Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 25-1176

Experimental Evidence on the Impact of Tutoring Format and Tutors: Findings from an Early Literacy Tutoring Program

Shirin Hashim
Harvard University

Katharine Pace Miles
Brooklyn College, CUNY

Erin Croke
City University of New York

This study presents the first within-program, within-tutor experimental evidence comparing the impact of in-person versus remote tutoring. Based on results from an early literacy tutoring initiative delivered by university students over Summer 2023, we find no statistically significant differences in students' literacy outcomes by instructional modality. However, students receiving in-person tutoring exhibited higher attendance rates and tutors reported closer relationships with their in-person students. Notably, we find substantial variation in students' outcomes due to differences among tutors, while these effects do not vary by modality. These findings suggest that while differences between in-person and remote tutoring may exist, the advantages of having a proficient tutor greatly outweigh these disparities. The study underscores the efficacy of remote tutoring, particularly when geographical constraints are a factor, and highlights the necessity of including interpersonal skills in tutor training, ensuring consistent attendance and program fidelity, and identifying and retaining highly effective tutors to maximize student learning.

VERSION: April 2025

Suggested citation: Hashim, Shirin, Katharine Pace Miles, and Erin Croke. (2025). Experimental Evidence on the Impact of Tutoring Format and Tutors: Findings from an Early Literacy Tutoring Program. (EdWorkingPaper: 25 -1176). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/zgy5-sr80

Experimental Evidence on the Impact of Tutoring Format and Tutors: Findings from an Early Literacy Tutoring Program

Shirin Hashim[a]

Katharine Pace Miles[b]

Erin Croke[c]

[a]Graduate School of Education, Harvard University, Cambridge, USA; [b]Early Childhood Education, Brooklyn College, CUNY, Brooklyn, USA; CUNY Central Office of Academic Affairs, City University of New York (CUNY), USA

Author Note

Dr. Shirin Hashim is a recent graduate from the Harvard Graduate School of Education, where she completed her Ph.D. in Education Policy and Program Evaluation. Dr. Katharine Pace Miles is an Associate Professor and Director of the Advanced Certificates in Reading Science at Brooklyn College, City University of New York (CUNY), the co-founder of CUNY Reading Corps, and the founder of The Reading Institute. Dr. Erin Croke is the Director of Literacy Initiatives at the CUNY Central Office of Academic Affairs.

Corresponding Author:
Katharine Pace Miles
Department of Early Childhood Education
Brooklyn College
2306b James Hall
Brooklyn, NY 11210.
kpmiles@brooklyn.cuny.edu

Abstract

This study presents the first within-program, within-tutor experimental evidence comparing the impact of in-person versus remote tutoring. Based on results from an early literacy tutoring initiative delivered by university students over Summer 2023, we find no statistically significant differences in students' literacy outcomes by instructional modality. However, students receiving in-person tutoring exhibited higher attendance rates and tutors reported closer relationships with their in-person students. Notably, we find substantial variation in students' outcomes due to differences among tutors, while these effects do not vary by modality. These findings suggest that while differences between in-person and remote tutoring may exist, the advantages of having a proficient tutor greatly outweigh these disparities. The study underscores the efficacy of remote tutoring, particularly when geographical constraints are a factor, and highlights the necessity of including interpersonal skills in tutor training, ensuring consistent attendance and program fidelity, and identifying and retaining highly effective tutors to maximize student learning.

Keywords: early literacy, tutoring, remote tutoring, virtual tutoring

Acknowledgments

The authors would like to thank the teachers and students who participated in the study, the CUNY Reading Corps implementation team, as well as Tom Kane, Doug Staiger, Jimmy Kim, attendees at the Society for the Scientific Study of Reading (SSSR) Conference, and attendees at the National Student Support Accelerator (NSSA) Conference for thoughtful comments and feedback. This work was supported in part by the Tiger Foundation, Heckscher Foundation for Young Children, Altman Foundation, The Benedict Silverman Foundation, Fund for Public School, Robin Hood Foundation, and NYC Summer Youth Employment Fund.

Introduction

Across the United States, more than two-thirds of children continue to fail to meet the National Assessment of Educational Progress (NAEP) proficiency standards in reading established by the Education Department's National Center for Education Statistics (NCES). While the issue was exacerbated by the COVID-19 pandemic, particularly in widening the proficiency gap along socioeconomic and racial lines (Callen et al., 2023), the downward trajectory of these dismal results was evident prior to the pandemic; in 2019, only 35 percent of fourth graders were proficient in reading, down from 37 percent in 2017. In fact, more than 60 percent of American fourth graders have not been proficient readers since testing began in the early 1990s, with the lowest achievers falling increasingly behind (National Assessment of Educational Progress, 2024). These patterns are especially concerning considering the long-term consequences of inadequate reading skills, such as lower high school graduation rates and higher rates of poverty (Fiester, 2010).

Recently, two key areas of research have garnered attention to address these pandemic-related learning setbacks and concerns about reading abilities: high-impact tutoring and the science of reading. Tutoring programs have proliferated across the nation to address learning disruptions over the last several years reflecting the substantial body of evidence demonstrating the large, positive gains that high-impact tutoring can produce, particularly for students from low-income families (Robinson & Loeb, 2021). According to recent reporting, states have dedicated $4.2 billion—more than any other academic recovery initiative—to tutoring and accelerating learning. This includes $700 million of federal relief dollars from state Elementary and Secondary School Emergency Relief (ESSER) funds, which were exclusively earmarked for tutoring expansion efforts (CCSSO, 2023). Yet, the case for tutoring in general is not a new one. The past challenges have been the high costs and staffing difficulties associated with in-person tutoring (Horn et al., 2014), as well as ongoing uncertainties about the efficacy of different programs and their characteristics (Robinson & Loeb, 2021). For example, while the pandemic-related shift toward technology, a greater investment in devices and internet services, and increased funding, created a prime opportunity for scholars and policymakers to push for the rapid deployment of tutoring services, both in-person and online, these investment decisions have been based on a narrow understanding about online tutoring.

Few studies to date have examined the impact of online tutoring, especially in the U.S. context (Kraft et al., 2022). Recently, Neitzel and Reilly (2024) and Neitzel and Storey (2024) found that first graders and first through 6th graders who received an online literacy tutoring program had significantly greater gains compared to students who did not receive tutoring (first graders effect size of +0.21; first to sixth graders effect size +0.12). However, we are not aware of any studies that directly measure the marginal impact of in-person tutoring over online tutoring of the same program delivered by the same tutors over the same period. This means that, until now, any conclusions drawn about the relative effect sizes of remote tutoring versus in-person tutoring are potentially conflated with differences in programs, curricula, tutoring skills, and settings. Furthermore, there is still much to learn about the characteristics of effective tutors, how tutoring programs foster strong student-tutor relationships, the multifaceted benefits of tutoring, and other relevant research areas (Robinson & Loeb, 2021).

Similarly, the science of reading literature dates back several decades but has only recently gained traction amongst policymakers and educators. This surge in interest is partly due to a series of stories and podcasts by APM Reports, which exposed the shortcomings in the widely used “balanced literacy” approach to reading instruction and the gap between existing research and classroom practices (Peak, 2022). In response, at least 40 states have enacted laws or new policies promoting evidence-based reading instruction in hopes that these mandates will align instruction to the science of reading. These measures typically require educators to receive professional development in evidence-based reading methods and to provide specific interventions for struggling readers (Schwartz, 2024). However, a key challenge for practitioners then becomes determining which interventions to implement and how to scale them effectively.

This study sits at the intersection of these two critical bodies of literature. We report findings from an early literacy tutoring initiative implemented in New York City Public Schools (NYCPS) during the summer of 2023. Tutoring was delivered by undergraduate student interns from the City University of New York (CUNY) who were trained on the Reading Ready and Reading Go programs, which are deeply rooted in the science of reading research and tailored for the most struggling readers. These interns were paired with predominantly low-income students entering grades 1 through 3 to provide one-on-one tutoring characterized as "high-dosage" (Fryer, 2016).

It is worth noting that the assignment process was not specifically designed for a randomized controlled trial, but rather aimed at an equitable distribution of resources and ensuring no students were excluded from receiving tutoring. Nonetheless, this method of randomly assigning students to different formats and randomly assigning tutors to students in both formats allows for important research inquiries, related to instructional format and tutor quality, which are critical for practitioners in designing, implementing, and scaling tutoring programs to optimize student outcomes. Specifically, we ask the following research questions:

RQ1. What is the impact of instructional modality (in-person versus remote) on students' literacy outcomes, attendance, and tutor-student relationships?

RQ2. How much of the overall variation in student outcomes can be attributed to differences between tutors?

RQ3. How do effects vary based on tutor characteristics, instructional modality, and across different outcomes?

Background

Research has consistently shown that well-designed, in-person tutoring programs can substantially improve learning across various subjects and student populations. A recent meta-analysis of experimental tutoring program evaluations spanning several decades revealed that tutoring, on average, led to an additional 3 to 15 months of learning across different grade levels (Nickow et al., 2020). Another comprehensive review of nearly 200 randomized and field experiments identified high-dosage tutoring as one of the few school-based interventions with demonstrably significant effects on student outcomes (Fryer, 2016).

The evidence supporting early elementary tutoring in reading is particularly robust, as highlighted by several recent meta-analyses indicating the potential for significant gains through individual literacy tutoring (Elbaum et al., 2000; Gersten et al., 2020; Neitzel et al., 2022). Combining insights from these studies, Cortes et al. (2023) observe that the most effective reading tutoring programs share characteristics with other high-dosage programs: they utilize evidence-based curricula and involve 20-60-minute sessions occurring multiple times each week with a consistent tutor.

