EdWorkingPaper No. 26-1595

Tutor Traits and Student Achievement: Evidence from a Randomized Experiment

Elizabeth Huffaker
University of Florida

Bess Levin Markel
Stanford University

Carly D. Robinson
Stanford University

Susanna Loeb
Stanford University

Tutors are increasingly central to school-based instruction, with the potential to personalize learning at scale. Little is known about which tutor characteristics predict effectiveness, limiting providers' and schools' recruiting, hiring, and retention strategies. Using random assignment of tutors in a high-dosage elementary-grade online literacy tutoring program, we examine how literacy development relates to demographic congruence and tutor preparation, experience, and demographic and cultural characteristics. Contrary to the teacher-student match research, we do not find that Black students benefit uniquely from assignment to Black tutors. Neither bachelor's degrees nor classroom teaching experience predicts superior gains. Instead, experience with the same provider is associated with 0.2 SD greater literacy growth, underscoring the importance of tutor development and retention for large-scale programs.

VERSION: September 2026

Suggested citation: Huffaker, Elizabeth, Bess Levin Markel, Carly D. Robinson, and Susanna Loeb. (2026). Tutor Traits and Student Achievement: Evidence from a Randomized Experiment. (EdWorkingPaper: 26-1595). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/jgzm-qp55

Tutor Traits and Student Achievement: Evidence from a Randomized Experiment

Elizabeth Huffaker[1*]

Bess Levin Markel[2]

Carly D. Robinson[2]

Susanna Loeb[2]

[1] College of Education, University of Florida, 1121 W 5th Ave, Gainesville, FL 32601
[2] Graduate School of Education, Stanford University, 507 Lasuen Mall, Stanford, CA, 94305
[*] Corresponding author. Email: ehuffaker@ufl.edu

September 2026

Abstract – Tutors are increasingly central to school-based instruction, with the potential to personalize learning at scale. Little is known about which tutor characteristics predict effectiveness, limiting providers' and schools' recruiting, hiring, and retention strategies. Using random assignment of tutors in a high-dosage elementary-grade online literacy tutoring program, we examine how literacy development relates to demographic congruence and tutor preparation, experience, and demographic and cultural characteristics. Contrary to the teacher-student match research, we do not find that Black students benefit uniquely from assignment to Black tutors. Neither bachelor's degrees nor classroom teaching experience predicts superior gains. Instead, experience with the same provider is associated with 0.2 SD greater literacy growth, underscoring the importance of tutor development and retention for large-scale programs.

Acknowledgments – We are grateful to our partners in this research for their invaluable contribution to project implementation. We thank Accelerate for their generous support of this research, as well as the Smith Richardson Foundation for their support of our full research program. We also received insightful feedback and research support from the National Student Support Accelerator team.

Tutor Traits and Student Achievement: Evidence from a Randomized Experiment

Abstract

Tutors are increasingly central to school-based instruction, with the potential to personalize learning at scale. Little is known about which tutor characteristics predict effectiveness, limiting providers' and schools' recruiting, hiring, and retention strategies. Using random assignment of tutors in a high-dosage elementary-grade online literacy tutoring program, we examine how literacy development relates to demographic congruence and tutor preparation, experience, and demographic and cultural characteristics. Contrary to the teacher-student match research, we do not find that Black students benefit uniquely from assignment to Black tutors. Neither bachelor's degrees nor classroom teaching experience predicts superior gains. Instead, experience with the same provider is associated with 0.2 SD greater literacy growth, underscoring the importance of tutor development and retention for large-scale programs.

The increasing emphasis on personalized learning, accelerated by efforts to address COVID-19 related learning loss, has led to tutors playing an increasingly integral role in school instruction. Following the onset of the COVID-19 pandemic, approximately 4.2 billion dollars of state-directed recovery funding was dedicated to tutoring and accelerated learning over the course of four school years (Council of Chief State School Officers, 2024). Even the sunsetting of these funding sources (i.e., the "ESSER cliff") has not reversed the evolution of tutoring from a peripheral intervention into a standard instructional activity in schools. According to the NCES pulse survey, the share of schools offering tutoring increased to a new high of 85% in June 2025 (National Center for Education Statistics, 2025), with 42% providing tutoring characterized as high-dosage (i.e., delivered in one-on-one or small-group sessions of at least 30 minutes, three or more times per week; Robinson & Loeb, 2021). This substantial dedication of school resources, both financial and in the form of scarce school-day instructional time, is justified by a large body of positive empirical evidence (Nickow et al., 2024).

However, the post-pandemic scale-up has faced implementation challenges, with several recent studies identifying challenges like dosage and staffing constraints that limit the full potential of ostensibly high-impact programs (Bhatt et al., 2025; Kraft et al., 2024; Huffaker et al., 2025a; Huffaker et al., 2025b; Ready et al., 2026). Amid high turnover, the challenge of maintaining a large and capable tutor pool is a substantial barrier to delivering transformative learning gains to students via tutoring (Makori et al., 2024), especially because the personal relationship component of tutor-student dynamics may be a key mechanism for academic and non-academic benefits (Bleiberg et al., 2026; Juel, 1996; Lee et al., 2026). To meet the growing demand for tutoring, providers have recruited tutors from a wide range of backgrounds, including teachers, volunteers, paraprofessionals, and college students, yielding a diverse but high-turnover workforce (Condliffe et al., 2026; Nickow et al., 2024; White et al., 2022).

While meta-analysis suggests that each type of paid tutor can be broadly effective with sufficient training and support (Nickow et al., 2024), only sparse research exists on the tutor characteristics, such as experience and demographics, that are associated with student achievement. This contrasts with the teacher literature, where decades of research have documented systematic associations between teacher characteristics–including demographic background, shared identity with students, professional experience, and training–and student achievement (e.g., Dee, 2005; Hanushek, 1986; Darling-Hammond, 2000; Yosso, 2005). However, tutors often work outside traditional school systems and occupy instructional roles that differ from classroom teachers, so it remains unclear whether insights from the teacher workforce generalize to tutors.

Existing theory based on teachers suggests several possible sources of variation in tutor effectiveness. One emphasizes congruence between educator and student identities. A second focuses on demographic and cultural characteristics of tutors that may shape relationships with students and approaches to instruction. A third highlights tutors' educational preparation and professional experience. Given tutoring's expanding role in the core instructional mission of public education and the diversity of the tutor workforce, understanding which tutor characteristics are associated with effectiveness is a practical priority for providers and districts alike. Clarifying which, if any, of these characteristics matter would inform how tutoring programs recruit, assign, train, and retain tutors.

This study uses the random assignment of tutors within a large-scale literacy tutoring study to examine three categories of tutor characteristics. First, we investigate demographic congruence between tutors and students, with a particular focus on Black tutor-student pairings. Second, we examine tutor demographic and cultural characteristics, including race, multilingual status, and parental education. Third, we consider indicators of educational preparation and professional experience, including bachelor's degree attainment, classroom teaching experience, and prior tutoring experiences. Identifying how instructor characteristics shape student learning is difficult for two reasons: educators are rarely assigned to students at random (e.g., less-proficient students are often matched to less-experienced instructors; Kalogrides & Loeb, 2013), and instructor traits are cross-correlated and either difficult (e.g., level of education) or impossible (e.g., racial identity) to assign at random. The random assignment in this study addresses the former, allowing us to estimate the effect of being assigned a tutor with a given characteristic free from student sorting. It does not address the latter: because tutor traits remain correlated, these estimates reflect an association with a bundle of characteristics rather than any single trait in isolation.

We find limited evidence that demographic congruence as measured by Black student-tutor pairings is associated with literacy growth. Students assigned to Black tutors demonstrate somewhat higher literacy growth on average, although this pattern does not appear to reflect race-specific matching effects. We also find little evidence that commonly emphasized indicators of educator preparation, including bachelor's degree attainment and prior classroom teaching experience, predict student outcomes. By contrast, prior tutoring experience with the focal provider is associated with substantially greater student literacy growth, approximately 0.2 standard deviations. This provider-specific experience effect is larger than any other tutor characteristic examined in the study and suggests that familiarity with a tutoring program may matter more for student learning than many conventional measures of educator qualifications.

