EdWorkingPaper No. 25-1326

Beyond the One-Teacher Model: Experimental Evidence on Using Embedded Paraprofessionals as Personalized Instructors

Elizabeth Huffaker
University of Florida

Monica G. Lee
Stanford University

Helen Zhou
Harvard University

Carly D. Robinson
Stanford University

Susanna Loeb
Stanford University

Using embedded paraprofessionals to provide personalized instruction is a promising model for differentiating instruction within the classroom. This study examines two randomized controlled trials of paraprofessional-led tutoring in early-grade math and literacy. However, intent-to-treat (ITT) analyses revealed no overall achievement impacts for either program. We then explore two mechanisms that have surfaced in the tutoring literature as central efficacy moderators—dosage and tailoring—as plausible explanations to these results. While dosage was low for both programs, we estimate significant benefits from treatment assignment at higher-dosage campuses in numeracy (i.e., up to 0.28 SD at 80% progression) but no effect at any level of observed dosage on literacy. Curricular analysis revealed the literacy program's rigid structure may have impeded adaptation to student proficiency while student skill did not predict differences in numeracy program impacts. Supplemented by tutor survey data, these findings suggest that successful implementation of para-tutoring may depend on role prioritization, instructional coordination, and the use of student data to provide responsive instruction.

VERSION: November 2025

Beyond the One-Teacher Model: Experimental Evidence on Using Embedded Paraprofessionals as Personalized Instructors

Elizabeth Huffaker[1,*]

Monica Lee[2]

Helen Zhou[3]

Carly Robinson[2]

Susanna Loeb[2]

[1]College of Education, University of Florida, 1121 SW 5th Ave, Gainesville, FL 32601
[2]Graduate School of Education, Stanford University, 520 Galvez Mall, Stanford, CA, 94305
[3]Department of Sociology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138
[*]Corresponding author. Email: ehuffaker@ufl.edu

November 2025

Abstract – Using embedded paraprofessionals to provide personalized instruction is a promising model for differentiating instruction within the classroom. This study examines two randomized controlled trials of paraprofessional-led tutoring in early-grade math and literacy. However, intent-to-treat (ITT) analyses revealed no overall achievement impacts for either program. We then explore two mechanisms that have surfaced in the tutoring literature as central efficacy moderators—dosage and tailoring—as plausible explanations to these results. While dosage was low for both programs, we estimate significant benefits from treatment assignment at higher-dosage campuses in numeracy (i.e., up to 0.28 SD at 80% progression) but no effect at any level of observed dosage on literacy. Curricular analysis revealed the literacy program's rigid structure may have impeded adaptation to student proficiency while student skill did not predict differences in numeracy program impacts. Supplemented by tutor survey data, these findings suggest that successful implementation of para-tutoring may depend on role prioritization, instructional coordination, and the use of student data to provide responsive instruction.

Acknowledgements – We are grateful to our partners in this research for their invaluable contribution to project implementation. We thank the Arnold Ventures Foundation for their generous support of this research. We also thank the Smith Richardson Foundation for their support of our full research program. We received insightful feedback and research support from the National Student Support Accelerator team, in particular Sarah Novicoff, Evan Bennet, Nancy Waymack and Cynthia Pollard.

Beyond the One-Teacher Model: Experimental Evidence on Using Embedded Paraprofessionals as Personalized Instructors

Students learn best from instruction that is adapted to their level of understanding. A century after (Vygotsky, 1978) introduced the "Zone of Proximal Development", the idea that teaching should occur at a student's "learning edge" is still widely cited as a teaching best practice (Rietmulder, 2022; Shubilla & Sturgis, 2012). However, the conventional one-teacher classroom is not designed to consistently facilitate bridging moments for every student. Even the most skillful instructor will struggle to simultaneously assess and tailor their teaching across all children, each at the precipices of their unique learning edges. Tutoring programs in schools during the school day can help to address this need (Robinson et al., 2024), though the promise of tutoring to improve student learning through individualized responsiveness to student need has historically been limited to families with the means to employ private tutors (Szuba, 2020).

In recent years, academic crisis, recovery funding, and the wide dissemination of promising research findings have combined to motivate efforts to make personalized instruction more available in public schools. Evidence that the academic penalties of the COVID-19 pandemic were steepest among *already*-struggling students intensified calls for instruction that can generate transformative learning gains: teaching that is individualized and situated within a caring student-educator relationship, defined broadly as tutoring (Nickow et al., 2024; White et al., 2022). Studies of tutoring have found impacts at the upper range of all educational interventions. However, rapid scaling has uncovered new barriers and, in some cases, diluted benefits (Kraft et al., 2024). Schools aiming to implement effective and enduring personalized instruction may struggle, for example, to ensure a reliable supply of qualified instructors (Groom-Thomas et al., 2023).

The experimental studies described in this paper explore one district's attempt to sustainably expand its provision of relationship-based, individualized math and reading instruction to high-need, early-grade students by moving beyond the one-teacher classroom model. They did so by leveraging instructional staff already employed by the district and embedded in student's classrooms—paraprofessionals.[1] This para-tutor approach may help to overcome some typical barriers to effective tutoring implementation, but it also may be subject to these same barriers. In this study we explore program implementation and how it varied across schools, potentially explaining variability in their effectiveness.

A para-tutor approach using preexisting school staff stands to address challenges to effectively scaling personalized, relationship-based instruction along three key dimensions: cost, dosage, and alignment. First, recruiting and compensating a qualified tutor labor force is costly.

[1] While terminology varies across districts, we use "paraprofessional" to encompass the titles of instructional aide, teaching assistant, and paraeducator.

Para-tutors are a promising option, being less costly than teacher-tutors and likely more effective than volunteer-based models (Nickow et al, 2024). Schools already regularly hire and employ paraprofessionals, who may transition into a structured tutoring role.

Second, insufficient tutoring *quantity* has persistently hampered implementation and effectiveness (Groom-Thomas et al., 2023). While offering tutoring during the school day increases attendance and allows educators to better reach their least engaged students (Bhatt et al., 2024; Robinson et al., 2025), even in-school programs are sometimes vexed by lower than expected session attendance (Ready et al, 2024). One way using para-tutors could improve dosage is if they motivate stronger school engagement; in-school tutoring was shown to boost student school attendance in at least one randomized experiment (Lee et al., 2024). Indeed, paras may be uniquely situated to develop strong relationships with tutored students as they are more likely than classroom teachers to share close links with students' communities (Basile et al., 2022). However, tutoring dosage could still fall below expectations if para-tutors are balancing many demands on their in-school time, as with teachers (Ready et al., 2024).

Third, tutoring programs may not be well aligned with classroom work and they also mis-align with student abilities. When content covered in tutoring sessions is not tightly coordinated with grade-level standards, this can undermine the efficacy of purportedly-individualized instruction (Huffaker et al., 2025; TNTP, 2025). Paraprofessionals, because they are embedded in the classroom and the school, may be better able to align the tutoring with classroom work. Similarly, if curricula of tutoring programs—which are often scripted to reduce variance in tutor quality (Cortes et al., 2024; Markovitz et al., 2022)—are *too* rigid, instructors may not be able to provide the flexible and responsive teaching, aligned to students' abilities, that undergirds the logic of tutoring.

To better understand the potential of paraprofessionals as tutors, providing individualized instruction in classrooms, this study includes a pair of randomized-controlled trials (RCTs), one in numeracy and and one in literacy, that evaluate the causal impacts of two para-tutoring programs on outcomes for early-grade students with below-grade-level proficiency.

Our experimental analysis identifies no overall impact from para-tutoring on student achievement, while our sequential exploratory analyses point to deficits in dosage and curricular design as moderators. There is evidence from the math program that para-tutoring can improve learning when tutors deliver a majority of the curriculum. However, dosage was not a driving factor for the literacy tutoring program. The absence of effects appears to be partly explained by a curriculum that was not responsive to students' individual abilities, thus forsaking the affordances of personalized instruction. We draw comparative insights from these findings and present practical implications for districts and providers interested in deploying embedded paraprofessionals to provide individualized teaching. This study contributes to the tutoring research literature and also speaks broadly to the promise and challenge of moving beyond a one-teacher model of instruction.

Two Paraprofessional-Led Interventions

This paper explores two programs adopted to serve students who *enter* schooling below grade-level. Like most districts, the large, urban, East-Coast district that implemented these programs entered the 2022-23 school year facing substantial and disproportionate declines in student proficiency relative to pre-pandemic benchmarks. In response, the District invested in scaling up relationship-based, personalized learning during the school day—often referred to as "high-impact tutoring". The District was particularly concerned about the academic progress of very young students (e.g., Copeland et al., 2024).

To leverage existing relationships and instructional capital, the District used paraprofessionals already embedded in early-grade classrooms to facilitate both interventions. Almost all Kindergarten classrooms already had a dedicated paraprofessional. These paras have varied responsibilities, including interpreting for families, providing behavioral support, and undertaking clerical tasks. The role is flexible, with schools and teachers having considerable discretion over their usage. Both interventions directed paraprofessionals' time towards receiving training and personalizing instruction to below-grade level students using externally-developed, highly structured programs. While the programs shared a common implementation structure, they differed in session length, cadence, and group size. We detail their distinct features and research contexts below.

The literacy program

The first program implemented was an early literacy pilot for Kindergarten and first graders at 13 schools in fall 2022. This intervention featured a newly-developed, highly-scripted early literacy curriculum meant to be easily implemented by para-tutors. The curriculum is grounded in science of reading principles, evidenced by its heavy emphasis on phonemic awareness and phonics (National Reading Panel, 2000). The program also sought to include evidence-based features of high-impact tutoring. For example, students were expected to receive 15-minute in-class, one-on-one early literacy sessions daily. The para-tutors received weekly synchronous coaching from the program developer, using tutoring session recordings and dashboards to monitor student progress. The extensiveness and rigidity of the program is notable: the complete curriculum encompassed 180 sessions with a target of 140 sessions. Regardless of baseline proficiency, students had the same entry point and followed the same sequence. Paras were not advised to repeat or skip topics based on student progress.

Due to the pilot nature of this program, implementation did not begin until November, three months into the school year. As the program's first-ever implementation with paraeducators, many human and organizational factors were still being refined, which should be considered when interpreting these impacts. While meta-analyses of early literacy tutoring are promising (i.e., 0.24-0.41 SD effect size ranges) (Elbaum et al., 2000; Gersten et al., 2020; Neitzel et al., 2022) this evaluation aims to contribute a evidence on whether novel delivery model—*external* early literacy tutoring programs implemented by *school-employed* paras—would be effective.

