The Impact of Tutor Gender Match on Girls' STEM Interest, Engagement, and Performance

Abstract

Gender disparities in STEM persist despite girls performing as well as boys academically, suggesting girls may benefit from role models who shape their perceptions of STEM. We examine whether female math tutors influence girls' STEM interest, attendance, and performance. We randomly assigned 422 ninth-grade students taking Algebra 1 to a same gender or opposite-gender tutor. Girls assigned to female tutors reported higher STEM interest (0.73 SD) and were more likely to pass the course with a C- or better (3.9-percentage points) than those with male tutors. We found no impact on attendance. Effects were stronger for students working with tutors in-person rather than virtually. We provide the first experimental evidence that female tutors can boost girls' STEM self-concept and academic outcomes.

Keywords:

Gender Studies, Mathematics Education. Self-Concept. Tutoring, Science Education

Introduction

Women are less likely to enter and persist in STEM majors and careers than men (Ceci et al., 2009; Cheryan et al., 2017) . This gender disparity cannot be explained away by ability—females and males perform equally well on STEM achievement assessments (Coley, 2001; Hyde et al., 2008; Keller et al., 2022). Rather, the disengagement of women may derive from their internalizing persistent, widely held misbeliefs about women's math abilities (Frome et al., 2006; Speer, 2023) or from assumptions about the appeal of math-based careers (Diekman et al., 2010). As Dasgupta and Stout (2014) noted, "gender gaps in science and math performance have been closing, but gaps in STEM self-concept and aspirations remain large."

Doubts about their STEM abilities and an absence of relatable role models in STEM fields can lead girls to opt out of pursuing higher level STEM courses and careers (Heybach & Pickup, 2017; Msambwa et al., 2024; M. T. Wang et al., 2013). Early experiences and cultural contexts often contribute to this phenomenon (Halpern et al., 2007; Makarova et al., 2019). For instance, social cognitive theory (Bandura, 1977) suggests that girls' STEM engagement is shaped by the interaction of personal beliefs, behaviors, and environmental influences, highlighting the potential for role models and supporting learning contexts to shape self-concept and motivation. School learning environments can play a pivotal part in either supporting or hindering girls' success and interest in STEM subjects (Msambwa et al., 2024).

Girls' interest in STEM often starts to diminish in high school (Sadler et al., 2012; Veldman et al., 2021), making ninth grade a particularly important time to intervene. Promising interventions that might enhance high school girls' STEM interest include those that provide exposure to STEM role models (González-Pérez et al., 2020; Kyte & Riegle-Crumb, 2017), opportunities to succeed in STEM (Luttenberger et al., 2019), or access to a caring educator who

can help with motivation and navigating stereotypes (Fox-Turnbull et al., 2023; N. Wang et al., 2023). These targeted supports can have a lasting impact on girls' engagement and persistence in STEM fields (Sadler et al., 2012).

Tutoring programs in STEM provide high school students with a unique opportunity to integrate exposure to role models, opportunities for success, and personalized support from a caring educator, making it a promising approach to sustaining and even reigniting high school girls' interest in STEM. Moreover, as tutoring gains popularity as an intervention to improve math test scores and boost students' Algebra 1 passing rates in high school (Guryan et al., 2023), the possibility of deploying STEM tutors in schools–particularly those focused on math gains–is becoming increasingly feasible.

The growing interest in school-based high school math tutoring programs opens the door to exploring how these programs can be tailored to maximize their impact, particularly by considering the potential benefits of exposing students to demographically similar tutors. The existing literature exploring the impacts of educator-student demographic matches suggests that students can benefit academically and psychologically when they share racial or gender characteristics with their teachers, although the results are mixed (Cleveland & Scherer, 2024; Dee, 2005; Ouazad, 2014; Paredes, 2014; Wright et al., 2017). Yet, some research shows that girl students who have female STEM teachers had higher test scores (Paredes, 2014) and were more likely to pursue further STEM coursework (Lim & Meer, 2017).

However, changing the demographic composition of the teacher workforce is not easy nor can it be rapid. Conversely, the tutor workforce is smaller and has a far larger proportion of new hires than teaching; it also does not require the intensity of preservice preparation that teaching does. As a result, students can gain access to demographically matched educators at a

far greater scale and far more quickly with tutors than they would be able to with teachers. We do not yet know whether working with a female STEM tutor will have similar impacts on girl students than working with a female math teacher. Teachers and tutors have different roles. However, because they are both adults working on STEM, they may both be effective role models.

This study examines the impact of assigning female high school students to same-gender math tutors on students' STEM beliefs, attendance, and math performance. Our research is guided by social cognitive theory (Bandura, 1977), role model theory (Morgenroth et al., 2015), and stereotype threat theory (Spencer et al., 1999), which collectively suggest that exposure to demographically similar role models—in this case, female math tutors—may positively shape girls' motivation and academic outcomes in STEM. We hypothesize that girls randomly assigned to female tutors would experience greater benefits compared to those paired with male tutors. We find that girls randomly assigned to female tutors report significantly higher interest in STEM (0.73 SD) than girls assigned to male tutors. Although we find no evidence that the female tutor-student gender match improved students' school attendance, girls assigned to female tutors performed better in their math course. Specifically, they were less likely to receive a grade of D or F grade (3.9 percentage points). As tutoring continues to gain traction as a strategy to accelerate student learning, these results highlight that program design features, such as exposing students to demographically similar tutors, can influence outcomes and that the benefits of tutoring may extend beyond academic achievement.

Background

The Importance of Secondary School for Decreasing Gender Disparities in STEM Fields

Girls and women continue to be underrepresented in STEM fields despite decades of efforts focused on understanding and changing this pattern (Charlesworth & Banaji, 2019). Gender disparities are commonly measured by comparing jobs obtained and college majors pursued, but the problem starts well before high school graduation. Scholars have identified a range of formative sociocultural, psychological, and structural barriers that contribute to this phenomenon (Halpern et al., 2007; Msambwa et al., 2024; M. T. Wang & Degol, 2017). Girls make consequential decisions about whether they intend to pursue careers in STEM fields by the time they enter the 9th grade (Kyte & Riegle-Crumb, 2017).

To pursue STEM fields in college and as careers, maintaining interest in math during high school is crucial—particularly for girls (Kyte & Riegle-Crumb, 2017). Yet, girls' STEM interest is more likely to diminish over the course of high school than their male classmates (Sadler et al., 2012; Speer, 2023). Girls' drop in STEM interest during high school may result from existing gender disparities in STEM participation in the workforce, leading girl students to have less exposure to same-gender STEM workers (Cheryan et al., 2017). This lack of exposure may result in a feedback loop in which girl students are less likely to pursue STEM majors and careers, resulting in comparably fewer female STEM role models for girls as they make critical decisions about their academic trajectory (Ceci et al., 2009).

Given these concerning trends, scholars have highlighted secondary school as a critical period for interventions to increase girls' interest and performance in STEM subjects (Chavatzia, 2017). Educators and peers in schools are important socializing agents who can promote or undermine girls' beliefs about STEM (N. Wang et al., 2023). Banerjee and colleagues (2016) found that female college student decisions to pursue STEM courses is, at least in part, due to their secondary teachers and classroom experiences, as well as their interest in, and the perceived

value of, STEM subjects. High school girl students exposed to females in STEM might believe that pursuing STEM in their future studies and professional careers is more viable, increasing their interest in doing so and altering their career expectations (Lv et al., 2022; Riegle-Crumb & Morton, 2017).

Schools and other educational organizations can strategically introduce female STEM role models for girl students. A brief role model-focused intervention, in which female role models described overcoming challenges and emphasized the importance of the subject, increased academic performance and persistence in one STEM course (Herrmann et al., 2016). Other interventions have emerged in recent years to increase the proportion of female students in STEM fields (González-Pérez et al., 2020; Prieto-Rodriguez et al., 2020). The *Girls Who Code* organization, for example, provides hands-on STEM learning experiences by pairing girls with female mentors in STEM fields to inspire girls to pursue STEM education (Piloto, 2023).

Background Literature on Teacher-Student Demographic Match

Interventions that focus on exposing girl students to female role models in STEM as a way to enhance STEM interest and aspirations build on many of the same theoretical assumptions as the educator-student match literature–that demographically similar role models can be important for changing the beliefs and trajectory for those who feel traditionally marginalized in a field (Blake-Beard et al., 2011). A few studies leverage random variation in student assignments to teachers to estimate the effects of these matches (Blazar, 2024), while most rely on quasi-experimental methods to estimate the impact of being assigned to a same-race or same-gender teacher.

Much of the evidence base on student-teacher match focuses on racial matching (Redding, 2022), which likely stems from the historic and ongoing efforts to address racial

inequality in education (Cabral-Gouveia et al., 2023; Coleman, 1968). Studies on teacher-student racial matching, such as those by Dee (2004, 2005) and Gershenson and colleagues (2021), demonstrate that students, particularly Black students, experience academic and social benefits when they are taught by teachers of the same race or ethnicity.

In contrast, the research on the effect of teacher-student gender matching on student outcomes is mixed (Cho, 2012; Dee, 2007; Ehrenberg et al., 1995; Sokal et al., 2007). One U.S.-based study found that female teachers were more effective, in general, and that they were particularly effective for increasing girl students' math performance (Hwang & Fitzpatrick, 2021), and several international studies show that girl students exposed to female math teachers score higher on math assessments (Lim & Meer, 2017; Paredes, 2014). These effects can persist beyond a single year: Lim and Meer (2020) found that the impact of having a female math teacher in middle school increased the likelihood girls took advanced STEM courses and aspired to do so in college. However, Cho (2012) finds weak evidence for a student-teacher match effect in a study of 15 OECD countries. Additionally, other studies in the U.S., Canada, and the U.K. show no significant impact of gender matching on student achievement (Ehrenberg et al., 1995; Francis et al., 2008; Sokal et al., 2007).

Most of the teacher-student demographic matching literature focuses on academic achievement, such as grades or test scores, but a few studies assess how race and gender matches improve students' perceptions of classroom relationships and aspirations. For instance, Scherer and Cleveland (2024) found that same-race, same-gender teachers can improve students' attendance and interpersonal self-management. Additionally, Egalite and Kisida (2015) showed that same-race, same-gender teacher-student matches increased the likelihood that middle school students considered attending college. The gender gap in future STEM aspirations already exists

by middle school (M. T. Wang & Degol, 2017), and the results from this study indicate that exposure to demographically similar teachers during secondary school matters.