This study contributes to a growing literature on tutoring implementation and effectiveness in three ways. First, it provides experimental evidence on demographic congruence in tutor-student relationships. Second, it extends research on educator characteristics into the tutoring workforce by examining demographic, cultural, educational, and experiential dimensions of tutor backgrounds. Third, and most importantly, it leverages the random assignment of tutors to provide credible evidence that provider-specific tutoring experience is associated with student achievement, suggesting that tutor effectiveness may depend as much on organizational learning and retention as on the characteristics tutors bring with them into the role.

Literature Review

Given the broad consensus that instructor quality varies substantially across individuals, a deep cross-disciplinary research base has emerged to identify which educator characteristics support student learning. We draw from three strands of this literature to examine variation in tutor effectiveness. The first emphasizes demographic congruence between instructors and students, including racial and ethnic match. The second highlights demographic and cultural characteristics that may shape how educators relate to students and approach instruction. The third focuses on educational preparation and professional experience.

Educator-Student Demographic Congruence

Drawing on theories of implicit bias, role modeling, cultural congruence, and stereotype threat, a large literature has examined whether students benefit from demographic congruence with educators (Dee, 2004; Ladson-Billings, 1995; Steele & Aronson, 1995; Gershenson et al., 2016). A meta-analysis of demographic match research found that achievement benefits are most consistently established for Black teacher-student pairings, while evidence for other pairings remains more limited (Redding, 2019). Proposed mechanisms include more favorable teacher perceptions, expectations, and discipline of students of color (Milner, 2005; McGrady & Reynolds, 2013; Oates, 2003; Capers, 2019), greater opportunities for identity affirmation, role modeling, and higher student perceptions (Irvine, 1989; Grissom et al., 2015; Cherng & Halpin, 2016), and the potential incorporation of culturally relevant instructional practices that draw on students' experiences and backgrounds (Mitchell, 1998; Lynn & Jennings, 2009: Milner, 2016).

Many of these pathways may be even more salient in tutoring, where there are small student-to-instructor ratios and sustained tutor-student relationships. Tutors, for instance, often engage in personal conversations with students, providing opportunities to leverage student experiences and interests during sessions (Chi et al., 2001; Fitzsimmons et al., 2021). Indeed, in one of the only extant identity alignment studies on tutors, Bleiberg et al. (2026) find that being randomly paired with a female math tutor increased girls' STEM interest by nearly three-quarters of a standard deviation.

At the same time, tutoring relationships may provide opportunities for students and tutors to connect around shared interests and experiences that are unrelated to demographic identities, potentially reducing the salience of demographic matching. Their comparatively lower position of authority may likewise reduce role model effects, as tutors may not be perceived as successful example figures—a key tenet of the theory (Gladstone & Cimpian, 2021; Downing et al., 2005). Lastly, tutors also have less influence over curriculum than classroom teachers, potentially limiting opportunities to draw on culturally relevant knowledge, which is often the mechanism by which teachers of color support their students of color (Irvine, 2003; Egalite et al., 2005). Whether these competing mechanisms ultimately strengthen or weaken the importance of demographic congruence in tutoring is therefore an empirical question. Because tutoring can draw on a more ethnoracially diverse labor pool than the teaching profession (Robinson & Loeb, 2021), evidence that demographic matches improve student outcomes could feasibly be prioritized, informing how tutors are recruited and assigned.

Tutor Demographic and Cultural Characteristics

Beyond demographic congruence, scholars have argued that educators of particular backgrounds bring personal and cultural resources that shape how they relate to students and support learning. Yosso's (2005) framework of community cultural wealth highlights forms of knowledge and experience often overlooked in conventional discussions of educator quality. These resources include linguistic, familial, navigational, resistance, aspirational, and cultural capital. In this study, we focus on two dimensions that are particularly relevant to the backgrounds of tutors in our sample: linguistic and familial capital.

Linguistic capital refers to the skills and perspectives associated with multilingualism. Research suggests that multilingual individuals often develop additional communication skills, literacy practices, and forms of expression (Martínez, 2010; Orellana & Reynolds, 2008; García & Kleifgen, 2020). Scholars have extended this concept to teachers, arguing that multilingual educators may draw on these experiences to connect with a wider range of students and support learning in diverse educational settings (Carreón et al., 2005; Higgins & Ponte, 2017). These advantages may also be relevant in tutoring, where relationship-building and individualized communication are common features of instruction. We therefore examine whether multilingual tutors are associated with stronger student literacy growth.

A second dimension of community cultural wealth is familial capital, which refers to knowledge, values, expectations, and resources acquired through family relationships (Bourdieu, 1993; Yosso, 2005). Family experiences shape educational and professional trajectories in a variety of ways, including beliefs about schooling, educational attainment, and familiarity with postsecondary and labor-market pathways (Henderson & Berla, 1994; Christenson et al., 1992; Conley, 2001). These experiences may also influence how tutors approach instruction and interact with students. In particular, tutors whose parents completed higher levels of education may have had greater exposure to educational institutions and pathways that shape their beliefs about learning and academic success. Parental education is also a core component of conventional socioeconomic status measures (Entwisle & Astone, 1994; Sirin, 2005), so estimates for this indicator capture familial capital together with broader socioeconomic background. We therefore examine parental education as one observable indicator of familial capital and explore whether it is associated with student reading growth.

Tutor Preparation and Experience

A third perspective emphasizes educational preparation and professional experience. Research on teacher effectiveness has long examined whether conventional measures of human capital, including educational attainment, credentials, and professional experience, predict student achievement (Hanushek, 1986; Podolsky et al., 2019). These characteristics are also plausible indicators of tutor quality. Given that tutors often have lower levels of educational attainment than teachers (ACS 2020-2024 5-year estimates; Ruggles et al., 2025) and tutoring-specific credentials remain uncommon, understanding whether the same characteristics that predict teacher effectiveness predict tutoring effectiveness has important implications for tutor recruitment, training, and retention.

Educational Background

Debate persists regarding whether formal educational attainment and credentials improve educator effectiveness. A longstanding finding in the teacher effectiveness literature is that additional educational attainment is only weakly associated with student achievement.

Hanushek's (1986) influential meta-analysis found little evidence that teachers' advanced degree attainment predicts student learning, and subsequent studies have similarly found that advanced degrees and the selectivity of teachers' undergraduate institutions are generally poor predictors of effectiveness (Chingos & Peterson, 2011; Nye et al., 2004; Ladd, 2008). However, some scholars argue that advanced degrees are unnecessary for teachers because of the pedagogical strength of traditional credentialing programs, which are associated with greater teacher effectiveness and lower turnover (Ladd, 2008; Boyd et al., 2009; Clotfelter et al., 2007; Darling-Hammond, 2000). Since tutors do not typically receive equivalent preparation, holding a bachelor's may be a more salient predictor of tutor quality.

While few tutors hold advanced degrees, a related question in the tutoring context is whether tutors benefit from possessing at least a bachelor's degree. As with advanced degrees for teachers, a B.A. is often not formally required for a tutoring job but is often strongly encouraged and incentivized. Conversely, some tutoring programs actively recruit college students as tutors. If formal educational attainment has a limited relationship with teacher effectiveness (Nye et al., 2004; Wilson & Floden, 2003), a similar pattern may emerge in tutoring contexts. If so, these findings would lend further support to recruiting college students as tutors (Kraft & Falken, 2021; Kraft et al., 2022; Robinson & Loeb, 2021; Robinson et al., 2021; Robinson et al., 2026).

Professional Experience

Unlike the returns to education and training, it is well established that teachers improve rapidly during their first few years of classroom experience, producing measurable gains to student achievement (Boyd et al., 2008; Podolsky et al., 2019). Like most professions, teaching involves a steep learning curve, and increased familiarity with instructional demands yields gains. Mid-to-late career returns on experience are much smaller (Clotfelter et al., 2007;

Gershenson, 2016; Papay & Kraft, 2015). In a high-turnover field like tutoring (Condliffe et al., 2026; White et al., 2022), long-run gains are unlikely to materialize regardless, but the extent to which tutors see a similar short-run growth trajectory is highly relevant to questions of strategies for tutor hiring resource allocation across recruitment and retention. Tutoring, being an often transitional or supplementary job (Nickow et al., 2024), raises additional questions about the transferability of experience across roles and contexts. For instance, given the narrower instructional role of tutoring, do tutors benefit from general educational experience (e.g., classroom teacher, paraeducator) that may emphasize a broader range of functions, such as small group management? And given the comparatively structured instructional environments of many tutoring programs, do gains from experience transfer across providers? Experience gained with a particular tutoring organization may be especially valuable, as tutors become familiar with the provider's curriculum, instructional routines, training systems, and supports. Whether gains from experience reflect broadly transferable instructional skills or familiarity with a particular tutoring model remains an open question.

