et al. 2023).

Assessments were conducted at the end of first grade, the beginning of second grade, and the end of second grade. The EGMA evaluates numeracy skills across a range of tasks of increasing difficulty, designed to measure progressively more advanced numerical skills. These tasks include number writing and reading, number comparison, addition and subtraction of one- and two-digit numbers, identification of numerical sequences, solving simple equations, and answering word problems framed in everyday language (e.g., "There are 7 children on a bus. If 2 get off and 1 gets on, how many are left?"). The EGRA evaluates literacy skills, including letter and sound recognition, reading of simple words and pseudo-words, passage reading, and comprehension.

Using children's answers to the EGMA and EGRA assessments, we construct summary indices of numeracy and literacy using principal component analysis (PCA) on the sub-tasks of EGMA (numeracy index) and EGRA (literacy index).[7] We use this approach as it provides an efficient measure of children's skills by giving greater weights to items that better capture variability in the underlying skills. We standardized these indices to have a mean of zero and a standard deviation of one for children in the control group. Panel A of Appendix Figure

[7]We estimate the PCA using children in the control group only in each survey wave. The resulting factor loadings are then used to predict index values for the full sample.

A2 plots the distribution of the raw numeracy scores, and Panel B plots the distribution of the numeracy index. These figures show that the test scores are well-distributed, with few students unable to answer any question or answering all questions correctly. As we show in Appendix Figure C1, our results are robust to using the simple sum of correct answers, instead of the index constructed using PCA.

Because both interventions specifically targeted numeracy skills, we focus on numeracy index as our primary outcome measure, while the literacy index is used to investigate potential spillover effects on non-targeted literacy outcomes. We further report disaggregated treatment impacts on different sub-scales of EGRA and EGMA. In particular, we report treatment effects separately for basic numeracy skills that children are expected to master by the end of first grade (such as being able to recognize and write one- and two-digit numbers), and higher-order mathematical skills (such as solving equations, sequences, and word problems).

4.2 Caregiver Interviews

We collect data from caregivers at the end of the parental engagement intervention, in February 2022, and eight months later, in October 2022. The survey collected information on caregiver-child interactions adapted from the HOME questionnaire (Bradley et al. 1988). We asked about the activities that caregivers performed with the child in the seven days preceding the interview, such as counting numbers, telling stories, or reading books. Interviews were conducted over the phone by trained enumerators, who were blind to caregivers' treatment status.

4.3 Teacher Interviews and Classroom Observation

We collected data on teachers' pedagogical practices and classroom behavior at the end of first and second grade using two complementary instruments. First, we conducted in-school interviews with teachers, in which we asked about their curricular knowledge and pedagogical practices, such as classroom and behavior management, learning expectations for students, and pedagogical content knowledge. To complement these self-reported measures, we carried out direct classroom observations using the Stallings Classroom Observation System (Stallings 1977). Trained enumerators visited study schools and spent approximately one hour observing from the back of the classroom without disrupting instruction. During these visits, enumerators captured a series of structured "snapshots" to document classroom dynamics, including the level of student engagement, their use of learning materials, the prevalence of off-task and disruptive behavior, the types of instructional resources used by both teachers and students, and pedagogical approaches adopted by the teachers.

4.4 Administrative Data

We complement these survey data with administrative records from SIMAT and SINEB. SIMAT provides student-level information on gender, date of birth, and household income, which we use as a proxy for socio-economic status (SES).[8] SIMAT also includes school-level information, such as school location (urban or rural), student enrollment and class size. SINEB compiles administrative records on teachers within Anexo3A and provides individual-level information on their demographic and professional characteristics, including date of birth (from which we construct age), gender, and educational attainment. Finally, we draw on a set of measures from *Predictores de Aprendizaje* (Predictors of Learning), an assessment developed by local education experts from the Department of Caldas and implemented by the Regional Family Compensation Fund (CONFA) to evaluate school readiness at the end of the Transición grade. We report additional details of this assessment in Appendix B.

4.5 Timeline

Appendix Figure A3 provides an overview of the study timeline and data collection activities. The parental engagement intervention was implemented between December 2021 and February 2022, during the summer break preceding children's enrollment in first grade. During this period, caregivers assigned to the intervention received the text messages. The teacher professional development intervention began in January 2022 and continued throughout the 2022 school year, targeting first grade teachers. In 2023, the program was extended to include second-

[8]Household income is recorded in six categories: 1 (very low income), 2 (low income), 3 (medium-low income), 4 (middle income), 5 (middle-high income), and 6 (high income). Using this variable, we classify families in high- vs low-SES, where the latter group includes families in the very low- or low-income categories. We use this variable in the heterogeneity analysis below.

grade teachers in treatment schools. As a result, children in treatment schools were assigned to the teacher development for two consecutive school years. Student achievement data were collected at three points in time: at the end of first grade, and at both the beginning and end of second grade. Data from parents were collected twice, immediately following the parental engagement intervention and again eight months later. Teacher interviews and direct classroom observations took place in October 2022 and 2023.

5 Results

This section first assesses the integrity of the experiment, then reports the intention-to-treat (ITT) estimates for numeracy, explores the potential mechanisms at play, and concludes with a discussion of the results and a cost-effectiveness analysis.

5.1 Integrity of the Experiment

Baseline Balance. We assess the internal validity of our experiment by examining whether pre-treatment characteristics of students and schools are balanced across experimental groups in Appendix Table A1. Panel A shows that, on average, children in the control group are 5.13 years old, 45 percent are female, and are on average from a low-income household. Students in the treatment groups are similar to children in the control group in terms of demographic characteristics. Moreover, there are no differences across groups in the cognitive, language, motor, and socio-emotional scores at the end of *Transición*, prior to the start of the two interventions.[9] Panel B examines whether the identity of the caregiver is balanced across experimental groups. The caregiver providing the phone number – most often the mother (73%) and occasionally the grandmother (8%) – served as the primary adult involved in the parental engagement intervention. The distribution of caregiver identities is well balanced between families who received text messages and those who did not, supporting the internal validity of the comparison. Finally, Panel C shows no systematic differences in school-level characteristics across the four experimental arms. Schools are mostly urban (91 %) with an average of 3 classes and 23 students per

[9] We describe the assessment used at the end of *Transición* in Section 4.

class. Overall, the evidence in Appendix Table A1 confirms that randomization successfully generated treatment and control groups that are comparable in terms of pre-treatment characteristics.[10]

Attrition. The student response rates were 71% in the control group and did not differ across experimental groups (Appendix Table A2). Appendix Table A3 shows that baseline characteristics remain balanced among non-attriting students. Caregiver response rates were approximately 71% and were comparable across treatment and control groups (Appendix Table A4). Teacher response rates were around 81% and did not vary across experimental conditions (Appendix Table A5). To assess sensitivity to missing outcome data, we re-estimate the main effects using inverse-probability weights (IPW) and compute Lee (2009) bounds. Both approaches yield estimates that remain positive and close to our preferred specification (Appendix Figure C1 and Appendix Table C1).

Compliance. Appendix Figure A4 documents compliance with the parental engagement intervention. As described above, to maintain contact with families regardless of assignment to the parental engagement intervention, we sent a small number of neutral, non-instructional text messages to all families. Families assigned to the parental engagement intervention received nearly the full set of program messages, while families randomized out of the parental engagement intervention received only these 7 neutral messages.[11] Appendix Figure A5 reports take-up of the teacher professional development program. Teachers in control schools did not attend any workshops, as expected by design. Participation among treated teachers was substantial: around 40% attended at least five workshops, and around 25% attended at least six workshops.

[10] Out of the 105 individual comparisons, none is statistically significant at the 1% level. Additionally, we perform 14 Wald joint tests, of which only the test for the number of classes is statistically significant at the 1% level.

[11] In 12 families (24 children), the caregiver received 54 messages – 47 from the treatment group and 7 from the control – because the children were twins and shared a single phone number. Our results remain robust when these children are excluded from the analysis (Appendix Figure C1).

5.2 Students' Achievement

We estimate Intention-to-Treat (ITT) effects on students' learning outcomes, using the following specification:

$$y_{i,s,b,t} = \alpha + \beta_1 T_i^P + \beta_2 T_i^T + \beta_3 T_i^{P+T} + X'_{i,s,t}\gamma + \delta_b + \omega_t + \epsilon_{i,s,b,t} \tag{1}$$

where $y_{i,s,b,t}$ is the outcome of student i, in school s and stratum b, measured in survey wave t. $X_{i,s,t}$ is a vector of enumerator-week fixed effects, δ_b are strata fixed effects, and ω_t are survey-wave fixed effects. The indicators T_i^P, T_i^T, and T_i^{P+T} denote assignment to the parental engagement intervention, the teacher development intervention, and the combined intervention, respectively.[12] The reference group includes children assigned to neither intervention. We cluster standard errors at the school level.

To benchmark the magnitude of the effects, we report the program impacts in terms of standard deviation (SD) units of the outcome variable in the control group. For outcomes corresponding to sub-indices, we also report Romano-Wolf step-down p-values to account for multiple hypothesis testing. Our main specification pools observations from the three survey waves. We also report heterogeneity in treatment impacts over time, as well as by child's baseline achievement (below or above the sample median of *Predictores de Aprendizaje*), child gender, household socio-economic status (low vs. high SES) and teacher characteristics.

Table 1 reports the treatment effect on children's numeracy skills. Column 1 presents the result for the summary index. Columns 2–4 disaggregate the results by skill level, moving from foundational to more advanced competencies. Specifically, Column 2 focuses on skills that children are expected to master by the end of first grade, such as reading, recognizing and writing one- and two-digit numbers. Column 3 reports effects on more advanced skills, including addition and subtraction of one- and two-digit numbers. Column 4 captures the highest-level skills in our assessment, such as identifying numbers that complete a sequence, solving simple equations, and problem solving. At the bottom of the table, we report p-values for the equality

[12]Equation (1) is observationally equivalent to a factorial specification with separate indicators for assignment to the parental intervention and the teacher intervention, plus their interaction. In the factorial specification, the interaction term equals the combined-treatment effect minus the sum of the two standalone effects.

Table 1: Intention-to-Treat: Students' Numeracy Skills

	Index	**Number write, read, compare**	**Add & subtract**	**Sequence, equations, Problems**
	(1)	(2)	(3)	(4)
Parental engagement	0.174***	0.145**	0.116	0.211***
	(0.064)	(0.062)	(0.076)	(0.057)
		[0.061]	[0.164]	[0.005]
Teacher development	0.203**	0.132	0.156*	0.246**
	(0.093)	(0.083)	(0.084)	(0.102)
		[0.164]	[0.162]	[0.052]
Combined treatment	0.175**	0.051	0.183**	0.206**
	(0.082)	(0.084)	(0.072)	(0.079)
		[0.538]	[0.033]	[0.033]
Observations	2485	2485	2485	2485
P-value: $T^P = T^T$	0.741	0.854	0.676	0.719
P-value: $T^P = T^{P+T}$	0.990	0.240	0.453	0.955
P-value: $T^T = T^{P+T}$	0.782	0.162	0.800	0.760
Share correct (control)	0.474	0.874	0.300	0.259

Notes: The table presents ITT effects on children's numeracy skills. All outcomes are standardized to have mean 0 and standard deviation 1 in the control group. Column 1 reports the effects on the numeracy index described in Section 4.1. Columns 2–4 report disaggregated treatment impacts on the outcomes indicated in the column headers; square brackets below the point estimates show Romano-Wolf step-down p-values. The p-values at the bottom of the table test whether coefficients are equal across experimental groups (parental engagement T^P, teacher development T^T, and combined T^{P+T}). Controls include survey wave, survey week, enumerator, and randomization strata fixed effects. Share correct (control) shows the proportion of correct answers on the task for children in the control group. Standard errors clustered at the school level are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

of treatment impacts across experimental groups, and the share of items correctly solved by children in the control group.

Column 1 shows that each of the two interventions in isolation improves children's numeracy skills compared to the control group. The point estimates imply that the numeracy scores of children who received only the parental engagement intervention are 0.174 SD higher compared to children in the control group (p-value<0.01). Similarly, children assigned only to the teacher development improved their test scores by 0.203 SD compared to the control group (p-value<0.05). The point estimates are similar in magnitude and not statistically different from each other (the p-value for the difference is 0.741).

We also find that the numeracy skills of children who received both teacher development and parental engagement intervention improved by 0.175 SD compared to children in the control group. The effects of the combined intervention are statistically indistinguishable from those of parental engagement alone (p-value = 0.990) and from the teacher development alone (p-value = 0.782), as indicated by the p-values reported at the bottom of the table. These results suggest that the effects of the two programs are largely *substitutes* rather than *complements*, as receiving both interventions does not yield additional gains beyond those achieved by either program individually.