This study addresses the effect of two key inputs into tutoring: tutor skill and delivery mode. Tutor skill is often discussed in terms of the types of tutors being hired, such as certified teachers, university students, family members, and so forth. Prior research has shown that tutoring is most effective when delivered by classroom teachers, though these programs can also be the most expensive to staff (Nickow et al. 2020). Relevant to our study, programs that employ university students tend to have more moderate effects on student learning compared to those that employ teachers or paraprofessionals (Robinson & Loeb, 2021). However, the relative effectiveness among different types of tutors is inferred from variances in effects across programs with diverse characteristics, rather than through direct comparisons. Moreover, as Robinson and Loeb (2021) note, it is often difficult to distinguish between the effects of tutor type and tutor training. It may be that tutors who undergo intensive training, are provided with ongoing support and feedback, and/or face accountability measures will be more effective regardless of their type.

Though limited research directly explores tutor effectiveness, insights can be drawn from the teacher value-added literature, which shows that effective teachers are best identified by their performance in the classroom rather than their observable characteristics, such as licensure or degree (Rockoff et al., 2011). While this study does not aim to estimate tutor value-added, we believe that understanding how much of the variation in student outcomes can be attributed to tutors, especially when holding constant other factors such as the amount of training, degree, and curriculum, can inform decisions around what inputs to prioritize when making decisions about programmatic characteristics.

The second input we examine in this study is the delivery mode. Before the rise of online tutoring, choices about delivery broadly pertained to programs' tutor-student ratios and where and when tutoring was going to take place. While there are still vast areas of unexplored research regarding these dimensions, there does appear to be consensus around the fact that high-impact tutoring programs tend to occur during the school day and take place in a one-on-one or small group setting of fewer than four students (Robinson et al., 2021).

With the rise of blended and online learning, the instructional format has become another critical component of the delivery mode, yet the literature on the impact of online tutoring is much more limited. Still, a few recent evaluations of virtual tutoring programs provide some preliminary evidence that they may be effective (Robinson & Loeb, 2021; Neitzel & Reilly, 2024; Neitzel & Storey, 2024). Also, Carlana & La Ferrara (2021) found that a virtual tutoring intervention that took place in Italy during the pandemic-related school closures of Spring 2020 increased middle school students' academic achievement by an effect size of 0.26 and improved students' social-emotional outcomes.

Another study in Spain, evaluating an intensive eight-week online tutoring program, found positive effects on middle school students' end-of-year math grades (0.48 standard deviations) and significant impacts on non-cognitive outcomes and grade retention (Gortazar et al., 2023). The authors posit that the disparity in effect sizes can be partly ascribed to differences in the delivery models between the two studies: the tutoring program in the former study was administered by volunteer university students, whereas in the latter, it was conducted by paid, qualified math teachers.

Also looking at online tutoring provided by student volunteers, Kraft et al. (2022) provide some of the only U.S.-based evidence, finding consistently positive but statistically insignificant effects of online tutoring on middle school students' math and reading outcomes. In a recent study, Loeb et al. (2023) found noteworthy results for younger students. Kindergarten through second-grade students who received virtual early literacy tutoring scored approximately 0.08 standard deviations higher on literacy tests compared to their counterparts not assigned to the program. This evidence is particularly critical since most prior research has centered around older students. This underscores the potential of virtual tutoring for young learners, especially in the realm of early literacy.

Importantly, none of these prominent studies provide within-program evidence comparing outcomes for students who received remote versus in-person tutoring delivered over the same time frame by the same group of tutors. The treatment effects captured by these studies estimate the impact of remote tutoring relative to no tutoring or the status quo. Some do attempt to infer the differences between in-person and virtual tutoring by comparing effect sizes across studies. However, the characteristics of tutoring programs and the environments in which they are implemented can drastically vary (Robinson & Loeb, 2021), making it challenging to make direct comparisons. For example, Kraft et al. (2022) note that their estimated effects for online tutoring are roughly one-third as large as comparable in-person tutoring effect sizes reported by Nickow et al. (2020). Loeb et al. (2023) acknowledge that the positive effects of the virtual model are more modest than similar early literacy tutoring programs delivered in-person, although they caution that differences might stem from various programmatic features beyond the mode of delivery. Still, others have argued that the impacts of online tutoring are comparable in magnitude to those of in-person programs (Carlana & Ferrara, 2021; Gortazar et al., 2023). As far as we know, this study provides the first experimental evidence comparing in-person versus remote tutoring for the same program being delivered by the same tutors, thus holding constant other program inputs that could also influence outcomes.

Method

Literacy Tutoring Initiative and Reading Programs

CUNY Reading Corps launched in the fall of 2020 through a collaborative effort between CUNY and NYCPS, the CRC initiative originated as a fieldwork requirement for preservice teachers enrolled in early literacy undergraduate and graduate courses at Brooklyn College, CUNY. Brooklyn College paired university students with NYCPS striving readers in need of one-on-one tutoring from schools with high economic need indices and low standardized test scores in English Language Arts. As part of the program, university students were mandated to complete the Reading Go training and conduct a total of 20 tutoring sessions, three to five times a week, over the semester. This embedded-course tutoring model has continued at Brooklyn College and is now underway at four additional CUNY colleges. It is also implemented at 25 universities across the country through a non-profit that grew out of this work.

Reading Go (formerly Reading Rescue) is an evidence-based reading intervention program intended for striving first and second graders developed in 1993. Each lesson, which is designed to take 30-45 minutes, addresses phonemic awareness, phonics, fluency, vocabulary, comprehension, and sentence writing through a five-step protocol. Reading Ready was developed based on research in 2020 as a precursor to Reading Go, but it can also stand alone as an early literacy preventative or intervention program for striving kindergarten and first-grade students. Each 20- to 30-minute Reading Ready lesson focuses on letter-sound knowledge, phonemic awareness, phonics, sentence reading, and decodable book reading.

Critically, both Reading Ready and Reading Go closely adhere to the research on effective instructional practices that address the essential components of literacy now known as the science of reading. Over the last 20 years, Reading Go has developed a substantial evidence base with effect sizes ranging from 0.33 to 0.48 (Muller, 2004; Ehri et al., 2007; Miles et al., 2018; Miles et al., 2022). While efficacy studies examining the impact of Reading Ready on students' foundational literacy skills are ongoing, a few studies present preliminary evidence. These utilize nationally normed pre/post assessments and curriculum-based measures incorporated into the program to monitor student progress. Miles and Fletcher (2023) present descriptive findings supporting the effectiveness of the program delivered through virtual tutoring, particularly with students in historically underserved schools in NYC. In a separate study, Belson et al. (2023) report their findings on the use of Reading Ready with preservice teachers from American University. The tutoring was conducted in-person with striving readers in high-needs schools in Washington, D.C. Both studies suggest a positive impact of Reading Ready on student achievement, although causal evidence is forthcoming.

Summer 2023 Implementation

In Summer 2023, CRC partnered with NYCPS to implement the Reading Ready and Reading Go programs across 12 Summer Rising sites, including one site in Manhattan, two sites in the Bronx, six sites in Brooklyn, and three sites in Queens. CRC staff followed a specific process to assign students to tutors and to each of the two instructional formats (in-person and remote.) First, students and their families enrolled in the Summer Rising program directly through NYCPS. At each of the 12 partner sites, school administrators identified 1st through 3rd-grade students who would receive early literacy tutoring provided by CRC, totaling 939 students across all sites. CRC staff then assigned tutors to do in-person tutoring at the nearest Summer Rising site based on their commuting address. In areas with a disproportionate number of in-person tutors or when there were multiple sites in close proximity, CRC staff prioritized distributing tutors across the sites. Out of 151 tutors, 56 were designated for remote tutoring only. This pool of remote tutors included those whose commute time to the nearest Summer Rising site would have been over 60 minutes and those who ended up with no in-person assignments due to 1 of the 12 Summer Rising sites dropping from the program after student and tutor assignment had taken place but before any tutoring sessions had occurred.

Importantly, the assignment process followed that of a cluster (site) randomized controlled trial (RCT). Students at each tutoring site were randomly assigned to receive either in-person or remote tutoring, and tutors were randomly assigned to students in both formats when possible. This allows us not only to identify the causal effect of the instructional format but to do so while considering potential variation across tutors, presenting the most stringent test possible for differences by format.

Participants

There were 607 students who enrolled in the NYCPS Summer Rising Program in Summer 2023 and completed at least one tutoring session through CRC. All student participants had completed kindergarten, first, or second grade and were enrolled in the NYCPS Summer Rising Program for the summer of 2023. A total of 446 students had complete pre-post test data on the Acadience assessment ($N = 446$); of those students, 186 were assigned to in-person tutors and 260 were assigned to remote tutors. A total of 426 students had complete Star Early Literacy data ($N = 426$); of those, 210 were assigned to in-person tutors and 216 were assigned to remote tutors. Table 2 provides descriptive statistics about the student population.

Supported by philanthropic funding and CUNY Career Launch, CRC successfully recruited 200 undergraduate interns from a pool of 500 applicants to serve as tutors. Out of this group, 196 tutors were hired and fulfilled all necessary training and clearance requirements. Among these, 151 tutors were assigned to tutor through Summer Rising, and after accounting for attrition, 139 tutors remained in the final analysis sample. Table 1 offers descriptive statistics about these tutors, revealing that they represented a relatively diverse group compared to the general teacher workforce. They also demonstrated strong academic performance, as indicated by their CUNY GPAs, and were predominantly female identifying. It is noteworthy that less than half of the tutors were enrolled in education-related majors, distinguishing them from CRC's usual tutor pool during the school year, which mainly consists of preservice teachers.