The numeracy program

In fall 2023, the district introduced a para-led numeracy intervention adapted from an established model and curriculum. Small-scale pilot evaluations of the program had found positive impacts in other districts (Clarke et al., 2016, 2020). The district adhered to most of the program features tested in those studies: paraprofessional instruction was to occur in 20-minute increments to small groups three times a week. The total curriculum encompassed 50 sessions. Unlike the literacy program, however, tutors were encouraged to be responsive to demonstrated proficiency. Students were re-evaluated at regular intervals to help tailor instruction and paraprofessionals were provided with strategies (e.g., using intentional seating and post-lesson reflections) to adapt support within the small groups.

To reduce costs, the district removed a one-on-one coaching element that had been present in the pilots. Instead, training included only two provider-led workshops. This study therefore measures whether a promising program remained impactful in a larger and more diverse district, with less intensive implementation supports. Additionally, it contributes to a relatively scant research base on early math tutoring (Nickow et al., 2024).

Data and Samples

We use district administrative data to examine the effects of assignment to each program on student outcomes. Data cover AY 2022-23 Kindergarteners and first graders for the literacy program, and AY 2023-24 Kindergartners for the numeracy program. They include demographic information, English learner and special education status, and test scores.

Implementation metrics like intervention dosage are captured using provider data and para-tutor surveys for the literacy program and the numeracy program, respectively. We draw from surveys of numeracy para-tutors reflecting on their experiences and role perceptions, to supplement our core analyses. Finally, for detailed insight into program content we accessed the complete literacy tutoring curriculum and a substantial subset of instructional materials for the numeracy intervention.

Measures

We preregistered end-of-year test scores as the main outcome for each study. For literacy this is a Dynamic Indicators of Basic Early Literacy Skills (DIBELS-8) summative score, and for numeracy it is an i-Ready Math Diagnostic (i-Ready) summative score. We standardized scores within grade-level using the mean and standard deviation of the control group score distribution.

We similarly standardized i-Ready sub-scores (i.e., algebra, geometry, measurement, and numeracy) for exploratory analysis.

Student demographic characteristics and baseline scores—standardized in the same manner as endline scores—are used as covariates in our main regression specification. Our secondary analysis also considers the quantity of tutoring received by students for both studies. See Appendix A for details on treatment of control variables and measures of tutoring quantity.

Sample Descriptions

We construct an intent-to-treat (ITT) sample for each experiment. Table A1 summarizes the composition of each, among students with endline scores (additional details in Appendix A). Both samples broadly reflect the District make-up. Because the math program included only Kindergartners, no students overlap across samples.

The literacy study sample includes 222 Kindergarten and 68 first grade students. The number of tutoring seats available varied across schools due to a limited number of school staff who could serve as tutors at each school. Within each classroom and baseline reading-level stratum, researchers randomly assigned 103 students to treatment and 182 students to the control. Ultimately, our analytic sample included 270 students with endline data—207 in Kindergarten and 63 in first grade.

The numeracy study sample includes 1,069 Kindergartners identified as tutoring-eligible across 37 schools and 94 classrooms. Within each classroom, researchers randomly assigned 384 students to the treatment group and 849 students to the control group. Of these students, 1,023 have endline scores and are captured in our confirmatory analysis.

Study Design

The experimental design for both studies stratified randomization within classrooms to assign eligible-students into either the treatment (i.e., paraprofessional-led tutoring) or control condition. Figure A1 summarizes the study design used to evaluate both interventions.

In both studies, the number of students randomized into treatment within each classroom was limited to a case load reasonable for one embedded paraprofessional to tutor, generally either five (in the literacy program) or four (in the numeracy program) students. Control students continued to receive all supportive services and interventions they usually would during their "business-as-usual" (BaU) course of schooling.

We test for balance across conditions with auxiliary regressions where pretreatment student traits are regressed on an indicator for assignment-to-treat and design controls. The results in Table A2 are consistent with successful randomization. Table A3 confirms that availability of endline data (i.e., attrition) is balanced across conditions.

Estimation Strategy

The following specification is used to estimate effects of assignment to treatment in each program:

$$Y_{ij} = \beta_0 + \beta_1 Treatment_{ij} + \beta_2 Baseline_i + \alpha_j + \gamma X_i + \varepsilon_{ij}$$

Where Y_{ij} is an outcome for student i in classroom j. The confirmatory outcome for the literacy and math programs are EOY standardized, composite DIBELS-8 and i-Ready scores, respectively. Other outcomes include sub-scores and session attendance. β_1 is the ITTeffect estimate. Additional student-level controls include an analogous BOY score ($Baseline_i$) and a vector of pre-treatment covariates capturing student gender, race/ethnicity, and English Learner and Special Education statuses. Randomization strata are controlled for with classroom fixed effects (α_j). Finally, we use heteroskedasticity robust standard errors (ε_{ij}). Because randomization was at the individual level, clustering standards errors would likely be overly conservative, increasing the risk of type II errors (Abadie et al., 2023). Still, we test the robustness of our results to allow for clustering at the classroom level classroom (Table A4). Appendix A notes on minor estimation differences between studies.

Moderator analysis

We unpack our topline findings by considering two central determinants of the efficacy of relationship-based, personalized learning: instructional *dosage* and *alignment*. To explore the role of dosage we estimate treatment-on-the-treated (TOT) effects using a two-stage least squares strategy and also look for heterogeneity in ITT effects by school-level average dose. For the latter strategy, we leverage the fact that our within-classroom randomization structure allows us to estimate internally valid school-level ITT effects by classifying schools based on the *average* quantity of tutoring received across all students assigned to treatment (e.g., 50% of intended, 60% of intended and so forth). We then estimate the effect of assignment across progressively higher levels of school dosage.

Next, we consider whether the degree of program adaptability to varied student academic needs (i.e., alignment between the content focus and students' proficiency) moderates the benefit of these programs. First, we assess whether the curricula and practice guides facilitate adaptable, tailored instruction or enforce a rigid, uniform structure (see Appendix A for more details on this process). We then explore adaptability-to-students empirically by interacting the treatment assignment indicator with baseline student score to predict tutoring efficacy by BOY proficiency. The estimated difference in the relationship between BOY and EOY scores between the treatment and control group captures variation in the benefits accrued from tutoring for differently-skilled students. We expect larger effects for students receiving tutoring that better aligns to their skill level.

Results

Table 1 summarizes the results of our main, preregistered analysis and indicates that, on average, assignment to neither program generated learning gains on summative EOY assessments. ITT estimates for both interventions have magnitudes below 0.05 standard deviations and are statistically insignificant. Heterogeneity analyses (Table A5) disaggregate impacts by pretreatment traits including grade-level and gender. We identify null results across these subsamples. Estimated effects on subscores for the numeracy study (Table A6, Column 1) are also null.

Dosage

Given the challenge many relationship-based, individualized interventions face in providing sufficient dosage (e.g., Makori et al., 2024) we use measures of tutoring *quantity* to begin unpacking our results. We observe substantial variation in the amount of tutoring that students assigned to treatment actually receive. Figure 1 presents these session count distributions by program and treatment assignment and Table A7 provides formal first stage estimates on the extensive (i.e., received any tutoring) and intensive (i.e., number of sessions) margins.

The literacy program curriculum was composed of 180 15-minute sessions, with a target of 140 sessions. However, zero students assigned to treatment (panel A) received the target dosage, and fewer than 20% received even 60% (i.e., 84 sessions) of the intended programming. This is partly explained by the delay of the start of implementation to November. On average, students assigned to treatment received 42 sessions of tutoring (i.e., 30% of program target) and students initially assigned to the control group received 1.9 sessions due to being pulled into treatment from the waitlist. Treatment on the treated (TOT) estimates for the literacy program were not statistically significant (Table A7).

Take-up was also much lower than the intended 50 sessions for the math numeracy program – nearly half the students assigned to treatment received zero tutoring. Conversations with our district partners indicated that classroom teachers in these schools may not have been consulted before administrators opted their campus into the study and thus they never intended to implement the program. Among all students assigned to treatment, the average number of sessions attended was only 18 (i.e., 36% of the program), yielding null average treatment on the treatment estimates (Table A7). Among the 65% of treatment group students that attended at least one session, the average number of sessions attended increased to 25—half the intended dose. Students assigned to the control group attended an average of one session of tutoring.

The average school-level tutoring dosage among students assigned to treatment is summarized in Figure 2. Among schools that implemented the early literacy program, only three schools completed 50% of the program, or 70 or more sessions. Among schools that implemented the numeracy program, about a third (i.e., 12 schools) had treatment students progress through at least half of the curriculum and only 4 schools had average dosage over 80%.

To examine ITT effects by school-level dosage we progressively restrict our sample to subsets observations from schools with treatment-group dosage at or above ascending thresholds (i.e., 40%, 50%, 60% of target dosage etc.). Because recommended and actual dosage varies meaningful across the literacy and numeracy programs—50 versus 140 sessions—we use different thresholds for each study.

Figure 3 plots the coefficients of interest from these regressions (see Table A8 for complementary tables). Point estimates from the literacy study remain very small (ES <0.1 SD) and imprecise. However, for the numeracy study, we observe a broadly linear increase in the magnitude of ITT effects as we restrict the numeracy study sample to higher levels of average school dosage. Including schools where students received between 1 and 24 sessions, ITT effects are precisely zero, however among only schools where treated students received at least 60% of the curriculum (i.e., 30 sessions or approximately 600 minutes of tutoring) the point estimate grows to 0.13 SD and approaches statistical significance.

The dosage level at which numeracy tutoring begins to show improvements in math achievement is substantially higher than the *average* first-stage effect estimated in Table A7 (i.e., 17 sessions of numeracy tutoring). This discrepancy helps explain why the estimated treatment effects among *the high-dosage sub-sample* are statistically significant positive even while the *full sample* ITT and TOT estimates are nonsignificant and slightly negative. Furthermore, when average dosage increases to 80%, that effect size grows to nearly three-tenths of a standard deviation (ES=0.28 SD, p<0.01), equivalent to roughly three months of math learning at the Kindergarten level (Bloom et al, 2011) and on par with RCT results from highly effective tutoring programs (Nickow et al, 2024). Consistent with the program emphasis, the same pattern emerges on the i-Ready numeracy and measurement subscores, but less so on the algebra subscore and not at all on the geometry subscore (Table A6).

Supplemental balance tests in the high-dosage schools support the *internal validity* of the positive estimated numeracy program impacts. However, the *generalizability* of these results is less clear. It is plausible, even likely, that high-dosage schools are different from low-dosage ones either in their tutoring implementation strategy or on pre-treatment dimensions.

