Closing the Gender Gap in STEM: Role of Performance Feedback and Advice*

Fulya Ersoy and Derek Rury

This version: November 20, 2024

Most Recent Version:

https://fulyaeg.github.io/ErsoyRury2024Closing.pdf

Abstract

Women remain underrepresented in STEM occupations, largely due to educational choices. This study investigates two interventions—performance feedback and advice—aimed at narrowing the gender gap in selecting math tasks over verbal tasks. In an online lab experiment, participants chose between completing a math task or a verbal task. Before making their choice, they received either performance feedback on previous tasks, advice on which task to choose, or both, with advisor gender varied. A significant gender gap was observed in math task choices in the group who did not receive the interventions. Performance feedback intervention eliminated this gap by increasing the proportion of females selecting the math task, while leaving choices of males unaffected. Being advised to choose math task increased math task selection only among females, while being advised to choose verbal task decreased math task selection only among males. The advisor's gender generally had no significant impact on the effectiveness of advice. This study contributes to our understanding how performance feedback and advice affect educational decisions, highlighting the complexity of these interventions.

*Fulya Ersoy, fulya@uchicago.edu, University of Chicago and Derek Rury, rury@uchicago.edu, University of Chicago.This project is registered at AEA RCT registry (https://doi.org/10.1257/rct.11317-2.0) and is approved by the University of Chicago Social and Behavioral Sciences IRB (protocol number: IRB23-0238). We thank Loyola Marymount University for funding. We thank participants of the Loyola Marymount University brownbag seminar for helpful comments.

1 Introduction

Despite substantial advancements in gender equality, women remain underrepresented in science, technology, engineering and mathematics (STEM) careers. Educational choices play an important role in this gender gap. Using data from 10 countries, Ahimbisibwe et al. (2024) shows that there is a large gender gap in applications to STEM majors among talented students. According to the National Center for Education Statistics, while women in the U.S. earn 57% of all bachelor's degrees, they account for only 21% of degrees in engineering and 19% of degrees in computer science. This disparity begins early with girls choosing different subjects in secondary school (Delaney and Devereux, 2019) and boys pursuing more advanced math and science courses in high school (Speer, 2023). This gap continues to grow in college and beyond (Speer, 2023; Delaney and Devereux, 2024). The implications of this gender disparity extend beyond individual career prospects, posing a substantial economic cost to society given the STEM workers are fundamental for innovation and productivity growth (Peri, Shih and Sparber, 2015). Hence, finding effective interventions to reduce this gender gap remains crucial.

In this study, we test the effectiveness of two interventions, providing information about previous test performance (performance feedback) and providing advice, on closing the gender gap in choosing math test instead of a verbal test through an online lab experiment. We design the experiment so that we can investigate the role of gender match between the advice giver and the advice taker as well as possible interactions between receiving advice and receiving performance feedback. There is a large literature in education that investigates how providing performance feedback affects students' academic performance (Azmat and Iriberri, 2010; Bandiera, Larcinese and Rasul, 2015) and individuals' competition entry decisions (Ertac and Szentes, 2011; Wozniak, Harbaugh and Mayr, 2016). Research in education also demonstrates that counselors and academic advisors can significantly influence students'

college-going decisions, college major choices, and college graduation rates (Canaan, Deeb and Mouganie, 2022; Mulhern, 2023). This influence is especially pronounced when there is a gender match between the student and the advisor (Carrell, Page and West, 2010; Canaan and Mouganie, 2023). Our interventions are informed and motivated by these strands of the literature.

For our experiment, we recruited participants from Prolific, an online research platform. Participants first took a short math and verbal assessment. A random half of the participants learned their math and verbal scores from the assessment whereas the other half did not receive any feedback on their performance. Then participants were randomly assigned to either not receive any advice or receive advice on what type of assessment (Math or Verbal) they should take in the second part of the study.[1] In the advice treatments, we also randomized the gender of the advisor (ungendered, female, or male).[2] Crucially, the gender of the advisor was orthogonal to the advice provided. In the second part of the study, subjects chose which assessment - math or verbal - to take. Subjects were paid either based on their performance in part 1 or part 2, randomly selected by the computer.