Researchers have proposed several mechanisms through which having a demographically similar educator might positively impact student outcomes. These mechanisms include exposure to role models, which can inspire students by showcasing paths of seemingly similar people (Dee, 2005; Steele, 1997), and mitigation of stereotype threat, where negative preconceptions about students' identity can hinder their confidence and performance (Beasley & Fischer, 2012). Additionally, shared experiences and cultural understanding between students and educators can foster a sense of belonging and support (Hart, 2020; Purdie-Vaughns et al., 2008). Finally, demographic similarity can contribute to the development of more positive and trusting relationships, which are often key to student engagement and success (Hughes et al., 2008).

Overall, the existing research highlights the potential for demographic alignment to shape not only academic outcomes, but also students' perceptions and aspirations. By extending the lens of research beyond the traditional classroom teacher, we can explore how similar dynamics might influence the effectiveness of other adults in students' lives, such as tutors, in fostering student interest and success in STEM. The present study considers, specifically, the extent to which girl students may benefit from being exposed to female math tutors in high school.

The Potential of Female Math Tutors to Increase Girls' STEM Interest and Performance

The primary role of tutors is to deliver personalized, relationship-based instruction, uniquely positioning them as educators who may be able to inspire students and boost their interest and performance in STEM. Tutors meet with students one-on-one or in small groups and, as a result, may have more intensive, personalized time with students than teachers. This dual role of tutors as both academic and motivational role models raises the question: could the

benefits of demographic matching often observed with classroom teachers extend to tutors, who occupy a distinct instructional role in students' educational experiences? To date, no research has estimated the effects of such demographic matching between tutors and students. In this section we discuss why high school math tutoring is a promising intervention point, whether a demographic match with a tutor may be more or less impactful than that with a teacher, and why girls might benefit from having female math tutors.

The Promise of High-Impact Tutoring for Boosting High School Math Achievement

The vast research base documenting the effectiveness of tutoring in accelerating student achievement (Dietrichson et al., 2017; Nickow et al., 2024) has helped tutoring become a popular approach for districts to combat stagnating student growth and widening achievement gaps as a result of the pandemic (Biden Administration, 2024). Evidence for math tutoring in early high school is particularly robust, with large scale randomized controlled trials demonstrating consistent gains for students (Bhatt et al., 2024; Guryan et al., 2023). While an early study focused just on high-impact tutoring for boy students (Guryan et al., 2023), additional research has shown that intensive tutoring is effective for girls as well (Bhatt et al., 2024).

The positive results from high school math tutoring initiatives come from studies evaluating programs that have features associated with "high-impact tutoring," which include high-dosage (three or more sessions per week), making use of formative assessments and other data to monitor student learning and drive instruction, aligning session materials with school curriculum, providing formal tutor training and ongoing supports, and prioritizing working with a consistent tutor (Cortes et al., 2025; Robinson & Loeb, 2021). This last feature—students working with a consistent tutor— facilitates relationship-building between a student and their tutor. As the tutor becomes an additional caring adult in a student's schooling experience, the

tutor-student demographic match may impact students. Districts may engage in other tutoring initiatives, like on-demand tutoring, however less evidence supports the effectiveness of these approaches (Huffaker et al., 2025).

Considering the Influence of Demographically Similar Tutors

Like teachers, tutors take on an instructional role for students, but differences between tutors and teachers might influence whether a student's demographic match with a tutor is more or less impactful than with their teacher. On one hand, having a demographically similar tutor could be less impactful because students tend to spend more time with classroom teachers over the course of the year than with their tutor, changing the "strength" of the treatment (Wedenoja et al., 2022). On the other hand, teachers often teach 30 students in one class at a time, and over 100 students over the course of a school year. Tutoring is, by definition, conducted one-on-one or in small groups. Thus, students and tutors meet frequently and have personalized interactions that might make the demographic match more salient (Dietrichson et al., 2017). Students also have many experiences with teachers over the course of their K-12 education (Havik & Westergård, 2020), whereas tutors may serve as a non-teacher role model, making their impact more novel. In addition, educators' perceptions matter and their own beliefs towards demographically similar students might influence their expectations (Gershenson et al., 2016) or positively bias their grading (Muntoni & Retelsdorf, 2018, 2018). We have little evidence on the difference in perceptions and beliefs between teachers and tutors.

How Female Math Tutors Can Empower Girls in STEM

The theorized mechanisms from many of the racial and gender match studies could explain *why* a girl student might benefit from having a math tutor who shares her gender. First, math tutors may act as a behavioral role model, demonstrating what is possible for high school

girls taking foundational math courses (Morgenroth et al., 2015). Female tutors may not only show what is possible for girls, but also that pursuing math can be exciting (Riegle-Crumb & Morton, 2017) . Second, girls in math may face stereotype threat, in which the stereotypes about gender and math ability can negatively impact girls' performance, confidence, and interest in math-related fields (Spencer et al., 1999). Even if stereotype threat does not impact K-12 girls' math achievement (Flore & Wicherts, 2015), girls may not pursue advanced math courses if they do not see themselves as "math people." In particular, they may lose confidence or interest in math as they start high school, and, as a result, interventions that provide girls at the secondary level with positive math role models—such as a female tutor—can improve their math self-concept and performance (Good et al., 2003). Finally, girl students may perceive that female tutors have higher expectations for them than male tutors (Gershenson et al., 2016) and they may develop more positive relationships with female tutors than with male tutors (Roorda & Jak, 2024), which may impact their subsequent STEM interest and engagement. These mechanisms provide a compelling rationale for investigating whether female math tutors can help mitigate barriers to STEM engagement and achievement for girl students.

The Present Study

The present study examines how same-gender tutors influence students' STEM beliefs, attendance, and math performance. Specifically, we focus on the implementation of a high school tutoring initiative targeting ninth grade students taking Algebra 1. For many districts across the United States, students must pass Algebra 1 to graduate from high school (Dee & Huffaker, 2024; Reys et al., 2007). Moreover, students' performance in Algebra 1 affects the STEM courses they can take in the future (Silver, 1997). Therefore, interventions that support

girl students' Algebra 1 performance and encourage positive experiences may have outsized impacts on their ability and interest to pursue STEM careers (Dee & Huffaker, 2024).

To conduct this study, we partnered with a large New England school district and Saga Education (Saga), a tutoring provider, for the 2021-22 school year. Saga is one of the few secondary school tutoring providers with robust evidence of the program success (Bhatt et al., 2024; Guryan et al., 2023). The district contracted with Saga to deliver tutoring to ninth grade students during math lab, an in-school math intervention block for all ninth-grade students. The program was designed for students to receive tutoring in pairs, however the group sizes varied based on the number of students in an intervention block and the number of tutors. Students were assigned to tutoring groups ranging from one to three students.

Saga's tutoring program operated at five high-school sites; school principals could opt-in to having the program at their school. Although the initial plan was for tutoring to be offered in-person, during the school day, the tutoring provider had trouble hiring enough local tutors. As a result, tutoring was ultimately offered in-person at two schools and virtually at three schools.

We ask: Do female students who are matched with a same-gender math tutor have:

1) more positive STEM-related beliefs than those who do not have a same-gender tutor?

2) have better attendance?

3) improved academic outcomes?

To answer these questions, we randomly assigned student tutoring groups to be matched with a female or male tutor. Given the variation in the program across schools, we also explore whether the effect of same gender math tutor on STEM interest varied by instructional modality. This analysis sheds light on whether demographic similarities hold the same weight in virtual classrooms as they do in brick-and-mortar classrooms. Similarities may be less obvious when

people are not physically in the same place, though many of the hypothesized mechanisms for the impact of demographic matching (e.g., perceived shared experiences, role modeling) still may hold (Darling-Aduana, 2021).

Methods

We preregistered our study design and research questions on the Registry for Effectiveness in Education Studies (REES, see Appendix B). Implementation challenges led us to deviate slightly from our preregistration, which we outline in Appendix C.

Sample

Saga delivered math tutoring in five high schools during the 2021-22 school year. All ninth-grade students taking Algebra 1 were eligible to participate in tutoring during math lab, a math intervention block. To be eligible for the study (i.e., be randomized to a tutor), students had to be present in math lab and be assigned to a tutoring group by November 1, 2021. Ultimately, 606 out of 1,327 ninth grade students enrolled at the five sites received tutoring. Of those 606 students, 146 (30.1%) started tutoring after November 1st and thus were not part of the initial randomization or had a non-binary tutor (see Appendix C).

Our final analytic sample included 422 students and 23 tutors. In the first column of Table 1, we provide descriptive statistics for students (Panel A) and tutors (Panel B). Half of the students were girls and 39 percent of tutors were female. Forty-five percent of students identified their race as White and 33 percent identified as Black, while 76 percent identified their ethnicity as Hispanic/Latino. Among the tutors, approximately 34 percent identified as White, 43 percent identified as Black, and about 2 percent identified as Asian. The race and ethnicity variables for tutors were constructed from an item that allowed tutors to report multiple race/ethnicities and are therefore not mutually exclusive. Approximately, 89 percent of students were eligible for

Free-and-Reduced-Priced Lunch (FRPL), 18 percent have Individualized Education Plans (IEP), and 37 percent are classified as having Limited English Proficiency (LEP).

Data

      The school district provided data on math grades, school attendance, student characteristics, and achievement data from the prior school year (i.e., 2020-21). The district did not administer an end of year exam in 2021-22 due to the pandemic. We received beginning- and end-of-year Foundational Skills Assessment (FSA) scores (Saga's assessment), tutoring attendance, tutor characteristics, and an end-of-year student survey from the tutoring provider.

Student and Tutor Gender

      We used data on student and tutor gender in the randomization procedure. We observe the gender of every student to be either male or female.[i] One tutor reported their gender was non-binary, however this tutor was excluded from the analytic sample because no sessions were observed with the non-binary tutor after November 1. To estimate our primary effect of interest, we compare girl students with female tutors to girl students with male tutors. The data also allow us to compare boy students with male or female tutors to girl students with male tutors, as well as boy students with a male tutor to boy students with a female tutor.

Student Outcomes

      We observed five outcomes: STEM interest, tutoring session attendance, school absences, end-of-year tutoring administered FSA, and earning a math grade of C or above. The STEM interest outcome is the standardized average of three items on the end-of-year student survey about the usefulness of math, interest in pursuing STEM in college, and interest in pursuing a STEM career ($\alpha = .53$). Table 2 includes the items and mean response for the sample. Appendix Table A1 provides more details on the response categories. Because the Cronbach's

alpha for the three-item scale was relatively low, we also examined the effect on each item individually (see Appendix Table A6).

We assess tutoring attendance and school absences after the official start of tutoring (November 1st). Collecting attendance data was challenging for many districts during the 2021-22 school year, the second year of the pandemic. While we include the results from our analyses on school attendance as it was one of our preregistered outcomes, we acknowledge that the quality of the data is questionable.