We use school-level data and tutor surveys to weigh explanations for the variance in observed dose. We consider three contextual dimensions: school-level characteristics independent to the intervention, experiences of paraprofessionals, and program implementation practices. First, it is plausible that school-level features could moderate detected effects. For instance, the average student at schools with a strong (or weak) attendance rate would naturally receive higher (or lower) program dosage. Overall attendance rates were slightly lower at high dosage schools (i.e., 60%+ of program delivered; N=10) versus low dosage schools (N=26). The share of high-dosage school students with more than 10 absences was 35%, compared with only 24% among low-dosage schools.

Tutoring impacts might also be attenuated at wealthier schools if parents use their resources to seek out-of-school tutoring comparable to the treatment. High-dosage campuses were indeed slightly poorer than low-dosage ones, though differences in economic disadvantage rate (58% vs. 54%) are not statistically significant. These results suggest, consistent with an opportunity equalizing motivation for providing in-school tutoring, that para-tutoring effects could be stronger where students are more likely to miss school (and therefore acquire gaps in classroom content learning) and less likely receive supplementary private instruction.

To better understand why these schools provided more tutoring, we turn to surveys fielded for the numeracy study (see Appendix A for details). While only a fifth of the 78 paraprofessionals who took the baseline survey responded to the end-of-year survey, limiting their role in our main analysis, the responses are qualitatively illustrative. First, para-tutors report balancing many “primary” responsibilities—just over four duties on average—among tutors at both high- *and* low-dosage schools. So, although schools implementing personalized learning with embedded paraprofessionals should be wary of over-burdening para-tutors, our data do not confirm that role *quantity* explains dosage. However, role type may be relevant. Three quarters of survey respondents at high-dosage schools indicated that 1:1 instruction was their primary role, while only half at low-dosage schools did.

This prioritization of para-led individualized instruction reflected intentional school practices for at least one high-dosage campus. According to district leaders, this exemplary school built in daily work time for the paraprofessional and classroom teacher to coordinate on instructional tasks such as reviewing student progress. Facilitating coordination between the Tier 1 and Tier 2 instructor aligns with recommended practice (TNTP, 2025) and contrasts with open-ended feedback from one para-tutor at a low-dosage campus: “It was hard to tell if the intervention was a priority because the teacher I worked with did not share/discuss any of the students performance data with me”. While embedded paras are well-situated to be powerful instructional resources, these perspectives from the field suggest successful implementation may require structured priority-setting and collaboration.

Adaptability and alignment

While a dosage explanation is consistent with our pattern of results for the numeracy program, it does not explain the null results of the literacy study. Despite sharing a setting and similar program model to the numeracy intervention, assignment to the literacy program generated no reading gains even at schools where treated students received over 50% of the 140-session target curriculum. Furthermore, the sub-set of the ten schools that adopted the numeracy program one year after adopting the literacy program are distinguished neither positively (i.e., evidence of year-to-year improvement in tutoring implementation) nor negatively (i.e., are particularly poor at facilitating para-tutoring) in year-two dosage. Therefore, results are unlikely to be due to site-specific implementation challenges and we explore the potential impact of program features.

The literacy program is characterized by an extensive (i.e., 140-session target) sequence that all students follow from the beginning, which likely limits the capacity for paraprofessionals to adapt the curriculum according to student proficiency. Our curricular analysis reveals that the program involves extensive repetition of foundational knowledge across many sessions. As a result, the introduction of all 44 English phonemes is not completed until the second half of the target curriculum, which many students never reached. Because all students were required to begin with lesson 1, many students were tutored on topics they had already mastered.

To empirically assess whether this program model constrained learning for higher achieving students, we regress EOY scores on an interaction between treatment and baseline student proficiency. Figure 4 presents the conditional predicted scores using the resulting margins. For the literacy program, predicted scores for lower achieving students assigned to treatment are indeed higher than those of control group students with the converse being true for higher achieving students. Specifically, the difference in slopes presented in Figure 4 is -0.270, $p<0.01$. For the numeracy program, however, we identify no relationship between BOY skill-level and treatment status with respect to EOY scores. These suggestive results are consistent with the interpretation that, despite using a one-to-one rather than small-group delivery model, the literacy program was less personalizable than the numeracy program, particularly for higher achieving students.

Discussion

We examine two experiments testing early-grade math and literacy instructional interventions, delivered by embedded paraprofessionals using a structured curriculum. Across both studies, we find no average effects on student achievement. However, the implications of these results for practice are limited without greater understanding of relevant program mechanisms and the contexts of these program implementations. We therefore investigated two key potential moderators—tutoring dosage and tailoring of instruction to student need—using rich implementation data. To motivate future research questions and contribute to continuous improvement of practice we also investigated these mechanisms qualitatively. While exploratory, we bring to bear a wealth of supplementary data to these analyses, including paraprofessional surveys, school-level administrative data, contextual detail from our District partners, and reviews of program curricula.

With respect to tutoring quantity, we find that when the average school-level dosage of the numeracy program exceeds approximately 50%, the magnitude of the ITT coefficients begin increasing. This suggests that, when implemented with moderate fidelity, the numeracy program boosts student learning. At least one high-dosage school cited the introduction of schedule time for instructional coordination between the teacher and the paraprofessional as an actionable strategy to boost implementation quality for in-classroom para-tutoring. Notably, this linearly increasing relationship between dosage and impacts resembles the pattern of results found across another series of tutoring RCTs evaluated by Bhatt et al., (2025).

The literacy program, however, improved reading skills for students with very low baseline proficiency but showed no measurable impacts for those with higher proficiency. Unlike the numeracy program where students were re-assessed at regular intervals and were not required to progress in a strictly linear fashion—students could skip topics, and tutors were also empowered to provide additional practice when students struggled—this program had a linear structure that unfolded at a slower pace. The curriculum content is evidence-based, as the approach seems to benefit students in need of foundational development. However, it is unlikely to challenge higher achieving students until well into the program, particularly given that prior meta-analytic research has also identified diminishing returns to reinforcement of foundational skills phonemic awareness (Erbeli et al., 2024). A simulation study of sequential tutoring strategies in math found that an adaptive rather than a foundations-for-all approach is likely to promote more learning (TNTP, 2025).

It is possible that higher achieving students may benefit from the literacy program if they reach more advanced curricula—especially since this was only the first year of implementation. The provider reports that subsequent implementations have improved dosage by an additional 20 sessions on average, and we anticipate that ongoing evaluations will shed further light on the program's evolution and effectiveness. However, in this sample, no schools reached 70% of target session completion, suggesting a supplemental instructional program that aims to span three quarters of standard school days may be unrealistic. Efforts to scale relationship-based, individualized interventions using non-teacher tutors (e.g., volunteers, paraprofessionals) must balance the benefits of clearly defining their program structure to reduce variance in instructional quality with the risks that, if taken too far, such a model quashes the transformative promise of tutoring to meet students where they are.

While these studies demonstrate several affordances of partnership-based, field-informed experiments in surfacing implementation considerations (e.g., the importance of coordination between paras and teachers; potential tensions between scripting versus tailoring), we also note limitations of this current work that ought to motivate future research. First, survey data were only available for the numeracy study, and response rates were too low for systematic analysis. Developing greater insight into the perspectives and experiences of para-tutors should be a priority of future research in this area. Second, the sample size of the literacy study was limited, so a larger-scale evaluation as the program continues to be iterated upon may be warranted.

Still, by conducting randomized evaluations of two interventions with partially overlapping features in a common setting, we have developed comparative insights relevant to developers and facilitators of similar programming. We conclude by highlighting three central takeaways. First, even when a designated tutor is already present within a student's classroom, dosage remains a major challenge for interventions aiming to scale in-school, relationship-based, personalized instruction. High cadence (i.e., daily) programs need to anticipate how the realities of the school day may interfere with implementation. Second, findings from the high-dosage campuses show that using paraprofessionals to move beyond the one-teacher model for classroom instruction is a promising strategy, but they must be intentionally integrated into the teaching and learning process. Schools should foster aligned supports, clear priorities, and structured collaboration with the classroom teacher. Third, even scripted intervention programs ought to use student achievement data to re-asses pacing and sequencing at reasonable intervals.

The results of our studies of para-led interventions in Kindergarten and first grade highlight the potential for paraprofessionals to help students learn on top of their many other contributions to students' classrooms and communities (e.g., as interpreters for families, as many cited in our survey). However, the absence of broad-based impacts highlight barriers to paras effective deployment as in-class tutors. The effectiveness of paraprofessionals as instructional leaders hinges on intentional program design and implementation, particularly in calibrating appropriate dosage and alignment with student skill progression.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When Should You Adjust Standard Errors for Clustering?*. *The Quarterly Journal of Economics*, *138*(1), 1–35. https://doi.org/10.1093/qje/qjac038

Basile, C. G., Maddin, B. W., & Audrain, R. L. (2022). *The Next Education Workforce: How Team-Based Staffing Models Can Support Equity and Improve Learning Outcomes*. https://rowman.com/ISBN/9781475867275/The-Next-Education-Workforce-How-Team-Based-Staffing-Models-Can-Support-Equity-and-Improve-Learning-Outcomes

Bhatt, M. P., Chau, T., Condliffe, B., Davis, R., Grossman, J., Guryan, J., Ludwig, J., Magnaricotte, M., Mattera, S., Momeni, F., Oreopoulos, P., & Stoddard, G. (2025). *Personalized Learning Initiative Interim Report: Findings from 2023-24*. https://educationlab.uchicago.edu/resources/personalized-learning-initiative-interim-report-findings-from-2023-24/

Bhatt, M. P., Guryan, J., Khan, S. A., LaForest-Tucker, M., & Mishra, B. (2024). *Can Technology Facilitate Scale? Evidence from a Randomized Evaluation of High Dosage Tutoring*. National Bureau of Economic Research.