Our analysis reveals a significant gender gap in choice of the math test, with only 11% of females in the control group (the group receiving neither advice nor performance feedback) opting for the math test compared to 35% of males. When dividing subjects into three categories based on their initial test scores—better at math, equally skilled, or worse at math—a gender gap in test choice is evident across all categories. For instance, among individuals who performed better in math, 16% of females and 32% of males opted to take

[1] Advice was created from an experiment we conducted with another set of Prolific subjects before conducting this experiment. The subjects in that experiment were asked to provide advice based on different score profiles; hence, advice were mostly aligned with subjects' scores. For some score combinations, we were able to randomly assign the advice.

[2] We included a treatment arm to study the effects of receiving advice from an ungendered source since advice giving can happen online without revealing one's identity.

the math test. In contrast, among those who were worse at math, only 7% of females chose the math test compared to 42% of males. Thus, gender gap persists even after controlling for the initial test scores. While the gap narrows with additional controls, suggesting that factors such as risk preferences contribute to the gender differences in the choice of math test (Croson and Gneezy, 2009), it remains significant.

Performance feedback intervention had a notable impact, particularly on female participants. While receiving performance feedback increased the likelihood of choosing the math test by 12.8 percentage points for the overall sample, the effect was driven primarily by females, whose math test choice increased by 24 percentage points, effectively closing the gender gap in this context. The effects was especially prominent for individuals who performed better at math. Within this group, 84% of females and 67% of males chose to take the math test.

Advice intervention did not significantly affect the math test choice or the gender gap in the math test choice, but the effect of advice intervention varied depending on its type (math vs verbal). Being randomly advised to take the math test compared to receiving no advice increased the likelihood of choosing the math test in the overall sample. This effect was driven by female participants. The effects for males were smaller, sometimes negative, and never statistically significant. Despite this, we cannot reject that the gender gap for those who were randomly assigned to receive Math advice is equal to the gender gap for those who did not receive any advice. Being randomly advised to take the verbal test compared to receiving no advice decreased the proportion choosing the math test (instead of the verbal test), solely among male subjects, marginally reducing the gender gap in the choice of math test.[3] Effects were generally not statistically different across the various advisor treatments

[3]One potential reason for why Verbal advice did not increase proportion of females choosing the verbal test is that a large proportion of them (89%) were already choosing the verbal test in the control group.

(ungendered, female, or male), suggesting advisor gender or having a gender match between the advisee and the advisor were not important factors in subjects' choice of math test in our context.

The experimental design allows us to investigate whether the performance feedback and advice interventions interact. In our setting, advice received and the message delivered through performance feedback were aligned for most of the subjects. For some score combinations, advice was perfectly aligned with the message delivered through performance feedback, e.g., subjects receive Verbal advice if their verbal score is greater than their math score or they receive Math advice if their math score is greater than their verbal score. For some other score combinations, we were able to randomly assign the advice; hence, advice either reinforced the message delivered through performance feedback or diluted it.[4] When looking across both genders, there is no evidence of that performance feedback and advice treatments act as complements or substitutes to each other in the choice of math task. There is suggestive evidence of that performance feedback and math advice are substitutes to each other for female subjects. In general, our results show that the effects of the performance feedback and advice interventions and the interactions between the two interventions differ by gender.

This paper contributes to two strands of the literature. First, this paper contributes to the literature that investigates how performance feedback affects individuals' decisions and whether there are gender differences in reactions to performance feedback. One strand of this literature shows that performance feedback positively affects effort (Eriksson, Poulsen and Villeval, 2009; Castagentti and Rury, 2024) and performance (Azmat and Iriberri, 2010;

[4]We notice that in other educational settings, advice and performance feedback might not be aligned. For example, if the advisor feedback is based on comparisons between the student's performance and that of their peers while the performance feedback is based on student's own scores, they might provide complementary information.