We also observe two measures of academic achievement. The tutoring provider assessed math competency with their FSA that they administer several times over the course of the year. We use district grade data to construct a binary measure equal to one if their ninth-grade math GPA in quarters 2, 3, and 4 is a C or above (i.e., GPA>1.7).[ii] We use grades from quarters 2, 3, and 4 because students were selected for tutoring after the first quarter grades were issued. Ninth grade Algebra 1 is often a "gateway" course that has implications for students' future academic trajectories and requires a minimum grade to demonstrate college and career readiness. Earning a C- or better in their 9th grade math course is also practically important because that is an indicator for eligibility for progressing through the high school math curriculum.

Missing Outcome Data

We observe math grades and student attendance (i.e., school, tutoring) for all students.[iii] However, only half of the students in the analytic sample completed the final student survey that included the STEM interest items (204 students; 48.3 percent) or the end-of-year FSA (195 students; 46.2 percent). District administrators and Saga staff attributed low completion rates to high rates of chronic absenteeism during the pandemic, students not consistently attending their

math intervention blocks, or students exiting the program early. Table A2 in the Appendix describes the proportion of the 422 students missing outcome data overall and by condition.

We find no evidence that completion of the tutoring provider's end-of-year student survey and FSA differed across the treatment arms for this study. Moreover, we even find little evidence that completion of the tutoring provider's end-of-year student survey and FSA differed based on student background characteristics (see Tables A3 and A4 in the Appendix).[iv] The joint F-tests testing for systematic missingness based on observable characteristics are not statistically significant.

Student Characteristics

Our primary model includes controls for student characteristics, including prior academic achievement, that we observe in the school administrative data and the program data. Student characteristics include binary indicators for whether a student was White, eligible for FRPL, had an IEP, and was classified as LEP. We create the prior achievement factor using achievement data measured prior to random assignment: students' math GPA in the first quarter of the 2021-22 school year, a binary indicator for whether a student received a low score on their 8th grade summative math assessment (STAR; 2020-21), beginning-of-year FSA, and first diagnostic FSA (measured prior to November 1st).[v] We create this factor because we do not observe 8th grade summative assessments for about 25 percent of students in 2020-21. We use Q1 math GPA and beginning-of-year FSA to impute missing values.

Randomization

The district enrolled students in the tutoring program on a rolling basis starting at the beginning of the 2021-22 school year in September through the end of October. Once students were enrolled in the tutoring program and assigned to a tutoring group, we used data on gender

(e.g., students, tutors), school, and math lab assignments (i.e., class in which tutoring took place) to randomly assign students to tutors. Because the tutoring provider hired tutors to work at specific schools, the randomization was stratified within school (for tutors) and by math lab classroom (for students). We structured the randomization procedure to maximize the likelihood that girl students had a 50% chance of being assigned to a female tutor, in order to maximize our power for estimating the same-gender effects for girl students. We assigned students to their primary tutor by November 1st, 2021, by which date almost all students who would be tutored had begun to participate in tutoring sessions.[vi]

Students in each condition – same gender or different gender – had similar characteristics and achievement prior to treatment. Columns 2-5 in Table 1 show the student characteristics and student achievement measured prior to the beginning of the program by condition. Students across conditions had statistically similar math grades in the first quarter, 8th grade math assessment scores in Spring 2021, prior achievement factor, and school attendance in 2020-21. The evidence of balanced prior achievement and attendance supports the validity of the randomization procedure.

Compliance with random assignment was high for students and tutors. Table 3 describes the proportion of sessions that students (i.e., boys, girls) had with tutors of their randomly assigned gender. Girls that were assigned a female tutor had 78.15 percent of their sessions with a female tutor, compared with girls who were assigned male tutors, who had 22.22 percent of their sessions with a female tutor. Student compliance with assignment was similarly high across the conditions, indicating that any non-compliance is unlikely to bias our estimates. Based on administrative records we can infer several reasons for the non-compliance that did happen, including tutors being absent, being reassigned to different schools, or leaving their positions. To

adjust for the non-compliance that did occur, we run treatment-on-the-treated (TOT) analyses in which we instrument the proportion of sessions students attended with tutors who shared their randomly assigned tutor gender with students' original condition assignment.

Estimation Strategy

We estimate effects of same-gender matching in a linear regression framework following the equation:

$$Y_{ics} = \beta_1 GStu \times FTut_i + \beta_2 BStu \times MTut_i + \beta_3 BStu \times FTut_i + X'_i\gamma + A'_i\omega + \delta_c + u_{ics}$$

where Y equals the outcome of interest for student i in class c in school s. *GStu×Ftut* is an indicator equal to one if a student is a girl and was randomly assigned a female tutor and zero otherwise. *BStu×Mtut* and *BStu×Ftut* are equal to one if a boy was randomly assigned a male or female tutor, respectively. The reference category is girl students who were assigned to male tutors. The primary effect of interest is β_1, which describes the effect of a female tutor on a girl student's outcomes compared to a girl with a male tutor. *X'* is a vector of exogenous student characteristics including binary indicators for White, FRPL, IEP, and LEP. The student characteristics control for elements of student's identities that are correlated with systematic differences in STEM interest. *A'* is a vector of prior achievement covariates including math Q1 grade, an indicator for low 2021 STAR test, and the prior math achievement factor. The baseline student achievement covariates accounts for underlying academic proficiency and the other outcomes (e.g., STEM interest, changes in student outcomes). δ is a classroom fixed effect for each math lab classroom c, which accounts for time-invariant characteristics including student sorting, curriculum, and teacher efficacy. u is the error term robust to clustering by school. Our primary analysis takes an intent-to-treat (ITT) approach using students' original tutor gender assignment to eliminate the likelihood that selection bias impacts our estimates.

We preregistered our model to focus specifically on whether girl students benefit more from working with female tutors than with male tutors. This specification does not test whether students in general perform better with same-gender tutors, though by changing the reference category we can estimate outcomes for boy students working with male versus female tutors (see Appendix Table A5). Nor does it address the related but distinct question of whether girl student-female tutor matches produce larger benefits than boy student-male tutor matches – that would require an interaction term between student gender and tutor gender. Our modeling choices reflect our primary interest in how female tutors affect girls' math outcomes.

In addition, the ITT analysis, we conduct a TOT analysis in which we instrument the proportion of tutoring sessions spent working with a female tutor with random assignment to work with a female tutor. We estimate 2SLS models following:

$$\widehat{PropC}_{ics} = \tau_1 GStu \times FTut_i + \tau_2 BStu \times MTut_i + \tau_3 BStu \times FTut_i + X'_i\gamma + A'_i\omega + \delta_c + \epsilon_{ics}$$

In the first stage, the endogenous variable is the proportion of sessions that girls who were assigned to female tutors actually occurred with female tutors ($\widehat{PropC}$). The instrument is a binary indicator for whether a student was a girl assigned to a female tutor ($Gstu \times FTut$).

$$Y_{ics} = \beta_1 \widehat{PropC}_{ics} + \beta_2 BStu \times MTut_i + \beta_3 BStu \times FTut_i + X'_i\gamma + A'_i\omega + \delta_c + u_{ics}$$

The second stage estimate (β_1) adjusts for imperfect compliance to treatment or the extent to which girls who were assigned female tutors instead participated in sessions with male tutors.

Finally, to test differences in the effects of same-gender tutors between in-person and virtual tutoring, we estimate the impacts of tutor-student gender matches separately by tutoring modality. The tutoring program delivery differed by school site so that some sites had in-person tutoring while others had virtual tutoring.

Results

Table 4 displays the results from our ITT analysis, where the reference category is girl students assigned to a male tutor. We find that girls assigned to female tutors developed an interest in STEM that was 0.73 standard deviations (SD) higher than girls assigned to male tutors. The results are consistent across the scale and each individual item (see Appendix Table A2).[vii] Figure 1 depicts the raw mean of students' STEM interest on the original scale by condition. As illustrated visually, we do not find evidence of a *general* gender match effect on interest in STEM – boy students do not report increased STEM interest when matched with a male tutor versus a female tutor (see Appendix Table A5). We also see an increase in math achievement as a result of the female gender match, where 51.8 percent of girl students working with male tutors earned an average math grade of C- or better compared to 55.6 percent of girl students working with female tutors (i.e., a four percentage point increase, 0.08 SD).[viii] We do not find evidence that the girl student-female tutor gender match increased end-of-year FSA scores, school attendance, or tutoring session attendance. The attendance estimates are imprecise, potentially due to low-quality data.

Table 5 presents the results of the TOT analysis, which estimates the impact on student outcomes for students if they received all sessions with tutors of their randomly assigned gender. The first stage F-statistic (i.e., Kleibergen-Paap) are statistically significant and greater than 37 for each outcome and that the Cragg-Donald Wald F statistics all exceed 100 (Stock & Yogo, 2002) indicating, not surprisingly, that randomization sufficiently affected gender matching for the two-stage approach to effectively estimate the effects. The estimated effects for girl students working with female tutors is a 0.98 SD increase in interest in STEM and a 5 percentage point

increase (0.10 SD) in the likelihood of earning an average of a C- or better in math for the rest of the school year than those working with male tutors.

Instructional Modality: In-Person vs. Virtual

Table 6 presents the results separately by whether tutoring sessions were held in-person or virtually. Panel A restricts the sample to in-person programs and Panel B restricts the sample to virtual programs. The positive impact of female tutors on girls' STEM interest is largely driven by students in the schools with in-person tutoring (0.62 SD) compared to those in schools with virtual tutoring (0.31 SD, *ns*). The same pattern holds for the impact of the female gender match on student math course grades: girls who worked with female tutors in-person were 2.7 percentage points (0.05 SD) more likely to earn at least a C- in math than girls who worked with male tutors, while girls who worked with female tutors virtually were only 1.4 percentage points (0.03 SD) more likely to earn at least a C- in math than girls who worked with male tutors.[ix] Although the female gender match effects appear to be driven by students in schools offering in-person tutoring, these results may be a function of unobserved sorting into schools. Students were not randomly assigned to in-person or virtual tutoring; instead, instructional modality was determined at the school level, and any differences in effects may reflect other school-level factors.

Discussion

Through a randomized controlled trial, we tested the impact of tutor gender match on female students' STEM beliefs, achievement, and attendance outcomes. We find that girl students benefit from being assigned to a female tutor. Female tutors substantially increased girls' interest in STEM relative to male tutors (SD=0.73). Additionally, we find evidence that the female tutor-tutee gender match confers academic benefits. Girls assigned to work with female

tutors were 3.9 percentage points more likely to earn at least a C- in math compared to girls assigned to male tutors. We do not find statistically significant effects on the tutoring provider assessment or attendance outcomes, though attendance was likely poorly measured. Overall, our results show that high school girls can benefit from working with female math tutors.