Data and Measures

Early Literacy Outcomes

The analyses include two different measures of reading performance: scaled scores from the Star Early Literacy assessment and overall composite scores on the Acadience Reading assessment. Star, developed by Renaissance Learning, is a computer-adaptive test designed to evaluate the literacy skills of early readers, usually in kindergarten through third grade. The test typically consists of 27 items of varying difficulty, determined by the student's responses. The questions assess students across 10 different subdomains: alphabetic principle, concept of word, visual discrimination, phonemic awareness, phonics, structural analysis, vocabulary, sentence-level comprehension, paragraph-level comprehension, and early numeracy (Renaissance Learning, 2023). The literacy-related subdomains are addressed through explicit and systematic instruction in the Reading Ready and Reading Go programs. It should be noted that the assessment does not measure oral reading fluency, but it does have strong convergent validity with other tests of oral reading fluency (Renaissance Learning, 2023). The test takes an average of nine minutes to complete and was administered by tutors as part of standard program implementation at the beginning and end of the program. The Star scaled scores were utilized after standardizing them using the grade-level means and standard deviations from the total sample to account for differences in the distribution of scores by grade. Normal curve equivalents (NCEs) were computed from the scaled scores using the conversion tables outlined in the Star Early Literacy technical manual. Owing to an error in the rostering process, norm-referenced scores such as student growth percentiles (SGPs) or NCE scores that are typically generated by Renaissance Learning, were not supplied by the test developer.

The Spring 2023 and Fall 2023 overall composite scaled scores on the Acadience Reading assessment, standardized within sample using grade-level means and standard deviations, were used as an additional measure of reading performance. Acadience is a required assessment in kindergarten through second grade across NYCPS. The test measures five core components of reading: phonemic awareness, phonics, reading fluency, vocabulary, and comprehension. Like Star Early Literacy, the test is largely aligned to the Common Core State Standards for English Language Arts (CCSS for ELA) and to the content of the Reading Ready and Reading Go programs. It is important to note that, unlike Star, the administration of this assessment is handled directly by the district itself, independent of the program, and the data were obtained by the district from administrative records.

While Acadience Reading and Star Early Literacy cover the same constructs of literacy, the advantage of using the Acadience data as a second academic measure is that they are more comprehensively available from administrative records. Additionally, they may capture longer-term trends since the post-test is conducted in the fall following the summer initiative, over two months after the end of the summer program. On the other hand, the Acadience pre-test is administered in the prior spring, further removed from students' baseline achievement levels at the beginning of the tutoring program. Post-test data are also not available for students who would have been entering the third grade in the fall since the assessment is only administered through the second grade. Thus, the Acadience data reflect only the rising first and second-grade participants in the program. This is critically important because it means that the Acadience sample skews younger compared to the Star sample.

Tutor-Student Relationship

To measure tutor-student relationships, the program implemented a slightly modified version of the student-teacher relationship scale (STRS) with all tutors. The short form of this survey typically consists of 15 items on a 5-point scale, which yields scores on Conflict and Closeness aimed at assessing a teacher's perception of their relationship with a particular student. The Closeness dimension evaluates the teacher's feelings of affection and open communication with the student (e.g. "I share an affectionate and warm relationship with this child"); Conflict measures the teacher's feelings of negativity and conflict with the student (e.g. "This child and I always seem to be struggling with each other") (Pianta, 2001). Research has shown that the full version of the scale is internally consistent (Pianta, 1992) and predictive of student achievement, behavior, and school retention (Hamre & Pianta, 2001; Pianta et al., 1995). Tutors who conducted at least one tutoring session were asked to complete the modified scale for each of their students as part of their paid internship experience. Appendix A reproduces the original STRS short form and details the modified version used by the program. In short, the modified version contained 12 items on a 5-point scale and was scored similarly to the original by averaging item-level scores pertaining to each dimension. However, it is important to note that there is limited research on the psychometric properties of the short form and no validation work was done to qualify the modifications made here.

Additional Data

In addition to assessment and survey data, we also use session completion records as captured by tutors as a measure of additional instructional time from the tutoring program. Lastly, we incorporate student demographic data provided by the NYC Department of Education, including students' race/ethnicity, economically disadvantaged status, disability status, and

English language learner status as controls. These data from the district identified which students participated in the summer program, their tutoring delivery format, and other pertinent program features, but all personally identifiable information was systematically removed.

Results

To answer RQ1 and RQ2 outlined above, we estimate the following model:

(1)
$$y_{ijk} = \beta_0 + \beta_1 x_i + \gamma Z_i + \mu_k + u_j + e_i$$

$$e_i \sim N(0, \sigma_e^2)$$

$$u_j \sim N(0, \sigma_u^2)$$

where y_{ijk} is the outcome of interest for student *i* with tutor *j* in site-by-grade k. x_i is an indicator variable equal to 1 if student *i* received in-person tutoring and 0 if they received remote tutoring. β_1 is our parameter of interest for RQ1, which is the mean difference between the effect of in-person and remote instruction. The model also includes site-by-grade fixed effects, μ_k, and a vector of student-level controls, Z_i. This includes students' baseline achievement levels and demographics, including race/ethnicity, English language learner status, disability status, and economic disadvantage classification. Finally, u_j is a random intercept, which captures the tutor-specific effect on average outcomes. This error term for tutors is distributed normally with a mean 0 and variance parameter σ_u^2. This variance parameter is our parameter of interest in addressing RQ2 as it captures the degree to which outcomes vary across tutors conditional on the other covariates.

RQ3 outlines three separate lines of inquiry. First, to investigate whether tutor characteristics are predictive of outcomes, we can include fixed effects for those tutor-level parameters in Model 1. Specifically, we include additional terms for tutors' cumulative grade point average, an indicator for whether the tutor was enrolled in an education-related major, and an indicator for whether the tutor was pursuing a bachelor's degree as opposed to an associate's degree. Secondly, to estimate the correlation between tutor effects on one outcome with tutor effects on another outcome, we fit a similar model to that of Model 1 but in which we estimate separate coefficients for each outcome, include a random tutor intercept for each outcome, and allow for correlation in the student-level errors between outcomes. Finally, to examine the heterogeneity in tutor effects by instructional modality as described in RQ3, we fit the following model:

(2) $$y_{ijk} = \beta_0 + \beta_1 x_i + \gamma Z_i + \mu_k + u_{0j} + u_{1j} + e_i$$

$$e_i \sim N(0, \sigma_e^2)$$

$$u_{0j} \sim N(0, \sigma_{u0}^2)$$

$$u_{1j} \sim N(0, \sigma_{u1}^2)$$

Here, the impact of in-person over remote tutoring is allowed to vary across tutors with the inclusion of an additional random effect. Typically, we would have added a random slope parameter to Model 1, which would have captured the difference between the tutor-specific effect of teaching in-person and teaching remotely. However, in Model 2, we have instead simply separated the overall random intercept into one specific to in-person (u_{1j}) and one specific to remote (u_{0j}). This allows us to estimate the variance of the tutor effect in each setting (σ^2_{u1} and σ^2_{u0}) along with the correlation between the in-person and remote tutor effects.

Descriptive Statistics

Table 2 presents descriptive statistics about the sample of students included in the study. Columns 1 through 3 reflect students who had complete pre-post test data on the Acadience assessment and Columns 4 through 6 reflect students who had complete Star Early Literacy assessment data. These samples are further separated into the students who received in-person and remote tutoring and covariate-adjusted differences in student characteristics between the two groups are given. In both assessment samples, despite random assignment, the in-person group tended to be more female-identifying students than the remote group. In the Acadience sample, the remote students included slightly more students with English language learner status than the in-person group. We also received marginally less administrative data (student demographics) from the district amongst the remote students versus the in-person students in the Star sample, but notably have complete data for the full Acadience sample.

Table 3 addresses these missing data in more detail and shows information on the full sample of students who signed up for the program (Columns 1-3), students who participated in at least one tutoring session (Columns 4-6), and information about those students who fell out of the assessed sample shown in Table 2 due to missing Acadience scores (Columns 7-9) or missing Star scores (Columns 10-12). From the first panel pertaining to the randomized sample, it is evident that there was a higher withdrawal rate amongst students assigned to receive remote instruction versus in-person instruction. While this is a significant difference, students were not informed of what form of tutoring they would receive in advance of the first session, so there is no apparent reason to attribute this difference to the assigned delivery format. However, the second panel for the remaining participant sample shows that there is also a statistically significant difference in the percentage of students who were missing either the Star pre-test or post-test. This difference may be related to the assigned delivery format. For example, it is possible that students assigned to remote tutoring were more difficult to track down for the post-assessment because of a lack of devices or connectivity or other reasons associated with the delivery format. This would align with the fact that in Table 4, we see that remote students attended fewer sessions relative to in-person students. Still, it does not appear that this attrition was systematic based on student characteristics. We find no statistically significant difference in the standardized Star pre-test scores (see Table 2), meaning the final analytic sample of students who received in-person tutoring looks similar to the corresponding sample of students who received remote tutoring. Moreover, we do not see differential attrition from the participant sample in terms of the Acadience assessment, nor do we see substantive differences in characteristics between the in-person and remote students amongst the students missing one more assessment. Importantly, one Summer Rising site entirely withdrew from the program after randomization, such that the participant sample includes 11 sites in total rather than 12.

Table 4 displays descriptive findings related to our outcomes of interest. The Star pre-test scores highlight that the population of students served by the program were materially underperforming relative to their grade level. Because scaled scores are challenging to interpret on their own, we include the corresponding normal curve equivalent (NCE) scores here. The NCEs are included for demonstrative purposes only, and we use the standardized scaled scores as described previously for all analyses. We see that students were scoring around 28 NCE units, on average, which corresponds to a percentile rank (PR) of about 15 using Renaissance Learning's

NCE to PR conversion table (Renaissance Learning, 2023). In other words, the students in our analytic sample had an average pretest score that was at the 15th percentile of the distribution of scores from the national norming population. The Acadience scores tell a similar story. Figures 2 and 3 present both the pre- and post-PR scores for Acadience and Star, respectively. Again, we can see that the distributions are severely right-skewed, indicating that the students being served were strikingly low achieving.