Bandiera, Larcinese and Rasul, 2015; Azmat, Bagues, Cabrales and Iriberri, 2019; Dobrescu, Faravelli, Megalokonomou and Motta, 2021; Bobba and Frisancho, 2022) at least in the short-run. Another strand of this literature focuses on decision to enter competition and finds that relative performance feedback shrinks the gender gap in competition entry, particularly for male-typed tasks (Ertac and Szentes, 2011; Wozniak, Harbaugh and Mayr, 2014; Berlin and Dargnies, 2016; Jeworrek, 2019; Coffman, Ugalde Araya and Zafar, 2024).[5] Our paper contributes to this literature by studying how *absolute* performance feedback impacts the choice of *task* (Math or Verbal). Papers that focus on the effects of performance feedback on task choice are scarce with the exception of Baier, Davis and Jaber-Lopez (2024) and Coffman, Ugalde Araya and Zafar (2024). Baier, Davis and Jaber-Lopez (2024) finds that there is no significant gender gap in the choice of male-typed task when receiving absolute performance feedback or absolute performance feedback combined with relative performance feedback.[6] Coffman, Ugalde Araya and Zafar (2024) is the closest paper to ours and finds that receiving *relative* performance feedback does not affect the gender gap in the choice of math test.[7]

Secondly, this paper contributes to the literature that investigates how advisors influence the decisions of their advisees (see Schotter (2003) for a review of the experimental economics literature on advice and Bonaccio and Dalal (2006) for a review of the organizational psychology literature), particularly in educational settings (Borghans, Golsteyn and Stenberg, 2015; Brandts, Groenert and Rott, 2015; Mulhern, 2023; Gentry, Meer and Serra, 2023).

[5]Most of the papers in this literature uses tasks like adding 2-digit numbers, ticking the symbols that adhere to a certain rule, forming words that begin with a specific letter, etc.

[6]Baier, Davis and Jaber-Lopez (2024) do not have a control group with no performance feedback hence they are unable to measure the effect of absolute performance feedback on task choice. Their math task is a summation task and their verbal task is a word puzzle. Subjects are the University of Innsbruck students.

[7]This outcome variable is not the main outcome variable studied in the paper and it is only presented in the supplementary appendix. Coffman, Ugalde Araya and Zafar (2024) mainly investigates the dynamic effects of relative performance feedback on tournament entry by providing a noisy but informative signal to the subjects. The design allows them to causally identify the effects of good news versus bad news.

Given the large gender gap in STEM fields and occupations, much of the work in advice regarding educational decisions has focused on the decision to pursue math fields in high school and beyond (Carlana, 2019; Welsch and Winden, 2018), some highlighting the importance of gender match between the student and the advisor/instructor (Carrell, Page and West, 2010; Canaan and Mouganie, 2023). However, in a field setting, students choose whether to seek advice and from whom to seek advice and they might receive different advice based on their gender (Heikensten and Isaksson, 2019; Gallen and Wasserman, 2023; Lordan and Lekfuangfu, 2023; Gallen and Wasserman, 2024; Coutts, Koh and Murad, 2024).[8]

Since both the advice seeking behavior and the advice received can be endogenous based on the gender of the advisee and the gender of the advisor, it is challenging to isolate the importance of advisor-advisee gender match in a field setting from other potential mechanisms. In our study, we shut down any advice seeking differences in gender by design. Moreover, the advice received is independent of the gender of the advisors.Our *null* results about the importance of gender match between the advisor and advisee suggests that it is not simply the case that girls follow the advice of female advisors more due to sharing the same gender identity, but there is another explanation, perhaps a role model effect (Porter and Serra, 2020; Patnaik, Pauley, Venator and Wiswall, 2023), for the studies finding importance of gender match between the student and the advisor in field settings (Carrell, Page and West, 2010; Canaan and Mouganie, 2023).

The rest of the paper is structured as follows. Section 2 provides details on the experimental design. Section 3 describes the data and our descriptive results. Section 4 presents

[8]Heikensten and Isaksson (2019) shows that males seek more advice than females, especially for verbal tasks. Gallen and Wasserman (2023) documents that female students disproportionately reach out to female mentors. Lordan and Lekfuangfu (2023) shows that career advice given by teenagers to a fictitious peer is affected by the gender of the fictitious peer. Gallen and Wasserman (2024) documents that the advice student receives is a function of their gender in a field experiment where college students seek information from working professionals. Coutts et al. (2024) finds that male advisors are more likely to obscure vague signals from female advisees than male advisees.

our results and Section 5 discusses the policy implications.