This research contributes to the literature on the benefits of an additional female STEM role model in girls' lives. Much research has focused on STEM mentors in college (M. Banerjee et al., 2018) or gender match between students and teachers (Blake-Beard et al., 2011; González-Pérez et al., 2020; Prieto-Rodriguez et al., 2020), however, as far as we know, this is the first study to assess the impact of demographically similar tutors on student outcomes. Moreover, we focus on a critical time period: the outset of high school, when girls may fall off track to pursuing STEM careers (Sadler et al., 2012; M. T. Wang & Degol, 2017). Our survey outcomes also suggest one of the mechanisms as to why a female STEM role model might increase girls' persistence in STEM–an increased valuing and interest in the subject.

Female tutors may be more readily able to increase girls' interest in STEM and their subsequent math performance when they interact in person. One potential reason for this difference is that the salience of the tutors' gender may be stronger in person. Another explanation might be that tutors are better able to connect with students and communicate in person. When it comes to tutoring in virtual settings, some hypothesize that virtual environments can increase stress for participants by increasing "close" eye contact and constant viewing of a video reflection (Bailenson, 2021), perhaps making connection or relationship-building more difficult.

Limitations

Limitations of the data and context presented a few challenges for this project. First, we only observe the STEM interest outcome for students who took the end-of-year survey. Although we do not find consistent differences between students for whom we observe and do not observe the STEM interest outcome on observable pre-intervention covariates, unobservable differences between these groups of students is possible. Thus, our STEM interest findings apply to a group of students engaged enough to complete an end-of-year survey and we do not know if they would extend to all students who received tutoring.

Second, our capacity to examine effects on achievement beyond grades are also limited because 9th graders in the district do not take a standardized assessment. We find trending positive effects of the student-tutor gender match for girls on tutoring provided assessments and math grades, however a standardized assessment could provide more precise and valid results. Third, We are also not able to estimate the long-term effects of the female student-tutor gender match. A next step for research on this topic might focus on showing how the female student-tutor gender match impacts course-taking behavior and labor market decisions.

Fourth, our results suggest that the positive effects of female tutors on girl students are driven by students attending the schools where instruction occurred in person. Ultimately, however, we cannot disentangle whether features of specific schools (e.g., the context or quality) or the modality drive the effects. Future research can formally test this by randomly assigning students to receive tutoring in-person versus virtually.

Fifth, our data allows us to focus narrowly on students' school-based math-related experiences when considering interest in STEM. While mathematics instruction and support from teachers and tutors are undoubtedly important, students' interest in STEM may also be shaped by their experiences in science classrooms. We find that the impact on STEM interest

holds when accounting for students' science teachers (see Appendix Table A7), which we would expect given the randomization. However future studies can broaden the scope to include the contributions of science instruction.

Finally, we acknowledge the limitations of quantitative research for operationalizing gender. The limitations of our data necessitate the description of gender as a binary. Although tutors had the opportunity to report a gender other than male or female, student administrative data is limited to male or female. As a result, our conceptualization of how gender perceptions impact STEM beliefs and achievement are limited, and we hope future research has the capacity to explore how tutoring influences the STEM beliefs of students who are non-binary or gender non-conforming, as well as those who belong to the LGBTQIA+ communities.

Implications for Policy

Educators and researchers have long recognized the benefits of role models for student development, particularly among student subgroups that experience persistent opportunity gaps (e.g., Black students; females in STEM). Yet, teacher labor markets are difficult to change. Teachers often stay in their jobs for many years – which is a benefit for students – and, as a result, only a small fraction of the workforce changes each year. Moreover, teachers need not only college degrees but substantial additional preparation to obtain the certifications needed for teaching.

The tutor workforce is far more flexible. By finding that gender-matched tutors can have positive effects on female students' interests and achievement, this study opens a faster and easy approach for obtaining the benefits of exposure to demographically similar educators. Education policies and practices can leverage a broader group of educators as role models – and as effective instructors – for students.

At the same time, we do not believe simply matching students and educators - including tutors - based on demographics is a viable policy solution. Although we observe that boy student outcomes were similar no matter their tutors' gender, there are potential negative externalities of students *only* being exposed to demographically similar educators. Instead, we propose that tutoring providers consider the communities they work in and strive to recruit a diverse set of tutors that reflect the backgrounds of the students they serve.

Our findings are consistent with social cognitive theory (Bandura, 1977), role model theory (Morgenroth et al., 2015), and stereotype threat theory (Spencer et al., 1999), which together suggest that exposure to female educators can strengthen self-efficacy, counter negative stereotypes, and enhance STEM motivation and achievement for girls. At the same time, future research should explore how to ensure that girls feel empowered to persist in STEM regardless of the gender of their tutor, so that supportive contexts and instructional practices—not only demographic match—drive equitable outcomes.

Putting our findings in the context of the gender imbalance in STEM fields, in particular, policies and practices could respond to the need by ensuring enough additional female tutors who are engaging in math or other STEM subjects. And if this type of high-impact tutoring is embedded into the school day in high school, girls would be exposed to the mentors they need to develop and sustain interest and achievement in STEM courses.

Our results are consistent with the idea that tutoring has beneficial effects beyond just achievement, and that tutors can serve as an additional caring educator in a students' learning trajectory. However, the scope of tutoring programs can vary greatly. This program prioritized students meeting frequently with a consistent tutor, which likely enhanced relationship development. For tutoring programs to be effective, they need to be strong programs.

No single solution will erase the gender gaps in STEM perpetuated by societal structures. Doing so will necessitate many strategies at different levels – we find that exposing girl students at the outset of their high school career to female math tutors may be one promising practice for mitigating gender inequality in our education system.

End Notes

[i] The school district only collects information on gender in two categories: male and female. As a result, our analyses are limited to this binary classification.

[ii] We set this threshold because 1.7 is approximately the median GPA (in addition to the practical implications of earning a C- or higher versus a D or lower)

[iii] There are three students who records suggest transfer out of the schools with the tutoring program. Outcomes for these students were logically imputed to have low attendance and a GPA below 1.7. The STEM interest results are robust to the exclusion of these students.

[iv] We find none of the pretreatment covariates are statistically significantly related to missing STEM interest. The pattern is similar for the FSA outcomes, with one exception (i.e., baseline achievement). Out of the 26 mean differences in Appendix Tables A3 and A4 only one is statistically significant at the $p < .05$ level, which is about what we would expect to find due to random chance.

[v] We define a low score on their 8th grade summative math assessment as 500 or lower (Range=0 to 1400). We impute prior achievement in a Full-Information Maximum Likelihood framework.

[vi] Four students in our analytic sample, or just under one percent of the sample, had their first session from November 2nd to November 8th.

[vii] The STEM interest results are robust to alternative definitions of the STEM interest outcome. Appendix Table A2 includes the main Intent to Treatment specification from Table 4, Panel B, Column 1 where the outcome is interest in math, wants to pursue STEM in college, wants to pursue a STEM career, and the average of wants to pursue STEM in and career.

[viii] High pre-treatment levels of STEM interest for boys could censor tutoring effects. However, in auxiliary results we find that boy's STEM interest is observed to be just above the middle value (approximately 3.1) on the 5 point Likert scale items. On average, boys reported they were "somewhat" rather than "extremely" interested in STEM.

[ix] In auxiliary analyses we find that TOT estimates have similar sign and significance to the ITT results in Table 6.

References

Bailenson, J. N. (2021). Nonverbal Overload: A Theoretical Argument for the Causes of Zoom Fatigue. *Technology, Mind, and Behavior*, *2*(1). https://doi.org/10.1037/tmb0000030

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191.

Banerjee, M., Schenke, K., Lam, A., & Eccles, J. S. (2018). The roles of teachers, classroom experiences, and finding balance: A qualitative perspective on the experiences and expectations of females within STEM and non-STEM careers. *International Journal of Gender, Science and Technology*, *10*(2), 287–307.

Banerjee, P. A. (2016). A systematic review of factors linked to poor academic performance of disadvantaged students in science and maths in schools. *Cogent Education*, *3*(1), 1178441. https://doi.org/10.1080/2331186X.2016.1178441

Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Social Psychology of Education*, *15*(4), 427–448. https://doi.org/10.1007/s11218-012-9185-3

Bhatt, M. P., Guryan, J., Khan, S. A., LaForest-Tucker, M., & Mishra, B. (2024). *Can Technology Facilitate Scale? Evidence from a Randomized Evaluation of High Dosage Tutoring*. National Bureau of Economic Research. https://www.nber.org/papers/w32510

Biden Administration. (2024). *FACT SHEET: Biden-Harris Administration Announces Improving Student Achievement Agenda in 2024*. https://www.whitehouse.gov/briefing-room/statements-releases/2024/01/17/fact-sheet-biden-harris-administration-announces-improving-student-achievement-agenda-in-2024/

Blake-Beard, S., Bayne, M. L., Crosby, F. J., & Muller, C. B. (2011). Matching by Race and

 Gender in Mentoring Relationships: Keeping our Eyes on the Prize. *Journal of Social*

 Issues, *67*(3), 622–643. https://doi.org/10.1111/j.1540-4560.2011.01717.x

Blazar, D. (2024). Why Black Teachers Matter. *Educational Researcher*, 0013189X241261336.

 https://doi.org/10.3102/0013189X241261336

Cabral-Gouveia, C., Menezes, I., & Neves, T. (2023). Educational strategies to reduce the

 achievement gap: A systematic review. *Frontiers in Education*, *8*.

 https://doi.org/10.3389/feduc.2023.1155741

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science:

 Sociocultural and biological considerations. *Psychological Bulletin*, *135*(2), 218–261.

 https://doi.org/10.1037/a0014412

Charlesworth, T. E., & Banaji, M. R. (2019). Gender in science, technology, engineering, and

 mathematics: Issues, causes, solutions. *Journal of Neuroscience*, *39*(37), 7228–7243.

Chavatzia, T. (2017). Cracking the code: Girls' and women's education in science, technology,

 engineering and mathematics (STEM). *UNESCO*.

 https://www.voced.edu.au/content/ngv:77679

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields

 more gender balanced than others? *Psychological Bulletin*, *143*(1), 1–35.

 https://doi.org/10.1037/bul0000052

Cho, I. (2012). The effect of teacher–student gender matching: Evidence from OECD countries.

 Economics of Education Review, *31*(3), 54–67.