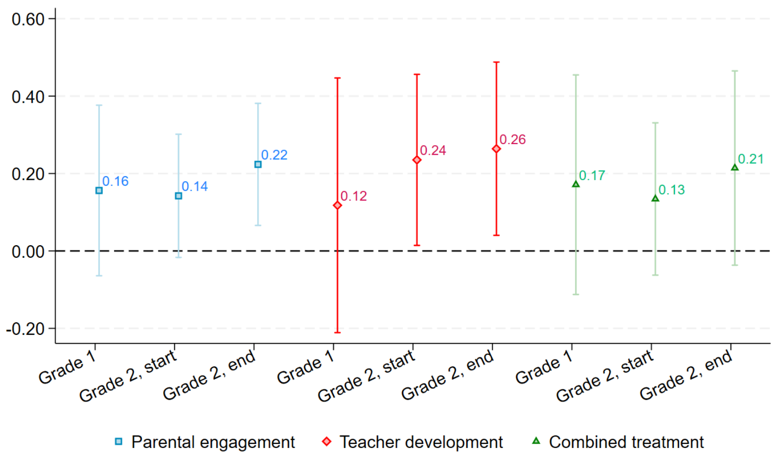
Unpacking the treatment impacts on the numeracy index in its different components, we find smaller effects on basic numeracy skills and larger effects on more advanced skills. When considering children's mastery of simple skills, such as writing, reading and comparing numbers (Column 2), the parental engagement intervention improves children's outcomes by 0.145 SD and the teacher development yields a similar effect of 0.132 SD. However, children in the control group were already performing well in these areas, correctly answering 87% of the relevant EGMA items, leaving limited room for improvement (this is shown in the bottom row of Table 1). In contrast, the largest treatment impacts are found in the most advanced skills that include completing numerical sequences, solving simple equations, and solving word problems (Column 4). In these tasks, children in the control group could correctly solve only 26% of test items on average, indicating significant scope for growth. On these dimensions, the parental engagement intervention improves achievement by 0.211 SD, the teacher development by 0.246 SD, and the combined intervention by 0.206 SD relative to children from the control group.

Our main findings are robust to a wide range of alternative specifications and diagnostic checks (see Appendix C). First, results are robust to using different sets of control variables, alternative constructions of the numeracy outcome, inverse probability-weighted (IPW) estimates for attrition, and when excluding potentially problematic observations such as twins assigned to both treatment and control messages (Appendix Figure C1). Second, to further address potential bias from differential attrition, we compute Lee (2009) bounds, which remain positive and close to the main estimates across treatment arms (Appendix Table C1). Third, randomization inference confirms that the estimated effects are not driven by parametric assumptions about the error distribution (Appendix Table C2). Fourth, we show that estimates are not affected by contamination bias arising from multiple treatment arms, with partially linear estimates closely aligned with alternative estimators (Appendix Table C3). Finally, we find no evidence that the interventions affected contemporaneous teachers' characteristics (Appendix Table C4) and school mobility in a way that could bias treatment assignment (Appendix Table C5).

5.3 Heterogeneity

We investigate heterogeneity in treatment impacts along several dimensions. First, Figure 1 presents treatment effects by survey wave. The effects of the parental engagement intervention are remarkably stable over time, with point estimates of roughly 0.15 SD at the end of first grade and at both the beginning and end of second grade. The impacts of the teacher development show a modest upward trend, consistent with increasing exposure: the estimated effects rise from 0.12 SD at the end of first grade to 0.24 SD at the beginning of second grade and 0.26 SD at the end of second grade. We observe a similar pattern in treatment effects for children receiving both programs.

Figure 1: Intention-to-Treat: Students' Numeracy Skills by Survey Wave



Notes: The figure presents heterogeneity in treatment effects on children's numeracy skills by study wave and their corresponding 95% level confidence intervals. The outcome variable is the index described in Section 4.1.

Second, Appendix Figure A6 explores heterogeneity in treatment effects by student characteristics in Panels A-C (baseline achievement, gender, and SES) and teacher characteristics Panels D-E (age and education). While differences across subgroups are not statistically distinguishable at conventional levels, the point estimates suggest some variation.

Panel A of Appendix Figure A6 shows that estimates are broadly similar for children below and above the median baseline score across all interventions, suggesting that the interventions benefit students across the achievement spectrum. Panel B suggests that the effects are larger for girls than for boys, particularly for the parental engagement and teacher development interventions. While confidence intervals overlap, this pattern is consistent with prior evidence documenting gender differences in responses to early childhood interventions (Evans et al. 2024).[13] Panel C shows broadly similar SES effects for the parental engagement intervention, whereas point estimates for the teacher development and combined interventions are larger for lower-SES children, though not significantly so. Panels D-E document that treatment effects do not vary substantially by observable teacher characteristics. In particular, the impact of the teacher development intervention is similar across teacher age and education levels.

Finally, Appendix Figure A7 presents quantile treatment effects, showing that the interventions generate broadly similar gains across the achievement distribution. This confirms that the programs benefit students at all points of the performance spectrum, without disproportionately favoring either lower- or higher-achieving children.

5.4 Mechanisms

5.4.1 Caregivers' Practices

To understand the mechanisms behind the effects of the parental engagement intervention, we examine its impact on self-reported caregiver–child interactions. In Table 2, we regress parents' behaviors on an indicator for whether the child was assigned to receive the parental engagement (either alone or combined with the teacher development), estimating:

$$y_{i,s,b,t} = \alpha + \beta_1 T_i^{P\ or\ P+T} + \omega_t + \delta_b + \epsilon_{i,s,b,t} \tag{2}$$

where $y_{i,s,b,t}$ is the outcome of parent i, in school s, stratum b, measured at time t (February or October), ω_t is a vector of survey-wave fixed effects, and $T_i^{P\ or\ P+T}$ is an indicator variable taking a value of one if parent i's child is assigned to the parental engagement program (with or without the teacher professional development program). The reference group therefore includes children in the control group and those assigned to the teacher development only.

Column 1 of Table 2 shows that the parental engagement intervention increases an overall

[13] Although findings are mixed overall (Magnuson et al. 2016), several studies show that girls benefit more from programs that expand cognitively stimulating inputs, particularly when they face less stimulating home environments (García et al. 2018).

Table 2: Intention-to-Treat: Caregiver-Child Interactions

	Index (1)	Count (2)	Read (3)	Story (4)
Parental engagement	0.104**	0.075	0.099**	0.075
	(0.049)	(0.048)	(0.049)	(0.049)
		[0.226]	[0.102]	[0.226]
Observations	1628	1628	1628	1628
Share (control)	0.638	0.684	0.676	0.559

Notes: The table presents ITT effects of the parental engagement intervention (alone or combined with the teacher professional development program) on caregiver–child interactions. The reference group consists of children in the control group and those assigned to the teacher development only. We use responses from the caregiver interviews from February and October 2022. Column 1 reports treatment effects on a summary index of parent–child interactions, constructed from the outcomes reported in columns 2–4; all outcomes are expressed in standard deviation units relative to the control group. Square brackets below the point estimates show Romano–Wolf step-down p-values. Controls include survey wave, and randomization strata fixed effects. Robust standard errors are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

index of parent-child interactions by 0.104 SD. Examining the individual components of the index in Columns 2-4, we find that treated caregivers are 0.075 SD more engaged in counting with their child (though this estimate is not statistically significant), 0.099 SD more engaged in reading, and 0.075 SD more engaged in storytelling.

Appendix Figure A8 plots the ITT estimates on the parent–child interaction index by caregiver survey wave (February and October 2022). Although the differences across waves are not statistically significant, the point estimates suggest that interactions increase during the intervention period (0.15 SD in February, immediately after program completion) but decline by the October follow-up (0.06 SD). This trajectory is consistent with a short-run boost in parent–child engagement while caregivers receive weekly messages, followed by a reversion toward baseline levels once the behavioral prompts are removed.

5.4.2 Teachers' Practices

We also examine whether the teacher development program translated into changes in classroom practices and teachers' behavior, using evidence from teacher interviews and in-class observations. Table 3 presents estimates from regressions of teacher outcomes on an indicator for whether the teacher's school was assigned to the teacher development program. Specifically, we use teacher-level data to estimate the following model:

$$y_{i,s,b,t} = \alpha + \beta_2 T_i^T + \omega_t + \delta_b + \epsilon_{i,s,b,t} \tag{3}$$

Table 3: Intention-to-Treat: Teachers' Practices

	Index (1)	Active class score (2)	Class preparation delivery (3)	Students' participation (4)	Class disruption (5)	Learning expectations (6)	Pedagogy knowledge (7)
Teacher development	1.082***	0.724***	0.215	0.176	-0.375**	0.556***	1.098***
	(0.178)	(0.195)	(0.130)	(0.171)	(0.156)	(0.139)	(0.213)
		[0.004]	[0.178]	[0.270]	[0.066]	[0.004]	[0.002]
Observations	118	118	118	118	118	118	118
Control group mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: The table presents ITT effects of the teacher development program on teachers' practices. The reference group consists of teachers in schools that did not receive the teacher professional development program. We use responses from the teacher survey and in-classroom observation from 2022 and 2023. Column 1 reports treatment effects on a summary index of teachers' practices, constructed from the outcome reported in columns 2-7, square brackets below the point estimates show Romano-Wolf step-down p-values. The exact components of each of the outcome variables in columns 2-7 are reported in Appendix Table A6. All variables have been standardized to have a mean of zero and standard deviation of one in control schools. Controls include year and randomization strata fixed effects. Standard errors clustered at the school level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

where $y_{i,s,b,t}$ is the outcome of teacher i, in school s, strata b, measured at wave t (end of first and second grade), ω_t is a vector of survey-wave fixed effects, and T_i^T is an indicator variable taking the value one if teacher i is in a treatment school (note that this is distinct from the variable T_i^T in equation (1) at the student level).

We find that the program changes teachers' practices. Column 1 reports a 1.082 SD improvement on a summary index of teachers' behavior, constructed using the information in Columns 2-7. The remaining columns decompose this effect into specific dimensions of practice (Appendix Table A6 reports the exact definition of each variable). Treated teachers lead more active and participatory lessons, with increased student–teacher interactions, as measured by the Stallings classroom observation tool (Column 2). Their instruction is also more structured: they are more likely to present lesson objectives at the start and summarize key points at the end (Column 3). These changes translate into higher student engagement and participation (Column 4) and fewer classroom distractions (Column 5). Finally, treated teachers are more likely to hold learning expectations aligned with the national curriculum (Column 6) and to demonstrate stronger understanding of active learning pedagogy (Column 7), reflecting the core principles emphasized in the professional development program. Appendix Figure A9 further shows that these impacts are persistent across survey waves, with ITT estimates on the teacher practice index remaining large (1.04 SD vs. 1.13 SD).

5.5 Spillovers

5.5.1 Spillover on Non-Target Subjects

We use data from the EGRA assessment to examine potential spillover effects on non-target subjects, specifically early literacy skills. Although the interventions were primarily designed to strengthen foundational numeracy, both programs included broader components that could plausibly enhance learning across domains. For example, the parental engagement program also aimed to build caregivers' confidence, motivation, and consistency in supporting their child's learning. Similarly, the teacher development intervention promoted active, student-centered pedagogical practices and better classroom management, which may benefit learning across multiple domains. These improvements in the home and classroom environments could generate meaningful spillovers onto literacy skills, even in the absence of direct instruction.

Column 1 of Appendix Table A7 shows the effects of the interventions on the overall literacy index. Both the parental engagement intervention (0.181 SD) and the teacher development intervention (0.146 SD) are associated with modest positive gains relative to the control group, although only the parental engagement estimate is statistically significant at the 10 percent level. By contrast, the estimate for the combined intervention is smaller (0.036 SD) and imprecisely estimated. The two standalone interventions are not statistically different from one another, while the combined intervention does not show clear evidence of additional gains in literacy. Overall, these results suggest modest spillovers from each program onto early literacy, but no evidence of additive benefits when both interventions are delivered together. Because literacy is a secondary, non-targeted outcome, these estimates should be interpreted cautiously. The study was powered to detect effects on numeracy rather than cross-domain spillovers, so the literacy estimates are less precise (with wider confidence intervals) and more sensitive to sampling variation.

5.5.2 Spillovers of the Parental Engagement Intervention on Untreated Students

Previous research in this context has documented that interventions can generate spillover effects among peers (Berlinski et al. 2023). This raises the possibility that children who were not directly assigned to the parental engagement program could still benefit indirectly from exposure to treated peers in their schools.