While tutoring modality was randomly assigned to participating students, there is no pure control group of students from whom tutoring was systematically withheld. Nevertheless, examining growth on a norm-referenced metric, such as NCE scores, can provide some context for the gains that students made over the summer. Table 4 shows that students who participated in the CRC summer program through Summer Rising grew approximately three NCE units, on average, on the Star Early Literacy assessment. This growth is statistically significant ($p < 0.01$) and worth highlighting considering the literature on the "summer slide." Although the findings are mixed, they broadly indicate that students often undergo learning losses (e.g., Atteberry & McEachin, 2021) or, at most, experience stagnation during the summer (Quinn & Polikoff, 2017; von Hippel et al., 2018). While the raw pre-post difference in the percent of students well-below benchmark on Acadience does not necessarily indicate the same type of growth over the summer, Figure 2 is suggestive of a distributional shift from Spring 2023 to Fall 2023 in terms of percentile ranks.

Alternatively, we can estimate the implied impact of tutoring by taking into account the estimated mean "summer slide" in literacy outcomes from a meta-analysis on summer impacts on achievements (Cooper et al., 1996). The authors find that summer loss equals about one tenth of a standard deviation relative to spring test scores, but is less detrimental in reading compared to math, and more detrimental for more economically disadvantaged students. In a separate analysis, they estimate the summer reading loss to be 0.04 standard deviation units in a sample of students that includes students of all socioeconomic backgrounds and is thus likely to be a conservative estimate of the losses typically experienced by low-income students relevant to this study. Incorporating this estimate, the implied gains over the summer tutoring program on Star scores would be about 0.07 standard deviation units and about 0.02 standard deviation units on Acadience. These effects are in line with the mean weighted impacts of RCTs of literacy interventions in lower elementary grades, which range from 0.02 to 0.04 (Kraft, 2020). In any case, due to the absence of a counterfactual, we are unable to attribute this growth to the tutoring program itself and these descriptive results should be interpreted as suggestive, rather than a causal impact estimate.

The bottom of Table 4 shows descriptive statistics for our non-academic outcomes of interest, including the number of sessions students attended and the two relationship measures captured by the tutor survey. Of note is the fact that students attended just over 13 sessions throughout the summer during which the program ran nearly every day for around six weeks. This indicates substantial absenteeism amongst participants, though this is not uncommon for summer programs (Augustine et al., 2016). Additionally, we see that students assigned to in-person tutoring attended an average of about 1.5 more sessions than those assigned to remote. While this does not take into account differences by site, it is a material difference. Finally, the table includes raw averages of the Closeness and Conflict dimensions from the STRS based on five-point scales. The tutor response rate on this survey was 64%, but because tutors were responding to the survey for each of their students, we have complete data for 81% of students. In general, we see that tutors reported feeling close to their students, on average, but more so when in-person. There is less variation in the average Conflict score, perhaps a reflection of the instrument itself, or the fact that the survey is designed for classroom teachers who are engaging with students for longer periods and would have more opportunities for conflictual interactions.

Finally, in Table 5, we present the pairwise correlations between the outcome variables outlined in Table 4. Of note is the moderate correlation between Acadience and Star. This is perhaps lower than we might expect for two assessments that cover the same constructs of literacy. One reason is that there is more variation in the Acadience scores compared to the Star scores. This could mean that Acadience has better discrimination ability, meaning it can more effectively differentiate between high and low performers, or it could mean that it has lower reliability driven potentially by the way in which Acadience is administered (by a proctor as opposed to being an adaptive online test.) In any case, the moderate correlation provides some further insight into why we observe average gains in the Star measure over the course of the summer but do not see the same for Acadience.

Average Effects by Modality

The top panel of Table 6 presents our estimates of the marginal impact of in-person tutoring over remote tutoring, β_1, from fitting Model 1. Columns 1 through 5 show the results for each of our five academic and non-academic outcomes—Fall 2023 Acadience and Star Early Literacy scores, number of sessions, and the Conflict and Closeness measures of tutor-student bond. From the estimates of β_0 in Columns 1 and 2, we can see that, on average, students gained a baseline of 0.31 to 0.35 standard deviation units over the summer across sites after controlling for student-level covariates. Importantly, this estimate should be considered a secular growth rate rather than the direct impact of the tutoring program since it is not relative to a comparison group that did not receive tutoring. The test scores are also standardized within the sample, so growth cannot necessarily be generalized to typical summer learning. However, because students were randomly assigned to instructional format, the estimates of β_1 represent the causal impact of receiving in-person over remote tutoring.

In general, we find no statistically significant difference in the impact on student achievement between the two formats. The estimated magnitude of the effect is also relatively small compared to the overall secular trend. For example, the in-person estimate on Acadience is 0.03 standard deviations whereas the estimated average growth overall is over ten times that at 0.31. Although we cannot rule out small differences that may be undetectable given the sample size, the findings still enable us to rule out large differences in learning outcomes based on format.

In Column 3, we consider the marginal impact of in-person over remote tutoring on the number of sessions students received, reflecting some degree of attendance behavior. From the estimate on β_0, we see that we expect students with an average tutor to attend 11.9 tutoring sessions and the marginal impact of being assigned to in-person tutoring is almost 2 additional sessions (the estimate for β_1) holding student covariates constant. The estimated impact of instructional format here is both statistically significant and large in terms of magnitude. It indicates that we would expect that a student assigned to in-person tutoring would receive over 15% more instructional time relative to a similar student with a similar tutor assigned to remote tutoring simply due to their assigned instructional format.

The top panel of Columns 4 and 5 of Table 6 present the marginal impact of instructional format on two measures of the tutor-student relationship captured by the relationship survey administered to tutors. Because Model 1 includes random intercepts for tutors, this removes any potential rater effects from the average in-person effect. Put differently, there may be some tutor-level variation in how tutors respond to the survey items regardless of instructional format, and these are taken into account when considering the average effect in-person versus remote on responses. This is particularly important in this case since these measures are self-reported by tutors. Overall, we find no statistically significant difference in how tutors perceive Conflict with their students based on instructional format. However, we see a statistically significant and large difference on the Closeness measure. On average, we see that the average tutor reports feeling closer to their in-person students by nearly 0.4 standard deviation units compared to their remote students.

Main Tutor Effects

Because Model 1 includes random effects at the tutor level, it can also tell us how much individual tutors affected outcomes. The bottom panel of Table 6 shows the estimates of the standard deviations of those random effects. The estimated tutor effect on Star is particularly of note; it suggests that the variation in how much students' scores differ based on their tutor is 0.22 standard deviation units. That is, a tutor whose effect is 1 standard deviation greater than the average tutor could have a relative effect of 0.22 standard deviation units on student achievement, on average, all else equal. This effect is large compared to the average secular growth of 0.352 and is materially larger than the estimated main effect of being in-person of 0.017. While the corresponding estimated tutor effect based on Acadience scores is smaller, 0.06, it is still significant, suggesting that having an effective tutor versus an average tutor could lead to nearly 20% more growth relative to the baseline of 0.308. Put another way, students at the same site who have the same baseline achievement and demographic characteristics could score 0.06 standard deviation units higher or lower depending on who their tutor is.

The estimated variation in tutor effects, σ_u, on the non-academic outcomes presented in Columns 3 through 5 are similarly large. The estimate for the number of sessions indicates that a tutor that is 1 standard deviation more effective in increasing the number of sessions relative to the average tutor would induce nearly 2.2 additional sessions. Again, this is a large effect compared to a baseline of 11.9 sessions and relative to the main in-person effect of 1.8 sessions. Columns 4 and 5 also show significant variation between tutors in terms of their perceived relationships with their students. Some tutors were more or less likely to report a conflictual or close relationship with their students in general relative to other tutors. Even in the case of the Closeness measure, in which we saw large and statistically significant impacts based on instructional modality, the estimated variation in tutor effects remained higher. Although interpreting tutor effects on tutor-reported measures is challenging, the results are consistent with the overall trend observed across all outcomes. Specifically, there was considerable variation in outcomes attributable to differences between tutors, and in each instance, the effect of having a tutor who was one standard deviation more impactful than the average tutor was greater than the effect of being in-person versus remote.

Heterogeneity by Tutor Characteristics

Given the significant variation we observe in tutor effects, a natural next question is whether there are any observable characteristics about tutors that are predictive of better student outcomes. The administrative data on tutors from CUNY allow us to examine the impact of three tutor-level characteristics: their cumulative GPA, whether they were enrolled in an education-related major, and whether they were pursuing a bachelor's degree versus an associate's degree. Table 7 presents the results from including fixed effects for these three variables into Model 1. We find no statistically significant impact of GPA or degree type on any of the outcomes. Of note, however, is the large and statistically significant impact of having a tutor who was pursuing an education-related major on students' academic outcomes. Specifically, we find that having a tutor who was pursuing an education-related major as opposed to some other major improved students' Acadience outcomes by 0.128 standard deviation units and students' Star outcomes by 0.156 standard deviation units, on average. We do not see similar impacts of major on non-academic outcomes.

Tutor Effects by Modality

Table 8 presents the results from fitting Model 2, in which we allow the tutor effect to vary based on instructional format by estimating the standard deviation of the tutor-level random intercept separately for in-person and remote, as well as the correlation between the tutor effects for each modality. In Column 1, we see that there is slightly more variation in the remote tutor effects compared to the in-person effects for Acadience, but they are nearly perfectly correlated. The results in Column 2 for the Star outcome display a similar trend. As we saw in Table 7, the tutor effects for Star are greater in magnitude overall than those of Acadience. However, like Acadience, the Star in-person and remote effects have starkly similar standard deviations and are perfectly correlated. In other words, while we observe significant variation in tutor effectiveness, there is nearly no difference in tutors' relative effectiveness based on instructional format. Tutors who are relatively more effective in the in-person setting tend to be relatively more effective in the remote setting and vice versa.