Cleveland, C., & Scherer, E. (2024). The Effects of Teacher-Student Demographic Matching on

 Social-Emotional Learning. *EdWorkingPaper*, *21–399*.

Coleman, J. S. (1968). Equality of Educational Opportunity. *Equity & Excellence in Education*, *6*(5), 19–28. https://doi.org/10.1080/0020486680060504

Coley, R. J. (2001). *Differences in the Gender Gap: Comparisons across Racial/Ethnic Groups in Education and Work. Policy Information Report.*

Cortes, K. E., Kortecamp, K., Loeb, S., & Robinson, C. D. (2025). A scalable approach to high-impact tutoring for young readers. *Learning and Instruction*, *95*, 102021.

Darling-Aduana, J. (2021). A Remote Instructor Like Me: Student–Teacher Congruence in Online, High School Courses. *AERA Open*, *7*, 23328584211018719. https://doi.org/10.1177/23328584211018719

Dasgupta, N., & Stout, J. G. (2014). Girls and Women in Science, Technology, Engineering, and Mathematics: STEMing the Tide and Broadening Participation in STEM Careers. *Policy Insights from the Behavioral and Brain Sciences*, *1*(1), 21–29. https://doi.org/10.1177/2372732214549471

Dee, T. S. (2004). Teachers, Race, and Student Achievement in a Randomized Experiment. *The Review of Economics and Statistics*, *86*(1), 195–210. https://doi.org/10.1162/003465304323023750

Dee, T. S. (2005). A Teacher like Me: Does Race, Ethnicity, or Gender Matter? *The American Economic Review*, *95*(2), 158–165.

Dee, T. S. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources*, *XLII*(3), 528. https://doi.org/10.3368/jhr.XLII.3.528

Dee, T. S., & Huffaker, E. (2024). *Accelerating Opportunity: The Effects of Instructionally Supported Detracking*. https://doi.org/10.26300/V492-1P91

Diekman, A. B., Brown, E. R., Johnston, A. M., & Clark, E. K. (2010). Seeking Congruity Between Goals and Roles: A New Look at Why Women Opt Out of Science, Technology, Engineering, and Mathematics Careers. *Psychological Science*, *21*(8), 1051–1057. https://doi.org/10.1177/0956797610377342

Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A.-M. (2017). Academic Interventions for Elementary and Middle School Students With Low Socioeconomic Status: A Systematic Review and Meta-Analysis. *Review of Educational Research*, *87*(2), 243–282. https://doi.org/10.3102/0034654316687036

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

Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do Teachers' Race, Gender, and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988. *ILR Review*, *48*(3), 547–561. https://doi.org/10.1177/001979399504800312

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, *53*(1), 25–44. https://doi.org/10.1016/j.jsp.2014.10.002

Fox-Turnbull, W. H., Docherty, P. D., Zaka, P., & Impey, T. (2023). Initial teacher education (ITE) students' perceptions of typical engineers: Assessing potential for bias in the formative career decision years. *International Journal of Technology and Design Education*, *33*(2), 439–456. https://doi.org/10.1007/s10798-022-09735-4

Francis, B., Skelton, C., Carrington, B., Hutchings, M., Read, B., & Hall, I. (2008). A perfect match? Pupils' and teachers' views of the impact of matching educators and learners by

gender. *Research Papers in Education*, *23*(1), 21–36.

https://doi.org/10.1080/02671520701692510

Frome, P. M., Alfeld, C. J., Eccles, J. S., & Barber, B. L. (2006). Why don't they want a male-dominated job? An investigation of young women who changed their occupational aspirations. *Educational Research and Evaluation*, *12*(4), 359–372.

https://doi.org/10.1080/13803610600765786

Gershenson, S., Hansen, M., & Lindsay, C. A. (2021). *Teacher Diversity and Student Success: Why Racial Representation Matters in the Classroom*. Harvard Education Press.

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. https://doi.org/10.1016/j.econedurev.2016.03.002

González-Pérez, S., Mateos de Cabo, R., & Sáinz, M. (2020). Girls in STEM: Is it a female role-model thing? *Frontiers in Psychology*, *11*, 564148.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, *24*(6), 645–662.

https://doi.org/10.1016/j.appdev.2003.09.002

Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M. V., Dodge, K., Farkas, G., Fryer Jr., R. G., Mayer, S., Pollack, H., Steinberg, L., & Stoddard, G. (2023). Not Too Late: Improving Academic Outcomes among Adolescents. *American Economic Review*, *113*(3), 738–765. https://doi.org/10.1257/aer.20210434

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The Science of Sex Differences in Science and Mathematics. *Psychological*

Science in the Public Interest, *8*(1), 1–51. https://doi.org/10.1111/j.1529-1006.2007.00032.x

Hart, C. M. D. (2020). An Honors Teacher Like Me: Effects of Access to Same-Race Teachers on Black Students' Advanced-Track Enrollment and Performance. *Educational Evaluation and Policy Analysis*, *42*(2), 163–187. https://doi.org/10.3102/0162373719898470

Havik, T., & Westergård, E. (2020). Do Teachers Matter? Students' Perceptions of Classroom Interactions and Student Engagement. *Scandinavian Journal of Educational Research*, *64*(4), 488–507. https://doi.org/10.1080/00313831.2019.1577754

Herrmann, S. D., Adelman, R. M., Bodford, J. E., Graudejus, O., Okun, M. A., & Kwan, V. S. Y. (2016). The Effects of a Female Role Model on Academic Performance and Persistence of Women in STEM Courses. *Basic and Applied Social Psychology*, *38*(5), 258–268. https://doi.org/10.1080/01973533.2016.1209757

Heybach, J., & Pickup, A. (2017). Whose STEM? Disrupting the Gender Crisis Within STEM. *Educational Studies*, *53*(6), 614–627. https://doi.org/10.1080/00131946.2017.1369085

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

Hughes, J. N., Luo, W., Kwok, O.-M., & Loyd, L. K. (2008). Teacher-student support, effortful engagement, and achievement: A 3-year longitudinal study. *Journal of Educational Psychology*, *100*(1), 1–14. https://doi.org/10.1037/0022-0663.100.1.1

Hwang, N., & Fitzpatrick, B. (2021). Student–Teacher Gender Matching and Academic
Achievement. *AERA Open*, *7*, 23328584211040058.
https://doi.org/10.1177/23328584211040058

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender
Similarities Characterize Math Performance. *Science*, *321*(5888), 494–495.
https://doi.org/10.1126/science.1160364

Keller, L., Preckel, F., Eccles, J. S., & Brunner, M. (2022). Top-performing math students in 82
countries: An integrative data analysis of gender differences in achievement, achievement
profiles, and achievement motivation. *Journal of Educational Psychology*, *114*(5), 966.

Kyte, S., & Riegle-Crumb, C. (2017). Perceptions of the social relevance of science: Exploring
the implications for gendered patterns in expectations of majoring in STEM fields. *Social
Sciences*, *6*(1), 19.

Lim, J., & Meer, J. (2017). The Impact of Teacher–Student Gender Matches. *Journal of Human
Resources*, *52*(4), 979. https://doi.org/10.3368/jhr.52.4.1215-7585R1

Lim, J., & Meer, J. (2020). Persistent Effects of Teacher–Student Gender Matches. *Journal of
Human Resources*, *55*(3), 809–835. https://doi.org/10.3368/jhr.55.3.0218-9314R4

Luttenberger, S., Steinlechner, P., Ertl, B., & Paechter, M. (2019). It takes more than one
swallow to make a summer: Measures to foster girls' and women's pathways into STEM.
Frontiers in Psychology, *10*, 1844.

Lv, B., Wang, J., Zheng, Y., Peng, X., & Ping, X. (2022). Gender differences in high school
students' STEM career expectations: An analysis based on multi-group structural
equation model. *Journal of Research in Science Teaching*, *59*(10), 1739–1764.
https://doi.org/10.1002/tea.21772

Makarova, E., Aeschlimann, B., & Herzog, W. (2019). The gender gap in STEM fields: The impact of the gender stereotype of math and science on secondary students' career aspirations. *Frontiers in Education*, *4*, 60. https://www.frontiersin.org/articles/10.3389/feduc.2019.00060/full

Morgenroth, T., Ryan, M. K., & Peters, K. (2015). The Motivational Theory of Role Modeling: How Role Models Influence Role Aspirants' Goals. *Review of General Psychology*, *19*(4), 465–483. https://doi.org/10.1037/gpr0000059

Msambwa, M. M., Daniel, K., Lianyu, C., & Antony, F. (2024). A Systematic Review Using Feminist Perspectives on the Factors Affecting Girls' Participation in STEM Subjects. *Science & Education*. https://doi.org/10.1007/s11191-024-00524-0

Muntoni, F., & Retelsdorf, J. (2018). Gender-specific teacher expectations in reading—The role of teachers' gender stereotypes. *Contemporary Educational Psychology*, *54*, 212–220. https://doi.org/10.1016/j.cedpsych.2018.06.012

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

Ouazad, A. (2014). Assessed by a Teacher Like Me: Race and Teacher Assessments. *Education Finance and Policy*, *9*(3), 334–372. https://doi.org/10.1162/EDFP_a_00136

Paredes, V. (2014). A teacher like me or a student like me? Role model versus teacher bias effect. *Economics of Education Review*, *39*, 38–49. https://doi.org/10.1016/j.econedurev.2013.12.001

Piloto, C. (2023, March 13). *The Gender Gap in STEM: Still Gaping in 2023*.

      https://professionalprograms.mit.edu/blog/leadership/the-gender-gap-in-stem/

Prieto-Rodriguez, E., Sincock, K., & Blackmore, K. (2020). STEM initiatives matter: Results

      from a systematic review of secondary school interventions for girls. *International*

      Journal of Science Education, *42*(7), 1144–1161.

      https://doi.org/10.1080/09500693.2020.1749909

Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Ditlmann, R., & Crosby, J. R. (2008). Social

      identity contingencies: How diversity cues signal threat or safety for African Americans

      in mainstream institutions. *Journal of Personality and Social Psychology*, *94*(4), 615–

      630. https://doi.org/10.1037/0022-3514.94.4.615

Redding, C. (2022). Is Teacher–Student and Student–Principal Racial/Ethnic Matching Related

      to Elementary School Grade Retention? *AERA Open*, *8*, 23328584211067534.

      https://doi.org/10.1177/23328584211067534

Reys, B. J., Dingman, S., Nevels, N., & Teuscher, D. (2007). High School Mathematics: State-

      Level Curriculum Standards and Graduation Requirements. *Center for the Study of*

      Mathematics Curriculum. https://eric.ed.gov/?id=ED535222

Riegle-Crumb, C., & Morton, K. (2017). Gendered Expectations: Examining How Peers Shape

      Female Students' Intent to Pursue STEM Fields. *Frontiers in Psychology*, *8*.

      https://doi.org/10.3389/fpsyg.2017.00329

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.