To investigate this, we exploit the within-school, individual-level randomization of the parental engagement intervention, which created exogenous variation in the proportion of schoolmates receiving the program. In Appendix Table A8, we restrict the sample to students who did not receive the parental engagement intervention and regress their numeracy scores on the within-school share of treated students receiving parental engagement (we include the main effect for the teacher development treatment).

The estimates are imprecise and not statistically significant from zero. The point estimate suggests that a full increase in peer exposure (from 0 to 100 percent of treated schoolmates) would reduce numeracy scores for non-treated students by 0.28 SD. However, observed variation in peers' exposure is substantially smaller, with an average within-school treatment share of around 50 percent.

5.6 Complementarity and Substitutability

Both the parental engagement and teacher development interventions independently improve children's numeracy skills, with effect sizes of approximately 0.17 and 0.20 SD, respectively. The combined intervention yields a 0.17 SD improvement – no larger than the effects of either program alone – suggesting that, in this setting, the two interventions operate as substitutes rather than complements.

When interpreting this result, it is important to distinguish between reduced-form policy impacts and structural parameters of the skill production function. The effects we estimate capture the overall impact of the interventions as implemented, including any behavioral responses by parents, teachers, or children themselves (Todd & Wolpin 2003). Limited complementarity in policy parameters does not rule out complementarities in the underlying structural technology of skill formation.

In our context, behavioral responses may attenuate complementarities in reduced form. For example, parents may reduce engagement when classroom instruction improves; teachers may reallocate attention toward lower-achieving students if some children receive additional support at home; or children who arrive better prepared may require fewer instructional inputs. We explore these potential behavioral mechanisms using the available data.

To assess whether parents crowd out investment in response to improvements in school quality, we return to the treatment effects on parent–child interactions studied in Section 5.4.1. We regress our measures of parental engagement on indicators for assignment to the parental engagement intervention only, the teacher development intervention only, or both, with the control group as the omitted category.

The results, reported in Appendix Table A9, provide no evidence of crowding out. Assignment to the teacher development intervention alone has no negative effect on parental engagement across any outcome, with coefficients close to zero throughout. The parental engagement intervention implemented in isolation yields positive effects across all measures of parent–child interactions, though these estimates are not statistically distinguishable from zero. In contrast, the combined treatment generates larger increases in parental engagement than either intervention implemented alone, with statistically significant effects on the summary index and on shared reading activities. Taken together, these patterns suggest that the reduced-form substitutability observed in child outcomes is unlikely to be driven by parental disengagement in response to improvements in classroom instruction.

Similarly, we explore whether teachers adapt their practices when some children receive additional support at home, by analyzing how the exogenous share of students assigned to the parental engagement and the teacher development treatment affects teachers' practices (Section 5.4.2). The results reported in Appendix Table A10 provide no evidence that greater exposure to exogenously treated parental engagement within a school affects teachers' classroom practices. Estimated coefficients are statistically insignificant across all outcomes, including the overall index and each individual domain of teaching practice. While most estimates are positive, they are imprecisely estimated and economically modest. In addition, these spillover coefficients are smaller than the direct effects reported in Table 3, particularly considering that the relevant variation in peer exposure is generated by only 50% of students being assigned to treatment within school.

A further explanation for the lack of additive effects is that the two interventions address largely overlapping pedagogical constraints. Both programs rely on structured, activity-based numeracy instruction that builds on children's intuitive understanding of quantities and connects it to formal mathematical concepts through guided practice. Because the parental and teacher interventions operate through similar instructional mechanisms – active engagement, structured routines, and a progression from concrete to symbolic learning – exposure to one may reduce the marginal gains from the other over the period we study. More generally, these patterns are consistent with diminishing returns to intensive investments in similar pedagogical inputs within a given time frame, such that once core instructional constraints are relaxed, additional exposure along the same dimension yields smaller incremental benefits.

Finally, dynamic complementarities in skill formation, as formalized by Cunha & Heckman (2007), Cunha et al. (2010), and Agostinelli & Wiswall (2025), may be more relevant for broader skill domains or longer developmental horizons than those examined here. Our interventions target early foundational numeracy over a relatively short period spanning pre-primary and early primary grades. In this setting, a single high-quality input may be sufficient to relax binding constraints and generate sizable gains, leaving limited scope for additional interventions to produce compounding effects within the same skill domain and time frame. Examining whether complementarities emerge across distinct skill domains or over longer horizons remains an important direction for future research.

5.7 Cost-Effectiveness Analysis

We conduct a cost-effectiveness analysis to compare the parental engagement and teacher professional development interventions from the perspective of a policymaker or program implementer. To compute program costs, we follow the framework introduced in Dhaliwal et al. (2012), which provides a standardized approach to measuring and comparing the cost-effectiveness of educational interventions. Our objective is not to estimate the full social returns to these programs, but rather to provide a transparent benchmark on the relative cost of generating improvements in children's numeracy skills across interventions. In line with this policy-oriented perspective, we measure the total implementation cost of each package, including the cost of the common pedagogical materials. Because these common materials were also provided in the control group, the resulting estimates should be interpreted as policy rollout cost per unit of impact, rather than as incremental experimental cost-effectiveness relative to the study's actual control condition.

Program Costs. For the parental engagement program, costs fall into two main categories: (i) program administration and staff, including a tutor supporting parents and a coordinator overseeing implementation; and (ii) implementation and materials, including workbook production and distribution and the delivery of SMS messages to caregivers. Total program costs amount to USD 20 per treated student.

For the teacher professional development program, costs include: (i) administration and staff, (ii) educational materials, and (iii) training. Administration and staff costs comprise tutor salaries, tutor training, tutor materials, and overall program administration. Educational materials include the production and distribution of teachers' and students' workbooks and manipulatives, incurred during the implementation years. Training costs include payments for replacement teachers during workshop participation and the per-teacher cost of each workshop, with each treated teacher attending six workshops over the course of the program. Total costs for the teacher development intervention amount to approximately USD 78 per treated student.

Cost-Effectiveness. To assess cost-effectiveness, we combine these cost estimates with the ITT effects on children's numeracy skills reported in Table 1. We compute the improvement in the numeracy index per USD 100 spent under each intervention. Based on these estimates, an expenditure of USD 100 on the parental engagement intervention increases children's numeracy skills by approximately 0.87 standard deviations. The corresponding improvement is 0.26 standard deviations for the teacher professional development program. For the combined intervention, the estimated gain is 0.18 standard deviations per USD 100 spent.

These results imply that over the time horizon studied, the combined intervention is dominated in cost-effectiveness terms: it delivers smaller gains per dollar than either the parental engagement intervention or the teacher professional development program in isolation. This finding mirrors the reduced-form evidence on learning outcomes, which suggests limited complementarity between the two programs.[14]

[14] As highlighted in Dhaliwal et al. (2012), comparative cost-effectiveness analyses should account for sampling

Amortizing Teacher Costs Across Multiple Cohorts. While the parental engagement intervention delivers benefits solely to the treated cohort, teacher professional development represents a one-time investment with potential spillovers to future cohorts taught by the same teachers. To capture this, we amortize the upfront training costs over an assumed number of subsequent student cohorts. Following standard guidance from the World Bank and International Monetary Fund (The World Bank 2013), we discount future benefits at 5% per year and account for potential depreciation in teachers' acquired skills over time.

Let B denote the ITT gain in numeracy per student for a single cohort taught by a trained teacher, then the present value (PV) of the cumulative numeracy gains per student from a single trained teacher is:

$$\text{PV} = \sum_{t=1}^{T} \frac{(1-\delta)^{t-1}\, B}{(1+0.05)^{t-1}}$$

where T is the number of future cohorts and δ is the annual depreciation rate of teacher skills. Considering $T = 5$ cohorts and $\delta = 0.05$, and using $B = 0.203$ based on the ITT estimates reported in Table 1, we get a PV of 0.839. The results are robust to alternative assumptions on the depreciation of teachers' skills and the discount rate, with the present value of cumulative numeracy gains ranging from 0.71 to 0.98 across specifications (Appendix Table A11). Our calculations assume a conservative horizon of $T = 5$ cohorts; a longer effective teaching horizon would raise the present value of the benefits.[15]

To relate this back to program costs, the teacher professional development program requires an investment of USD 78 per student in the first cohort. For the initial cohort alone, the teacher development program generates 0.26 SD per USD 100 spent, compared with 0.87 SD for the parental engagement program. When we account for benefits to future cohorts by amortizing teacher training costs, the present value of cumulative gains is 0.84 SD for the original USD 78 investment, equivalent to approximately 1.08 SD per USD 100 spent. Under these assumptions,

variability: point estimates of program impact may differ but not be statistically distinguishable. In our case, the numeracy gains of the parental engagement (0.174 SD), teacher development (0.203 SD), and combined (0.175 SD) interventions are not statistically different from one another. Importantly, even if we treat the impacts as identical across interventions, the combined intervention remains dominated because it delivers smaller gains per dollar spent than either program implemented alone.

[15] This should be interpreted as a conservative benchmark. On the one hand, we do not account for possible sibling spillovers, which could increase the total benefits of the teacher development and parental engagement. On the other hand, there may be additional recurrent costs related to teacher development that are not incorporated in this calculation.

the teacher development program becomes at least as cost-effective as, and potentially more cost-effective than, the parental engagement intervention.

6 Conclusion

We study the relative effectiveness and cost-effectiveness of family- and school-based learning interventions using a randomized controlled trial in Colombia that follows children as they transition from pre-primary to primary school. We randomly assign children to a parental engagement program, a teacher professional development program, both, or a control group. Both interventions share a child-centered pedagogical philosophy that places children at the center of the learning process, promotes active exploration, and uses structured, play-based learning activities and concrete manipulatives to connect children's intuitive understanding of quantities, patterns, and relationships to formal mathematical concepts.

The parental engagement program, implemented during the summer preceding first grade, encouraged caregivers to engage children in simple, developmentally appropriate activities embedded in daily routines, focusing on numeracy while also supporting broader early cognitive skills. The teacher professional development program, delivered during first and second grade, combined in-person workshops, weekly classroom coaching, and ongoing remote support to strengthen teachers' pedagogical skills and improve mathematics instruction.

Both programs independently improve early numeracy skills measured via standardized tests: children in the parental engagement program score 0.17 SD higher on a summary numeracy index, while children in the teacher professional development program improve by 0.20 SD. The gains are greatest for more advanced skills, such as solving sequences, equations, and word problems, where the performance of the control group is much lower than in the easier numeracy tasks.

Consistent with the programs' pedagogical philosophy, we observe substantial changes in the behaviors of both caregivers and teachers. Caregivers in the parental engagement program increase learning-oriented interactions at home, while teachers assigned to professional development adopt more active and structured teaching methods, increase student participation, reduce classroom distractions, and better align instruction with child-centered pedagogy.

Combining the two interventions yields no additional learning gains compared to receiving either of the two programs in isolation, suggesting that, over the short term and within the numeracy domain we study, the programs function as substitutes. This pattern is consistent with both interventions targeting largely overlapping pedagogical constraints through similar instructional mechanisms – structured and activity-based practice with a progression from concrete to symbolic math representations. Once this constraint is relaxed by exposure to one program, additional instruction along the same pedagogical dimension appears to generate diminishing marginal returns within the time frame we examine.

In terms of cost-effectiveness, the parental engagement program yields larger gains per dollar in the short run, reflecting its low per-child cost, while the teacher professional development program becomes at least as cost-effective as the parental engagement program, and may be more cost-effective once fixed costs are spread over time. The absence of additive effects from the combined intervention, together with its higher per-child cost, implies that it is less cost-effective than either program alone, suggesting that strategically concentrating resources on a single binding constraint – at home or in school – maximizes short-run learning gains per dollar spent.

While our results are internally valid and robust within the context we study, they do not rule out potential complementarities between earlier and later interventions, as emphasized by theoretical models of early skill formation (Cunha & Heckman 2007). Such complementarities may arise across broader skill domains, over longer developmental horizons, or through interactions between diverse interventions and inputs, including educational, health, or social programs. Understanding when interventions act as complements or substitutes is an important avenue for future research, with direct implications for how public funds should be allocated to programs that most effectively promote human capital accumulation.

References

Adhvaryu, A., Molina, T., Nyshadham, A. & Tamayo, J. (2024), 'Helping Children Catch Up: Early Life Shocks and the Progresa Experiment', *The Economic Journal* **134**(657), 1–22.

Agostinelli, F. & Wiswall, M. (2025), 'Estimating the Technology of Children's Skill Formation', *Journal of Political Economy* **133**(3), 846–887.