Professional experience may influence tutor effectiveness through two distinct pathways: the development of general instructional expertise and familiarity with a specific tutoring organization's instructional model. Several features of the contemporary tutoring landscape suggest that experience may be particularly context dependent. Tutors often work within structured instructional models that rely on provider-specific curricula, training systems, progress-monitoring tools, and routines for supporting students. As tutors accumulate experience, they may become more proficient in implementing these practices and integrating them into their instruction. If these skills are closely tied to a particular tutoring model, experience gained within one organization may not transfer fully across providers.

Distinguishing between general and organization-specific returns therefore has implications for how organizations approach recruitment, training, and retention. This study contributes experimental evidence on these questions.

Program Background

We implemented this study in partnership with an established curriculum and assessment company that launched a virtual reading tutoring program following the onset of the COVID-19 pandemic. The curriculum, instructional materials, and nationally normed assessment used in the tutoring program drew from the company's existing diagnostic and intervention products. The provider managed tutor recruitment, training, coaching, and ongoing implementation support for participating districts. The provider's standardized approach is particularly well suited to the aims of this study because curriculum, training, coaching, and implementation supports were largely consistent across tutors and school sites. This consistency reduces variation in program design and allows greater focus on variation in tutor characteristics. We emphasize that this study was not designed to identify the overall impacts of the program, but instead to explore the relative efficacy of participating tutors.

The program was designed to align with widely cited features of high-impact tutoring. Specifically, tutoring was delivered to small groups (i.e., 3-4 students) at least three times per week by a consistent tutor over the course of a semester (e.g., Robinson et al., 2024). To tailor instruction to student learning needs, baseline achievement data were used to group students with similar skill profiles. Tutors then targeted lessons at the group-specific learning gaps. Because these practices align with widely adopted principles of high-impact tutoring (Robinson et al., 2024), the findings may generalize to many contemporary school-based tutoring programs.

By the 2021-22 academic year, the provider had served approximately 6,200 K-5 students across 95 schools in the two states included in this study. Tutoring was delivered exclusively online. Region 1 is located in the South Atlantic Census Division, while Region 2 is located in the West South-Central Census Division. The provider also operated tutoring programs and implementation support services in additional regions beyond those included in this study.

Methods

The study was designed to leverage the random assignment of tutors to small groups to examine how tutor characteristics relate to student outcomes. We address two research questions:

- RQ1: Does being matched with a Black tutor affect the effectiveness of high-impact tutoring for Black students?
- RQ2: How are tutor demographic, cultural, educational, and professional characteristics associated with student literacy growth?

RQ1 is our preregistered[1] confirmatory analysis. Although we initially planned to examine additional dimensions of demographic congruence, including Hispanic identity and gender, sample composition limited our ability to do so. We therefore focus on Black tutor-student pairings because this comparison offers the strongest statistical power in our data and aligns with the primary focus of the student-teacher congruence literature (Redding, 2019).

RQ2 is motivated by the educator effectiveness, human capital, and cultural capital literatures reviewed above. This exploratory analysis uses the random assignment of tutors to examine whether tutor demographic, cultural, educational, and experiential characteristics are associated with student literacy growth.

Overall, RQ1 examines demographic congruence between tutors and students, while RQ2 examines whether broader tutor characteristics are associated with student literacy growth. In the following sections, we detail the data, sample, randomization structure, and estimation strategies used to answer these questions.

Data and Measures

We draw on four primary sources of data collected by the provider: student achievement and demographic records, tutoring attendance records, and tutor surveys. These deidentified student- and tutor-level records were linked using shared tutor group identifiers.

Student achievement files include scores from the DIBELS-8 reading assessment administered at the beginning, middle, and end of the school year. We standardized scores within grade and assessment administration. These files also included basic student demographic information (i.e., student gender, language learner status, and race/ethnicity) collected by the assessment platform. Additional demographic variables were available for only a subset of students and were therefore excluded from the analysis. Race/ethnicity categories are mutually exclusive and allow students to identify as American Indian or Alaska Native, Asian, Native Hawaiian or Pacific Islander, Hispanic/Latino, Black/African American, White, or Multiracial. Students could also choose not to specify their race/ethnicity.

Tutoring attendance files included counts of assigned and attended tutoring sessions and the resulting attendance rate. Because many students were reassigned to different tutors after the winter assessment window, we focus on attendance measures from October through December, which correspond most closely to the period of initial tutor assignment. Attendance was automatically recorded in the online platform.

The research team designed student and tutor surveys to support the study and collect information not available in administrative records (see Appendix for survey instruments). Because student surveys were administered at only one site, they are excluded from the present analyses. The tutor survey collected measures corresponding to the three categories of tutor characteristics examined in this study: demographic congruence, demographic and cultural characteristics, and educational preparation and professional experience. Key measures included tutors' race/ethnicity, gender, multilingual status, parental education, educational attainment, prior tutoring experience, and classroom teaching experience. The literature review informed survey items on educator characteristics and effectiveness (e.g., Hanushek, 1986).

To measure race/ethnicity, the survey included both closed-response categories and an open-ended identity item. This approach allowed us to identify tutors reporting multiple racial or ethnic identities that would not have been captured by the closed-response categories alone. Although all tutors were asked to complete the baseline survey, item nonresponse varied across measures.

Sample

The provider shared data on 1,952 students who received tutoring during the 2022-23 academic year across the two study regions. Because the analyses require linking student outcomes to randomly assigned tutors with observed characteristics, we impose several restrictions to construct the analytic sample. Consistent with our preregistered design, we restrict the sample to students assigned to tutoring groups at the beginning of the school year and focus on the first tutor randomly assigned to each group. This excludes 314 students who first entered tutoring during the spring semester. The resulting sample contains 1,638 students assigned to 512 tutoring groups and linked to 155 unique tutors. Of these tutors, 122 (79%) completed at least one item on the tutor survey, corresponding to 1,412 students across 420 tutoring groups. We examine the implications of survey nonresponse in supplementary analyses.

Because Black and non-Black tutors were alternated on the assignment list used for randomization, analyses of Black tutor-student pairings are restricted to the portion of the sample in which assignment to a Black tutor remained possible. This design-based restriction reduces the sample by 403 students. We additionally exclude observations in which student or tutor race/ethnicity is missing or recorded as "prefer not to say" for analyses requiring racial or ethnic classification. Supplementary analyses demonstrate that results are robust to alternative coding decisions for these cases.

The final analytic sample contains 1,032 randomized student-tutor matches in which race/ethnicity is known for both individuals. Of these students, 943 have both baseline (i.e., beginning-of-year DIBELS-8 composite) and endline (i.e., middle-of-year DIBELS-8 composite) literacy scores and comprise our primary ITT sample. Table 1 presents summary statistics for these students. Students span kindergarten through fifth grade, with roughly two-thirds enrolled in grades 3-5, and are balanced by gender (51% female). The sample is predominantly Black (76%), with smaller shares of White (15%) and Hispanic (6%) students, and 16% are classified as English language learners. Ninety-two percent of students were enrolled in the primary geographic region (i.e., Region 1). Baseline and endline reading scores are standardized within grade and assessment administration across all students served by the provider. On average, students attended 33 tutoring sessions during the fall semester, corresponding to approximately 43% of assigned sessions.

Table 2 summarizes the 86 tutors included in the match sample. All tutors report prior tutoring experience, and 85% had previously tutored for the study program. This variation in provider-specific experience allows us to examine whether experience acquired within a particular tutoring organization is associated with student outcomes. The tutor workforce is overwhelmingly female (92%) and racially diverse, with Black (43%) and White (41%) tutors comprising the majority and the remaining 16% identifying as another race. Seventeen percent of tutors report speaking more than one language. This composition broadly resembles other large-scale tutoring samples. Across 80 Personalized Learning Initiative (PLI) sites, tutors were 38% Black, 45% White, 27% Hispanic, and 38% multilingual (Condliffe et al., 2026).