Roorda, D. L., & Jak, S. (2024). Gender match in secondary education: The role of student
    gender and teacher gender in student-teacher relationships. *Journal of School Psychology*,
    107, 101363. https://doi.org/10.1016/j.jsp.2024.101363

Sadler, P. M., Sonnert, G., Hazari, Z., & Tai, R. (2012). Stability and volatility of STEM career
    interest in high school: A gender study. *Science Education*, *96*(3), 411–427.
    https://doi.org/10.1002/sce.21007

Silver, E. A. (1997). On My Mind: "Algebra for All"- Increasing Students' Access to Algebraic
    Ideas, Not Just Algebra Courses. *Mathematics Teaching in the Middle School*, *2*(4), 204–
    207. https://doi.org/10.5951/MTMS.2.4.0204

Sokal, L., Katz, H., Chaszewski, L., & Wojcik, C. (2007). Good-bye, Mr. Chips: Male teacher
    shortages and boys' reading achievement. *Sex Roles*, *56*, 651–659.

Speer, J. D. (2023). Bye bye Ms. American Sci: Women and the leaky STEM pipeline.
    Economics of Education Review, *93*, 102371.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype Threat and Women's Math
    Performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28.
    https://doi.org/10.1006/jesp.1998.1373

Steele, C. M. (1997). How Stereotypes Shape Intellectual Identity and Performance. *American
    Psychologist*, *52*(6), 613–629.

Stock, J. H., & Yogo, M. (2002). *Testing for weak instruments in linear IV regression*. National
    Bureau of Economic Research Cambridge, Mass., USA.
    https://www.nber.org/papers/t0284

Veldman, J., Van Laar, C., Thoman, D. B., & Van Soom, C. (2021). "Where will I belong
    more?": The role of belonging comparisons between STEM fields in high school girls'

STEM interest. *Social Psychology of Education*, *24*(5), 1363–1387.

https://doi.org/10.1007/s11218-021-09663-6

Wang, M. T., & Degol, J. L. (2017). Gender Gap in Science, Technology, Engineering, and

Mathematics (STEM): Current Knowledge, Implications for Practice, Policy, and Future

Directions. *Educational Psychology Review*, *29*(1), 119–140.

https://doi.org/10.1007/s10648-015-9355-x

Wang, M. T., Eccles, J. S., & Kenny, S. (2013). Not Lack of Ability but More Choice: Individual

and Gender Differences in Choice of Careers in Science, Technology, Engineering, and

Mathematics. *Psychological Science*, *24*(5), 770–775.

https://doi.org/10.1177/0956797612458937

Wang, N., Tan, A.-L., Zhou, X., Liu, K., Zeng, F., & Xiang, J. (2023). Gender differences in

high school students' interest in STEM careers: A multi-group comparison based on

structural equation model. *International Journal of STEM Education*, *10*(1), 59.

https://doi.org/10.1186/s40594-023-00443-6

Wedenoja, L., Papay, J., & Kraft, M. A. (2022). Second time's the charm? How sustained

relationships from repeat student-teacher matches build academic and behavioral skills.

Education Working Paper, 22–590.

Wright, A., Gottfried, M. A., & Le, V.-N. (2017). A Kindergarten Teacher Like Me: The Role of

Student-Teacher Race in Social-Emotional Development. *American Educational

Research Journal*, *54*(1_suppl), 78S-101S. https://doi.org/10.3102/0002831216635733

Figure 1. STEM Interest by Condition
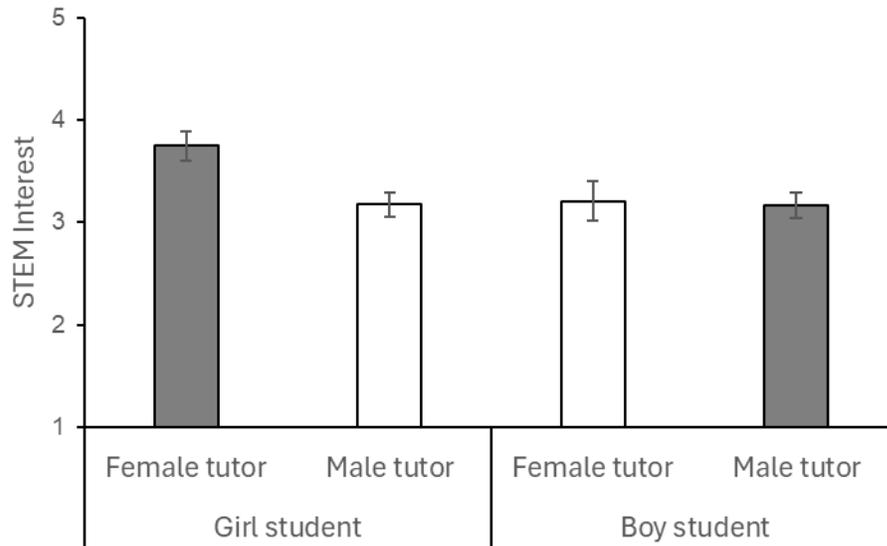
Note: The bars depict the mean of the raw STEM interest variable by condition. The error bars represent the standard error of the mean.

Table 1. Student and Tutor Characteristics

Characteristics	Full Sample Mean	Girl Student Female Tutor	Girl Student Male Tutor	Boy Student Female Tutor	Boy Student Male Tutor
Panel A. Student Characteristics for Full Sample and by Condition					
Student Gender	0.505	1	1	0	0
White	0.450	0.440	0.442	0.442	0.470
Black	0.329	0.333	0.364	0.299	0.311
Hispanic	0.758	0.750	0.791	0.753	0.735
FRPL	0.893	0.869	0.899	0.896	0.902
IEP	0.118	0.095	0.093	0.169	0.129
LEP	0.374	0.417	0.380	0.403	0.326
Algebra 1 GPA Q1	2.136	1.921	2.252	2.020	2.226
	(1.429)	(1.446)	(1.408)	(1.392)	(1.454)
STAR Math Spring 2021	0.050	0.048	0.039	0.065	0.053
	(0.218)	(0.214)	(0.194)	(0.248)	(0.225)
Achievement 2021	0.046	0.016	0.128	-0.062	0.050
	(0.583)	(0.59)	(0.59)	(0.535)	(0.592)
School Attendance 2021	30.261	29.179	30.891	29.130	30.992
	(6.004)	(7.032)	(5.731)	(5.561)	(5.66)
N	422	84	129	77	132
Panel B. Tutor Characteristics for Full Sample					
	Full Sample Mean	Girl Student Female Tutor	Girl Student Male Tutor	Boy Student Female Tutor	Boy Student Male Tutor
Tutor Female	0.382	1	0	1	0
Tutor White	0.341	0.262	0.434	0.221	0.371
Tutor Black	0.434	0.452	0.372	0.494	0.447
Tutor Asian	0.175	0.238	0.132	0.260	0.129
N	23	9	14	9	14

Note: Unit of analysis in Panel A is student and Panel B is tutor. In auxiliary regression joint F-tests show that variation in the covariates is not explained by the independent variables (see first row of Table 1).

Table 2. STEM Interest Items

Item	M	SD
How useful do you think math will be to you in the future?	3.108	1.518
How interested are you in pursuing a STEM major in college?	3.338	1.324
How interested are you in pursuing a STEM-related career?	3.368	1.360
Average of STEM Interest items	3.271	1.008

Note: The response options for each item varied from 1-5. See Appendix Table A1 for details. M=Mean; SD=Standard Deviation.

Table 3. Compliance with Treatment

Independent Variable Assignment	Compliance
Girl Student Female Tutor	0.7815
	(0.2971)
Male Student Male Tutor	0.7566
	(0.3234)
Male Student Female Tutor	0.7909
	(0.2995)
Girl Student Male Tutor	0.7778
	(0.3282)
Average	0.7767
	(0.3391)

Note: Each row is the proportion of tutoring sessions that a girl or boy student had with their randomly assigned tutor gender. Standard deviations in parentheses.

Table 4. Intent to Treat Estimates

	(1)	(2)	(3)	(4)	(5)
	STEM Interest	FSA	Math C or above	Tut Sessions	Sch Absences
Panel A. Classroom Fixed Effects					
Girl student Female tutor	0.7624**	-0.59036	0.0437	-1.8942	3.2062
	(0.0895)	(0.35436)	(0.0395)	(1.9118)	(4.7592)
Boy student Male tutor	0.0939	-0.51929	-0.0965	0.8259	1.8078
	(0.1805)	(0.58465)	(0.0887)	(1.8014)	(1.6647)
Boy student Female tutor	0.0608	0.21324	-0.0014	-1.4686	7.2728
	(0.1723)	(0.57303)	(0.1084)	(4.6678)	(4.2405)
Panel B. Classroom Fixed Effects and Controls					
Girl student Female tutor	0.7329**	0.25949	0.0388*	-1.5349	2.4930
	(0.1285)	(0.33300)	(0.0089)	(2.2299)	(4.1917)
Boy student Male tutor	0.0565	-0.13215	-0.0942+	1.3567	0.4427
	(0.1963)	(0.25361)	(0.0356)	(1.5658)	(1.5798)
Boy student Female tutor	0.0303	0.13714	-0.0314	-1.4969	6.7334
	(0.1955)	(0.47089)	(0.0885)	(4.9667)	(4.8112)
Girl student Male tutor (Mean)	-0.16	71.81	0.56	27.58	27.64
N	204	195	422	422	422
Student Characteristics	X	X	X	X	X
Classroom Fixed Effect	X	X	X	X	X
Prior Achievement	X	X	X	X	X

Note: The reference category for each estimate is Girl student Male tutor. Sample size and average outcomes girls with male tutor are the same for Panels A and B. Independent and control variables are observed for all cases. Student characteristics include binary indicators for FRPL, IEP, LEP, and White. Prior achievement includes math Q1 grade, indicator for low 2021 STAR test, and prior math achievement factor. Standard errors are clustered by school. $+<0.10$, $*p < 0.05$, $** p <0.01$, $***p <0.001$.