Alan, S. & Mumcu, I. (2024), 'Nurturing Childhood Curiosity to Enhance Learning: Evidence from a Randomized Pedagogical Intervention', *American Economic Review* **114**(4), 1173–1210.

Alcaldía de Manizales (2017), 'Balance Educativo: Resultados de Las Pruebas Saber 2016'. Documento oficial.
URL: *https://manizales.gov.co/RecursosAlcaldia/201904092249175840.pdf*

Almond, D., Currie, J. & Duque, V. (2018), 'Childhood Circumstances and Adult Outcomes: Act II', *Journal of Economic Literature* **56**(4), 1360–1446.

Alvarez Marinelli, H., Berlinski, S. & Busso, M. (2024), 'Remedial Education: Evidence from a Sequence of Experiments in Colombia', *Journal of Human Resources* **59**(1), 141–174.

Angrist, J. D. & Lavy, V. (2001), 'Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools', *Journal of Labor Economics* **19**(2), 343–369.

Angrist, N., Djankov, S., Goldberg, P. K. & Patrinos, H. A. (2021), 'Measuring Human Capital Using Global Learning Data', *Nature* **592**(7854), 403–408.

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y. & Schady, N. (2016), 'Teacher Quality and Learning Outcomes in Kindergarten', *The Quarterly Journal of Economics* **131**(3), 1415–1453.

Ashraf, N., Banerjee, A. & Nourani, V. (2021), 'Learning to Teach by Learning to Learn', *Ms. University of Chicago and Makerere University* .

Athey, S. & Imbens, G. W. (2017), The Econometrics of Randomized Experiments, *in* 'Handbook of Economic Field Experiments', Vol. 1, Elsevier, pp. 73–140.

Banerjee, A. V., Bhattacharjee, S., Chattopadhyay, R., Duflo, E., Ganimian, A. J., Rajah, K. & Spelke, E. S. (2025), 'Children's Arithmetic Skills Do Not Transfer between Applied and Academic Mathematics', *Nature* pp. 1–9.

Banerjee, A. V., Cole, S., Duflo, E. & Linden, L. (2007), 'Remedying Education: Evidence from Two Randomized Experiments in India', *Quarterly Journal of Economics* **122**(3), 1235–1264.

Barrera-Osorio, F., Gonzalez, K., Lagos, F. & Deming, D. J. (2020), 'Providing Performance Information in Education: An Experimental Evaluation in Colombia', *Journal of Public Economics* **186**, 104185.

Berlinski, S. & Busso, M. (2017), 'Challenges in Educational Reform: An Experiment on Active Learning in Mathematics', *Economics Letters* **156**, 172–175.

Berlinski, S., Busso, M. & Giannola, M. (2023), 'Helping Struggling Students and Benefiting All: Peer Effects in Primary Education', *Journal of Public Economics* **224**, 104925.

Bold, T., Filmer, D., Martin, G., Molina, E., Rockmore, C., Stacy, B., Svensson, J. & Wane, W. (2017), 'What Do Teachers Know and Do? Does It Matter?', *Policy Research Working Paper* **7956**.

Bradley, R. H., Caldwell, B. M., Rock, S. L., Hamrick, H. M. & Harris, P. (1988), 'Home Observation for Measurement of the Environment: Development of a Home Inventory for Use with Families Having Children 6 to 10 Years Old', *Contemporary Educational Psychology* **13**(1), 58–71.

Carneiro, P., Cruz-Aguayo, Y., Pachon, R. H. & Schady, N. (2022), Dynamic Complementarity in Elementary Schools: Experimental Estimates from Ecuador, Technical report, Working Paper.

Carneiro, P. & Heckman, J. (2003), 'Human Capital Policy', *IZA Discussion Paper* .

Carneiro, P. & Heckman, J. J. (2002), 'The Evidence on Credit Constraints in Post-Secondary Schooling', *The Economic Journal* **112**(482), 705–734.

Chetty, R., Friedman, J. N. & Rockoff, J. E. (2014), 'Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood', *American Economic Review* **104**(9), 2633–2679.

Cilliers, J., Fleisch, B., Prinsloo, C. & Taylor, S. (2020), 'How to Improve Teaching Practice? An Experimental Comparison of Centralized Training and In-Classroom Coaching', *Journal of Human Resources* **55**(3), 926–962.

Cunha, F. & Heckman, J. (2007), 'The Technology of Skill Formation', *American Economic Review* **97**(2), 31–47.

Cunha, F., Heckman, J. J. & Schennach, S. M. (2010), 'Estimating the Technology of Cognitive and Noncognitive Skill Formation', *Econometrica* **78**(3), 883–931.

Currie, J. & Almond, D. (2011), Human Capital Development before Age Five, Vol. 4 of *Handbook of Labor Economics*, Elsevier, pp. 1315–1486.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K. & Sundararaman, V. (2013), 'School Inputs, Household Substitution, and Test Scores', *American Economic Journal: Applied Economics* **5**(2), 29–57.

De Barros, A., Fajardo-Gonzalez, J., Glewwe, P. & Sankar, A. (2024), 'The Limitations of Activity-Based Instruction to Improve the Productivity of Schooling', *The Economic Journal* **134**(659), 959–984.

Dehaene, S. (2011), *The Number Sense: How the Mind Creates Mathematics*, OUP USA.

Dhaliwal, I., Duflo, E., Glennerster, R. & Tulloch, C. (2012), 'Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries', *Abdul Latif Jameel Poverty Action Lab, Massachusetts Institute of Technology, Cambridge, MA* .

Dillon, M. R., Kannan, H., Dean, J. T., Spelke, E. S. & Duflo, E. (2017), 'Cognitive Science in the Field: A Preschool Intervention Durably Enhances Intuitive but Not Formal Mathematics', *Science* **357**(6346), 47–55.

Doss, C., Fahle, E. M., Loeb, S. & York, B. N. (2019), 'More than Just a Nudge: Supporting Kindergarten Parents with Differentiated and Personalized Text Messages', *Journal of Human Resources* **54**(3), 567–603.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J. et al. (2007), 'School Readiness and Later Achievement', *Developmental Psychology* **43**(6), 1428.

Duncan, G., Kalil, A., Mogstad, M. & Rege, M. (2023), 'Investing in Early Childhood Development in Preschool and at Home', *Handbook of the Economics of Education* **6**, 1–91.

Evans, D. K., Jakiela, P. & Acosta, A. M. (2024), The Impacts of Childcare Interventions on Children's Outcomes in Low- and Middle-Income Countries: A Systematic Review, *in* 'AEA Papers and Proceedings', Vol. 114, American Economic Association, pp. 463–466.

García, J. L., Heckman, J. J. & Ziff, A. L. (2018), 'Gender Differences in the Benefits of an Influential Early Childhood Program', *European Economic Review* **109**, 9–22.

Gerber, A. S. & Green, D. P. (2012), *Field Experiments: Design, Analysis, and Interpretation*.

Gilraine, M. (2016), 'School Accountability and the Dynamics of Human Capital Formation', *URL: http://tinyurl.com/Gilraine-JMP* .

Glewwe, P., Kremer, M. & Moulin, S. (2009), 'Many Children Left Behind? Textbooks and Test Scores in Kenya', *American Economic Journal: Applied Economics* **1**(1), 112–135.

Glewwe, P., Kremer, M., Moulin, S. & Zitzewitz, E. (2004), 'Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya', *Journal of Development Economics* **74**(1), 251–268.

Glewwe, P. & Muralidharan, K. (2016), Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications, *in* 'Handbook of the Economics of Education', Vol. 5, Elsevier, pp. 653–743.

Goff, L., Malamud, O., Pop-Eleches, C. & Urquiola, M. (2025), 'Interactions between Family and School Environments: Access to Abortion and Selective Schools', *Journal of Human Resources* **60**(3), 907–949.

Goldsmith-Pinkham, P., Hull, P. & Kolesár, M. (2024), 'Contamination Bias in Linear Regressions', *American Economic Review* **114**(12), 4015–4051.

Granger, E. M., Bevis, T. H., Saka, Y., Southerland, S. A., Sampson, V. & Tate, R. (2012), 'The Efficacy of Student-Centered Instruction in Supporting Science Learning', *Science* **338**(6103), 105–108.

Grönqvist, E., Öckert, B. & Rosenqvist, O. (2025), 'Does the "Boost for Mathematics" Boost Mathematics? A Large-Scale Evaluation of the "Lesson Study" Methodology on Student Performance', *American Economic Journal: Economic Policy* **17**(3), 345–372.

Gunnsteinsson, S., Molina, T., Adhvaryu, A., Christian, P., Labrique, A., Sugimoto, J., Shamim, A. A. & West Jr, K. P. (2022), 'Protecting Infants from Natural Disasters: The Case of Vitamin A Supplementation and a Tornado in Bangladesh', *Journal of Development Economics* **158**, 102914.

Hanushek, E. A., Schwerdt, G., Wiederhold, S. & Woessmann, L. (2015), 'Returns to Skills around the World: Evidence from PIAAC', *European Economic Review* **73**, 103–130.

Heckman, J. J. (2006), 'Skill Formation and the Economics of Investing in Disadvantaged Children', *Science* **312**(5782), 1900–1902.

ICFES (2017), 'Resultados Nacionales Saber 5° y 9° – 2016. Lenguaje y Matemáticas', *Instituto Colombiano para la Evaluación de la Educación (ICFES)* .

Imbens, G. W. & Rubin, D. B. (2015), *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge University Press.

Jackson, C. K. (2018), 'What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes', *Journal of Political Economy* **126**(5), 2072–2107.

Jervis, P., Giannola, M., Cardona-Sosa, L., Day, M., Grantham-McGregor, S., Meghir, C., Rubio-Codina, M. & Attanasio, O. (2024), Early Childhood Intervention for the Poor: Long Term Outcomes, Technical report, National Bureau of Economic Research.

Johnson, R. C. & Jackson, C. K. (2019), 'Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending', *American Economic Journal: Economic Policy* **11**(4), 310–349.

Kalil, A., Michelini, M. & Ramos, P. (2025), 'The Promise of Digital Technology and Generative AI for Supporting Parenting Interventions in Latin America', *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2025-123).

Lee, D. S. (2009), 'Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects', *Review of Economic Studies* **76**(3), 1071–1102.

Magnuson, K. A., Kelchen, R., Duncan, G. J., Schindler, H. S., Shager, H. & Yoshikawa, H. (2016), 'Do the Effects of Early Childhood Education Programs Differ by Gender? A Meta-Analysis', *Early Childhood Research Quarterly* **36**, 521–536.

Mayer, S. E., Kalil, A., Oreopoulos, P. & Gallegos, S. (2019), 'Using Behavioral Insights to Increase Parental Engagement: The Parents and Children Together Intervention', *Journal of Human Resources* **54**(4), 900–925.

Muralidharan, K. (2017), Field Experiments in Education in Developing Countries, *in* 'Handbook of Economic Field Experiments', Vol. 2, Elsevier, pp. 323–385.

Muralidharan, K., Singh, A. & Ganimian, A. J. (2019), 'Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India', *American Economic Review* **109**(4), 1426–1460.

OECD (2016), 'Reviews of National Policies for Education: Education in Colombia', *OECD Publishing* .

Popova, A., Evans, D. K., Breeding, M. E. & Arancibia, V. (2022), 'Teacher Professional Development around the World: The Gap between Evidence and Practice', *The World Bank Research Observer* **37**(1), 107–136.

Rege, M., Størksen, I., Solli, I. F., Kalil, A., McClelland, M. M., Ten Braak, D., Lenes, R., Lunde, S., Breive, S., Carlsen, M. et al. (2024), 'The Effects of a Structured Curriculum on Preschool Effectiveness: A Field Experiment', *Journal of Human Resources* **59**(2), 576–603.

Reynolds, A. J. & Temple, J. A. (2019), *Sustaining Early Childhood Learning Gains: Program, School, and Family Influences*, Cambridge University Press.

Rossin-Slater, M. & Wüst, M. (2020), 'What Is the Added Value of Preschool for Poor Children? Long-Term and Intergenerational Impacts and Interactions with an Infant Health Intervention', *American Economic Journal: Applied Economics* **12**(3), 255–286.

RTI-International (2009), 'Early Grade Reading Assessment Toolkit', *World Bank Working Paper, Office of Human Development* .

Schleicher, A. (2018), 'Insights and Interpretations', *PISA 2018* **10**.

Stallings, J. (1977), *Learning to Look: A Handbook on Classroom Observation and Teaching Models*, Wadsworth Publishing Company.

The World Bank (2013), Staff Guidance Note on the Application of the Joint Bank-Fund Debt Sustainability Framework for Low-Income Countries, Technical Report 82566, The World Bank.