Clarke, B., Doabler, C. T., Smolkowski, K., Baker, S. K., Fien, H., & Strand Cary, M. (2016). Examining the Efficacy of a Tier 2 Kindergarten Mathematics Intervention. *Journal of Learning Disabilities*, *49*(2), 152–165. https://doi.org/10.1177/0022219414538514

Clarke, B., Doabler, C. T., Turtura, J., Smolkowski, K., Kosty, D. B., Sutherland, M., Kurtz-Nelson, E., Fien, H., & Baker, S. K. (2020). Examining the efficacy of a kindergarten mathematics intervention by group size and initial skill: Implications for practice and policy. *The Elementary School Journal*, *121*(1), 125–153. https://doi.org/10.1086/710041

Clarke, B., Turtura, J., Lesner, T., Cook, M., Smolkowski, K., Kosty, D., & Doabler, C. T. (2022). A Conceptual Replication of a Kindergarten Math Intervention Within the Context of a Research-Based Core. *Exceptional Children*, *89*(1), 42–59. https://doi.org/10.1177/00144029221088938

Copeland, K. A., Porter, L., Gorecki, M. C., Reyner, A., White, C., & Kahn, R. S. (2024). Early Correlates of School Readiness Before and During the COVID-19 Pandemic Linking Health and School Data. *JAMA Pediatrics*, *178*(3), 294–303. https://doi.org/10.1001/jamapediatrics.2023.6458

Cortes, K., Kortecamp, K., Loeb, S., & Robinson, C. (2024). *A Scalable Approach to High-Impact Tutoring for Young Readers: Results of a Randomized Controlled Trial* (Working Paper No. 32039). National Bureau of Economic Research. https://doi.org/10.3386/w32039

Elbaum, B., Vaughn, S., Tejero Hughes, M., & Watson Moody, S. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, *92*(4), 605–619. https://doi.org/10.1037/0022-0663.92.4.605

Erbeli, F., Rice, M., Xu, Y., Bishop, M. E., & Goodrich, J. M. (2024). A Meta-Analysis on the Optimal Cumulative Dosage of Early Phonemic Awareness Instruction. *Scientific Studies of Reading*, *28*(4), 345–370. https://doi.org/10.1080/10888438.2024.2309386

Gersten, R., Haymond, K., Newman-Gonchar, R., Dimino, J., & Jayanthi, M. (2020). Meta-Analysis of the Impact of Reading Interventions for Students in the Primary Grades. *Journal of Research on Educational Effectiveness*, *13*(2), 401–427. https://doi.org/10.1080/19345747.2019.1689591

Groom-Thomas, L., Leung, C., Loeb, S., Pollard, C., Waymack, N., & White, S. (2023). Challenges and Solutions: Scaling Tutoring Programs. *IDB Publications*. https://doi.org/10.18235/0005070

Heinrich, C. J., Burch, P., Good, A., Acosta, R., Cheng, H., Dillender, M., Kirshbaum, C., Nisar, H., & Stewart, M. (2014). Improving the Implementation and Effectiveness of Out-of-School-Time Tutoring. *Journal of Policy Analysis and Management*, *33*(2), 471–494. https://doi.org/10.1002/pam.21745

Huffaker, E., Robinson, C. D., Bardelli, E., White, S., & Loeb, S. (2025). *When interventions don't move the needle: Insights from null results in education research*. EdWorking Papers. https://doi.org/10.26300/58DD-6D02

Korbey, H. (2024, May 2). Kindergarten math is often too basic. Here's why that's a problem. *The Hechinger Report*. http://hechingerreport.org/kindergarten-math-is-often-too-basic-heres-why-thats-a-problem/

Kraft, M. A., Schueler, B. E., & Falken, G. (2024). *What Impacts Should We Expect from Tutoring at Scale? Exploring Meta-Analytic Generalizability*. edworkingpapers.com. https://doi.org/10.26300/ZYGJ-M525

Lee, M. G., Loeb, S., & Robinson, C. D. (2024). *Effects of High-Impact Tutoring on Student Attendance: Evidence from the OSSE HIT Initiative in the District of Columbia*. https://doi.org/10.26300/WGHB-4864

Makori, A., Burch, P., & Loeb, S.. (2024). *Scaling High-impact tutoring: School Level Perspectives on Implementation Challenges and Strategies*. https://doi.org/10.26300/H8Z5-T461

Markovitz, C. E., Hernandez, M. W., Hedberg, E. C., & Whitmore, H. W. (2022). Evaluating the Effectiveness of a Volunteer One-on-One Tutoring Model for Early Elementary Reading Intervention: A Randomized Controlled Trial Replication Study. *American Educational Research Journal*, *59*(4), 788–819. https://doi.org/10.3102/00028312211066848

National Reading Panel. (2000). *Teaching Children To Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. National Institute of Child Health and Human Development.

Neitzel, A. J., Lake, C., Pellegrini, M., & Slavin, R. E. (2022). A Synthesis of Quantitative Research on Programs for Struggling Readers in Elementary Schools. *Reading Research Quarterly*, *57*(1), 149–179. https://doi.org/10.1002/rrq.379

Nickow, A., Oreopoulos, P., & Quan, V. (2024). The Promise of Tutoring for PreK–12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence. *American Educational Research Journal*, *61*(1), 74–107. https://doi.org/10.3102/00028312231208687

Parker, D. C., Nelson, P. M., Zaslofsky, A. F., Kanive, R., Foegen, A., Kaiser, P., & Heisted, D. (2019). Evaluation of a Math Intervention Program Implemented With Community Support. *Journal of Research on Educational Effectiveness*, *12*(3), 391–412. https://doi.org/10.1080/19345747.2019.1571653

Ready, Douglas D., McCormick, Sierra G., & Shmoys, Rebecca J. (n.d.). *The Effects of In-School Virtual Tutoring on Student Reading Development: Evidence from a Short-Cycle Randomized Controlled Trial*. https://doi.org/10.26300/569P-WZ78

Rietmulder, J. (2022). The Learning Edge. *The Circle School*. https://www.circleschool.org/more-to-explore/writings/the-learning-edge/

Robinson, C. D., Bisht, B., & Loeb, S. (2025). The Inequity of Opt-in Educational Resources and an Intervention to Increase Equitable Access. *Educational Researcher*, *54*(6), 328–338. https://doi.org/10.3102/0013189X251331518

Robinson, C., Pollard, C., Novicoff, S., White, S., & Loeb, S. (2024). The Effects of Virtual Tutoring on Young Readers: Results From a Randomized Controlled Trial. *Educational Evaluation and Policy Analysis*. https://journals.sagepub.com/doi/10.3102/01623737241288845

Shubilla, L., & Sturgis, C. (2012). Supporting Student Success in a Competency-Based Learning Environment. *CompetencyWorks Issue Brief.*

Szuba, L. (2020). Tutor and Tutoring in the History of Education (to the Great French Revolution). *21st Century Pedagogy*, *4*(1), 49–59. https://doi.org/10.2478/ped21-2020-0008

TNTP. (2025). *Unlocking Algebra*. https://tntp.org/publication/unlocking-algebra/

Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-Analyses of the Effects of Tier 2 Type Reading Interventions in Grades K-3. *Educational Psychology Review*, *28*(3), 551–576. https://doi.org/10.1007/s10648-015-9321-7

White, S., Groom-Thomas, L., & Loeb, S. (2022). *Undertaking Complex but Effective Instructional Supports for Students: A Systematic Review of Research on High-Impact Tutoring Planning and Implementation. EdWorkingPaper No. 22-652*. Annenberg Institute for School Reform at Brown University. https://eric.ed.gov/?id=ED625876

Table 1: Intent-to-Treat (ITT) Effects on EOY Performance

	Literacy Program: DIBELS Summative (SD)		Numeracy Program: i-Ready Summative (SD)	
	(1)	(2)	(3)	(4)
ITT	0.0549	0.0513	-0.0397	-0.0222
	(0.1109)	(0.0643)	(0.0579)	(0.0527)
Female		-0.2489*		0.0013
		(0.0995)		(0.0494)
White		-0.4767+		0.2079+
		(0.2764)		(0.1111)
Black		-0.3641		-0.3599**
		(0.2497)		(0.1181)
Hispanic		-0.2244		-0.2121
		(0.3236)		(0.1305)
English Learner		0.1519		0.0032
		(0.1887)		(0.1055)
Special Education		-0.2648		-0.1262
		(0.1617)		(0.0802)
Baseline score (sd)		0.4895***		0.3483***
		(0.0628)		(0.0310)
Constant	0.0000	0.4832+	0.0369	0.3752+
	(0.0763)	(0.2622)	(0.1764)	(0.2046)
N	270	270	1023	1023

Note: Heteroskedasticity robust standard errors in parentheses. Models examining ITT effects for the reading program sample (Cols 1 and 2) include a variable for floor scorers, an indicator that equals one if a student scored at the minimum possible value in their baseline early literacy score; coefficients for this variable are omitted from display. All regressions control for randomization strata. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Figure 1. Student-Level Tutoring Attendance