Column 3 presents these statistics for the number of sessions outcome. Again, we see that the variation in the in-person effect and in the remote effect are similarly large. Interestingly, however, the moderate correlation between the two parameters suggests that tutors who are more effective at raising attendance in-person are not necessarily more effective at raising attendance remotely. Being in-person may be more important for increasing attendance amongst students assigned to tutors who are relatively worse at inducing attendance in general.

Finally, we also find that the in-person and remote tutor effects on their perceived relationship with their students is also nearly perfectly correlated as shown in Columns 4 and 5. This is not surprising given the self-reported nature of these measures. What is noteworthy, however, is that we see substantially more variation between tutors in the remote setting than in the in-person setting in terms of how they perceive Conflict with their students.

Tutor Effects Across Outcomes

Table 9 presents the estimates for the pairwise correlations between tutor effects on one outcome with tutor effects on another outcome. These estimates help us gain a better understanding of whether the tutors who are most effective at improving literacy are also more effective at, say, promoting attendance. In short, while we find a near perfect correlation between the tutor effects on the two literacy outcomes, none of the other outcomes are strongly correlated. We observe a positive, weak correlation between the tutor effect on number of sessions and each of the measures of tutor-student relationship, suggesting that tutors who were more effective in promoting attendance were also relatively more likely to perceive conflict and closeness in their relationships with their students. More interestingly, we do not see a notable correlation between the tutor effects on Star and on number of sessions. This implies that it is not necessarily the case that tutors who are relatively more effective in improving literacy are also more effective in promoting attendance.

Discussion

This study presents the first experimental evidence comparing the marginal impact of in-person versus remote literacy tutoring, or any tutoring, for the same program delivered by the same tutors. Given the random assignment of students to instructional format and tutors, we are also able to explore the variation in student outcomes due to differences between tutors. In sum, we find no statistically significant main effect of being in-person versus remote on students' literacy outcomes. Additionally, while we observe substantial variation in student learning outcomes attributable to differences across tutors, we find no evidence that tutor effects differ by modality. As such, the format of instruction appears to be a less critical factor than selecting and supporting effective tutors in order to drive students' literacy outcomes.

However, a couple caveats are worth noting. First, the estimated variance in tutor effects is large and significant both with regards to Acadience and Star, but it is notably larger for Star. It may be that the Star assessment is a more sensitive measure to differences between tutors. It may also be that tutor effects fade out, such that Acadience, which is administered 2-3 months after the end of the summer program, is not capturing short-term differences in the impact between tutors. Secondly, while the estimated average effect of being in-person versus remote is small and not statistically significant, they are in line with the mean weighted impacts of RCTs of literacy interventions in lower elementary grades, which range from 0.02 to 0.04 (Kraft, 2020). More relevantly, a large-scale national RCT of a district-led summer learning program found a positive 0.08 standard deviation impact on math, but no statistically significant impact on literacy after six weeks (Augustine et al., 2016). This highlights the challenge in trying to move literacy outcomes in a short period, which makes it even harder to show differences in literacy outcomes based on instructional format. Thus, while we can likely rule out large differences in literacy outcomes between in-person and remote, there may still be small differences that are undetectable given the scale of the study.

Moreover, we find that students who were assigned to in-person tutoring exhibited higher attendance rates as reflected in the number of tutoring sessions they received, and tutors reported closer relationships with their in-person students. These results suggest that the modality of instruction facilitates attendance and relationships before it impacts literacy. Because increased instructional time and mentorship bonds are potential mechanisms by which tutoring impacts literacy, program developers may still want to consider instructional format when designing interventions, integrate interpersonal dimensions into tutor training and curriculum development, and implement oversight measures for program attendance and efficacy. This is further supported by the fact that we do not observe strong correlations between tutor effects on literacy outcomes and on attendance. This could be because attendance is conflated with logistical factors, such as working devices, disruption caused by summer travel plans, or changing summer school schedules. It may be that tutors who are better at navigating these logistics are not necessarily the ones who are most effective in improving literacy and vice versa.

In any case, given the substantial role of tutors in impacting student outcomes, it becomes of paramount importance that we can identify effective tutors early on. While there is no evidence that tutors' GPAs or undergraduate degree types affect outcomes, we do find a substantial, statistically significant improvement in students' literacy outcomes, both Acadience and Star, when they are tutored by individuals pursuing education-related majors. These tutors may have been more engaged or invested in training or program implementation due to its relevance to their career advancement. Alternatively, we may be observing the impact of having a tutor who is adequately trained in the science of reading, and is thereby more effective in improving literacy outcomes, especially given the program's alignment in its underlying approach to literacy instruction. These findings offer valuable insights for literacy tutoring providers. As more districts and states move towards adopting evidence-based approaches to literacy instruction, understanding how to scale tutoring programs and identifying and retaining tutors capable of effectively implementing those programs becomes crucial. In theory, tutoring providers are potentially more nimble compared to schools in terms of being able to hire, retain, and dismiss personnel. Thus, exploring additional tutor characteristics that predict effectiveness could be a promising avenue for future research. While existing literature in the context of teacher effectiveness may offer some insights, tutoring presents unique dynamics that warrant further investigation.

Finally, the generalizability of our findings depends on various factors, including the context of the tutoring program. Our evaluation took place over a summer, which presents unique characteristics such as primarily serving low-achieving, low-income students, limited time for observing impacts or building strong tutor-student relationships, and potentially combatting "summer slide," making it challenging to demonstrate growth. The summer context also meant that tutors were recruited from a general pool of undergraduate students interested in paid summer internship programs, potentially resulting in a less specialized talent pool compared to the school year when tutors are recruited from schools of education. Additionally, as part of their program practice, CUNY Reading Corps employs a site coordinator to oversee each tutoring site. This individual is tasked with managing both in-person and remote tutoring sessions conducted at their site. This oversight by the site coordinator may partially explain why we did not observe significant differences between in-person and remote tutoring outcomes. However, it is worth noting that despite the presence of the site coordinator, we still observe significant differences in attendance between the two formats. It is plausible that these differences could have been influenced by connectivity issues experienced during remote sessions, but further investigation is needed. Finally, given that tutors were tutoring in both formats, it's also possible that the lack of observed differences could be because tutors gained experience in both modalities. Perhaps tutoring in-person makes one a more effective tutor remotely or vice versa. This suggests that a tutor who only does remote or only in-person tutoring might be more or less effective than another tutor who is proficient in both formats.

In sum, while in-person tutoring is not always feasible, remote instruction emerges as a promising alternative. It is worth highlighting that remote tutoring is not synonymous with remote instruction as we have come to know it during the COVID-19 pandemic. Tutoring provides "short bursts of instruction," which are more appropriate for young students' short attention spans (Cortes et al., 2023). As such, we conclude that the focus should not solely be on the delivery mode, but rather on ensuring the presence of effective tutors and high-quality instructional materials. Remote instruction, when implemented thoughtfully, can serve as a valuable tool in supporting student learning, particularly in early literacy interventions. Therefore, future efforts should prioritize the identification, support, and retention of effective tutors while also emphasizing the importance of comprehensive program design and implementation.

References

Acadience Learning, Inc. (2021). Benchmarks and Composite Score. Retrieved from: https://acadiencelearning.org/wp-content/uploads/2021/11/Acadience-Reading-K-6-Benchmark-Goals-handout_2021_color.pdf.

Augustine, C. H., McCombs, J. S., Pane, J. F., Schwartz, H. L., Schweig, J. D., McEachin, A., Siler-Evans, K., & Wallace Foundation, commissioning body. (2016). Learning from Summer: Effects of Voluntary Summer Learning Programs on Low-Income Urban Youth. RAND Corporation.

Battistich, V., Schaps, E., & Wilson, N. (2004). Effects of an Elementary School Intervention on Students' “Connectedness” to School and Social Adjustment During Middle School. *The Journal of Primary Prevention*, 24(3), 243–262.

Belson, S.I., Sodani, D.G., Hammeed, R., Miles, K.P., Anderson, A., Smith, M. Dennis, T., Ocheze, J., & Parker, C. (2023). *Consequential high intensity tutoring: Developing literary identities within foundational reading intervention through culturally responsive practices* [Conference Presentation]. American Education Research Association, Chicago, Illinois.

Birch, S.H., & Ladd, G.W. (1997). The teacher–child relationship and children's early school adjustment. *Journal of School Psychology*, 35(1), 61–79.

Callen, I., Carbonari, M.V., DeArmond, M., Dewey, D., Dizon-Ross, E., Goldhaber, D., Isaacs, J., Kane, T.J., Kuhfeld, M., McDonald, A., McEachin, A., Morton, E., Muroga, A., & Staiger, D. O. (2023). Summer School As a Learning Loss Recovery Strategy After COVID-19: Evidence from Summer 2022. Research Report. Cambridge, MA: Center for Education Policy Research, *Harvard University/CALDER Working Paper No. 2901-0823*.

Carlana, M., & Ferrara, E.L. (2021). Apart but Connected: Online Tutoring and Student Outcomes during the COVID-19 Pandemic. *Annenberg Institute at Brown EdWorkingPaper*: 21-350.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5–51.

Chen, J., Jiang, H., Justice, L.M., Lin, T., Purtell, K.M., & Ansari, A. (2020). Influences of Teacher-Child Relationships and Classroom Social Management on Child-Perceived Peer Social Experiences During Early School Years. *Frontiers in Psychology*, 11, 586991–586991.

Connor, C.M., Alberto, P.A., Compton, D.L., O’Connor, R.E. (2014). *Improving Reading Outcomes for Students with or at Risk for Reading Disabilities: A Synthesis of the Contributions from the Institute of Education Sciences Research Centers (NCSER 2014-3000)*. National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education.