Todd, P. E. & Wolpin, K. I. (2003), 'On the Specification and Estimation of the Production Function for Cognitive Achievement', *The Economic Journal* **113**(485), F3–F33.

UNESCO (2025), UNESCO Institute for Statistics Processed by Our World in Data "Share of Students Achieving Minimum Math Proficiency by the End of Primary Education", Technical report, UNESCO Institute for Statistics. Dataset. Retrieved August 8, 2025 from https://archive.ourworldindata.org/20250730-091307/grapher/

share-of-students-at-end-of-primary-education-achieving-minimum-math-pr
html. Archived on July 30, 2025.

World Bank (2018), *World Development Report 2018: Learning to Realize Education's Promise*, World Bank, Washington, DC.

York, B. N., Loeb, S. & Doss, C. (2019), 'One Step at a Time: The Effects of an Early Literacy Text-Messaging Program for Parents of Preschoolers', *Journal of Human Resources* **54**(3), 537–566.

Appendix to "Effective Families or Effective Schools? Experimental Evidence on Fostering Children's Numeracy"

The appendix is organized as follows. Appendix A presents additional tables and figures for the paper, Appendix B describes the school readiness assessment collected prior to randomization, and Appendix C presents robustness checks for the ITT estimates reported in Table 1.

A Appendix Tables and Figures

Table A1: Baseline Balance

	Control		Treatment Comparisons							Joint Wald p-value	Obs
	Mean	SD	Any T vs C	T^P vs C	T^T vs C	T^{P+T} vs C	T^P vs T^T	T^P vs T^{P+T}	T^T vs T^{P+T}		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Child covariates											
Age	5.127	0.378	0.014	0.014	-0.008	0.036	-0.021	0.023	0.048	0.606	1160
			(0.024)	(0.027)	(0.035)	(0.032)	(0.033)	(0.031)	(0.037)		
Gender (female)	0.450	0.498	0.000	0.000	0.000	0.000	0.000	0.000	0.000	–	1160
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		
SES	1.974	0.857	0.043	0.079	0.007	0.032	-0.065	-0.041	0.006	0.458	1160
			(0.075)	(0.050)	(0.122)	(0.111)	(0.117)	(0.112)	(0.078)		
Cognitive	0.581	0.252	0.030	0.017	0.040	0.035	0.024	0.020	-0.005	0.551	1160
			(0.022)	(0.022)	(0.028)	(0.029)	(0.023)	(0.026)	(0.015)		
Language	0.712	0.186	-0.007	-0.017	-0.001	0.000	0.013	0.014	-0.001	0.719	1160
			(0.017)	(0.017)	(0.020)	(0.020)	(0.018)	(0.017)	(0.011)		
Motor	0.634	0.219	-0.005	0.009	-0.001	-0.027	-0.013	-0.038	-0.026	0.236	1160
			(0.017)	(0.013)	(0.027)	(0.026)	(0.028)	(0.026)	(0.015)		
Socio-emotional	0.848	0.181	0.009	0.003	0.015	0.012	0.014	0.010	-0.004	0.929	1160
			(0.016)	(0.014)	(0.024)	(0.021)	(0.020)	(0.016)	(0.015)		
Panel B: Caregiver-child relationship (phone owner)											
Mother	0.732	0.444	0.030	0.007	0.059	0.030	0.058	0.030	-0.027	0.544	1160
			(0.030)	(0.034)	(0.041)	(0.039)	(0.043)	(0.041)	(0.033)		
Grandmother	0.078	0.269	-0.007	0.028	-0.023	-0.036	-0.051	-0.064	-0.014	0.060	1160
			(0.015)	(0.021)	(0.022)	(0.019)	(0.026)	(0.023)	(0.020)		
Father	0.052	0.223	0.009	0.009	-0.001	0.020	-0.013	0.007	0.019	0.785	1160
			(0.014)	(0.019)	(0.014)	(0.021)	(0.019)	(0.024)	(0.023)		
Other	0.138	0.345	-0.032	-0.044	-0.035	-0.013	0.007	0.027	0.021	0.496	1160
			(0.026)	(0.032)	(0.031)	(0.031)	(0.029)	(0.030)	(0.030)		
Panel C: School covariates											
Multigrade	0.033	0.180	-0.005	0.005	-0.013	-0.010	-0.018	-0.016	0.000	0.479	1160
			(0.005)	(0.005)	(0.009)	(0.007)	(0.011)	(0.010)	(0.000)		
Rural	0.089	0.286	0.031	-0.013	0.065	0.055	0.082	0.074	-0.003	0.656	1160
			(0.035)	(0.012)	(0.064)	(0.059)	(0.074)	(0.070)	(0.003)		
# Classes	2.751	2.002	-0.178	0.040	-0.300	-0.335	-0.423	-0.460	-0.037	0.007	1160
			(0.236)	(0.030)	(0.374)	(0.369)	(0.367)	(0.365)	(0.013)		
Class size	23.242	4.923	0.365	0.167	0.268	0.719	0.127	0.521	0.272	0.198	1160
			(0.532)	(0.178)	(0.896)	(0.911)	(0.999)	(0.986)	(0.133)		

Notes: The table shows baseline balance across the four experimental groups. Panel A shows child-level characteristics. Panel B shows the relationship between the caregiver (phone owner) and the child. Panel C shows school-level characteristics. Columns 1 and 2 show the summary statistics for the control group at baseline. Column 3 shows the coefficient from regressing the baseline variable on an indicator for any treatment. Columns 4, 5, and 6 show the coefficient from regressing the baseline variable on separate indicators for being in group T^P, T^T, or T^{P+T}. Columns 7 and 8 show the coefficient from regressing the baseline variable on separate indicators for being in group T^T or T^{P+T}, omitting the control group (coefficients are relative to T^P). Column 9 shows the coefficient from regressing the baseline variable on an indicator for being in group T^{P+T}, omitting the control group and T^P (coefficients are relative to T^T). Column 10 reports the p-value from a joint Wald test that the coefficients on T^P, T^T, and T^{P+T} are jointly equal to zero in the specification of columns 4–6. Standard errors are clustered at the school level. All regressions include randomization strata fixed effects.

Table A2: Response Rate: Student Assessment

	Interviewed
	(1)
Parental engagement	-0.008
	(0.028)
Teacher development	0.009
	(0.033)
Combined treatment	0.037
	(0.029)
Observations	3480
Control group mean	0.709

Notes: The outcome is an indicator for whether the child's numeracy scores were observed. The regression includes randomization strata fixed effects. Standard errors clustered at the school level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A3: Baseline Balance Excluding Attritors

	Control		Treatment Comparisons							Joint Wald p-value	Obs
	Mean	SD	Any T vs C	T^P vs C	T^T vs C	T^{P+T} vs C	T^P vs T^T	T^P vs T^{P+T}	T^T vs T^{P+T}		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Child covariates											
Age	5.171	0.382	-0.022	-0.050	-0.032	0.023	0.018	0.073	0.056	0.139	849
			(0.027)	(0.039)	(0.036)	(0.030)	(0.035)	(0.033)	(0.035)		
Gender (female)	0.429	0.496	0.000	0.000	0.000	0.000	0.000	0.000	0.000	–	849
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		
SES	2.040	0.799	0.031	0.121	-0.042	-0.011	-0.165	-0.135	0.017	0.029	849
			(0.069)	(0.043)	(0.116)	(0.102)	(0.110)	(0.101)	(0.067)		
Cognitive	0.604	0.246	0.032	0.031	0.027	0.039	-0.002	0.012	0.015	0.592	849
			(0.025)	(0.027)	(0.029)	(0.031)	(0.026)	(0.030)	(0.023)		
Language	0.724	0.183	-0.003	-0.008	0.000	0.001	0.005	0.007	0.002	0.899	849
			(0.014)	(0.013)	(0.018)	(0.019)	(0.016)	(0.016)	(0.013)		
Motor	0.654	0.209	0.000	0.023	-0.011	-0.017	-0.034	-0.039	-0.006	0.175	849
			(0.021)	(0.012)	(0.031)	(0.030)	(0.030)	(0.029)	(0.016)		
Socio-emotional	0.859	0.172	0.015	0.013	0.012	0.019	0.001	0.010	0.010	0.720	849
			(0.016)	(0.015)	(0.023)	(0.020)	(0.020)	(0.017)	(0.014)		
Panel B: Caregiver-child relationship (phone owner)											
Mother	0.737	0.441	0.039	0.037	0.054	0.026	0.021	-0.004	-0.023	0.517	849
			(0.034)	(0.035)	(0.043)	(0.045)	(0.041)	(0.041)	(0.029)		
Grandmother	0.081	0.273	-0.008	0.032	-0.037	-0.029	-0.068	-0.059	0.008	0.111	849
			(0.019)	(0.029)	(0.025)	(0.023)	(0.031)	(0.029)	(0.023)		
Father	0.051	0.220	0.006	0.007	-0.005	0.015	-0.012	0.008	0.019	0.818	849
			(0.016)	(0.019)	(0.018)	(0.023)	(0.020)	(0.024)	(0.023)		
Other	0.131	0.339	-0.037	-0.076	-0.012	-0.013	0.060	0.055	-0.004	0.124	849
			(0.032)	(0.044)	(0.034)	(0.034)	(0.027)	(0.028)	(0.033)		
Panel C: School covariates											
Multigrade	0.020	0.141	-0.008	0.004	-0.016	-0.014	-0.020	-0.019	0.000	0.482	849
			(0.007)	(0.005)	(0.011)	(0.010)	(0.013)	(0.012)	(0.000)		
Rural	0.061	0.239	0.039	0.009	0.054	0.062	0.048	0.056	0.011	0.347	849
			(0.033)	(0.014)	(0.058)	(0.058)	(0.070)	(0.070)	(0.013)		
# Classes	2.904	2.017	-0.199	0.022	-0.328	-0.351	-0.441	-0.467	-0.036	0.044	849
			(0.252)	(0.032)	(0.405)	(0.390)	(0.399)	(0.388)	(0.014)		
Class size	23.722	4.222	0.205	0.031	0.065	0.559	0.062	0.521	0.330	0.206	849
			(0.473)	(0.197)	(0.840)	(0.836)	(0.976)	(0.966)	(0.148)		

Notes: The table shows baseline balance across the four experimental groups for non-attriting children. Panel A shows child-level characteristics. Panel B shows the relationship between the caregiver (phone owner) and the child. Panel C shows school-level characteristics. Columns 1 and 2 show the summary statistics for the control group at baseline. Column 3 shows the coefficient from regressing the baseline variable on an indicator for any treatment. Columns 4, 5, and 6 show the coefficient from regressing the baseline variable on separate indicators for being in group T^P, T^T, or T^{P+T}. Columns 7 and 8 show the coefficient from regressing the baseline variable on separate indicators for being in group T^T or T^{P+T}, omitting the control group (coefficients are relative to T^P). Column 9 shows the coefficient from regressing the baseline variable on an indicator for being in group T^{P+T}, omitting the control group and T^P (coefficients are relative to T^T). Column 10 reports the p-value from a joint Wald test that the coefficients on T^P, T^T, and T^{P+T} are jointly equal to zero in the specification of columns 4–6. Standard errors are clustered at the school level. All regressions include randomization strata fixed effects.

Table A4: Response Rate: Caregiver Survey

	Interviewed
	(1)
Parental engagement	-0.003
	(0.020)
Observations	2308
Control group mean	0.708

Notes: The outcome variable is an indicator for whether the caregiver was interviewed in the caregiver survey. The regression includes randomization strata fixed effects. Standard errors clustered at the school level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A5: Response Rate: Teacher Survey

	Interviewed (1)
Teacher development	0.051 (0.058)
Observations	143
Control group mean	0.806

Notes: The outcome variable is an indicator for whether the teacher was interviewed in the teacher survey. The regression includes randomization strata fixed effects. Standard errors clustered at the school level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Variable Definition: Teacher Practices

Variable name	Questionnaire items
Active class score	Stallings classroom observation (activity initiated by student or teacher): Read out loud, explain/demonstrate, Discussion/Q&A, Manipulative activity in small group, exercises, verbal instruction, summary of class
Class preparation/delivery	Stallings classroom observation: (i) teacher presented the learning objectives at the beginning of the lesson, (ii) teacher gave instructions before every activity, (iii) teacher did a closure at the end of the class summarizing learning, (iv) material was prepared.
Students' participation	Stallings classroom observation: (i) students ask questions, (ii) students answer questions, (iii) students follow instructions.
Class disruption	Stallings classroom observation: (i) the teacher interrupts classroom to get students' attention, (ii) students are distracted by other activities, (iii) class starts late to get students' attention.
Learning expectations	Teachers' survey: (i) all of your students can learn mathematics, (ii) students find it fun to learn mathematics, (iii) all of your students look forward to math class, (iv) math does not cause distress for many students, (v) some students have aptitude for mathematics, (vi) by end of the year most students will achieve the expected math level.
Pedagogy knowledge	Teachers' survey: students should: (i) learn only quantities of a maximum of one hundred, (ii) not memorize additions and subtractions, (iii) learn to vertically add and subtract, (iv) learn to add and subtract 3-digit numbers, (v) learn how to solve expressions with missing numbers, (vi) learn how to solve unfamiliar problems, (vii) learn additions and subtractions, (viii) use their fingers to learn mathematics, (ix) use concrete materials to learn mathematics.