Only seven tutors identified as male, preventing planned analyses of gender congruence and illustrating the pronounced gender imbalance within the tutor workforce. Sixty-five percent of tutors held a bachelor's degree. This sits between the 70% reported among PLI tutors (Condliffe et al., 2026) and the 46% reported among tutors in the 2020-2024 American Community Survey extract (Ruggles et al., 2025). The remaining third reflects the provider's practice of hiring current college students, many of whom reported attending some college but had not yet completed a bachelor's degree. Fifty-five percent of tutors reported having at least one parent with a bachelor's degree (among the 82 who completed that field).

Half of tutors report prior classroom teaching experience, and a similar share report experience as teaching assistants. Tutors led an average of 5.3 tutoring groups, with an average group size of 2.7 students.

Randomization

Our randomization strategy adapts established methods for instructor-level randomization (e.g., Antecol et al., 2015; Bleiberg et al., 2026) to the provider's tutor assignment process. Tutors were randomly assigned to small groups of students within tutor region-by-recruitment-wave strata. Among the 86 tutors in the analytic sample, most were in the first (N=76) and in the primary region (N=63). Several additional features of the assignment process warrant discussion.

First, the research team generated separate randomized assignment lists for each region, ensuring that tutors were assigned only to groups within their designated region. To support analyses of demographic congruence, Black and non-Black tutors were alternated on the assignment lists, yielding an approximately 50 percent probability of assignment to a Black tutor within the analytic sample. Because tutors were recruited throughout the year, separate assignment lists were generated for each recruitment wave. Three recruitment waves were ultimately randomized, although 88 percent of tutors included in the analysis were recruited during the first wave (93 percent of students).

Second, student groups were not randomly ordered before tutor assignment. The provider formed within-grade tutoring groups using beginning-of-year achievement data, including both composite scores and skill-specific assessment information, and ordered groups from lower to higher proficiency. To account for this feature of the assignment process, all analyses control for baseline achievement and a floor-score indicator. The results are robust to alternative specifications, including models that directly control for group order (Table A3).

We also assess the validity of the randomization procedure. For RQ1, we test whether baseline achievement differs across tutor-student match types after controlling for preregistered design fixed effects, including randomization strata and school fixed effects. We find no statistically significant differences in baseline achievement between students assigned to Black versus non-Black tutors within racial groups, consistent with successful randomization (Table A2). We additionally assess differential attrition and find no statistically significant relationship between treatment status and endline score availability (Table A2, Column 2), suggesting that attrition is unlikely to threaten internal validity.

Finally, analyses addressing RQ2 must account for variation in tutor caseload. Tutors could choose how many groups they wished to tutor, resulting in workloads ranging from one to nine groups, with an average of approximately five. If tutor caseload is correlated with both tutor characteristics and effectiveness, for example if more successful tutors may take on large caseloads, estimates of tutor characteristics may be biased. We therefore control for a tutor's caseload, measured as the number of tutoring groups assigned to each tutor, in all RQ2 analyses. This adjustment is particularly important because tutors with prior experience in the focal program instructed, on average, more groups than tutors without prior program experience.

Estimation Strategies

We leverage the random assignment of tutors to examine how tutor characteristics relate to student literacy growth. We first estimate the effects of Black tutor and Black student pairings (RQ1). We then examine whether broader tutor demographic, cultural, educational, and professional characteristics are associated with student outcomes (RQ2).

Student-tutor racial congruence [RQ1]

To address RQ1, we estimate whether pairing a Black tutor with a Black student is associated with student literacy growth using the following specification.

$$Y_{ijks} = \beta_0 + \beta_1 BlackTutorXBlackStudent_{is} + \beta_2 BlackTutorXNonBlackStudent_{is} + \beta_3 NonBlackT \quad (1)$$

Y_{ijks} is the outcome of interest for student *i* in grade *j* of school *k* in region-by-entry cohort strata *s*. We use winter DIBELS-8 composite literacy scores as our primary outcome. We focus on winter outcomes because tutors and groups were non-randomly reassigned after the winter assessment window based on formative assessment data, making end-of-year outcomes less clearly attributable to the initial random assignment. We also estimate this specification using fall semester attendance as the outcome.

β_1 is the coefficient of interest for our main outcome while β_2 and $\beta_3$capture outcomes for the remaining tutor and student race combinations. α_j, α_k, and α_s represent grade-level, school, and randomization strata (tutor region-by-entry-cohort) fixed effects. X_i is a vector of pretreatment student level covariates, including gender, race/ethnicity, and beginning-of-year composite reading scores. We also present results estimated using a more parsimonious equation that excludes demographic controls. The residual term is ε_{ijks}. Because "treatment" status (i.e., demographic match) is determined by student characteristics, we use heterogeneity robust standard errors in the main analysis. In supplementary analyses we cluster standard errors at the tutor level.

One complication in interpreting Equation 1 is that any overall advantage associated with Black tutors may appear alongside demographic congruence effects. If Black tutors are, on average, more effective than non-Black tutors, both the Black tutor and Black student pairing coefficient and the Black tutor and non-Black student coefficient may be positive. To distinguish demographic congruence from a more general tutor race association, we also estimate Equation 2:

$$Y_{ijks} = \beta_0 + \beta_1 BlackTutorXBlackStudent_{is} + \beta_2 BlackStudent_{is} + \beta_3 NonBlackTutor_{is} + \alpha_j \quad (2)$$

Here, the coefficient of interest remains β_1, while β_3 captures the average association between assignment to a Black tutor and student outcomes. Equations 1 and 2 capture the same underlying relationships, but use different parameterizations. Equation 2 separates demographic congruence effects from overall differences between Black and non-Black tutors, making interpretation more straightforward.

Tutor characteristics [RQ2]

To address RQ2, we estimate the relationship between tutor characteristics and student literacy growth using the following specification, estimated separately for each tutor characteristic:

$$Y_{ijkst} = \beta_0 + \beta_1 TutorTrait_t + \beta_3 BaselineScore_i + BlackTutor_t + NGroups_t + \alpha_j + \alpha_k + \alpha_s + X_i$$

(3)

Here Y_{ijkst} is the standardized literacy composite score for student *i* in grade *j* of school *k* in region-by-entry cohort strata *s* who was assigned tutor *t*. β_1 is the coefficient of interest and captures the association between the tutor characteristic under study and student literacy growth. α_j, α_k, and α_s represent grade, school, and randomization strata (region-by-entry-cohort) fixed effects. X_i is a vector of pretreatment student-level traits, including gender and race/ethnicity. We also present results from a more parsimonious specification that excludes student demographic controls. Relative to Equations 1 and 2, we additionally control for whether a tutor is Black and the number of tutoring groups assigned to the tutor. The first reflects the tutor assignment procedure, while the second accounts for variation in tutor caseloads. The residual term is ε_{ijkst}. Because tutor characteristics vary only at the tutor level, standard errors are clustered by tutor.

We examine tutor characteristics from two broad categories. The first captures educational preparation and professional experience, including bachelor's degree attainment, classroom teaching experience, teaching assistant experience, and prior experience with the tutoring provider. The second captures demographic and cultural characteristics, including race and ethnicity, multilingual status, and parental education. These categories correspond directly to the conceptual framework developed in the literature review and allow us to examine whether characteristics highlighted in prior research on educator effectiveness are associated with student literacy growth in tutoring contexts.

Interpreting these estimates requires distinguishing two challenges that complicate research on how instructor characteristics shape student learning. The first arises when educators are non-randomly assigned to students so that estimates reflect the students an instructor tends to teach. The second arises because instructor traits are bundled: individuals who possess one characteristic tend to differ systematically along other, correlated characteristics. The random assignment of tutors to groups addresses the first challenge directly: after conditioning on the design controls, tutor characteristics are unrelated to student characteristics, so our estimates are not driven by the selective matching of students to tutors. However, because tutor characteristics remain correlated with one another (Figure A1), each coefficient should be interpreted as the effect of being assigned a tutor who possesses the characteristic in question, together with its correlates. In other words, these estimates capture the effect of assigning students to tutors who possess a given characteristic as it naturally occurs in the workforce, rather than the isolated causal effect of that characteristic alone. Still, because hiring decisions are based on these observable packages of traits, we view these bundled estimates as the most relevant for policy and practice.