Cook, P. J., Dodge, K., Farkas, G., Fryer, R. G., Guryan, J., Ludwig, J., Mayer, S., Pollack, H., & Steinberg, L. (2015). Not Too Late: Improving Academic Outcomes for Disadvantaged Youth. *Policy File*, Institute for Policy Research, Northwestern University.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review. *Review of Educational Research*, 66(3), 227–268.

Cortes, K., Kortecamp, K., Loeb, S., & Robinson, C.D. (2023). A Scalable Approach to High-Impact Tutoring for Young Readers: Results of a Randomized Controlled Trial. Retrieved from: https://studentsupportaccelerator.org/sites/default/files/Scalable%20Approach%20to%20HighImpact%20Tutoring.pdf

de Graaff, S., Bosman, A.M.T., Hasselman, F., & Verhoeven, L. (2009). Benefits of Systematic Phonics Instruction. *Scientific Studies of Reading*, 13(4), 318–333.

Ehri, L. C., Dreyer, L. G., Flugman, B., & Gross, A. (2007). Reading Rescue: An Effective Tutoring Intervention Model for Language-Minority Students Who Are Struggling Readers in First Grade. *American Educational Research Journal*, 44(2), 414–448.

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.

Espinas, D.R., & Fuchs, L.S. (2022). Data-Based Individualization in Reading. *The Reading League Journal*, *3*(1), 4–13.

Fahle, E.M., Kane, T.J., Patterson, T., Reardon, S.F., Staiger, D.O., & Stuart, E.A. (2023) *School District and Community Factors Associated With Learning Loss During the COVID-19 Pandemic*. Center for Education Policy Research, Harvard University.

Fiester, L. (2010). Early warning! Why reading by the end of third grade matters. Research Report. Baltimore, MD: Annie E. Casey Foundation.

Filderman, M.J., Toste, J.R., Didion, L.A., Peng, P., & Clemens, N.H. (2018). Data-based decision making in reading interventions: A synthesis and meta-analysis of the effects for struggling readers. *The Journal of Special Education*, *52*(3), 174–187.

Fryer R.G. Jr. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *Quarterly Journal of Economics*, 129(3), 1355–1407.

Fryer, R.G. Jr. (2016). The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments. *NBER Working Paper Series*, 22130.

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.

Gortazar, L., Hupkau, C., & Roldan, A. (2023). Online tutoring works: Experimental evidence from a program with vulnerable children. *Annenberg Institute at Brown EdWorkingPaper*: 23-743.

Horn, M., Staker, H., & Christensen, C. (2014). Blended: Using Disruptive Innovation to Improve Schools. *The Christensen Institute*.

Karcher, M.J., Davis, C., & Powell, B. (2002). The Effects of Developmental Mentoring on Connectedness and Academic Achievement. *The School Community Journal*, 12(2), 35.

Kim, J.S., & Quinn, D.M. (2013). The Effects of Summer Reading on Low-Income Children's Literacy Achievement From Kindergarten to Grade 8: A Meta-Analysis of Classroom and Home Interventions. *Review of Educational Research*, 83(3), 386–431.

Kim, J.S., Guryan, J., White, T.G., Quinn, D.M., Capotosto, L., & Kingston, H. C. (2016). Delayed Effects of a Low-Cost and Large-Scale Summer Reading Intervention on Elementary School Children's Reading Comprehension. *Journal of Research on Educational Effectiveness*, 9(sup1), 1–22.

Kraft, M.A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, 49(4): 241-253.

Kraft M.A., List J.A., Livingston J.A., & Sadoff S. (2022). Online Tutoring by College Volunteers: Experimental Evidence from a Pilot Program. *American Economic Review: Papers & Proceedings*, 112, 614-618.

Loeb, S., Novicoff, S., Pollard, C., Robinson, C., & White, S. (2023) The Effects of Virtual Tutoring on Young Readers: Results from a Randomized Controlled Trial. Retrieved from: https://studentsupportaccelerator.org/sites/default/files/Effects%20of%20Virtual%20Tutoring%20on%20Young%20Readers.pdf

Lu, M.S., Whittaker, J.E., Ruzek, E., Pianta, R.C., & Vitiello, V.E. (2023). Fostering Early Motivation: The Influence of Teacher-Child Relationships and Interactions on Motivation in the Kindergarten Classroom. *Early Education and Development*, 34(3), 648–665.

McCombs, J.S., Augustine, C.H., Schwartz, H.L., Bodilly, S.J., McInnis, B., Lichter, D.S., & Cross, A.B. (2011). *Making Summer Count: How Summer Programs Can Boost Children's Learning.* RAND Corporation.

Miles, K.P., & Fletcher, A. (2023). Improving Vulnerable Populations' Emergent Reading Outcomes by Training Preservice Teachers in an Evidence-Based Program. *Journal of Research in Childhood Education*, 37(3), 442–462.

Miles, K., Lauterbach, M.D., Murano, D.M., & Dembek, G.A. (2018). Reading Rescue: A follow-up on effectiveness of an intervention for struggling readers. *Journal of Educational Research*, *112*(2), 255–269.

Miles, K.P., McFadden, K. E., Colenbrander, D., & Ehri, L. C. (2022). Maximising access to reading intervention: Comparing small group and one-to-one protocols of Reading Rescue. *Journal of Research in Reading*, *45*(3), 299–323.

McIntosh, K., & Goodman, S. (2016). *Integrated multi-tiered systems of support: Blending RTI and PBIS*. Guilford Publications.

National Institute of Child Health Human Development (NICHD). (2000). *Report of the National Reading Panel: teaching children to read : an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: reports of the subgroups*. National Institute of Child Health and Human Development, National Institutes of Health.

National Student Support Accelerator. (n.d.). Research Priority: Identifying the Characteristics of Effective Tutoring. Retrieved from: https://studentsupportaccelerator.org/research/research-priorities.

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.

Neitzel, A. J., & Reilly, J. (2024). An evaluation of Ignite Reading virtual literacy tutoring in Massachusetts. Center for Research and Reform in Education, The Johns Hopkins University.

Neitzel, A.J., & Storey, N. (2024). Air Reading: A randomized evaluation of a virtual tutoring model. Center for Research and Reform in Education, Johns Hopkins University.

Nickow, A., Oreopoulos, P. & Quan, V. (2020). The Impressive Effects of Tutoring on PreK-12 Learning: A Systematic Review and Meta-analysis of the Experimental Evidence. *NBER Working Paper*, 27476.

Peak, C. (2022). How legislation on reading instruction is changing across the country. APM Reports. Retrieved from: https://www.apmreports.org/story/2022/11/17/reading-instruction-legislation-state-map.

Robinson, C.D. & Loeb, S. (2021). High-Impact Tutoring: State of the Research and Priorities for Future Learning. *Annenberg Institute at Brown EdWorkingPaper*: 21-384.

Robinson, C., Kraft, M., Loeb, S., Schueler, B. (2021). Accelerating Student Learning with High-Dosage Tutoring. *EdResearch for Recovery*. Annenberg Institute at Brown University. Retrieved from: https://annenberg.brown.edu/sites/default/files/EdResearch_for_Recovery_Design_Principles_1.pdf

Rockoff, J.E., Jacob, B.A., Kane, T.J., & Staiger, D.O. (2011). Can You Recognize an Effective Teacher When You Recruit One? *Education Finance and Policy*, 6(1), 43-74.

Saadvera A., Rapaport A., & Silver D. (2020). Matching supports to students needs: Survey results highlight where schools, policymakers can help. *CRPE Evidence Project.* Retrieved from: https://www.crpe.org/sites/default/files/matching_supports_to_student_needs_ep.pdf

Schueler, B.E., & Rodriguez-Segura, D. (2023). A Cautionary Tale of Tutoring Hard-to-Reach Students in Kenya. *Journal of Research on Educational Effectiveness*, 16(3), 442–472.

Schwartz, S. (2024). Which States Have Passed ‘Science of Reading’ Laws? What’s in Them? EdWeek. Retrieved from: https://www.edweek.org/teaching-learning/which-states-have-passed-science-of-reading-laws-whats-in-them/2022/07.

Solari, E., Hall, C., & McGinty, A. (2021). Brick by Brick: A Series of Landmark Studies Pointing to the Importance of Early Reading Intervention. *The Reading League Journal*, *2*(1), 18-21.

Spear-Swerling, L. (Ed.). (2022). *Structured literacy interventions: Teaching students with reading difficulties, grades K-6*. Guilford Publications.

St. Martin, K., Vaughn, S., Troia, G., Fien, & H., Coyne, M. (2020). Intensifying literacy instruction: Essential practices. MiMTSS Technical Assistance Center, Michigan Department of Education.

Terry, N. P., Gerido, L. H., Norris, C. U., Johnson, L., & Little, C. (2022). Building a framework to understand and address vulnerability to reading difficulties among children in schools in the United States. *New Directions for Child and Adolescent Development*, *2022*(183-184), 9-26.

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2024 Reading Assessment.

Vagos, P. & Carvalhais, L. (2022). Online Versus Classroom Teaching: Impact on Teacher and Student Relationship Quality and Quality of Life. *Frontiers in Psychology*, 13, 828774–828774.

Watson, R., Neuman, S.B., & Dickinson, D.K. (2001). Literacy and Oral Language: Implications for Early Literacy Acquisition'. *Handbook of early literacy research*, 43-53.

Workman, J., von Hippel, P.T., & Merry, J. (2023). Findings on Summer Learning Loss Often Fail to Replicate, Even in Recent Data. *Sociological Science*, 10(8), 251–285.

Tables & Figures

Table 1: CUNY Reading Corps Tutor Characteristics

N	139
Female	79%
Asian	36%
Black	16%
Hispanic	27%
White	14%
% Education-Related Major	47%
Avg. Recent Cumulative GPA	3.39

Notes: The sample includes tutors who had at least one student attend at least one section. Education-related majors include Early Childhood Education, Childhood Education, Education, Developmental Psychology, teaching majors, and subject-specific majors for grades 1-6 and 7-12. An additional 10% of tutors declared majors related to Psychology and Sociology.