Table A7: Spillovers on Literacy Skills

	Literacy Index	**Letter initial sound**	**Simple pseudo words**	**Passage reading & comprehension**
	(1)	(2)	(3)	(4)
Parental engagement	0.181*	0.124	0.169*	0.150*
	(0.100)	(0.088)	(0.096)	(0.087)
		[0.369]	[0.232]	[0.232]
Teacher development	0.146	0.108	0.112	0.157*
	(0.091)	(0.094)	(0.085)	(0.087)
		[0.497]	[0.397]	[0.218]
Combined treatment	0.036	0.079	-0.022	0.050
	(0.097)	(0.083)	(0.091)	(0.081)
		[0.558]	[0.759]	[0.631]
Observations	2485	2485	2485	2485
P-value: $T^P = T^T$	0.636	0.857	0.387	0.930
P-value: $T^P = T^{P+T}$	0.089	0.524	0.017	0.191
P-value: $T^T = T^{P+T}$	0.111	0.625	0.069	0.171
Share correct (control)	0.442	0.524	0.563	0.354

Notes: The table presents ITT effects on children's literacy skills. All outcomes are standardized to have mean 0 and standard deviation 1 in the control group. Column 1 reports the effects on the literacy index described in Section 4.1. Columns 2–4 report disaggregated treatment impacts on the outcomes indicated in the column headers; square brackets below the point estimates show Romano-Wolf step-down p-values. The p-values at the bottom of the table test whether coefficients are equal across experimental groups (parental engagement T^P, teacher development T^T, and combined T^{P+T}). Controls include survey wave, survey week, enumerator, and randomization strata fixed effects. Share correct (control) shows the proportion of correct answers on the task for children in the control group. Standard errors clustered at the school level are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A8: Spillovers from the Parental Engagement Intervention

	Numeracy Index
% treated parents (within school)	-0.281
	(0.629)
Observations	1195

Notes: The table presents estimates of spillover effects from the parental engagement intervention on untreated children. We regress the numeracy index of children not assigned to the parental engagement intervention on the within-school share of children assigned to the intervention. Controls include survey wave, survey week, enumerator, and randomization strata fixed effects. We include the main effect for the teacher development intervention. Standard errors clustered at the school level are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A9: Intention-to-Treat: Caregiver-Child Interactions
Full Model Corresponding to Eq (1)

	Index (1)	Count (2)	Read (3)	Story (4)
Parental engagement	0.087	0.076	0.058	0.079
	(0.069)	(0.069)	(0.069)	(0.070)
		[0.802]	[0.828]	[0.802]
Teacher development	0.003	0.032	-0.020	-0.002
	(0.072)	(0.073)	(0.073)	(0.073)
		[0.950]	[0.966]	[0.992]
Combined treatment	0.125*	0.108	0.124*	0.069
	(0.072)	(0.073)	(0.072)	(0.073)
		[0.577]	[0.459]	[0.828]
Observations	1628	1628	1628	1628
Share (control)	0.638	0.684	0.676	0.559

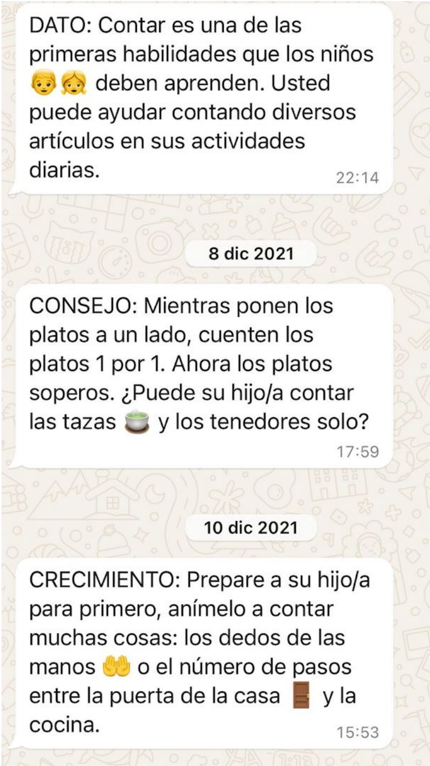
Notes: The table presents ITT effects on caregiver–child interactions. The reference group consists of children in the control group. We use responses from the caregiver interviews from February and October 2022. Column 1 reports treatment effects on a summary index of parent–child interactions, constructed from the outcomes reported in columns 2–4; all outcomes are expressed in standard deviation units relative to the control group. Square brackets below the point estimates show Romano–Wolf step-down p-values. Robust standard errors are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A10: Teachers' practices on share (exogenous) treated parental engagement

	Index (1)	Active class score (2)	Class preparation delivery (3)	Students' participation (4)	Class disruption (5)	Learning expectations (6)	Pedagogy knowledge (7)
Share treated parental engagement (within school)	0.566	0.539	0.657	0.768	0.560	0.282	-0.042
	(0.604)	(0.731)	(0.761)	(0.460)	(0.605)	(0.510)	(0.593)
		[0.864]	[0.864]	[0.525]	[0.864]	[0.864]	[0.932]
Observations	118	118	118	118	118	118	118
Control group mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000

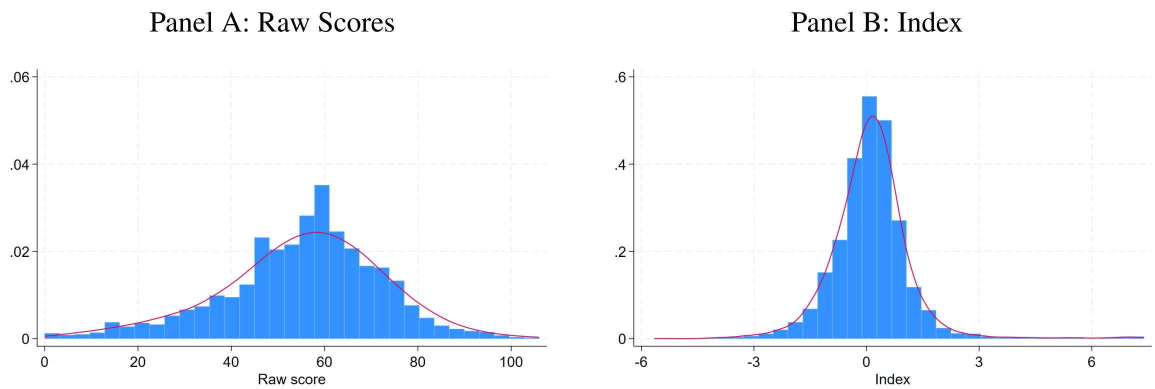
Notes: The table reports the effects of share (exogenous) treated parental engagement and the teacher development treatment on teachers' practices. Each column is a different regression for the effect on the share (exogenous) treated parental engagement and the teacher development treatment on the outcome in the column header. The reference group consists of teachers in schools that did not receive the teacher professional development program. We use responses from the teacher survey and in-classroom observation from 2022 and 2023. Column 1 reports treatment effects on a summary index of teachers' practices, constructed from the outcome reported in columns 2-7, square brackets below the point estimates show Romano-Wolf step-down p-values. The exact components of each of the outcome variables in columns 2-7 are reported in Appendix Table A6. All variables have been standardized to have a mean of zero and standard deviation of one in control schools. Controls include year and randomization strata fixed effects. Standard errors clustered at the school level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure A1: Example of Text Messages



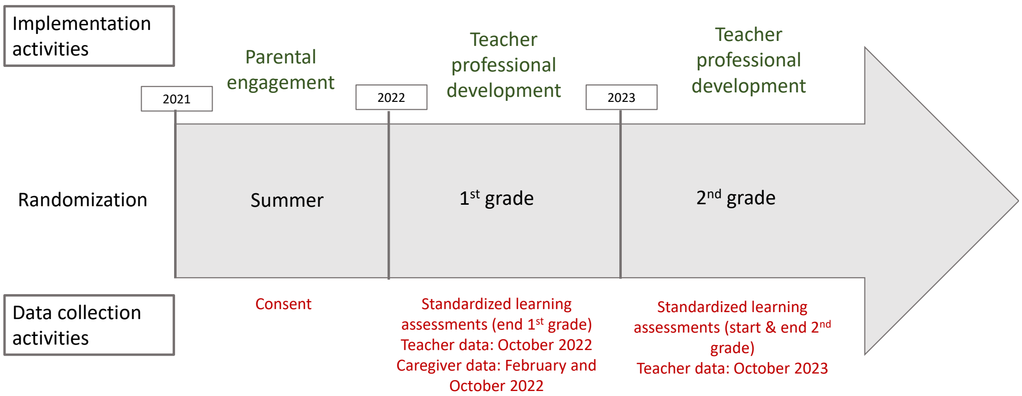
Notes: The figure shows the three text messages that were sent to caregivers in the parental engagement intervention in a particular week. The first message is a *FACT: Counting is one of the first math skills that children learn. You can help by counting items during everyday activities.* The second one is a *TIP: As you put the dishes away, count the plates 1-by-1 with your child. Try it again with bowls. Can your child count the cups by himself (herself)? Forks?* The third is a *GROWTH: Keep counting. You're preparing your child for first grade! You can count many things together - the fingers on your hand, or the number of steps from the front door to the kitchen.*

Figure A2: Distribution of Numeracy Scores



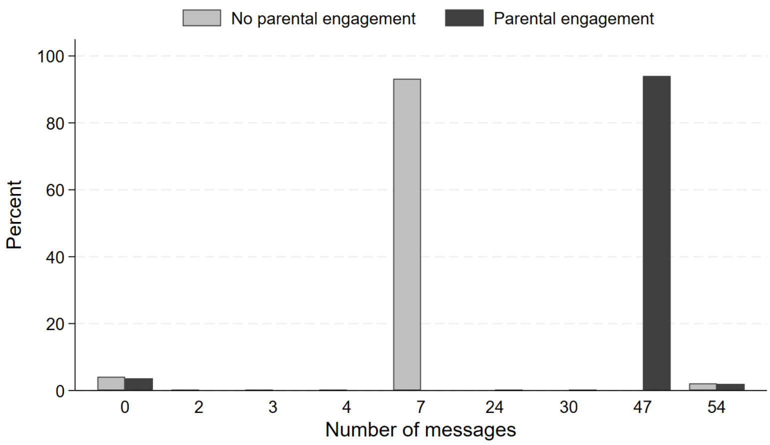
Notes: The figure plots the distribution of children's numeracy scores. Panel A plots the raw scores, and Panel B plots the index described in section 4.1.

Figure A3: Timeline



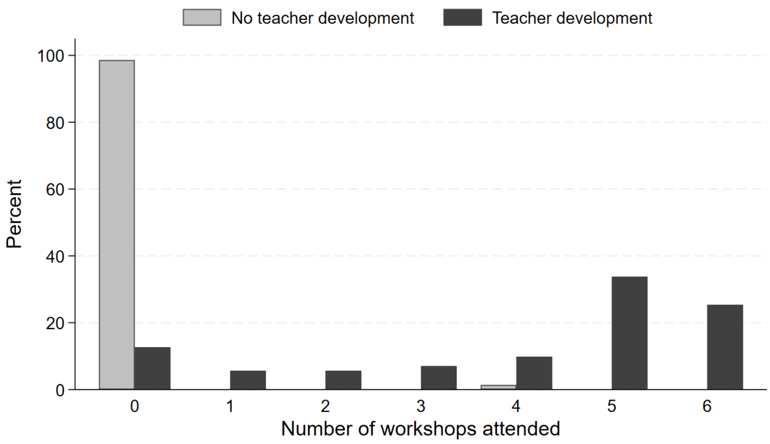
Notes: The figure plots the timeline of the interventions and data collection activities. The parental engagement intervention lasted from December 2021 to February 2022, during the summer break preceding children's enrollment in first grade. The teacher professional development intervention took place over the 2022 and 2023 school years. First-grade teachers were trained during the 2022 school year, while second-grade teachers were trained in the 2023 school year. Student achievement data were collected at three points in time: at the end of first grade, and at both the beginning and end of second grade. Data from parents were collected twice, immediately following the parental engagement intervention, in February 2022, and again eight months later, in October 2022. Teacher interviews and direct classroom observations took place in October 2022 and October 2023.