Additional Analyses

In addition to the robustness checks described throughout the main analyses, we conduct four sets of supplemental analyses. First, we assess the sensitivity of results to alternative sample restrictions, including whether to retain respondents who declined to report their race/ethnicity and whether to retain tutors whose race/ethnicity was first reported on the second fall survey administration (Appendix B). Second, we examine heterogeneity by student age by estimating our primary models separately for students in grades K-3 and grades 4-5. These analyses assess whether the relationship between tutor characteristics and student outcomes varies across developmental stages. Third, we test the sensitivity of results to the inclusion of additional group-level controls, including group size, group order, average baseline achievement, and the share of students identifying as Black within the tutoring group. Although these characteristics should be unrelated to tutor assignment after conditioning on the randomization strata, they may nevertheless shape tutoring interactions and student learning. These analyses therefore provide both a robustness check and insight into whether group composition moderates the relationships examined in this study. Fourth, because tutor characteristics are correlated with one another (Figure A1), we assess whether the provider-experience association is robust to conditioning on other tutor traits by estimating a single specification that enters all tutor characteristics simultaneously (Table A8).

Results

We present results for our two research questions. First, we examine whether Black tutor and Black student pairings are associated with student literacy growth (RQ1). Second, we examine whether tutor demographic, cultural, educational, and professional characteristics are associated with student outcomes (RQ2). We conclude by presenting supplemental and sensitivity analyses.

Tutor-Student Demographic Congruence

Table 3 presents estimates for RQ1, which examines whether Black tutor-student pairings are associated with student literacy growth. Across specifications, we find no evidence that, on average, Black students assigned to Black tutors demonstrate greater literacy growth than Black students assigned to non-Black tutors. Under our preregistered specification (Panel A), the coefficient for the Black tutor-student pairing is positive in magnitude (approximately 0.06 standard deviations) but statistically insignificant across models that include design and baseline controls (Column 1), student demographic controls (Column 2), and group-level controls (Column 3). Although the confidence intervals rule out large congruence effects, they cannot rule out smaller effects. Panel B presents estimates from an alternative specification that includes Black tutor and Black student main effects alongside a single interaction term for Black tutor-student pairings.

Results across different measurement approaches are substantively similar. The estimated demographic congruence effect is generally robust across alternative specifications and sample restrictions (Table A3). Results remain essentially unchanged when using alternative clustering approaches, excluding tutors who reported race/ethnicity only in the second survey administration, including tutors who selected "prefer not to say," and controlling for additional tutor- and group-level characteristics. One exception emerges in the K-3 subsample, where the Black tutor-student pairing coefficient is positive and statistically significant (0.12 SD). No comparable effects are observed among students in grades 4-5. Because we did not hypothesize heterogeneous treatment effects by grade level, we view this finding as exploratory and interpret it cautiously. Additional research is needed to determine whether it reflects a meaningful developmental difference or chance variation.

The absence of a significant demographic congruence effect does not appear to be explained by differences in tutoring attendance. As shown in Table A4, students assigned to Black and non-Black tutors attended tutoring sessions at similar rates after accounting for student characteristics. More broadly, estimates of generic tutor-student race match effect, using both the preregistered framework and a within-tutor fixed effects approach (i.e., because collinearity issues are avoided in race-neutral match analysis) yield no evidence of systematic demographic congruence effects on literacy growth (Table A5). The within-tutor approach (Panel B) does find that matched pairs exhibit slightly stronger total session attendance (i.e., plus 1.6 sessions per semester) with marginal significance.

Tutor Characteristics

We next examine RQ2, which investigates whether tutor demographic, cultural, educational, and professional characteristics are associated with student literacy growth. Figures 1 and 2 present coefficient estimates from models that control for randomization features, baseline achievement, and student demographic characteristics. Corresponding regression estimates are reported in Appendix Tables A6 and A7. All tutor characteristics are measured using binary indicators, so coefficients represent comparisons between tutors possessing a given characteristic and tutors who do not.

Demographic and Cultural Characteristics

Figure 2 and Table A7 present estimates for tutor demographic and cultural characteristics. Among these, the only positive predictor we identify, with marginal statistical significance, is that students assigned to Black tutors in our sample demonstrate slightly higher literacy growth on average than students assigned to non-Black tutors (i.e., 0.07 SD). By contrast, students assigned to multilingual tutors and tutors classified in the "other race" category demonstrate somewhat lower literacy growth than comparison students, by approximately 0.12 and 0.11 standard deviations, respectively. These estimates should be interpreted cautiously because they are based on a relatively small and heterogeneous – by both second language and race – group of tutors. Only 17% of tutors in the sample were multilingual, and 16% were classified in the "other race" category (Table 2). We find no evidence that tutor parental education — our indicator of familial capital — predicts student literacy growth, though the estimate is imprecise. We were unable to examine tutor gender because only seven tutors in the sample identified as male.

Educational Preparation

Figure 1 and Table A6 present estimates for educational preparation. We find little evidence that degree attainment predicts tutor effectiveness. Students assigned to tutors holding bachelor's degrees do not demonstrate greater literacy growth than students assigned to tutors without bachelor's degrees. Because most tutors without bachelor's degrees were current college students, these estimates compare degree holders to individuals who are generally still progressing through higher education rather than to adults without postsecondary experience.

Professional Experience

We next examine whether professional experience is associated with tutor effectiveness. Prior classroom teaching experience and teaching assistant experience are not associated with student literacy growth. Students assigned to tutors with these backgrounds perform similarly to students assigned to tutors without such experience.

In contrast, prior experience with the focal tutoring provider is associated with greater literacy growth. Students assigned to returning tutors demonstrate approximately 0.2 standard deviations greater literacy growth than students assigned to tutors without prior experience in the program. Because every tutor in the analysis sample reported some prior tutoring experience, tutors without provider-specific experience had tutored in other settings. This estimate therefore captures the association of experience within the focal program over and above general tutoring experience. This estimate is substantially larger than any other tutor characteristic examined in the study and remains similar across alternative model specifications, including models with and without student demographic controls (Table A6). This association also remains the largest and statistically significant coefficient when all tutor characteristics are entered jointly (Table A8).

Discussion

Using random assignment of tutors within a large-scale literacy tutoring program, this study examined two questions: whether demographic congruence between Black students and Black tutors is associated with literacy growth, and whether broader tutor characteristics are associated with student outcomes. Three findings stand out. First, we find limited evidence that Black tutor-student pairings improve literacy outcomes. Second, we find no evidence that requiring tutors to hold a bachelor's degree improves student outcomes; current college students performed similarly to bachelor's degree holders. Third, prior experience with the tutoring provider is associated with substantially greater literacy growth. These findings suggest that tutor continued experience with a tutoring program may be more consequential for student learning than other characteristics commonly emphasized in tutor recruitment and hiring, such as traditional credentials or demographic matching.

Key Findings and Future Directions for Research

Unlike much of the analogous literature on teacher demographic congruence (Dee, 2004; Redding, 2019), we do not find evidence that racial congruence between tutors and students is meaningfully associated with student achievement – nor are bachelor's degree attainment or prior classroom teaching experience. Among the characteristics examined, prior experience with the tutoring provider exhibits the strongest association with literacy growth, corresponding to a fifth of a standard deviation greater literacy growth than that observed for students assigned to tutors without program-specific experience.

Understanding why provider-specific experience is associated with stronger student outcomes is an important direction for future research. At least two mechanisms could plausibly produce this pattern, and the present design cannot fully distinguish between them. One possibility is that returning tutors have genuinely learned. They have received more program-specific training and professional development, and they accumulate expertise through repeated experience delivering instruction within the same curriculum, routines, and instructional expectations. This interpretation is consistent with research showing that educator effectiveness improves during the early years of practice (Boyd et al., 2008; Podolsky et al., 2019), and with emerging evidence that returns to experience may not fully transfer across educational settings (Kraft et al., 2026). It would also help explain why prior classroom teaching and instructional assistant experience, acquired outside this provider's model, are not associated with student outcomes in our analyses. An equally plausible alternative is that this relationship reflects selection. More effective tutors may be more likely to remain with the organization because successful instructional experiences strengthen self-efficacy, which is itself associated with educator retention (Xiao & Zheng, 2025). Providers may also selectively retain stronger tutors through reappointment decisions or targeted retention efforts. Because tutor characteristics are not themselves randomly assigned, the present study cannot adjudicate between these accounts. What random assignment does eliminate is the key confound that complicates most educator-effectiveness research: that students are selectively assigned to educators.