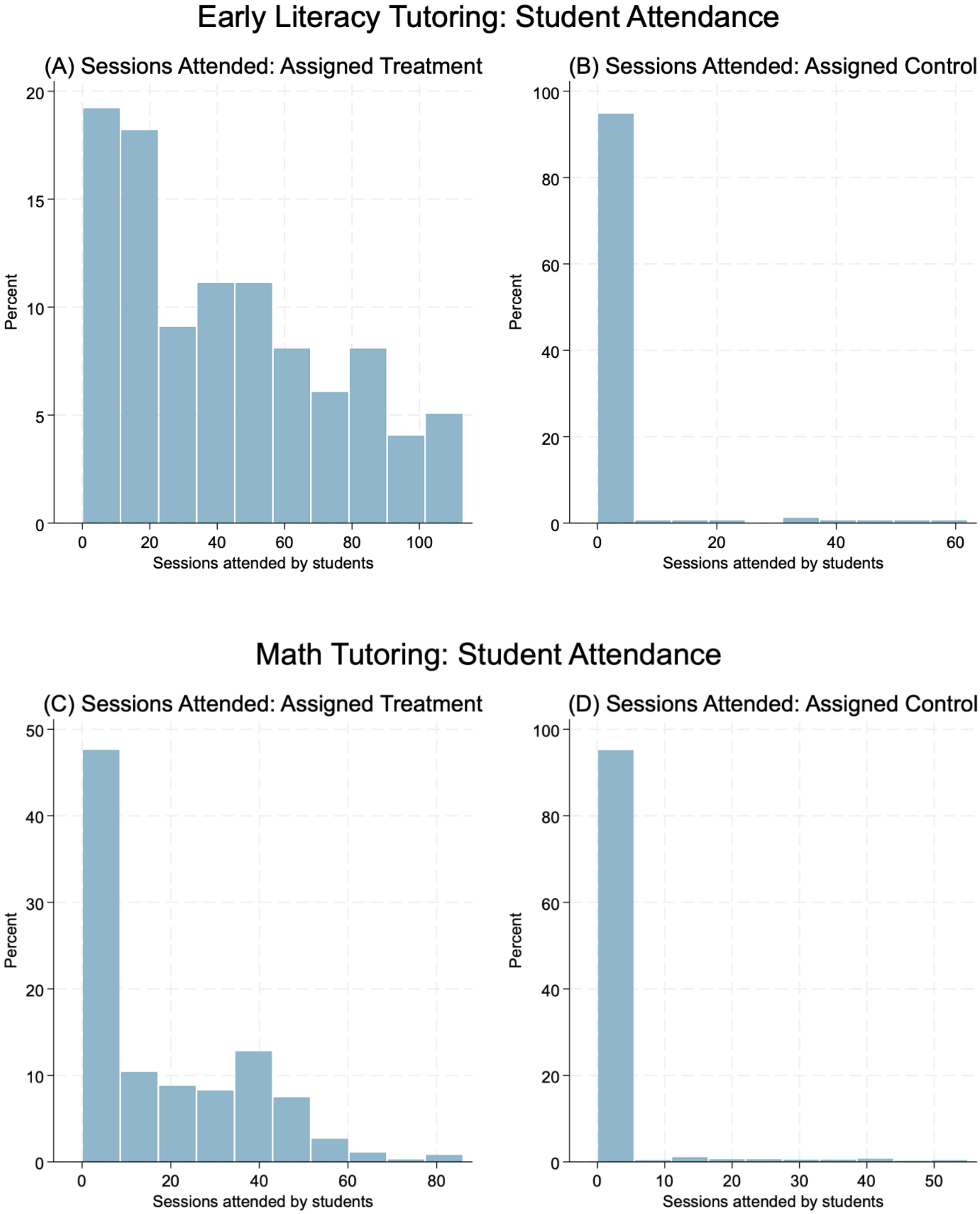
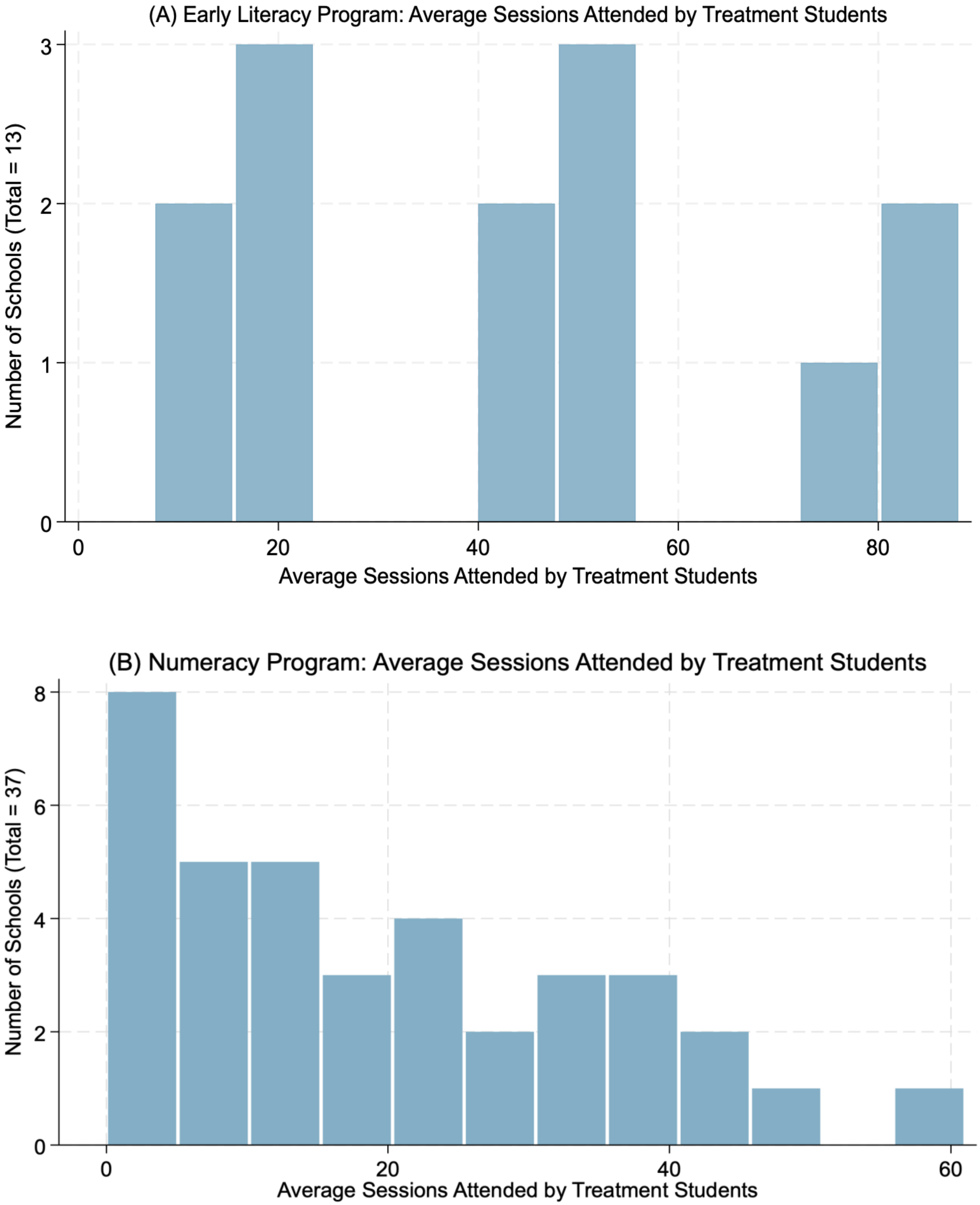
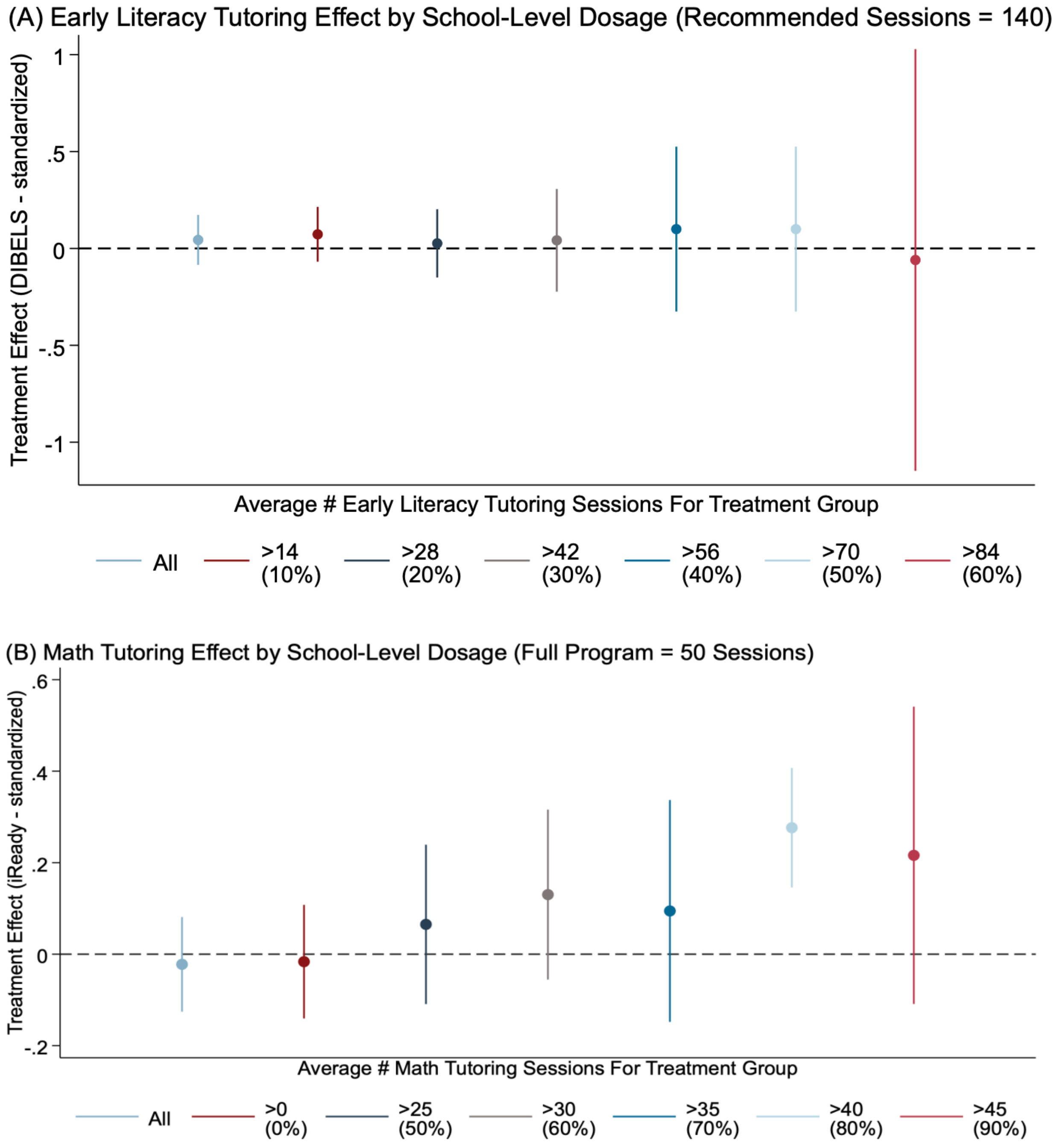


Figure 2. School Level Tutoring Attendance


(A) Early Literacy Program: Average Sessions Attended by Treatment Students

(B) Numeracy Program: Average Sessions Attended by Treatment Students

Figure 3. Intent-to-Treat Effects by School-Level Dosage



(A) Early Literacy Tutoring Effect by School-Level Dosage (Recommended Sessions = 140)

(B) Math Tutoring Effect by School-Level Dosage (Full Program = 50 Sessions)