Figure A4: Number of Messages Received



Notes: The figure shows the percentage of messages received by treatment group (no parental engagement vs. parental engagement). A total of 47 texts were scheduled to families in the parental engagement group and a total of 7 non-instructional text messages were scheduled to families in the control group.

Figure A5: Teacher Attendance at In-person Workshops



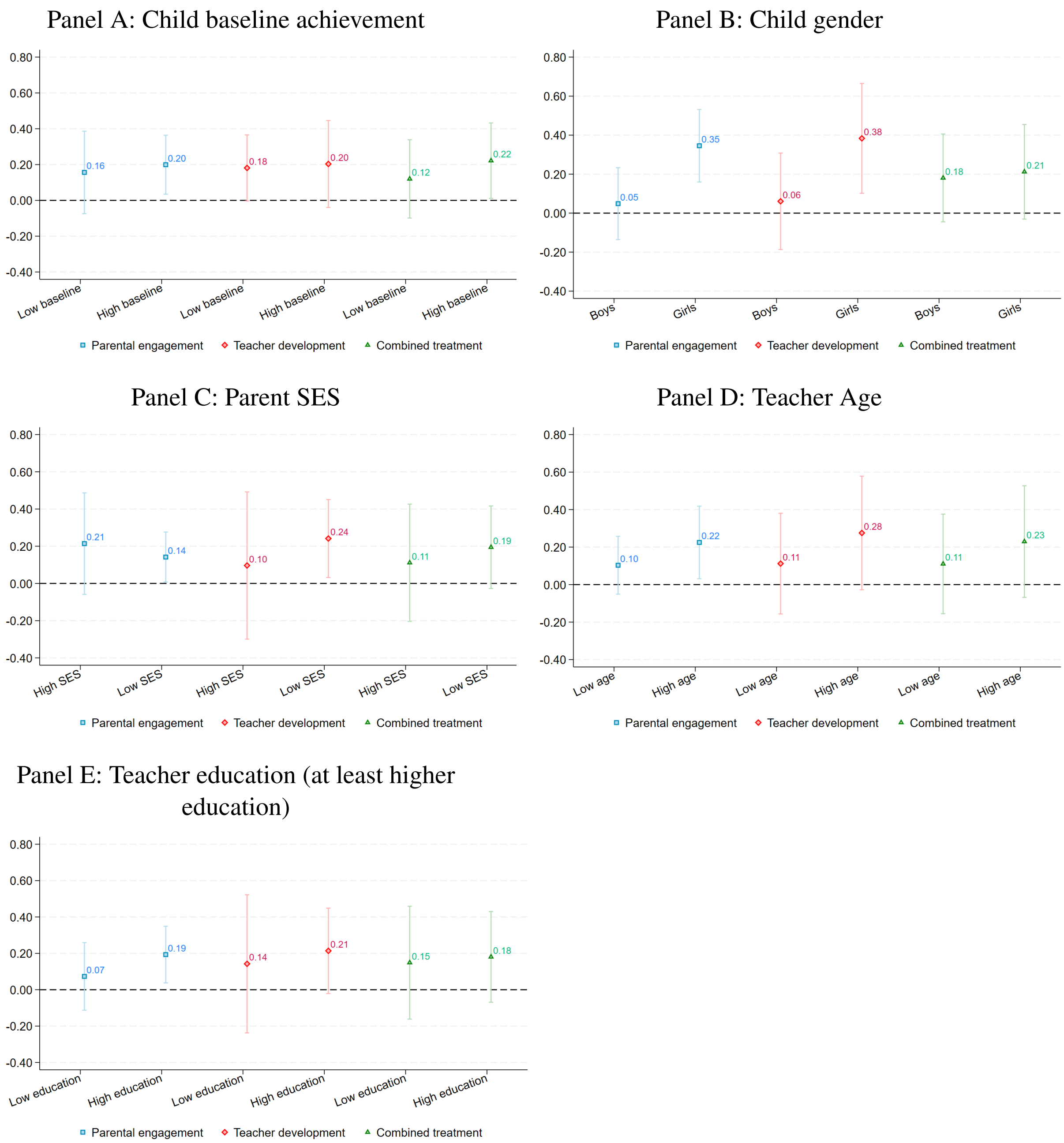
Notes: The figure shows the percentage of in-person workshops attended by teachers by school treatment status (no teacher development vs. teacher development).

Table A11: Sensitivity of the present value of numeracy gains to alternative assumptions on skill depreciation and discounting for the teacher development intervention

Depreciation (δ)	Discount rate (r)	PV	Cost effectiveness per 100 USD
0.01	0.05	0.905	1.160
0.025	0.05	0.880	1.128
0.05	**0.05**	**0.839**	**1.075**
0.075	0.05	0.800	1.026
0.10	0.05	0.764	0.980
0.05	0.01	0.901	1.155
0.05	0.025	0.877	1.124
0.05	0.075	0.805	1.032
0.05	0.10	0.773	0.991
0.01	0.01	0.976	1.251
0.025	0.025	0.921	1.181
0.075	0.075	0.769	0.986
0.10	0.10	0.707	0.906

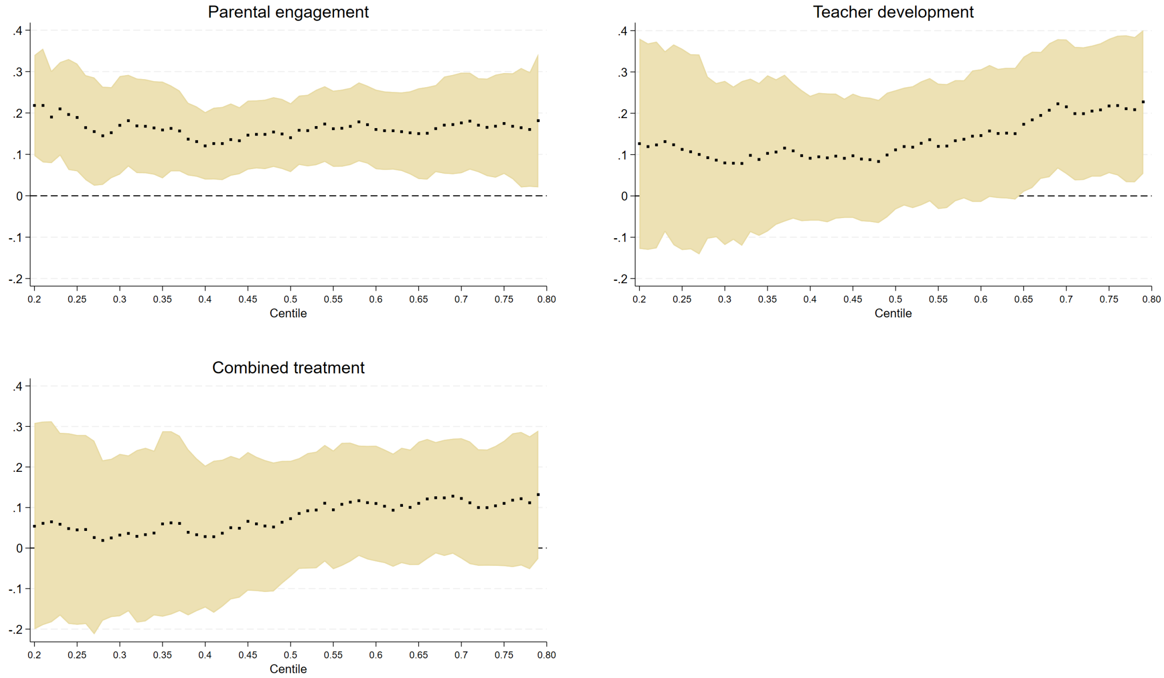
Notes: Table reports the present value (PV) of the cumulative numeracy gains generated by a trained teacher when teacher professional development is treated as a one-time investment with benefits accruing to subsequent cohorts. The PV is computed as $\sum_{t=0}^{T-1} \frac{(1-\delta)^t B}{(1+r)^t}$, where $B = 0.203$ is the ITT effect for one cohort, $T = 5$ is the number of cohorts, δ is the annual depreciation rate of teachers' acquired skills, and r is the discount rate. Following standard practice in cost-effectiveness analysis, future learning gains are discounted at rate r and allowed to decay at rate δ to reflect potential erosion of instructional effectiveness over time. Cost-effectiveness per 100 USD is computed as $(PV/78) \times 100$, where 78 USD is the total program cost per treated student.

Figure A6: Intention-to-Treat: Numeracy Skills by Student and Teacher Characteristics



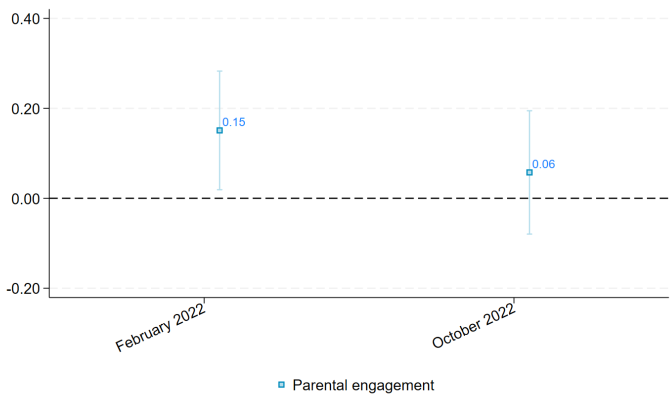
Notes: The figures present heterogeneity by student and teacher characteristics in treatment impacts on numeracy skills and their corresponding 95% level confidence intervals. Panel A shows heterogeneity by child baseline achievement in the *Predictores de Aprendizaje* assessment, splitting the sample at the median. Panel B shows heterogeneity by gender. Panel C shows heterogeneity by SES, splitting the sample in lower- and higher-SES (as defined in the main text). Panel D shows heterogeneity by teacher's age (median split of the sample). Panel E shows heterogeneity by teacher's education (at least higher education). In each panel, the coefficients are from a model similar to (1) where we interact the treatment indicators with an indicator for the heterogeneity category (e.g., SES). The outcome variable is the index described in Section 4.1.

Figure A7: Quantile Treatment Effects



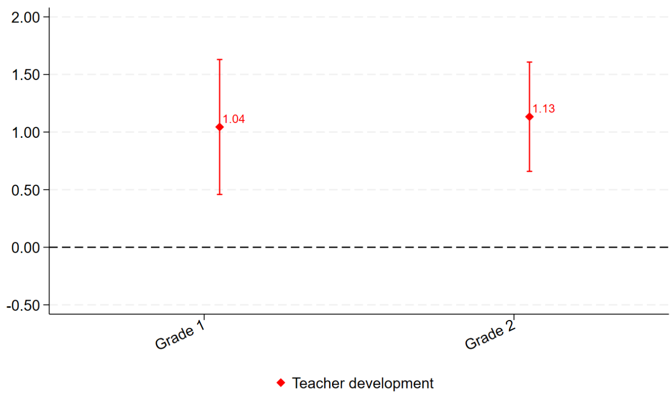
Notes: The figures present the quantile treatment effects and their corresponding 95% level confidence intervals. The outcome variable is the index described in Section 4.1.

Figure A8: Intention-to-Treat: Caregiver-Child Interactions by Survey Wave



Notes: The figure presents ITT effects of the parental engagement intervention (alone or combined with the teacher professional development program) on parent–child interactions by wave (February or October 2022) and their corresponding 95% level confidence intervals. The reference group consists of children in the control group and those assigned to the teacher development only. The outcome variable is the index in column 1 of Table 2.

Figure A9: Intention-to-Treat: Teachers' Practices by Survey Wave



Notes: The figure presents ITT effects of the teacher professional development intervention on teachers' practices by wave (October 2022 or October 2023) and their corresponding 95% level confidence intervals. The reference group consists of teachers in schools that did not receive the teacher professional development program. The outcome variable is the index in column 1 of Table 3.

B Predictores de Aprendizaje

As part of the Predictores de Aprendizaje assessment, kindergarten teachers completed a structured questionnaire for each student, reporting on the attainment of developmental milestones across four domains: cognitive, language, motor and socio-emotional. The questionnaire included concrete indicators such as whether the child “maintains attention during proposed activities” and “understands temporal concepts such as yesterday and tomorrow" (cognitive skills); “articulates consonant combinations” and “describes what is observed in a picture” (language skills); “walks backwards on tiptoes and heels” and “completes simple figures” (motor skills); and “shows empathy for others” and “resolves everyday social situations” (socio-emotional skills). These data were collected prior to the randomization of schools and students. We sum item-level responses to construct standardized indices for each domain and an overall school readiness index.