Future work that combines experimental designs with measures of tutor training, instructional practice, and retention decisions may be particularly valuable for disentangling these pathways. More detailed data on tutor tenure and experience would also help distinguish learning and selection explanations. For example, a learning account would be more consistent with improvements that accumulate gradually with additional experience, in contrast, a selection account would be more consistent with differences that emerge primarily between returning and non-returning tutors.

Implications for Policy and Practice

The findings point to retention and organizational support as important priorities for tutoring providers and districts. Online tutoring can lower barriers to entry by allowing tutors to work remotely and with flexible schedules. However, these same features may also lower barriers to exit, making turnover an ongoing challenge for providers seeking to build stable tutor workforces. Prior experience with the focal provider is strongly associated with student literacy growth, and districts and providers may benefit from strategies that support tutor retention. At the same time, the mechanisms underlying this relationship remain uncertain. Whether provider-specific experience reflects accumulated learning, selective retention of effective tutors, or some combination of the two, the findings suggest that how tutoring organizations develop, identify, and retain tutors deserves greater attention from both researchers and practitioners.

The findings also speak to tutor recruitment and selection. As high-impact tutoring expands, recruitment strategies may emphasize traditional indicators of educator preparation, including degrees, credentials, and classroom teaching experience. In this study, bachelor's degree attainment and prior classroom teaching experience are weak predictors of student outcomes. This pattern suggests that these conventional criteria should be interpreted cautiously in tutoring contexts, especially when providers use structured curricula, training systems, and ongoing support. The results also provide evidence relevant to the widespread practice of hiring college students as tutors. In this setting, tutors without bachelor's degrees produced student outcomes comparable to those of degree holders. Because most non-degree-holding tutors in our sample were current college students, these findings support the use of college students as a tutor workforce in structured tutoring programs. However, because our estimates identify the effect of being assigned a tutor with a given characteristic rather than the effect of the characteristic itself, we caution against reading them as hiring criteria, particularly for tutor race. Finally, the findings suggest that tutor effectiveness in structured programs may depend partly on proficiency within a specific instructional model. Provider-specific experience may capture knowledge and skills that develop through participation in the program itself, including familiarity with curriculum, routines, technology systems, and supports. This interpretation shifts attention toward onboarding, coaching, professional development, and retention as central features of tutor workforce strategy. Future research should examine which organizational supports are most effective in helping tutors develop the knowledge and skills associated with stronger student outcomes.

Limitations

While we can draw several valuable insights from this study, several limitations of the experiment are worth noting. First, tutoring occurred exclusively online, which may attenuate the underlying mechanisms that make demographic matches relevant for students' academic outcomes (Hashim et al., 2025). Second, the tutor characteristics examined in RQ2 are not themselves randomly assigned, and our estimates should be read accordingly. Because tutors were randomly assigned to student groups, these estimates are not confounded by the selective matching of students to tutors that complicates observational studies of instructor characteristics; they recover the effect of being assigned a tutor who possesses a given characteristic. What random assignment cannot do is isolate the independent contribution of any single trait. This distinction matters most for the provider-experience finding random assignment rules out that returning tutors simply taught stronger students, but it cannot distinguish whether these tutors became more effective through experience (learning) or whether more effective tutors were retained (selection). We therefore interpret the 0.2 SD estimate more narrowly as the effect of assignment to a returning tutor in this sample rather than as the causal return to experience per se.

Third, the intervention lasted only 10 weeks, which is shorter than the match period for teacher-student congruence studies. Fourth, comparisons involving several dyads and tutor subgroups of interest, particularly tutors who identify as Hispanic and/or male, were severely underpowered. Finally, since we do not have data on classroom teacher race, potential interactions between tutor and teacher demographics remain unknown. It is plausible that identity match effects are moderated by external student relationships. The teaching force of the focal district for this study is nearly 50% Black, many times the national 7% Black teacher average. Having an assigned tutor be Black may not be high salience to students who are frequently instructed by Black classroom teachers.

Conclusion

Tutors are increasingly core providers of children's education and recipients of substantial resource allocations by schools and districts. Yet little is known about the characteristics, skills, and experiences that make tutors effective. This study provides experimental evidence on this question. The findings suggest that, when tutoring providers use research-based curricula, ongoing coaching, and aligned professional development, tutors from varied backgrounds can support student learning. They also suggest that organizational supports and tutor retention may be important components of effective tutoring systems, especially in structured tutoring programs.

Endnotes

1. Pre-analysis plan [ID number blinded], Open Science Framework registry.

References

Antecol, H., Eren, O., & Ozbeklik, S. (2015). The effect of teacher gender on student achievement in primary school. *Journal of Labor Economics*, *33*(1), 63–89.

Bhatt, M. P., Chau, T., Condliffe, B., Davis, R., Grossman, J., Guryan, J., ... & Stoddard, G. (2025). *Personalized Learning Initiative Interim Report: Findings from 2023-24*. MDRC.

Bleiberg, J., Robinson, C. D., Bennett, E., & Loeb, S. (2026). The impact of tutor gender match on girls' STEM interest, engagement, and performance. *American Educational Research Journal*, *63*(3), 514–543.

Bourdieu, P. (1993). *Sociology in question* (Vol. 18). Sage.

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management, 27*(4), 793–818. https://doi.org/10.1002/pam.20377

Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, *31*(4), 416–440.

Capers, K. J. (2019). The role of desegregation and teachers of color in discipline disproportionality. *The Urban Review*, *51*(5), 789–815.

Carreón, G. P., Drake, C., & Barton, A. C. (2005). The importance of presence: Immigrant parents' school engagement experiences. *American Educational Research Journal*, *42*(3), 465–498.

Cherng, H.-Y. S., & Halpin, P. F. (2016). The importance of minority teachers: Student perceptions of minority versus White teachers. *Educational Researcher*, *45*(7), 407–420.

Chi, M. T., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, *25*(4), 471–533.

Chingos, M. M., & Peterson, P. E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review*, *30*(3), 449–465.

Christenson, S. L., Rounds, T., & Gorney, D. (1992). Family factors and student achievement: An avenue to increase students' success. *School Psychology Quarterly*, *7*(3), 178–193.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007a). How and why do teacher credentials matter for student achievement? (NBER Working Paper No. 12828). National Bureau of Economic Research.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007b). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673–682.

Condliffe, B., Davis, R., Mattera, S. K., & Grossman, J. (2026). *Insights into the tutor workforce: Tutors' experiences, diversity, and retention in the Personalized Learning Initiative*. MDRC. https://www.mdrc.org/

Conley, D. (2001). Capital for college: Parental assets and postsecondary schooling. *Sociology of Education*, *74*(1), 59–72.

Council of Chief State School Officers. (2024). *State approaches to high-impact tutoring: Implementation and lessons learned*. https://ccsso.org/resource-library/state-approaches-high-impact-tutoring

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education Policy Analysis Archives*, *8*(1), 1–44.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, *95*(2), 158–165.

Downing, R. A., Crosby, F. J., & Blake-Beard, S. (2005). The perceived importance of developmental relationships on women undergraduates' pursuit of science. *Psychology of Women Quarterly*, *29*(4), 419–426.

Egalite, A. J., Kisida, B., & Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review*, *45*, 44–52. https://doi.org/10.1016/j.econedurev.2015.01.007

Entwisle, D. R., & Astone, N. M. (1994). Some practical guidelines for measuring youth's race/ethnicity and socioeconomic status. *Child Development,* 65(6), 1521–1540. https://doi.org/10.1111/j.1467-8624.1994.tb00833.x

Fitzsimmons, W., Trigg, R., & Premkumar, P. (2021). Developing and maintaining the teacher-student relationship in one to one alternative provision: The tutor's experience. *Educational Review*, *73*(4), 399–416.