Figure 4. Predicted EOY Scores by Baseline Achievement and Treatment Status

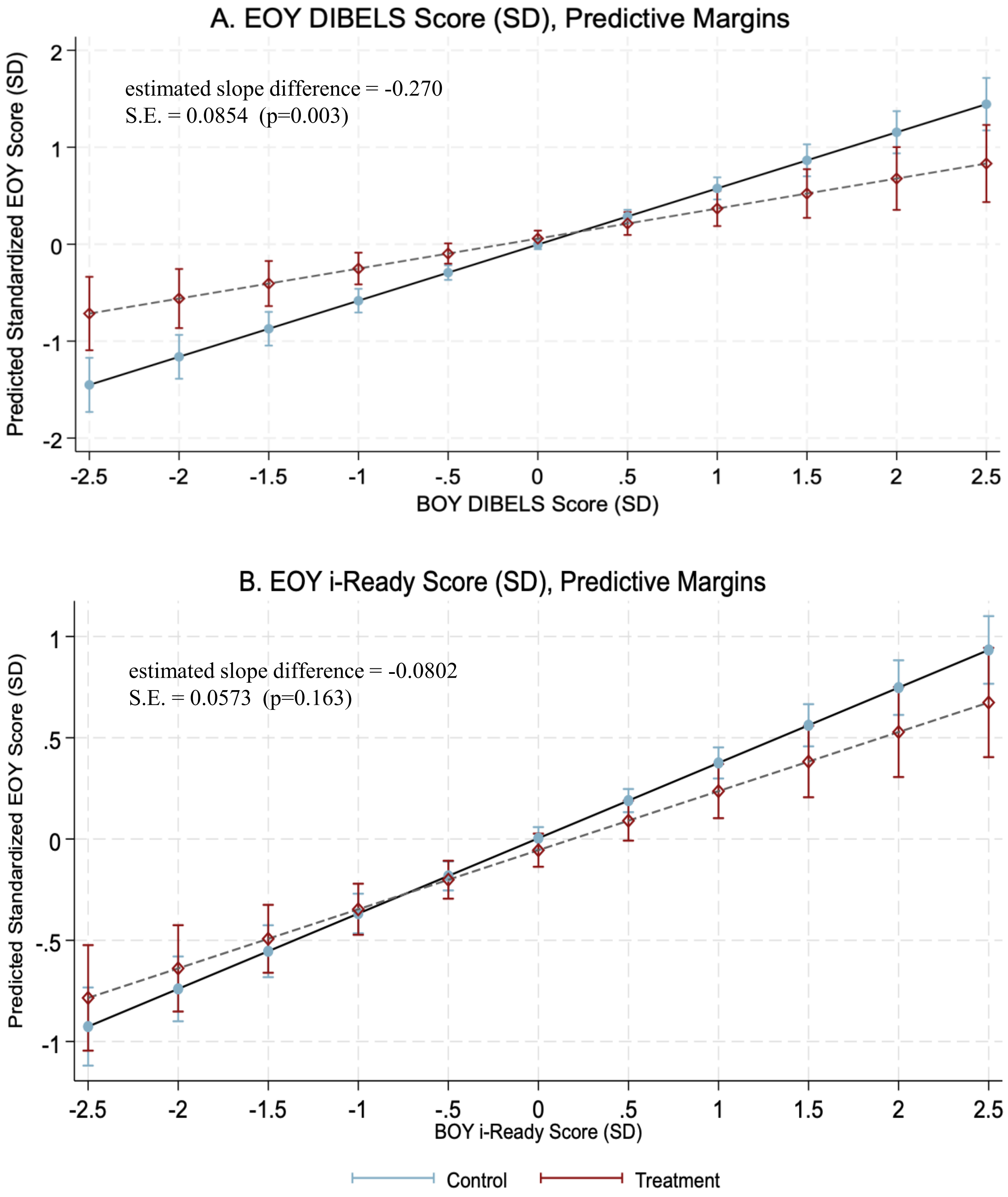


Figure A1. Summary of study designs

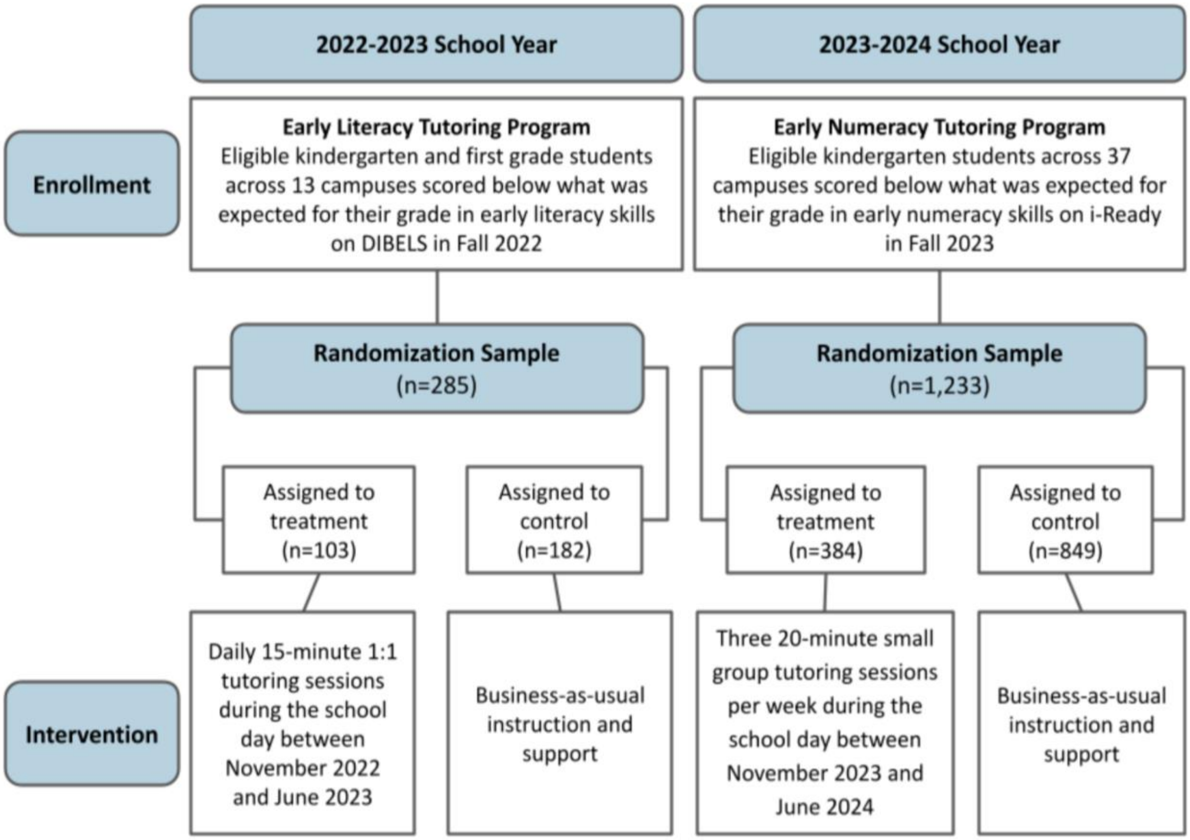

2022-2023 School Year
2023-2024 School Year
Enrollment
Early Literacy Tutoring Program
Eligible kindergarten and first grade students across 13 campuses scored below what was expected for their grade in early literacy skills on DIBELS in Fall 2022
Early Numeracy Tutoring Program
Eligible kindergarten students across 37 campuses scored below what was expected for their grade in early numeracy skills on i-Ready in Fall 2023
Randomization Sample (n=285)
Randomization Sample (n=1,233)
Assigned to treatment (n=103)
Assigned to control (n=182)
Assigned to treatment (n=384)
Assigned to control (n=849)
Intervention
Daily 15-minute 1:1 tutoring sessions during the school day between November 2022 and June 2023
Business-as-usual instruction and support
Three 20-minute small group tutoring sessions per week during the school day between November 2023 and June 2024
Business-as-usual instruction and support

Table A1: Summary Statistics for Intent-to-Treat Samples (Among Students with Endline Data)

	Literacy Tutoring Sample		Numeracy Tutoring Sample	
	Mean	SD	Mean	SD
(A) Baseline Characteristics				
Female	0.49		0.50	
Black	0.69		0.61	
Hispanic	0.26		0.19	
White	0.04		0.05	
Other Ethnic/Racial Identity	0.01		0.15	
Special Education	0.13		0.15	
English Learner	0.24		0.16	
Kindergarten	0.77		1.00	
First Grade	0.23		0.00	
(B) Pre-Intervention Test Scores (Standardized)				
K BOY i-Ready composite	-0.07	(0.99)	-0.03	(0.97)
K BOY DIBELS composite	0.02	(1.03)	.	.
1 BOY i-Ready composite	-0.12	(1.01)	.	.
1 BOY DIBELS composite	-0.06	(1.00)	.	.
(C) Post-Intervention Test Scores (Standardized)				
K EOY i-Ready composite	-0.05	(0.95)	-0.02	(1.01)
K EOY DIBELS composite	0.04	(0.92)	.	.
1 EOY i-Ready composite	-0.08	(0.98)	.	.
1 EOY DIBELS composite	-0.05	(0.93)	.	.
Number of Students	290		1023	
Number of Schools	16		37	
Number of Classrooms	46		92	

Notes: BOY denotes "Beginning-of-year", EOY denotes "end-of-year"

Table A2: Baseline Characteristics by Tutoring Assignment Among ITT Sample

	Literacy Program			Numeracy Program		
	Control	Treatment	Strata Adjusted Difference	Control	Treatment	Strata Adjusted Difference
	(1)	(2)	(3)	(1)	(2)	(3)
Female	0.500	0.475	0.0329	0.498	0.500	0.0113
	(0.030)	(0.037)	(0.0470)	(0.019)	(0.027)	(0.0341)
White	0.042	0.055	-0.0341	0.150	0.166	0.0026
	(0.012)	(0.017)	(0.0256)	(0.013)	(0.020)	(0.0199)
Black	0.683	0.646	0.0157	0.618	0.590	-0.0205
	(0.028)	(0.036)	(0.0576)	(0.018)	(0.027)	(0.0259)
Hispanic	0.261	0.276	0.0458	0.181	0.203	0.0285
	(0.026)	(0.033)	(0.0478)	(0.014)	(0.022)	(0.0231)
Other Ethnic/Racial Identity	.	.	.	0.051	0.041	-0.0106
	.	.	.	(0.008)	(0.011)	(0.0136)
English Learner	0.239	0.249	0.0568	0.159	0.172	0.0153
	(0.025)	(0.032)	(0.0399)	(0.014)	(0.020)	(0.0211)
Special Education	0.127	0.138	-0.0420	0.148	0.148	-0.0023
	(0.020)	(0.026)	(0.0497)	(0.013)	(0.019)	(0.0225)
Beginning-of-year i-Ready composite (SD)	.	.	.	-0.011	-0.094	-0.0794
	.	.	.	(0.037)	(0.051)	(0.0568)
Beginning-of-year DIBELS composite (SD)	0.004	-0.000	-0.0317	.	.	.
	(0.060)	(0.074)	(0.0872)	.	.	.
Joint P-Value			0.521			0.691
N Students	182	103	285	725	344	1069
Number of classroom randomization blocks	25	15	40	94	94	94

Note: This table reports baseline charctertistics by treatment condition and program for the sample of students identified as eligible for tutoring in each study cohort. Columns 1, 2, 4, and 5 report the variable means and standard deviations for each group as indicated. Column 3 and 6 reports the adjusted difference between the group of students assigned to the control group versus students assigned to treatment group, controlling for randomization strata. The demographic categories used are what is reported in the district's administrative data set and we recognize are not representative of the full range of student identities and experiences. We do not know how students are placed into these categories. All students are marked as either Male or Female. No students are listed in more than one racial/ethnic category. * p<0.10, ** p<0.05, *** p<0.010.