Table 5. Instrumental Variables: Estimates Adjusted for Compliance with Treatment

	(1)	(2)	(3)	(4)	(5)
	STEM Interest	FSA	Math C or above	Tut Sessions	Sch Absences
Girl student Female tutor	0.9788*	3.3011	0.0500*	-2.2832	3.2133
	(0.3021)	(4.4610)	(0.0118)	(2.9557)	(5.3199)
Girl student Male tutor (Mean)	-0.16	71.81	0.56	27.58	27.64
N	202	193	422	422	422
Kleibergen-Paap	37.7	43.0	48.7	48.7	48.7
Cragg-Donald	531.3	671.1	1570.9	1570.9	1570.9
Student Characteristics	X	X	X	X	X
Classroom Fixed Effect	X	X	X	X	X
Prior Achievement	X	X	X	X	X

Note: The endogenous variable is the proportion of sessions that girls who were assigned to female tutors had with female tutors and the instrument is a binary indicator for whether a student was a girl assigned to a female tutor. Independent and control variables are observed for all cases. Standard errors are clustered by school. $+ < 0.10$, $*p < 0.05$, $** p < 0.01$, $***p < 0.001$.

Table 6. The Effect by Instructional Modality: In-Person vs. Virtual

	(1)	(2)	(3)	(4)
	Panel A. In Person			
	STEM Interest	FSA	Math C or above	Tut Sessions
Girl student Female tutor	0.6156*	0.12931	0.0273+	-1.8494
	(0.0251)	(0.18065)	(0.0042)	(5.7609)
Boy student Male tutor	0.0022	-0.32882	-0.0544	3.5885
	(0.2510)	(0.30600)	(0.0318)	(4.1110)
Boy student Female tutor	-0.0460*	-0.28860	0.0622	1.1287
	(0.0012)	(0.09764)	(0.0224)	(9.0903)
Girl student Male tutor (Mean)	-0.08090	63.21	0.3089	41.46
N	121	116	249	248
	Panel B. Virtual			
Girl student Female tutor	0.3135	0.04603	0.0141	2.5295
	(0.3345)	(0.18062)	(0.0794)	(3.5193)
Boy student Male tutor	-0.1468	0.01741	-0.1806	-1.4359
	(0.2008)	(0.09319)	(0.0806)	(1.7636)
Boy student Female tutor	-0.0186	-0.62294	-0.2128	1.0327
	(0.2229)	(0.41785)	(0.1386)	(2.3925)
Girl student Male tutor (Mean)	-0.2273	79.58	0.8361	30.08
N	83	79	173	169

Note: All models include classroom fixed effects and controls. Virtual=Virtual tutoring; In Person= In-person tutoring. All models include student characteristics, prior achievement, and classroom fixed effects. Standard errors are clustered by school. +<0.10, *p < 0.05, ** p <0.01, ***p <0.001.

Appendix Table A1. STEM Interest Questionnaire Items

Description	Item
Math usefulness	How useful do you think math will be for your future?
STEM College interest	How interested are you in pursuing a STEM major in college?
STEM Career interest	How interested are you in pursuing a STEM-related career?

Likert	Math usefulness	STEM College interest	STEM Career interest
1	Not at all useful	Not at all interest	Not at all interest
2	Slightly useful	Slightly interested	Slightly interested
3	Somewhat useful	Somewhat interested	Somewhat interested
4	Quite useful	Quite interested	Quite interested
5	Extremely useful	Extremely interest	Extremely interest

Appendix Table A2. Missing Outcomes by Student and Tutor Gender

Outcome	All	Girl Student Female Tutor	Girl Student Male Tutor	Boy Student Male Tutor	Boy Student Female Tutor
STEM Interest	0.517	0.595	0.519	0.595	0.462
Math C or above Q2-Q4	0	0	0	0	0
SA	0.538	0.643	0.543	0.643	0.454
School Absences	0	0	0	0	0
Tutoring Attendance	0	0	0	0	0

Note: Each cell describes the proportion of students who were missing an outcome.

Appendix Table A3. Mean Characteristics Conditional Observing STEM Interest

Variable	(1) STEM Interest Observed	(2) STEM Interest Missing	(3) p-Value
Student Female	0.537	0.471	0.413
Tutor Female	0.413	0.348	0.235
Algebra 1 GPA Q1	0.385	0.328	0.342
STAR Math Spring 2021	0.041	0.059	0.508
Achievement 2021	-0.026	0.124	0.215
Baseline Absences 2021	5.120	3.253	0.070
FRPL	0.908	0.877	0.388
IEP	0.133	0.103	0.080
LEP	0.358	0.392	0.251
White	0.427	0.475	0.332
Black	0.372	0.284	0.346
Hispanic	0.766	0.750	0.914
Instructional Modality	0.587	0.593	0.320

Note: Columns 1 and 2 describe the mean of the variable described in the row conditional on the STEM interest being observed (Column 1) and STEM Interest not being observed (Column 2). Column 3 includes the p-value from a model where we regress an indicator for missing STEM interest on the pre-treatment characteristic described in the row with classroom fixed effects and standard errors clustered by school. The joint F-test for missing STEM interest (F-test = 1.77, p = 0.297) is not statistically significant. The mean difference for Algebra 1 GPA Q1, STAR Math Spring 2021, prior math achievement factor, and prior school attendance are standardized. SD=Standard Deviation

Appendix Table A4. Mean Characteristics Conditional Observing End-of-Year FSA

Variable	(1) EOY FSA Observed	(2) EOY FSA Missing	(3) p-Value
Student Female	0.456	0.545	0.185
Tutor Female	0.328	0.429	0.106
Algebra 1 GPA Q1	0.328	0.375	0.445
STAR Math Spring 2021	0.051	0.049	0.223
Achievement 2021	0.152	-0.040	0.021
Baseline Absences 2021	3.122	5.081	0.055
FRPL	0.872	0.911	0.061
IEP	0.108	0.125	0.255
LEP	0.374	0.371	0.519
White	0.462	0.446	0.997
Black	0.287	0.366	0.413
Hispanic	0.744	0.768	0.645
Instructional Modality	0.595	0.580	0.339

Note: Columns 1 and 2 describe the mean of the variable described in the row conditional on the EOY FSA being observed (Column 1) and EOY FSA not being observed (Column 2). Column 3 includes the p-value from a model where we regress an indicator for missing EOY FSA on the pre-treatment characteristic described in the row with classroom fixed effects and standard errors clustered by school. The joint F-test for missing end-of-year FSA (F-test = 3.00, $p = 0.156$) is not statistically significant. The mean difference for Algebra 1 GPA Q1, STAR Math Spring 2021, prior math achievement factor, and prior school attendance are standardized. SD=Standard Deviation. EOY=End of Year.

Appendix Table A5. Intent to Treat Estimates with Alternative Reference Category: Boy student Female tutor

	(1)	(2)	(3)	(4)	(5)
	STEM Interest	FSA	Math C or above	Tut Sessions	Sch Absences
Boy student Male tutor	0.0261	-2.6929	-6.2906	2.6581	-6.2906
	(0.1649)	(3.0186)	(5.2013)	(3.6463)	(5.2013)
Girl student Male tutor	-0.0303	-1.3714	-6.7334	1.4338	-6.7334
	(0.1955)	(4.7089)	(4.8112)	(5.2879)	(4.8112)
Girl student Female tutor	0.7025*	1.2235	-4.2403	-0.3375	-4.2403
	(0.1607)	(1.4929)	(8.6561)	(3.5561)	(8.6561)
Boy student Female tutor (Mean)	-0.13	61.06	0.45	41.10	40.54
N	204	195	422	422	422
Student Characteristics	X	X	X	X	X
Classroom Fixed Effect	X	X	X	X	X
Prior Achievement	X	X	X	X	X

Note: Independent and control variables are observed for all cases. Student characteristics include binary indicators for FRPL, IEP, LEP, and White. Prior achievement includes math Q1 grade, indicator for low 2021 STAR test, and prior math achievement factor. Standard errors are clustered by school.$+<0.10$, $*p < 0.05$, $** p <0.01$, $***p <0.001$.

Appendix Table A6. STEM Interest Effect by Components

	(1)	(2)
Outcome	Math Usefulness	STEM College Interest
Girl student Female tutor	0.7165+	0.6808***
	(0.2855)	(0.0553)
Boy student Male tutor	0.2913	0.0858
	(0.4523)	(0.1464)
Boy student Female tutor	0.5928	-0.2043
	(0.4757)	(0.1762)
N	204	204
Student Characteristics	X	X
Classroom Fixed Effect	X	X
Prior Achievement	X	X
Outcome	STEM Career Interest	STEM College/Career Interest
Girl student Female tutor	0.7363**	0.7086**
	(0.1314)	(0.0901)
Boy student Male tutor	-0.2127	-0.0635
	(0.1796)	(0.1486)
Boy student Female tutor	-0.3001	-0.2522
	(0.1529)	(0.1633)
N	204	204
Student Characteristics	X	X
Classroom Fixed Effect	X	X
Prior Achievement	X	X

Note: See Appendix Table B1 for complete survey items. Student characteristics include binary indicators for FRPL, IEP, LEP, and White. Prior achievement includes math Q1 grade, indicator for low 2021 STAR test, and prior math achievement factor. Standard errors are clustered by school. $+<0.10$, $*p < 0.05$, $** p < 0.01$, $***p < 0.001$.

Appendix Table A7. Intent to Treat Estimates with Science Teacher Fixed Effects

	(1)	(2)	(3)	(4)	(5)
	STEM Interest	FSA	Math C or above	Tut Sessions	Sch Absences
Girl student Female tutor	0.9013*	0.3963	0.0357	-0.4241	0.8646
	(0.2008)	(4.5035)	(0.0223)	(1.7697)	(4.2758)
Boy student Male tutor	0.1241	-3.7011	-0.0908+	1.2049	-1.3112*
	(0.1793)	(2.9910)	(0.0403)	(1.4406)	(0.4720)
Boy student Female tutor	0.1463	-1.1032	-0.0041	-1.2704	5.0995
	(0.1958)	(6.0320)	(0.0956)	(5.2736)	(4.0004)
Girl student Male tutor (Mean)	-0.16	71.81	0.56	27.58	27.64
N	204	195	422	422	422
Student Characteristics	X	X	X	X	X
Science Teacher Fixed Effect	X	X	X	X	X
Prior Achievement	X	X	X	X	X

Note: Student characteristics include binary indicators for FRPL, IEP, LEP, and White. Prior achievement includes math Q1 grade, indicator for low 2021 STAR test, and prior math achievement factor. Standard errors are clustered by school. $+<0.10$, *$p < 0.05$, ** $p <0.01$, ***$p <0.001$.