García, O., & Kleifgen, J. A. (2020). Translanguaging and literacies. *Reading Research Quarterly*, *55*(4), 553–571.

Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, *11*(2), 125–149.

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, *52*, 209–224.

Gladstone, J. R., & Cimpian, A. (2021). Which role models are effective for which students? A systematic review and four recommendations for maximizing the effectiveness of role models in STEM. *International Journal of STEM Education*, *8*(1), Article 59.

Grissom, J. A., Kern, E. C., & Rodriguez, L. A. (2015). The "representative bureaucracy" in education: Educator workforce diversity, policy outputs, and outcomes for disadvantaged students. *Educational Researcher*, *44*(3), 185–192.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, *24*(3), 1141–1177.

Hashim, S., Miles, K. P., & Croke, E. (2025). *Experimental evidence on the impact of tutoring format and tutors: Findings from an early literacy tutoring program* (EdWorkingPaper No. 25-1176). Annenberg Institute for School Reform at Brown University.

Henderson, A. T., & Berla, N. (1994). *A new generation of evidence: The family is critical to student achievement*. National Committee for Citizens in Education.

Higgins, C., & Ponte, E. (2017). Legitimating multilingual teacher identities in the mainstream classroom. *The Modern Language Journal*, *101*(S1), 15–28.

Huffaker, E., Lee, M. G., Zhou, H., Robinson, C. D., & Loeb, S. (2025a). *Beyond the one-teacher model: Experimental evidence on using embedded paraprofessionals as personalized instructors*. [Manuscript submitted for publication].

Huffaker, E., Lee, M., Zhou, H., Robinson, C., & Loeb, S. (2025b). *The promise and pitfalls of paraprofessional tutors: Evidence from a pair of randomized controlled trials*. Society for Research on Educational Effectiveness.

Irvine, J. J. (1989). Beyond role models: An examination of cultural influences on the pedagogical perspectives of Black teachers. *Peabody Journal of Education*, *66*(4), 51–63.

Irvine, J. J. (2003). *Educating teachers for diversity: Seeing with a critical eye*. Teachers College Press.

Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly*, *31*(3), 268–289.

Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher*, *42*(6), 304–316.

Kraft, M. A., & Falken, G. T. (2021). A blueprint for scaling tutoring and mentoring across public schools. *AERA Open*, *7*, 1–15. https://doi.org/10.1177/23328584211042858

Kraft, M. A., List, J. A., Livingston, J. A., & Sadoff, S. (2022, May). Online tutoring by college volunteers: Experimental evidence from a pilot program. In *AEA Papers and Proceedings* (Vol. 112, pp. 614–618). American Economic Association.

Kraft, M. A., Papay, J. P., James, J. K., & Monti-Nussbaum, M. (2026). *Is teacher effectiveness fully portable? Evidence from the random assignment of transfer incentives* (Working Paper No. 34845). National Bureau of Economic Research.

Kraft, M. A., Schueler, B. E., & Falken, G. T. (2024). What impacts should we expect from tutoring at scale? Exploring meta-analytic generalizability. *Review of Educational Research*. Advance online publication. https://doi.org/10.3102/00346543261446660

Ladd, H. F. (2008). Teacher effects: What do we know. In *Teacher quality: Broadening and deepening the debate* (pp. 3–26). Harvard Education Press.

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, *32*(3), 465–491.

Lee, M. G., Loeb, S., & Robinson, C. D. (2026). The effects of high-impact tutoring on student attendance: Evidence from a state initiative. *AERA Open*, *12*, 1–16. https://doi.org/10.1177/23328584261448132

Lynn, M., & Jennings, M. E. (2009). Power, politics, and critical race pedagogy: A critical race analysis of Black male teachers' pedagogy. *Race Ethnicity and Education*, *12*(2), 173–196.

Makori, A., Burch, P., & Loeb, S. (2024). *Scaling high-impact tutoring: School level perspectives on implementation challenges and strategies*. [Manuscript submitted for publication].

Martínez, R. A. (2010). Spanglish as literacy tool: Toward an understanding of the potential role of Spanish-English code-switching in the development of academic literacy. *Research in the Teaching of English*, *45*(2), 124–149.

McGrady, P. B., & Reynolds, J. R. (2013). Racial mismatch in the classroom: Beyond black-white differences. *Sociology of Education*, *86*(1), 3–17.

Milner, H. R. (2005). Stability and change in US prospective teachers' beliefs and decisions about diversity and learning to teach. *Teaching and Teacher Education*, *21*(7), 767–786.

Milner, H. R. (2006). Preservice teachers' learning about cultural and racial diversity: Implications for urban education. *Urban Education*, *41*(4), 343–375.

Mitchell, A. (1998). African American teachers: Unique roles and universal lessons. *Education and Urban Society*, *31*(1), 104–122.

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.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257.

Oates, G. L. S. C. (2003). Teacher‐student racial congruence, teacher perceptions, and test performance. *Social Science Quarterly*, *84*(3), 508–525.

Orellana, M. F., & Reynolds, J. F. (2008). Cultural modeling: Leveraging bilingual skills for school paraphrasing tasks. *Reading Research Quarterly*, *43*(1), 48–65.

Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, *130*, 105–119.

Podolsky, A., Kini, T., & Darling-Hammond, L. (2019). Does teaching experience increase teacher effectiveness? A review of US research. *Journal of Professional Capital and Community*, *4*(4), 286–308.

Ready, D. D., McCormick, S. G., & Shmoys, R. J. (2026). The effects of in-school virtual tutoring on student reading development: Evidence from a short-cycle randomized controlled trial. *Journal of Education for Students Placed at Risk (JESPAR)*, 1–21. Advance online publication.

Redding, C. (2019). A teacher like me: A review of the effect of student–teacher racial/ethnic matching on teacher perceptions of students and student academic and behavioral outcomes. *Review of Educational Research*, *89*(4), 499–535.

Robinson, C. D., Kraft, M. A., Loeb, S., & Schueler, B. E. (2021). *Accelerating student learning with high-dosage tutoring* (EdResearch for Recovery Design Principles Series). EdResearch for Recovery Project.

Robinson, C. D., & Loeb, S. (2021). High-impact tutoring: State of the research and priorities for future learning. *National Student Support Accelerator*, *21*(284), 1–53.

Robinson, C. D., Meyer, K., Bailey-Fakhoury, C., Zandieh, A., & Loeb, S. (2026). Answering the call: How changes to the salience of job characteristics affect college students' decisions. *Social Science Research*, 137, 103351. doi: 10.1016/j.ssresearch.2026.103351

Robinson, C. D., Pollard, C., Novicoff, S., White, S., & Loeb, S. (2024). *The effects of virtual tutoring on young readers: Results from a randomized controlled trial* (EdWorkingPaper No. 24-955). Annenberg Institute for School Reform at Brown University.

Ruggles, S., Flood, S., Sobek, M., Backman, D., Cooper, G., Richards, S., Rodgers, R., Schroeder, J., & Williams, K. C. W. (2025). *IPUMS USA: Version 16.0* [Dataset]. IPUMS. https://doi.org/10.18128/D010.V16.0

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. Review of Educational Research, 75(3), 417–453. https://doi.org/10.3102/00346543075003417

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797–811.

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.

Wilson, S. M., & Floden, R. E. (2003). *Creating effective teachers: Concise answers for hard questions. An addendum to the report "Teacher preparation research: Current knowledge, gaps, and recommendations"*. AACTE Publications.

Xiao, Y., & Zheng, L. (2025, July). The relationship between self-efficacy and job satisfaction: A meta-analysis from the perspective of teacher mental health. In *Healthcare* (Vol. 13, No. 14, p. 1715). MDPI.

Yosso, T. J. (2005). Whose culture has capital? A critical race theory discussion of community cultural wealth. *Race Ethnicity and Education*, *8*(1), 69–91.