Table A3: Availability of EOY Scores by Treatment Status (Attrition)

	Strata-Adjusted Mean Difference	
	Program	Program (i-
	(1)	(2)
Treat	0.0179	0.0095
	(0.0329)	(0.0132)
Control Mean	0.940	0.953
N Students	285	1,069

Note: This table reports availability of endline scores - DIBELS for the literacy program and i-Ready for the Numeracy program - by treatment condition and program for the sample of students identified as eligible for tutoring. Cells report the adjusted difference between the group of students assigned to the control group versus students assigned to treatment group, controlling for randomization strata.

Table A4: Intent-to-Treat (ITT) Effects on EOY Performance; Alternative Standard Errors

	Literacy Program: DIBELS Summative (SD)		Numeracy Program: i-Ready Summative (SD)	
	(1)	(2)	(3)	(4)
ITT	0.0513	0.0513	-0.0222	-0.0222
	(0.0855)	(0.0699)	(0.0527)	(0.0565)
Female	-0.2489**	-0.2489*	0.0013	0.0013
	(0.0862)	(0.1082)	(0.0494)	(0.0497)
White	-0.4767	-0.4767	0.2079+	0.2079+
	(0.3257)	(0.3005)	(0.1111)	(0.1164)
Black	-0.3641	-0.3641	-0.3599**	-0.3599**
	(0.2893)	(0.2715)	(0.1181)	(0.1366)
Hispanic	-0.2244	-0.2244	-0.2121	-0.2121+
	(0.3263)	(0.3518)	(0.1305)	(0.1246)
English Learner	0.1519	0.1519	0.0032	0.0032
	(0.1877)	(0.2051)	(0.1055)	(0.1186)
Special Education	-0.2648+	-0.2648	-0.1262	-0.1262
	(0.1499)	(0.1758)	(0.0802)	(0.0811)
Baseline score (sd)	0.4895***	0.4895***	0.3483***	0.3483***
	(0.0550)	(0.0683)	(0.0310)	(0.0354)
Classroom Clustered Errors	NO	YES	NO	YES
Constant	0.5091	0.5091+	0.3752+	0.3752**
	(0.3375)	(0.2874)	(0.2046)	(0.1246)
N	270	270	1023	1023

Note: Estimates with heteroskedasticity robust standard errors are presented in columns 1 and 3, while standard errors that allow for clustering at the classroom level are presented in columns 2 and 4. Models examining ITT effects for the reading program sample (Cols 1 and 2) include a grade level variable and a variable for floor scorers, an indicator that equals one if a student scored at the minimum possible value in their baseline early literacy score; coefficients for these variables are omitted from display. All regressions control for randomization strata. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table A5: Intent-to-Treat (ITT) Effects on EOY Composite Score Performance by Student Trait

	Full Sample	Male	Female	Black	Hispanic	White	English Learner	Not English Learner	Special Education	Not Special Education	Kindergarten	1st Grade
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A. Literacy Program: DIBELS Summative (SD)												
ITT	0.0440	0.1858+	-0.0946	-0.0359	0.4206**	.	0.6418***	-0.0651	-0.0255	0.0590	0.0301	0.0721
	(0.0637)	(0.1034)	(0.1108)	(0.0834)	(0.1229)	.	(0.1557)	(0.9177)	(0.0850)	(0.0574)	(0.0651)	(0.1882)
Control Mean	0.0000	0.0057	-0.0063	-0.1277	0.0715	0.8366	0.0562	-0.3623	-0.0177	0.0591	0.0000	0.0000
N	270	136	134	186	68	12	63	34	207	236	207	61
B. Numeracy Program: i-Ready Summative (SD)												
ITT	-0.0222	-0.0019	0.0003	0.0456	0.0273	-0.0901	-0.0006	-0.0059	-0.2989	-0.0065	-0.0222	.
	(0.0527)	(0.0788)	(0.0799)	(0.0684)	(0.1733)	(0.1079)	(0.1715)	(0.0555)	(0.2569)	(0.0570)	(0.0527)	.
Control Mean	0.0019	-0.0229	0.0270	-0.1560	0.0670	0.4122	0.0481	-0.0065	-0.2346	0.0419	0.0019	
N	1023	516	507	622	194	157	165	858	151	872	1023	

Robust standard errors in parentheses. The Literacy Program did not include a insufficient number of observaitons for White students to estimate the regression in column 6. All regressions control for baseline student characteristics and randomization strata. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table A6: Numeracy Program Intent-to-Treat (ITT) Effects on EOY Sub-Score Performance by School-Level Dosage

	Average % of Program Complete (Max. 50 Sessions)						
	-	>0%	>50%	>60%	>70%	>80%	>90%
Numeracy Program	Any (N=37)	(N=32)	Sessions	Sessions	Sessions	Sessions	Sessions
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. i-Ready EOY Numeracy Sub-Score							
ITT	-0.0292	-0.0236	0.1122	0.1449	0.1264	0.3019+	0.2884
	(0.0584)	(0.0630)	(0.1077)	(0.1213)	(0.1287)	(0.1708)	(0.3437)
B. i-Ready EOY Measurement Sub-Score							
ITT	0.0298	0.0497	0.0852	0.1433	0.1569	0.3195+	0.8035**
	(0.0572)	(0.0612)	(0.0954)	(0.1068)	(0.1297)	(0.1747)	(0.2243)
C. i-Ready EOY Algebra Sub-Score							
ITT	-0.0424	-0.0390	0.0209	0.0877	0.0350	0.2569	-0.0941
	(0.0566)	(0.0601)	(0.0980)	(0.1141)	(0.1387)	(0.1998)	(0.3205)
D. i-Ready EOY Geometry Sub-Score							
ITT	-0.0283	-0.0395	0.0038	0.0705	0.0241	0.0468	-0.1151
	(0.0552)	(0.0588)	(0.0948)	(0.1018)	(0.1266)	(0.1748)	(0.3320)
Student Observations	1,023	880	281	227	150	85	32

Robust standard errors in parentheses. All regressions include controls for classroom randomization strata, BOY composite test score and baseline characteristics. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table A7: Two-Stage Least Squares Estimates of Treatment Take-up & Effects

	Literacy Program		Numeracy Program	
	Ever Tutored	No. of sessions	Ever Tutored	No. of sessions
	(1)	(2)	(3)	(4)
A. First Stage				
	0.8783***	36.0531***	0.6459***	16.9048***
	(0.0483)	(5.1283)	(0.0237)	(0.9130)
Control Mean	0.0526	1.9006	0.051	1.0260
B. Average Treatment on the Treated				
	0.0392	0.0010	-0.0341	-0.0013
	(0.0889)	(0.0022)	(0.0769)	(0.0029)
N	270	270	1012	1012

Robust standard errors in parentheses. All regressions control for baseline student characteristics and randomization strata. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table A8: Intent-to-Treat (ITT) Effects on EOY Performance by School-Level Dosage

	Average % of Program Complete (Max. 140 Sessions)						
	-	>10%	>20%	>30%	>40%	>50%	>60%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Literacy Program	Any (N=15)	Sessions	Sessions	Sessions	Sessions	Sessions	Sessions
A. DIBELS EOY Score (with controls for baseline student traits)							
ITT	0.0440	0.0805	0.0261	0.0419	0.0997	0.0997	-0.0597
	(0.0637)	(0.0750)	(0.0853)	(0.1261)	(0.1844)	(0.1844)	(0.3919)
Control Mean	-0.0076	0.1049	0.0979	0.0432	0.0430	0.0430	0.2214
Student Observations	270	202	166	119	43	43	25
	Average % of Program Complete (Max. 50 Sessions)						
	-	>0%	>50%	>60%	>70%	>80%	>90%
Numeracy Program	Any (N=37)	(N=32)	Sessions	(N=10)	Sessions	Sessions	(N=2)
B. i-Ready EOY Composite Score (with controls for baseline student traits)							
ITT	-0.0222	-0.0164	0.0651	0.1303	0.0945	0.2764**	0.2161
	(0.0527)	(0.0633)	(0.0888)	(0.0948)	(0.1237)	(0.0666)	(0.1658)
C. i-Ready EOY Composite Score (without controls for baseline student traits)							
ITT	-0.0124	-0.0047	0.0569	0.1175	0.0955	0.2761***	0.2871**
	(0.0539)	(0.0649)	(0.0873)	(0.0902)	(0.1168)	(0.0828)	(0.1015)
Control Mean	-0.0100	-0.0295	-0.1623	-0.2207	-0.1288	-0.3698	-0.1534
Student Observations	1,023	880	281	227	150	85	32

Robust standard errors in parentheses. All regressions include controls for classroom randomization strata, BOY students test scores and, as indicated, baseline characteristics. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Appendix A. Data Notes

Additional details on samples

Literacy program

Our study of the AY 2022-23 literacy program includes 222 Kindergarten and 68 first grade students The number of tutoring seats available varied across schools due to a limited number of school staff who could serve as tutors at each school. Generally, there was a cap of approximately five students who could receive tutoring in each classroom. Eligibility for the study was limited to students who scored one or more grade levels below what was expected for their grade in early literacy skills in the baseline DIBELS assessment. Within each classroom and baseline reading level band strata, researchers randomly assigned 103 students to the treatment group and 182 students to the control group.

Among this intent-to-treat (ITT) sample, five students left the district during tutoring implementation, and 15 students were the missing demographic and/or endline test score data necessary for inclusion in our analysis. Therefore, 270 students—207 in Kindergarten and 63 in first grade – are included in our confirmatory analysis.

Numeracy program

Our study of the AY 2023-24 numeracy program includes 1,069 Kindergartners identified as tutoring-eligible across 37 schools and 94 classrooms. Seven of these participating schools received the literacy tutoring program the prior year. Of these students, 1,023 have endline test scores and are captured in our confirmatory analysis.

The ITT sample deviates from the planned sample in two minor ways: First, while eligibility was supposed to be limited to students with "Below Grade Level" achievement on the beginning-of-year (BOY) i-Ready assessment, 40 of these 1,023 Kindergartners scored on-level at baseline. Second, 40 schools had planned to participate in the RCT. However, three withdrew before randomization occurred and are therefore excluded from this study. As in the literacy study, the composition of the remaining sample is balanced on gender and demographic shares broadly reflect District make-up (Table A1).

Additional details on study design

Because these interventions took place during class time, students in the control group received instruction and support from their classroom teacher at the same time as the paraprofessional facilitated tutoring for the treatment group. Because tutors worked with only one to three students at a time the experience of students with their teacher was not substantially affected by whether the tutoring was happening.