Table 2: Balance Table

	(1)	(2)	(3)	(4)	(5)	(6)
	Acadience Sample (N=446)			Star Sample (N=426)		
	In-Person	Remote	*p-value*	In-Person	Remote	*p-value*
N	186	260	--	210	216	--
Rising 1st Grade	0.452	0.538	--	0.352	0.472	--
Rising 2nd Grade	0.548	0.462	--	0.419	0.431	--
Rising 3rd Grade	--	--	--	0.229	0.097	--
Avg. Standardized Pre-Test	-0.196	-0.207	0.975	0.024	-0.062	0.525
Missing STRS Response	0.151	0.196	0.171	0.167	0.102	0.187
Has Admin. Data	1.000	1.000	--	0.962	0.954	0.080*
Amongst those with data:						
Female	0.511	0.438	0.014**	0.520	0.417	0.008***
Asian	0.242	0.269	0.215	0.238	0.277	0.193
Black	0.290	0.285	0.582	0.262	0.257	0.496
Hispanic	0.333	0.296	0.759	0.406	0.311	0.875
White	0.091	0.131	0.471	0.074	0.150	0.313
ELL	0.328	0.431	0.089*	0.332	0.417	0.225
ED	0.882	0.812	0.270	0.856	0.816	0.980
SWD	0.220	0.215	0.920	0.252	0.223	0.789

Notes: P-values (* p<.10 ** p<.05 *** p<0.01) are not shown for differences in the percent of students by grade-level because all estimates are calculated with site-by-grade fixed effects. The Acadience sample includes students who attended at least one session and for whom we received Spring 2023 and Fall 2023 assessment data. Third graders are not included because the district does not administer Acadience in 3rd grade. The Star sample includes students who attended at least one session and who took the Star Early Literacy assessment both at the beginning and at the end of the program. STRS refers to the student-tutor relationship survey administered to tutors.

Table 3: Attrition Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Randomized Sample (N=939)			Participant Sample (N=607)			Missing Acadience (N=73)			Missing Star (N=181)		
	In-Person	Remote	*p-value*	In-Person	Remote	*p-value*	In-Person	Remote	*p-value*	In-Person	Remote	*p-value*
N	401	538	--	273	334	--	32	41	--	63	118	--
Withdrawal (from program)	0.319	0.379	0.034**	--	--	--	--	--	--	--	--	--
Rising 1st Grade	0.359	0.448	--	0.374	0.488	--	0.562	0.561	--	0.444	0.517	--
Rising 2nd Grade	0.416	0.390	--	0.425	0.413	--	0.438	0.439	--	0.444	0.381	--
Rising 3rd Grade	0.224	0.162	--	0.201	0.099	--	--	--	--	0.111	0.102	--
Missing Star Pre-Test	--	--	--	0.011	0.054	0.005***	--	--	--	0.048	0.153	0.003***
Missing Star Post-Test	--	--	--	0.231	0.353	0.003***	--	--	--	1.000	1.000	--
Avg. Std. Star Pre-Test	--	--	--	0.020	-0.070	0.461	--	--	--	0.007	-0.086	0.552
Missing Spring Acadience	--	--	--	0.092	0.073	0.943	0.625	0.537	0.336	--	--	--
Missing Fall Acadience	--	--	--	0.128	0.123	0.448	0.875	0.902	0.186	--	--	--
Avg. Std. Spring Acadience	--	--	--	-0.281	-0.255	0.972	-0.804	-0.325	0.455	--	--	--
Missing STRS Response	--	--	--	0.183	0.195	0.397	0.312	0.171	0.829	0.238	0.364	0.020**
Has Admin. Data	0.958	0.950	0.044**	0.960	0.958	0.067*	0.781	0.756	0.789	0.952	0.966	0.812
Amongst those with data:												
Female	0.487	0.456	0.160	0.504	0.438	0.006***	0.480	0.355	0.903	0.450	0.474	0.681
Asian	0.180	0.190	0.137	0.214	0.250	0.330	0.080	0.065	0.494	0.133	0.202	0.600
Black	0.271	0.335	0.385	0.271	0.291	0.458	0.320	0.355	0.954	0.300	0.351	0.933
Hispanic	0.430	0.339	0.851	0.401	0.322	0.864	0.560	0.484	0.954	0.383	0.342	0.448
White	0.081	0.114	0.244	0.084	0.122	0.578	0.040	0.097	0.494	0.117	0.070	0.519
ELL	0.312	0.356	0.068*	0.324	0.403	0.138	0.400	0.226	0.747	0.300	0.377	0.226
ED	0.857	0.826	0.722	0.866	0.806	0.395	0.760	0.774	0.491	0.900	0.789	0.238
SWD	0.258	0.243	0.851	0.256	0.228	0.839	0.120	0.226	0.106	0.267	0.237	0.678

Notes: P-values (* p<.10 ** p<.05 *** p<0.01) are not shown for differences in the percent of students by grade-level because all estimates are calculated with site-by-grade fixed effects. The participant sample includes all students who attended at least one session. The missing Acadience and missing Star samples are subsets of the participant sample who were missing one or both pre-post assessments. The missing Acadience sample does not include rising 3rd graders because the district does not administer this assessment in 3rd grade. STRS refers to the student-tutor relationship survey administered to tutors.

Table 4: Outcomes of Interest

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Total			In-Person			Remote		
	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N
Acadience Reading:									
Spring 2023 RCS	113.657	84.475	446	116.022	87.551	186	111.965	82.333	260
Fall 2023 RCS	112.303	75.372	446	119.048	79.483	186	107.477	72.056	260
Spring 2023 % Well Below	0.439	--	446	0.435	--	186	0.442	--	260
Fall 2023 % Well Below	0.469	--	446	0.430	--	186	0.481	--	260
Star Early Literacy:									
Pre-Test SS	630.211	133.131	426	643.086	132.652	210	617.694	132.706	216
Post-Test SS	634.190	133.130	426	648.871	140.204	210	619.917	136.023	216
Pre-Test NCE	28.248	21.609	426	28.016	20.439	210	28.473	22.735	216
Post-Test NCE	31.073	21.809	426	31.204	21.166	210	30.946	22.466	216
Non-Academic:									
Number of Sessions	13.114	6.424	607	13.949	6.331	273	12.431	6.332	334
Conflict Score	1.709	0.989	492	1.636	1.013	223	1.769	0.967	269
Closeness Score	3.911	0.957	492	4.133	0.847	223	3.727	1.004	269

Notes: Acadience Reading and Star Early Literacy scores are shown for students with complete pre-post assessment data. RCS refers to reading composite scores. SS refers to raw scaled score. NCE refers to normal curve equivalent and % Well Below is the percent of students deemed as "well below" Acadience Reading's criterion-referenced benchmark. NCEs are scores that have been scaled so that they have a normal distribution with a mean of 50 and are similar to percentile ranks (PRs) in that they indicate where a score falls in a distribution of scores (i.e. a national norm population). However, NCEs are on an equal-interval scale, which enables them to be averaged and to be utilized for growth estimates. To interpret them, we can convert them back to PRs. NCE and the benchmark categories are not primary outcomes of interest but are included to contextualize the raw scores. Conflict and Closeness are two dimensions captured by a survey administered to tutors about their perceived relationships with their students.

Table 5: Pairwise Correlations Between Outcomes

	(1)	(2)	(3)	(4)	(5)
	Acadience	Star	Number of Sessions	Conflict	Closeness
Acadience	1.000	0.543	0.051	-0.143	0.103
Star	0.543	1.000	0.110	-0.265	0.120
Number of Sessions	0.051	0.110	1.000	0.067	0.048
Conflict	-0.143	-0.265	0.067	1.000	-0.391
Closeness	0.103	0.120	0.048	-0.391	1.000

Notes: Correlations are estimated using standardized scaled scores for Acadience Reading and Star Early Literacy and standardized Likert scale scores for Conflict and Closeness.

Table 6: Marginal Impact of In-Person Over Remote Tutoring

	(1) Acadience	(2) Star	(3) Sessions	(4) Conflict	(5) Closeness
Fixed parameters:					
β_1 (in-person)	0.030	0.017	1.831***	-0.093	0.396***
	(0.050)	(0.082)	(0.532)	(0.099)	(0.092)
β_0 (constant)	0.308***	0.352*	11.872***	0.639**	-0.172
	(0.130)	(0.202)	(1.460)	(0.276)	(0.242)
Random parameters:					
σ_{u0} (tutor intercept s.d.)	0.060	0.220	2.234	0.191	0.496
	(0.068)	(0.064)	(0.352)	(0.101)	(0.058)
σ_e (student-level residual s.d.)	0.467	0.683	5.131	0.912	0.718
	(0.018)	(0.029)	(0.177)	(0.035)	(0.028)
Number of Students	446	408	565	467	467
Number of Tutors	129	127	137	117	117

Notes: Standard errors are shown in parentheses. Statistical significance (* $p<.10$ ** $p<.05$ *** $p<0.01$) is shown only for the fixed parameters. All estimates are calculated with school-by-grade fixed effects and include controls for students' baseline achievement and demographic characteristics, including race/ethnicity, English language learner status, disability status, and economically disadvantaged classification. For all non-academic outcomes, the baseline achievement included is for Star. Results are slightly bigger in magnitude and more precisely estimated if Acadience pre-test scores are used instead. However, using Acadience results in smaller sample sizes and would exclude rising third grade students.