Table 1 - Student Sample Summary Statistics

	Tutor-Student Match Sample	
	Mean	N
Student Baseline Characteristics		
Baseline Reading Score (Std.)	-0.031	943
	(0.968)	
Kindergarten	0.081	941
1st Grade	0.094	941
2nd Grade	0.146	941
3rd Grade	0.210	941
4th Grade	0.230	941
5th Grade	0.240	941
Female	0.506	943
ELL Status	0.159	943
American Indian / Alaska Native	0.007	943
Asian	0.002	943
Black	0.759	943
Hispanic	0.056	943
Native Hawaiian / Pacific Islander	0.013	943
Multiracial	0.014	943
White	0.148	943
Region 1	0.919	943
Region 2	0.081	943
Outcomes		
Endline Reading Score (Std.)	-0.007	943
	(0.989)	
Tutoring Sessions Attended (Fall)	32.507	938
	(10.695)	
Tutoring Attendance Rate (Fall)	0.431	938
	(0.297)	

Notes: Standard deviations of continuous variables presented in parentheses under the mean.

Table 2 - Tutor Sample Summary Statistics

	Tutor-Student Match Sample	
	Mean	N
Tutor Traits and Background		
Female	0.919	86
Multilingual	0.174	86
Tutor Race: Black	0.430	86
Tutor Race: White	0.407	86
Tutor Race: Other	0.163	86
Tutor Has BA or Higher	0.647	85
At Least One Parent Has BA	0.549	82
Prior Experience as Tutor	1.000	85
Prior Experience as a Tutor for Study Program	0.847	85
Prior Experience as a Teacher	0.500	84
Prior Experience as a Teaching Assistant	0.506	85
Tutoring Load		
Number of Tutor Groups	5.337	86
	(3.411)	
Average Group Size	2.699	86
	(0.651)	

Notes: Some tutors elected not to answer certain survey questions; analyses include all available responses provided by each tutor. Standard deviations of continuous variables presented in parentheses under the mean.

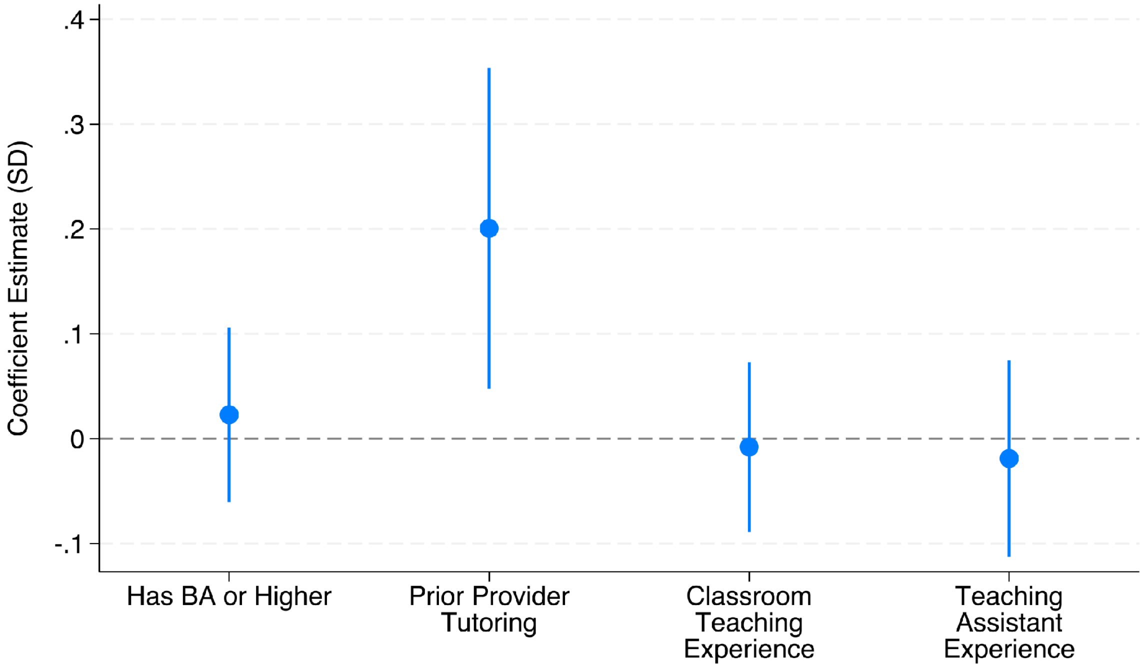
Table 3: Intent-to-Treat (ITT) Effects of Student-Tutor Race Match on Reading Scores

	Standardized Reading Scores		
	(1)	(2)	(3)
A. Equation 1 Estimates			
Black Tutor-Black Student	0.056	0.057	0.056
	(0.042)	(0.042)	(0.042)
Black Tutor-Non-Black Student	0.117	0.114	0.106
	(0.076)	(0.117)	(0.117)
Non-Black Tutor- Non-Black Student	0.010	0.020	0.011
	(0.072)	(0.098)	(0.098)
Conditional Control Mean	-0.079	-0.079	-0.079
B. Equation 2 Estimates			
Black Student	-0.010	-0.020	-0.011
	(0.072)	(0.098)	(0.098)
Black Tutor	0.107	0.095	0.095
	(0.083)	(0.081)	(0.080)
Black Tutor X Black Student	-0.051	-0.038	-0.040
	(0.092)	(0.090)	(0.090)
Conditional Control Mean	-0.079	-0.079	-0.079
N	941	941	941
Controls for student characteristics	NO	YES	YES
Controls for group size and average baseline scores	NO	NO	YES

Notes: All models include controls for baseline student achievement, as well as grade, school, and tutor wave X region fixed effects. Baseline student achievement controls include a standardized measure of baseline achievement and a dummy variable for the minimum score. Student characteristic controls include indicators of student/race ethnicity, gender and English Learner status. Group controls include are baseline group achievement and group size. Sample size was cut off after the last Black Tutor was assigned. Heteroscedasticity robust standard errors are in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

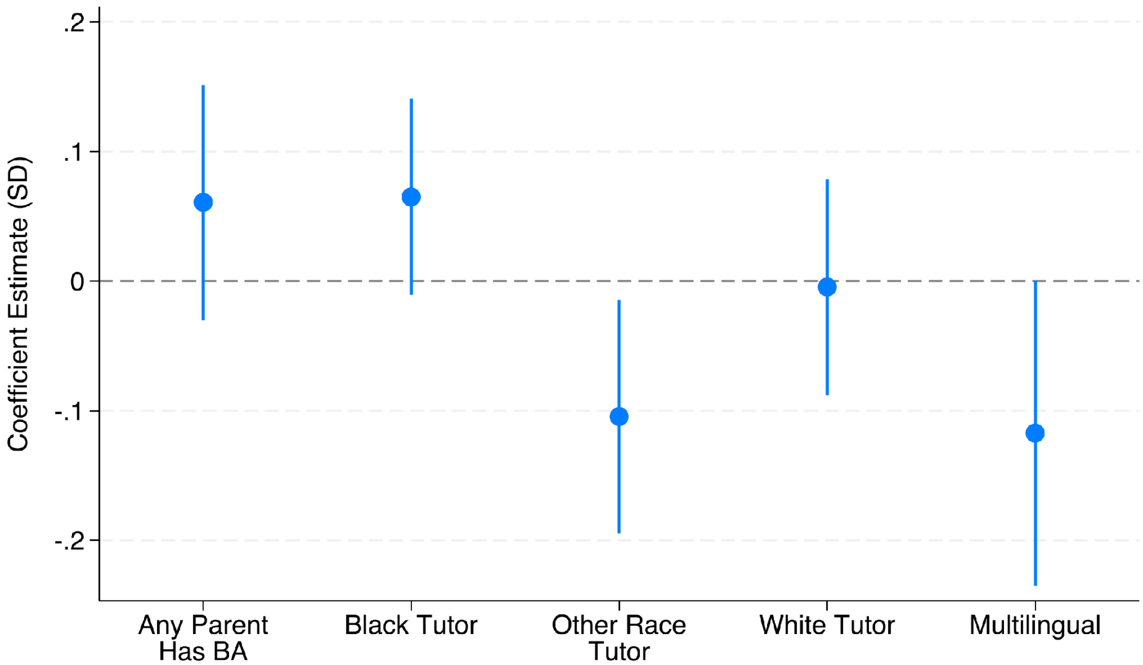
Figures

Figure 1. Relationships Between Tutor Background and Student Reading Growth



Notes: All models include student demographic controls.

Figure 2. Relationship Between Tutor Demographic Traits and Student Reading Outcome



Notes: All models include student demographic controls.