Additional details on control measures

We use the mutually exclusive District administrative categories for student race/ethnicity to create “Black”, “White”, and “Hispanic” indicators. Due to small counts of tutoring-eligible students in other categories, we collapse “multiracial”, “Native Hawaiian or Other Pacific Islander,” “American Indian or Alaska Native,” and “Asian” ethnoracial identities into a single “other” category. We also use binary indicators for whether a student is labeled as an English Learner, or is in Special Education. Finally, we create a binary variable for female gender that is coded as 1 for students identified as female in administrative data and zero otherwise. Fewer than five tutoring-eligible students in our focal cohorts have an “unknown” gender in the administrative data.

Additional details on tutoring dosage

We measure literacy tutoring dosage using comprehensive literacy provider participation metrics including start/end dates, time in tutoring, session attendance, curriculum progress and tutor assignments. For the numeracy program we surveyed tutors using an online platform to collect session-level data on student tutoring attendance, topic taught, time-in-tutoring, and tutor demographic information. We particularly lean on the session attendance data to characterize student engagement with and progress through each tutoring program. To contextualize the generalizability of our dosage-dependent results we use publicly-available school-level attendance and achievement data.

Differences between ITT estimation in the literacy vs. numeracy study

There are two small differences in the estimating equations of the studies. First, because DIBELS-8 suffers from floor effects in early grades, an indicator variable for minimum score is included in the literacy study specification. Second, no students are indicated in the “other” race/ethnicity category in the literacy sample, so the demographic reference category for that study is white students. However, for the math study the reference category is students classified as “other” race/ethnicity.

Additional details on tutor surveys

For the numeracy program, we supplement our analysis of student engagement and outcomes with beginning (N=92) and end-of-year (N=17) para-tutors surveys that provided opportunities for them to share open-ended feedback and response to closed-response and likert-scaled questions about their experiences (see Appendix B for the full survey). We use responses to an item requesting para-tutors indicate each of their work responsibilities from a list to the breadth of para-tutor workload. Due to the limited sample size on the end-of-year survey, we do not make comparisons across administrations.

Details of curricular analyses

We reviewed the curricular materials and tutor guidance that was made available to us from both the numeracy and the literacy tutoring program. From the numeracy program, only a subset of lessons was provided but we were also given their directions for tutors, which provide specific guidance about strategies to differentiate instruction among small groups. For the literacy program, we accessed the complete curriculum and tutor materials. The analysis involved close reading of materials for approximately a dozen lessons at spaced intervals (i.e., 1-3, 10, 15, 16, 31, 44, 50, 53, 54, 68, 80, 102, 132, 180) as well as checks for specific elements (i.e., the progression of phonemic awareness activities) across the curriculum. Because few students progressed beyond the first half of the curriculum, students mostly received instruction on foundational reading skills emphasized in the early lesson blocks. The alignment of the curriculum with the DIBELS-8 assessment was also tracked, as was the prevalence of relationship building activities and indicators for within-session pacing.

Appendix B. Survey Instruments

Final Paraprofessionals Survey

Introduction: Thank you for all of your invaluable contributions to the ROOTS program! We are hoping to learn more about the educators who work with students on the ROOTS program. Please take just a few minutes to answer this survey. It will help us learn more about how to support students!

Information**:**
(name): Please enter your first and last name:

- First Name:
- Last Name:

(email_dcps) Please enter your DCPS email address:

- ____

(employee_number) If available, please enter your employee number:

- ____

(tutor_type) Please select which describes you best (check all that apply):

- I am a full-time paraprofessional/aide
- I am a part-time paraprofessional/aide
- I am a volunteer
- I am a former teacher
- I am a teacher
- I am a PTA employee
- I am an instructional coach
- Other ____

(homeroom_teacher) Who is your homeroom teacher?

- ____

(school) Please select the school you work in:

- ▼ Amidon-Bowen ES (1) ... Whittier ES (44)

Final items (paraprofessionals' perceptions):
(intro_items) *On the next few pages, we are going to ask you some questions about how you felt about teaching the ROOTS program. We are interested in how your experiences change over the course of a year. The questions in this survey are quite nuanced and a lot of them may seem the same, but there are subtle differences. Bear in mind that what might seem the same for you may be different for other people, and we are trying to understand those differences.*

(rse) How confident are you that you can build positive relationships with the students you work with in small groups?

- Not at all confident (1)

- Slightly confident (2)
- Somewhat confident (3)
- Quite confident (4)
- Extremely confident (5)

(tsr1) How positive do you think your relationships with students are?

- Not positive at all (1)
- Slightly positive (2)
- Somewhat positive (3)
- Quite positive (4)
- Extremely positive (5)

(se) *Here are a few more questions to answer about how confident you feel about teaching students math this semester.*

	Not at all confident (1)	Slightly confident (2)	Somewhat confident (3)	Quite confident (4)	Extremely confident (5)
How confident are you that you can help your students understand the material in a ROOTS math session? (se_understand)	○	○	○	○	○
How confident are you that you can engage students during a ROOTS math session? (se_engage)	○	○	○	○	○
How confident are you that you can help students improve their understanding of numbers? (se_grade)	○	○	○	○	○

(learn) How much do you think your students will learn from you?

- Almost nothing (1)
- A little bit (2)
- Some (3)
- Quite a bit (4)
- A tremendous amount (5)

(bel) How much can you do to help students feel like they belong at school?

- Almost nothing (1)
- A little bit (2)
- Some (3)

- Quite a bit (4)
- A tremendous amount (5)

(value) How much can you do to help students value math/numbers?

- Almost nothing (1)
- A little bit (2)
- Some (3)
- Quite a bit (4)
- A tremendous amount (5)

(cult_comp1) How easy do you find interacting with people who are from different cultural backgrounds than your own?

- Not easy at all (1)
- Slightly easy (2)
- Somewhat easy (3)
- Quite easy (4)
- Extremely easy (5)

(intro) *Instructions: Please answer the following questions about your beliefs about your own experiences and opinions.*

(se_tutor) How confident are you that you can be an effective educator for your students?

- Not at all confident (1)
- Slightly confident (2)
- Somewhat confident (3)
- Quite confident (4)
- Extremely confident (5)

(enjoy) How much do you think you will enjoy teaching ROOTS?

- Do not enjoy at all (1)
- Enjoy a little bit (2)
- Enjoy somewhat (3)
- Enjoy quite a bit (4)
- Enjoy a tremendous amount (5)

(role_para) What are the top two most important roles of a paraprofessional?

	Select your first and second choice.				
	To accelerate students' learning (1)	To serve as a role model (2)	To have high expectations for students (3)	To be a caring adult in a student's life (4)	To help students thrive in all aspects of their lives (5)

Most important role (1)	○	○	○	○	○
2nd most important role (2)	○	○	○	○	○

(teach_interest) How interested are you in becoming a classroom teacher?

- Not interested at all (1)
- Slightly interested (2)
- Somewhat interested (3)
- Quite interested (4)
- Extremely interested (5)

(ed_interest) Overall, how likely are you to pursue a teacher license?

- Not likely at all (1)
- Slightly likely (2)
- Somewhat likely (3)
- Quite likely (4)
- Extremely likely (5)

Paraprofessional role and professional development opportunities:

(para_roles) As a paraprofessional, your duties may vary. Below is a list of roles that might be part of your job. Please sort each role into the following categories based on how frequently you perform them in your current position:

	Not my role (1)	Rare (2)	Secondary (3)	Primary (4)
Behavioral and social support	○	○	○	○
Instructional support	○	○	○	○
Delivering small group instruction	○	○	○	○
Delivering 1:1 instruction	○	○	○	○
Supervising students	○	○	○	○
Clerical duties	○	○	○	○

Providing information between school and parents	○	○	○	○
Developing lesson plans	○	○	○	○
Interpreting for families	○	○	○	○

(para_roles_more_less) Please sort each role into the following categories based on whether you would like to do it more, about the same, or less than you currently do:

	Less of (1)	About the same (2)	More of (3)
Behavioral and social support	○	○	○
Instructional support	○	○	○
Delivering small group instruction	○	○	○
Delivering 1:1 instruction	○	○	○
Supervising students	○	○	○
Clerical duties	○	○	○
Providing information between school and parents	○	○	○
Developing lesson plans	○	○	○
Interpreting for families	○	○	○

(teachers) How often are you in direct communication with classroom teachers?

- Never
- Rarely
- Sometimes
- Often
- Very often

(pd_types) What forms of professional development would you like to receive more of, about the same, or less of from your school or district?

	Less of (1)	About the same (2)	More of (3)
School-based mentoring from certificated teachers (1)	○	○	○
School-based mentoring from my peer paraeducators (4)	○	○	○
Regular observations and feedback from my principal, other administrators, or department chair (5)	○	○	○
Scheduled time to collaborate with certificated teachers in my school (e.g., common planning time, peer observation and feedback) (6)	○	○	○
Scheduled time to collaborate with peer paraeducators (7)	○	○	○
School wide professional development (9)	○	○	○
District professional development (10)	○	○	○

Other: (11)	○	○	○

Demographics:

(intro_demos) *Almost done, [first name]! Please answer the following items so that we can accurately describe the people who take the survey, in general.*

(female) How do you identify?

- Man
- Woman
- Non-Binary
- Other ____
- Prefer not to answer

(age) How old are you?

- Under 18
- 18-24 years old
- 25-34 years old
- 35-44 years old
- 45-54 years old
- 55-64 years old
- 65+ years old
- Prefer not to answer

(tutor_prior) Have you taught the ROOTS curriculum before?

- Yes, at DCPS
- Yes, somewhere else
- No
- Not sure or can't remember

(female) How do you identify?

- Man
- Woman
- Non-Binary
- Other ____
- Prefer not to answer

(language) Please indicate the primary language spoken in your childhood home. (Please select only one)

- Arabic
- Chinese
- English
- Farsi
- French

- German
- Japanese
- Korean
- Portuguese
- Spanish
- Other ____
- Prefer not to answer

(para_ed) Please select the highest level of education you have completed.

- ▼ Did Not Attend School (0) ... Completed graduate school (18)

(mother_ed) Please select the highest level of education completed by your mother.

- ▼ Did Not Attend School (0) ... N/A (19)

(father_ed) Please select the highest level of education completed by your father.

- ▼ Did Not Attend School (0) ... N/A (19)

(selfcontained_class) Are you working in a self-contained classroom?

- Yes
- No

(ell) Are you working with English Language Learners?

- Yes
- No

(comments) If there is anything else you feel that we should know regarding this survey, please leave us a note below.

- ____
  ____
  ____
  ____