C ITT Effects: Robustness

In this section we report several robustness checks for the results shown in Table 1.

In Appendix Figure C1, we show that the treatment impacts are robust to: (i) a different set of control variables (marked with diamonds and triangles in the figure), (ii) an alternative outcome specification, where the numeracy score is constructed as the raw sum of correct items (marked with dots in the figure), (iii) inverse probability-weighted (IPW) estimates for attrition (marked with Xs in the figure), and (iv) after excluding the 24 twins who received both control and treatment text messages (marked with plus signs).

Second, we assess the sensitivity of the results to differential attrition using Lee (2009) bounds. This approach trims the outcome distribution of the group with the higher selection rate to equalize selection across treatment and control, thereby providing worst-case bounds for the treatment effects under monotonicity. As shown in Appendix Table C1, the lower and upper bounds remain positive and close in magnitude to the main estimates across treatment arms. This indicates that the results are not driven by selective attrition and that the conclusions are robust even under conservative assumptions about sample selection.

Third, Appendix Table C2 presents randomization inference results. This non-parametric method reassigns treatment labels across units to generate a reference distribution of the test statistic under the sharp null hypothesis of no treatment effect. The findings align with those from standard parametric tests, indicating that our results are robust and not driven by specific distributional assumptions such as normality (Gerber & Green 2012, Imbens & Rubin 2015).

Fourth, we assess contamination bias using the diagnostic proposed by Goldsmith-Pinkham et al. (2024). In settings with multiple randomly assigned treatments, regression-based estimates may reflect not only the effect of a given treatment arm itself, but also components induced by heterogeneity in the effects of other treatment arms. We compare the partially linear (PL) estimator – i.e. the standard uninteracted regression estimand – with the own-treatment effect component (OWN), which isolates the treatment arm's effect purged of contamination from other arms, as well as with alternative estimands implied by different weighting schemes, namely the unweighted ATE, the easiest-to-estimate weighted (EW), and the common-weighted (CW) estimators. Appendix Table C3 reports these pairwise differences for both the full sample and the overlap sample. The overlap sample restricts attention to observations with common support across treatment arms, so that all estimators are identified and the comparisons are less sensitive to limited overlap. The key diagnostic for contamination bias is the difference between PL and OWN: when OWN is close to PL, contamination from other treatment arms is limited. Instead, PL−ATE, PL−EW, and PL−CW capture differences between PL and alternative weighting schemes under treatment effect heterogeneity. In our setting, these differences are small and statistically insignificant across treatment arms, with similar results in the overlap sample, indicating limited contamination bias.

Fifth, Appendix Table C4 examines whether the interventions affected contemporaneous teacher characteristics. Estimates are close to zero and statistically insignificant across specifications, indicating that the interventions did not alter the composition of teachers in treated schools. The control means indicate that teachers in the control group are on average about 53 years old. The majority are female (89%), permanently employed (91%), and hold a higher education degree (76%).

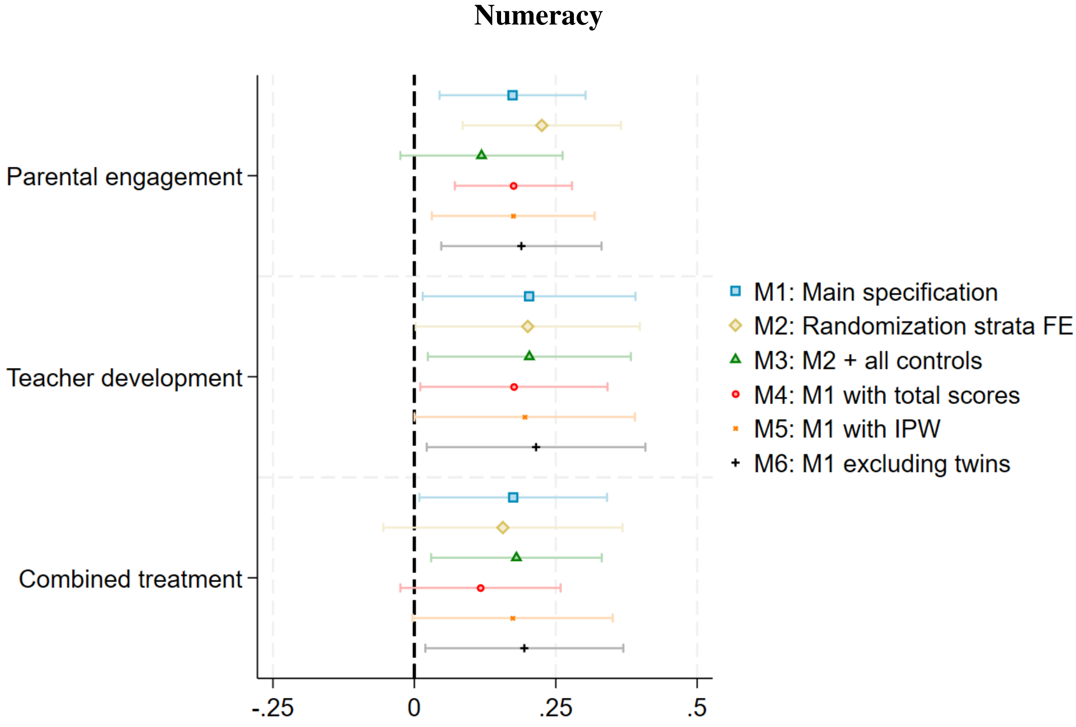
Sixth, Appendix Table C5 examines whether the interventions affected school mobility in a way that could alter children's treatment status. Mobility is rare: only about 6% of children initially enrolled in control schools subsequently transferred to a different school. The estimated effects of the interventions on the likelihood of switching schools are small and statistically indistinguishable from zero. Thus, there is no evidence that the interventions induced selective school mobility, and endogenous changes in school enrollment are unlikely to bias our estimates.

Table C1: Intention-to-Treat: Students' Numeracy Skills - Lee (2009) bounds

Treatment	Lee bounds		Trimming rate
	Lower	Upper	
Parental engagement	0.107 (0.078)	0.241*** (0.066)	0.026
Teacher development	0.157* (0.094)	0.225*** (0.084)	0.010
Combined treatment	0.040 (0.070)	0.270*** (0.068)	0.045

Notes: The table presents Lee (2009) bounds for the treatment effects on the numeracy index from the benchmark model (Table 1) to account for differential attrition between treatment and control groups. The lower and upper bounds are obtained by trimming the outcome distribution of the group with the higher selection rate so that selection rates are equal across groups. The trim rate indicates the proportion of observations removed from the group with higher selection. Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure C1: Intention-to-Treat: Robustness Checks



Notes: The figure presents robustness checks for the estimates in Table 1. Squares represent the specification reported in Column 1 of Table 1. Diamonds show estimates controlling only for randomization strata fixed effects. Triangles show estimates controlling for the balanced covariates (Appendix Table A1) and randomization strata fixed effects. Dots present estimates from the main specification using the standardized sum of correct answers as the outcome instead of the numeracy index. Xs present inverse probability-weighted (IPW) estimates for attrition. Pluses present estimates from the main specification excluding the 24 twins that received control and treatment text messages. Standard errors clustered at the school level. Confidence intervals at 95% level.

Table C2: Intention-to-Treat: Students' Numeracy Skills - Randomization Inference p-values

	Index	**Number write, read, compare**	**Add & subtract**	**Sequence, equations, Problems**
	(1)	(2)	(3)	(4)
Parental engagement	0.174***	0.145**	0.116	0.211***
	(0.064)	(0.062)	(0.076)	(0.057)
	[0.030]	[0.056]	[0.118]	[0.016]
Teacher development	0.203**	0.132	0.156*	0.246**
	(0.093)	(0.083)	(0.084)	(0.102)
	[0.080]	[0.208]	[0.140]	[0.010]
Combined treatment	0.175**	0.051	0.183**	0.206**
	(0.082)	(0.084)	(0.072)	(0.079)
	[0.104]	[0.616]	[0.058]	[0.040]
Observations	2485	2485	2485	2485
P-value: $T^P = T^T$	0.741	0.854	0.676	0.719
P-value: $T^P = T^{P+T}$	0.990	0.240	0.453	0.955
P-value: $T^T = T^{P+T}$	0.782	0.162	0.800	0.760
Share correct (control)	0.474	0.874	0.300	0.259

Notes: The table presents ITT effects on children's numeracy skills, with randomization-inference p-values in square brackets below the point estimates. All outcomes are standardized to have mean 0 and standard deviation 1 in the control group. Column 1 reports the effects on the numeracy index described in Section 4.1. Columns 2–4 report disaggregated treatment impacts on the outcomes indicated in the column headers. Controls include survey wave, survey week, enumerator, and randomization strata fixed effects. Share correct (control) shows the proportion of correct answers on the task for children in the control group. Standard errors clustered at the school level are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C3: Contamination Bias Test (Goldsmith-Pinkham et al. 2024)

Panel A: full sample	PL−OWN	PL−ATE	PL−EW	PL−CW
Parental engagement			-0.039	-0.024
			(0.023)	(0.052)
Teacher development			-0.044	-0.008
			(0.068)	(0.087)
Combined treatment			0.001	0.079
			(0.072)	(0.078)

Panel B: overlap sample	PL−OWN	PL−ATE	PL−EW	PL−CW
Parental engagement	0.025	0.032	0.007	0.026
	(0.016)	(0.029)	(0.010)	(0.022)
Teacher development	0.015	0.008	0.008	0.022
	(0.025)	(0.031)	(0.021)	(0.021)
Combined treatment	0.014	-0.007	0.006	0.010
	(0.022)	(0.027)	(0.019)	(0.019)

Notes: The table reports differences between the partially linear (PL) estimator, i.e. the standard uninteracted regression estimand, and alternative estimators, with standard errors clustered at the school level in parentheses (Goldsmith-Pinkham et al. 2024). "OWN" denotes the own-treatment effect, i.e. the treatment effect purged of contamination from exposure to other treatment arms. "ATE," "EW," and "CW" denote alternative estimands that differ in the weighting of treatment effects. Accordingly, the column PL−OWN provides the main diagnostic for contamination bias, while the columns PL−ATE, PL−EW, and PL−CW capture differences between PL and alternative weighting schemes under treatment effect heterogeneity. Panel A reports results for the full sample, while Panel B reports results for the overlap sample, which restricts attention to observations with common support across treatment arms so that all estimators are identified and comparisons are less sensitive to limited overlap

Table C4: Intention-to-Treat: Teachers' characteristics

Teacher	**Age** (1)	**Female** (2)	**Permanent** (3)	**Higher education** (4)
Parental engagement	0.028	0.009	0.017	0.022
	(0.492)	(0.011)	(0.011)	(0.019)
	[0.930]	[0.810]	[0.525]	[0.647]
Teacher development	-1.554	-0.039	-0.048	-0.114
	(1.885)	(0.053)	(0.043)	(0.074)
	[0.802]	[0.810]	[0.701]	[0.561]
Combined treatment	-0.931	-0.026	-0.056	-0.142
	(1.797)	(0.055)	(0.044)	(0.085)
	[0.894]	[0.894]	[0.627]	[0.517]
Observations	2766	2766	2766	2766
P-value: T^P=T^T	0.419	0.377	0.130	0.062
P-value: T^P=T^{T+P}	0.609	0.541	0.096	0.051
P-value: T^T=T^{T+P}	0.114	0.247	0.665	0.304
Mean control	52.677	0.891	0.909	0.759

Notes: The table presents ITT effects on teachers' contemporaneous characteristics. Controls include survey wave and randomization strata fixed effects. Square brackets below the point estimates show Romano-Wolf step-down p-values. Standard errors clustered at the school level are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C5: Intention-to-Treat: School Mobility

	Changed school
	(1)
Parental engagement	0.005
	(0.023)
Teacher development	0.012
	(0.040)
Combined treatment	-0.016
	(0.040)
Observations	2485
P-value: $T^P = T^T$	0.855
P-value: $T^P = T^{P+T}$	0.545
P-value: $T^T = T^{P+T}$	0.173
Share changed school (control)	0.059

Notes: The table presents ITT effects on children's mobility. The outcome variable is an indicator variable for whether the child changed school over the study period. Controls include survey wave, survey week, enumerator, and randomization strata fixed effects. p-values at the bottom of the table test the equality of coefficients (parental engagement T^P, teacher development T^T, and combined T^{P+T}). Standard errors clustered at the school level in parentheses. Share changed school is the control-group mean. *** p<0.01, ** p<0.05, * p<0.1.