Appendix B. Preregistration

Registry of Efficacy and Effectiveness Studies

Study Title:
Effect of Tutor-Student Demographic Matching

Registry ID: 15300.1v1

Section I: General Study Information

PI name: Carly Robinson
PI affiliation: Brown University

Co-PI name: Josh Bleiberg
Co-PI affiliation: University of Pittsburgh

Co-PI name: Susanna Loeb
Co-PI affiliation: Stanford University

Primary Funding Source(s):
-

Award Number(s):
-

IRB Name:
Brown University

IRB Approval Date:
2021-03-16

IRB Approval Number:
2102002929

Other Registration Name:
-

Other Registration Date:
-

Other Registration Number:

-

Study Start Date:

2021-08-01

Study End Date:

2023-08-31

Intervention Start Date:

2021-11-01

Timing of entry:

Prior to collection of outcome data

Brief Abstract:

This randomized controlled trial seeks to explore the gender matching effect in a tutoring program on student beliefs, engagement, and academic outcomes in a large school district in New England. All Algebra 1 students at 5 sites were eligible to receive tutoring, the majority of whom were 9th graders.

Keywords: tutoring;
randomized controlled trial

Comments:

-

Section II starts on the next page.

Section II: Description of Study

Type of Intervention:
Practice

Topic Area of Intervention:
Mathematics and Science
Education

Number of intervention arms:
1

Target school level:
9

**Target school
type:** Urban

**Location of
Implementation:** *United
States* : Northeast

Further description of location:
-

Brief Description of Intervention Condition:
Student received small-group math tutoring from a tutor of their same gender

Brief Description of Comparison Condition:
Students received small-group math tutoring from a tutor that is not their same gender.

Comparison condition:
Other

Comments:
-

Section III: Research Questions

Confirmatory research questions:
Question 1:
Do students who have a same-gender math tutor have more positive math-related beliefs than those who do not
have a same-gender tutor?

Question 2:

Do students who have a same-gender math tutor have better attendance than those who do not have a same-gender tutor?

Question 3:

Do students who have a same-gender math tutor have improved academic outcomes than those who do not have a same-gender tutor?

Exploratory research questions:

Question 1:

Do students and tutors who have similar beliefs about tutoring at the beginning of the year have more positive mathrelated beliefs, better attendance, and improved academic outcomes than those who have less similar beliefs?

Question 2:

What is the distribution of tutor effectiveness?

Question 3:

Do students with high levels of self-efficacy have better academic outcomes or attendance?

Question 4:

Do students who value math skills have better academic outcomes or attendance?

Comments:

-

Section IV-A: Study Design (Selection)

Study Design:
Randomized Trial
(RT)

Comments:

-

Section IV-B: Study Design (Input)

Study Design: Input

Unit of random assignment of intervention:
Student

Assignment within sites or blocks:

Yes

Define the sites or blocks:

School

Probability of assignment to treatment the same across sites or blocks: No

Probability of assignment to treatment:

0.395

Unit outcome data measured: Student

Intermediate clusters between unit of random assignment and unit of measurement:

Yes

Description of intermediate clusters:

The randomization was also stratified by course period, and then further stratified by whether a student needed to be assigned to a tutor who spoke Spanish and the gender breakdown of the student group (100% female, 100% male, both female & male)

Comments:

Unweighted probability of being assigned to a match: 0.41

Weighted for likelihood of being assigned to a match across strata: 0.40

Site 1: 0.53

Site 2: 0.39

Site 3: 0.23

Site 4: 0.41

Design Classification

Based on the responses above, this study has been classified as: RT: Multisite (Blocked)

Section V: Sample Characteristics

Approximate number of students in the comparison condition within each school: 69

Approximate number of students in the intervention condition within each school: 48

Number of schools: 4

Were there certain students that were targeted for the study?
No

Were there certain students that were excluded from the study?
Yes - Due to the small number of students identifying as non-binary, for privacy concerns (i.e., within certain strata, students could be identifiable), we excluded them from the analysis.

Were there certain schools that were targeted for the study?
No

Were there certain schools that were excluded from the study?
No

Comments:
We will run two sets of models: one where we exclude students who had multiple primary tutors and one where they're included.

Section VI: Outcomes (Input)

Confirmatory question 1: Outcome Measure 1
Outcome domain: Student Social, Emotional, & Behavior - Attitudes
Minimum detectable effect size: .171
Outcome measure: Beliefs about math
Scale of outcome measure: Continuous
Normed or state test: No
Test-retest reliability:
Internal consistency:
Inter-rater reliability:
Same outcome measure in treatment and comparison groups: Yes

Confirmatory question 1: Outcome Measure 2
Outcome domain: Student Social, Emotional, & Behavior - Attitudes
Minimum detectable effect size: .171
Outcome measure:
Scale of outcome measure:
Normed or state test:
Same outcome measure in treatment and comparison groups:

Confirmatory question 1: Outcome Measure 3
Outcome domain: Student Social, Emotional, & Behavior - Attitudes

Minimum detectable effect size: .171

Outcome measure: Perceptions of tutor-student relationship

Scale of outcome measure: Continuous

Normed or state test: No

Test-retest reliability:

Internal consistency:

Inter-rater reliability:

Same outcome measure in treatment and comparison groups: Yes

Confirmatory question 2: Outcome Measure 1

Outcome domain: Student Social, Emotional, & Behavior - Attendance

Minimum detectable effect size: .171

Outcome measure: Attendance in tutoring

Scale of outcome measure: Continuous

Normed or state test: No

Test-retest reliability:

Internal consistency:

Inter-rater reliability:

Same outcome measure in treatment and comparison groups: Yes

Confirmatory question 2: Outcome Measure 2

Outcome domain: Student Social, Emotional, & Behavior - Attendance

Minimum detectable effect size: .171

Outcome measure: Attendance in school

Scale of outcome measure: Continuous

Normed or state test: No

Test-retest reliability:

Internal consistency:

Inter-rater reliability:

Same outcome measure in treatment and comparison groups: Yes

Confirmatory question 3: Outcome Measure 1

Outcome domain: Student Achievement - Mathematics

Minimum detectable effect size: .171

Outcome measure: Provider Assessment

Scale of outcome measure: Continuous

Normed or state test: No

Test-retest reliability:

Internal consistency:

Inter-rater reliability:

Same outcome measure in treatment and comparison groups: Yes

Confirmatory question 3: Outcome Measure 2

Outcome domain: Student Achievement - Mathematics

Minimum detectable effect size: .171

Outcome measure: Passing math course

Scale of outcome measure: Binary

Normed or state test: No

Test-retest reliability:

Internal consistency:

Inter-rater reliability:

Same outcome measure in treatment and comparison groups: Yes

Section VII: Analysis Plan

Baseline data collected prior to start of intervention:

Yes

Description of baseline data:

2020/21 School Year

- State Standardized Test Scores
- Provider Math Assessment
- Days Absent/Enrolled

Fall 2021

Student and Tutor Survey Fall 2021

- Baseline Perceptions of math
- Baseline anticipated relationships

Covariates you plan to include in the model:

English Language Learner Status, Free and Reduced Lunch Status, Gender, Grade, Vector of tutor

characteristics, Race, Special Education Status, Student Pretest **Analytic model:**

$$Y_{ijk} = \beta_0 + \beta_1 FstuFtut_i + \beta_2 MstuMtut_i + \beta_3 MstuFtut_i + \gamma mathb \int X_i + \eta T_j + \alpha_k + \epsilon_{ijk}$$

To explore the impact on student gender match, we will create indicators for student-tutor gender matches (i.e.,

Female student / Female tutor; Female student / Non-Female tutor; Male student / Male tutor; Male student / NonMale tutor). Due to the small number of students in our data identifying as non-binary, we exclude them to protect privacy.

We regress the student outcome of interest, Y, on these indicators, with Female student / NonFemale tutor groups left out as a reference category.

β_1 will provide the causal estimate of a female student being assigned to receive tutoring from a female tutor
as opposed to a non-female tutor.

$\beta_3 - \beta_2$ will provide the causal estimate of a male student being assigned to receive tutoring from a
male tutor as opposed to a non-male tutor.

$\mathbf{X}$ is a vector of student characteristics for each student i. $\mathbf{T}$ is a vector of tutor characteristics for each tutor j.
α_k is a school fixed effect. ϵ_{ijk} is an error term. Standard errors are clustered at the class level to account for nesting of students within a math intervention period.

Tutors exited and joined the program after the initial randomization. Thus, to better understand the impact of being assigned to a tutor with a similar gender, we also will instrument a students' initial treatment assignment (i.e., the initial tutor's gender) with the proportion of scheduled sessions with a tutor who shares the gender of the students was initially assigned to.

Plan to handle cases with missing outcome data:
Delete cases with missing data for the outcome being analyzed

Planned multiple comparisons adjustment, confirmatory question 1 (Student Social, Emotional, & Behavior):
Yes

Number of planned comparisons to adjust, confirmatory question 1 (Student Social, Emotional, & Behavior):

3

Correction for multiple comparisons, confirmatory question 1 (Student Social, Emotional, & Behavior): other

Planned multiple comparisons adjustment, confirmatory question 2 (Student Achievement - Mathematics):

No

Planned multiple comparisons adjustment, confirmatory question 3 : No

Comments:

We worked with the tutoring provider to create a randomization protocol to preserve randomization when students were added to the program and/or had to leave a specific tutoring period due to scheduling changes.

Section VIII: Additional Information

Links:

No links have been added yet.

Files:

File Name:

Description: Power calculation

Comments:

-

Appendix C. Deviations from the Preregistration

We made several deviations from our publicly posted preregistration, which was before we had access to the outcome data. In this section, we outline the changes that were made.

Sample:
In our preregistration, we note that "We worked with the tutoring provider to create a randomization protocol to preserve randomization when students were added to the program and/or had to leave a specific tutoring period due to scheduling changes".

In practice, we did this to create a process for randomly assigning tutors to students after November 1. Ultimately, however, we excluded these students who joined tutoring after the official start date because we found they were all tutored for several months by tutor supervisors (who were predominantly women) before being assigned a new tutor later in the school year. The supervisors did not tutor any students in the initial randomization, suggesting that the randomization protocol was not implemented as planned. 483 students were included in the original randomization. 17 students were excluded because they were tutored by a non-binary tutor. 26 were assigned to an "Open Slot" who were ultimately tutored by a supervisor.

Measures:
We had noted we would use the "Student Pretest" as a covariate in our primary model.

We intended to use the FSA diagnostic scores, however many students in our sample were missing FSA diagnostic scores and would result in these students being dropped from our primary analysis. To address this, we used the continuous Q1 math GPA and 8th grade summative assessments to impute missing values to create a "prior math achievement factor."

Model:
In our preregistered model, we noted that we would include a school fixed effect.

This was incorrect, as the randomization was stratified at the classroom level. Our analysis includes a fixed effect for the students' math class. The classroom fixed effects control for time-invariant characteristics including student sorting, curriculum, and aspects of teacher efficacy.