Table 7: Impact Estimates by Tutor Characteristics

	(1) Acadience	(2) Star	(3) Sessions	(4) Conflict	(5) Closeness
Fixed parameters:					
β_1 (in-person)	0.036	0.025	1.785***	-0.091	0.413***
	(0.053)	(0.083)	(0.581)	(0.110)	(0.100)
β_2 (cumulative GPA)	-0.046	0.103	0.758	-0.074	0.133
	(0.049)	(0.080)	(0.612)	(0.098)	(0.113)
β_3 (education major)	0.128**	0.156*	0.952	-0.084	0.079
	(0.059)	(0.092)	(0.753)	(0.123)	(0.145)
β_4 (bachelor's degree)	0.059	-0.072	-0.198	0.000	-0.200
	(0.076)	(0.134)	(0.982)	(0.158)	(0.188)
β_0 (constant)	0.319	0.617	8.768***	1.013**	-0.370
	(0.220)	(0.429)	(2.618)	(0.461)	(0.472)
Random parameters:					
σ_{u0} (tutor intercept s.d.)	0.067	0.148	2.294	0.108	0.477
	(0.066)	(0.086)	(0.373)	(0.186)	(0.061)
σ_e (student-level residual s.d.)	0.470	0.653	5.133	0.933	0.712
	(0.019)	(0.030)	(0.190)	(0.039)	(0.030)
Number of Students	397	351	487	396	396
Number of Tutors	112	111	117	98	98

Notes: Standard errors are shown in parentheses. Statistical significance (* p<.10 ** p<.05 *** p<0.01) is shown only for the fixed parameters. All estimates are calculated with school-by-grade fixed effects and include controls for students' baseline achievement and demographic characteristics, including race/ethnicity, English language learner status, disability status, and economically disadvantaged classification. For all non-academic outcomes, the baseline achievement included is for Star. Results are slightly bigger in magnitude and more precisely estimated if Acadience pre-test scores are used instead. However, using Acadience results in smaller sample sizes and would exclude rising third grade students.

Table 8: Tutor Random Effects for In-Person and Remote

	(1)	(2)	(3)	(4)	(5)
	Acadience	Star	Sessions	Conflict	Closeness
σ_{u0} (tutor remote intercept s.d.)	0.071	0.222	2.431	0.363	0.509
σ_{u1} (tutor in-person intercept s.d.)	0.050	0.217	2.632	0.054	0.478
σ_{e} (student-level residual s.d.)	0.466	0.683	5.005	0.892	0.718
$\rho_{u1,u0}$ (corr. remote, in-person)	1.000	1.000	0.540	1.000	1.000
Number of Students	446	408	565	467	467
Number of Tutors	129	127	137	117	117

Notes: All estimates are calculated with school-by-grade fixed effects and include controls for students' baseline achievement and demographic characteristics, including race/ethnicity, English language learner status, disability status, and economically disadvantaged classification. Fixed effects are suppressed for succinctness. For all non-academic outcomes, the baseline achievement included is for Star. Results are slightly bigger in magnitude and more precisely estimated if Acadience pre-test scores are used instead. However, using Acadience results in smaller sample sizes and would exclude rising third grade students.

Table 9: Correlations of Tutor Effects Between Outcomes

	(1)	(2)	(3)	(4)	(5)
	Acadience	Star	Number of Sessions	Conflict	Closeness
Acadience	1.000	1.000	--	--	0.092
Star	1.000	1.000	0.014	-0.100	0.005
Number of Sessions	--	0.014	1.000	0.306	0.169
Conflict	--	-0.100	0.306	1.000	-0.188
Closeness	0.092	0.005	0.169	-0.188	1.000

Notes: All estimates are calculated with school-by-grade fixed effects and include controls for students' baseline achievement and demographic characteristics, including race/ethnicity, English language learner status, disability status, and economically disadvantaged classification. Fixed effects and other random effect estimates are suppressed for succinctness. In some cases, the joint model between two outcomes did not converge, thus the correlation between the tutor effects for those outcomes cannot be estimated. In those cases, we cannot reject that the correlation is 0. For all non-academic outcomes, the baseline achievement included is for Star. Results are slightly bigger in magnitude and more precisely estimated if Acadience pre-test scores are used instead. However, using Acadience results in smaller sample sizes and would exclude rising third grade students.

Figure 2: Distribution of Pre- and Post-Percentile Rank Scores on Acadience Reading

Notes: Data are shown for students in the analytic sample who had complete pre-post data for the Acadience Reading assessment (N=446).

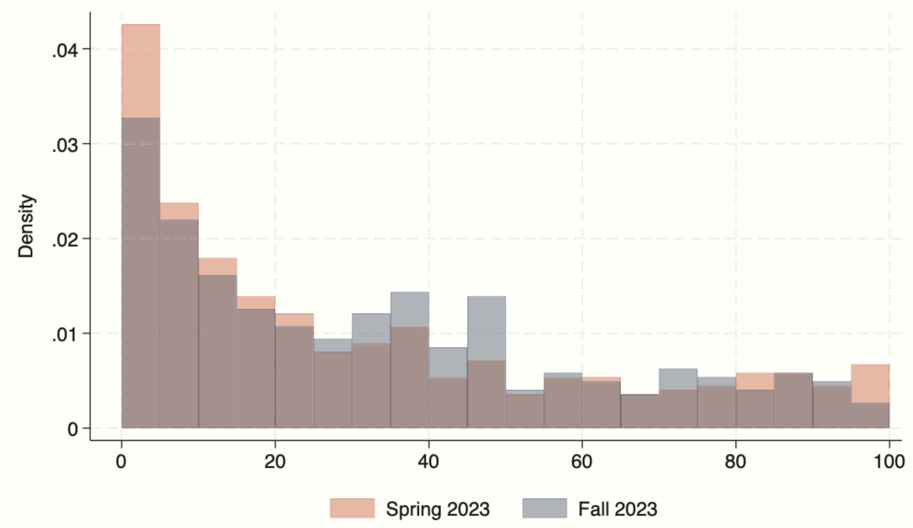
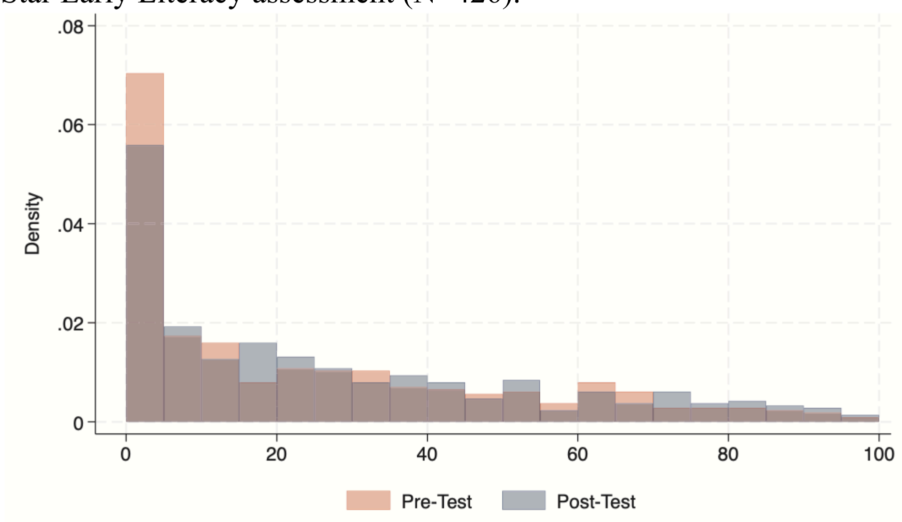


Figure 3: Distribution of Pre- and Post-Percentile Rank Scores on Star Early Literacy

Notes: Data are shown for students in the analytic sample who had complete pre-post data for the Star Early Literacy assessment (N=426).



Appendix A: Student-Tutor Relationship Scale

	Original Version: **Student-Teacher Relationship Scale - Short Form** **(Pianta, 1992)**	**Modified Version:** **Student-Tutor Relationship Scale**
Response Categories	1 - Definitely does not apply 2 - Not really 3 - Neutral, not sure 4 - Applies somewhat 5 - Definitely applies	1 - Definitely does not apply 2 - Not really 3 - Neutral, not sure 4 - Applies somewhat 5 - Definitely applies
Questions	1. I share an affectionate, warm relationship with this child. 2. This child and I always seem to be struggling with each other. 3. If upset, this child will seek comfort from me. 4. This child is uncomfortable with physical affection or touch from me. 5. This child values his/her relationship with me. 6. When I praise this child, he/she beams with pride. 7. This child spontaneously shares information about himself/herself. 8. This child easily becomes angry with me. 9. It is easy to be in tune with what this child is feeling. 10. This child remains angry or is resistant after being disciplined. 11. Dealing with this child drains my energy. 12. When this child is in a bad mood, I know we're in for a long and difficult day. 13. This child's feelings toward me can be unpredictable or can change suddenly. 14. This child is sneaky or manipulative with me. 15. This child openly shares his/her feelings and experiences with me.	1. I share an affectionate, warm relationship with this child. 2. This child and I always seem to be struggling with each other. 3. If upset, this child will seek comfort from me. 4. This child values his/her relationship with me. 5. When I praise this child, they beam with pride. 6. This child spontaneously shares information about themselves. 7. This child easily becomes angry with me. 8. It is easy to be in tune with what this child is feeling. 9. Dealing with this child drains my energy. 10. When this child is in a bad mood, I know we're in for a long and difficult session. 11. This child's feelings toward me can be unpredictable or can change suddenly. 12. This child openly shares their feelings and experiences with me.
Scoring	Closeness: 1, 3, 4R, 5, 6, 7, 9, 15 Conflict: 2, 8, 10, 11, 12, 13, 14 Subscale scores are the mean of included items. Item 4 is reverse scored.	Closeness: 1, 3, 4, 5, 6, 8, 12 Conflict: 2, 7, 9, 10, 11 Subscale scores are the mean of included items.

Summary of Modifications: All binary gender pronouns were replaced with gender-neutral pronouns. Question 4 was removed since it refers to "physical touch," which is not relevant in the remote learning environment. Questions 10 and 14 were removed because they were not relevant to the limited and structured student-tutor interaction. The word "day" was replaced with "session" in Question 12.