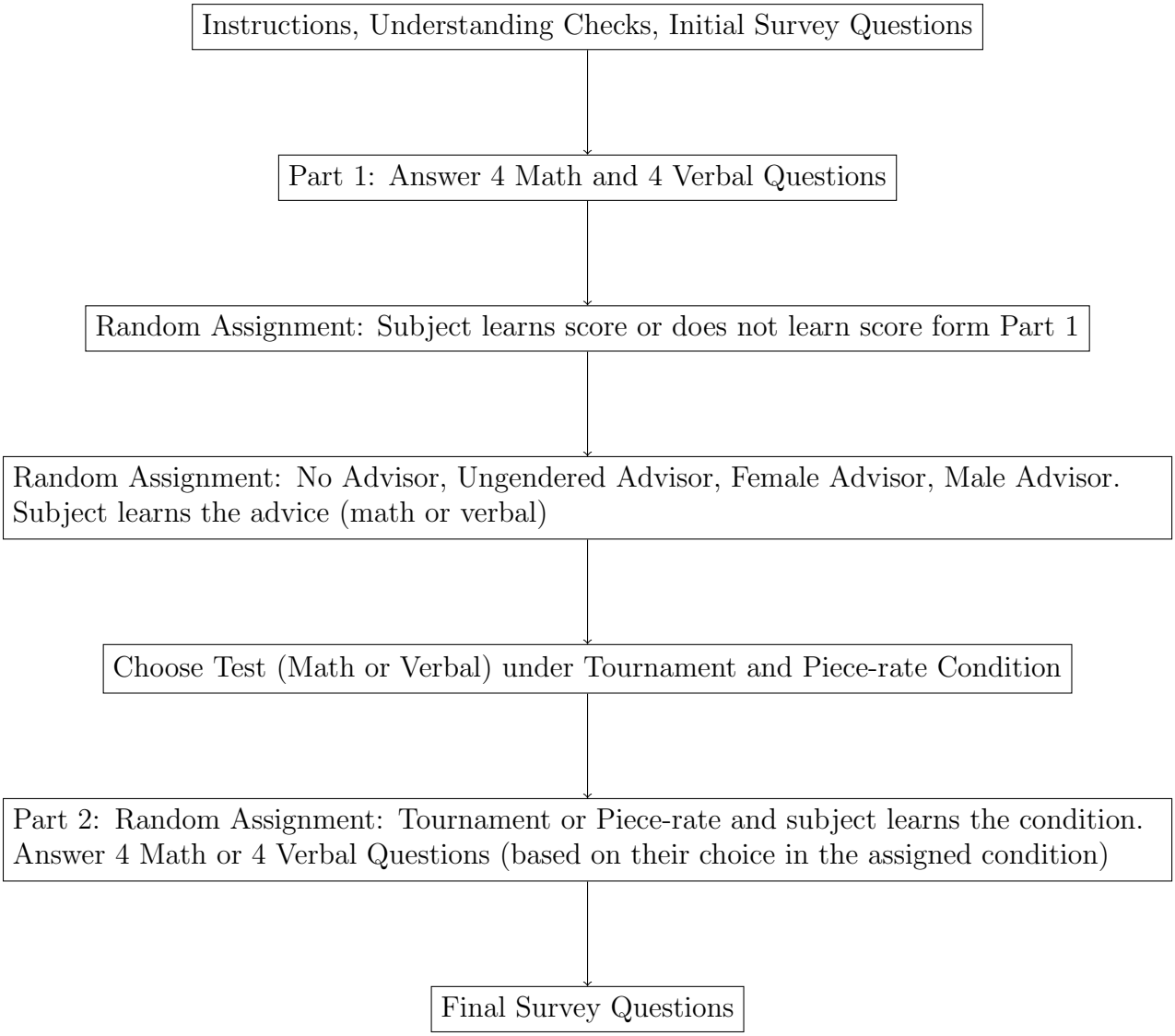
2 Experimental Design

2.1 Design Overview

Subjects recruited from Prolific read the instructions, completed comprehension checks, and provided demographic information. In Part 1, they took assessments where they answer four math and four verbal questions. After completing these assessments, some subjects were randomly assigned to receive performance feedback, informing them of the number of questions they answered correctly on the math and verbal assessments. They, then, were randomly assigned into one of the following treatments: no advisor, ungendered advisor, female advisor, or male advisor. Those who were in the advice treatments received advice ("choose math" or "choose verbal") about which test to choose in the next part of the study. Advice was independent of the advisor gender. Subjects then selected which test they preferred to complete under a piece-rate condition and under a tournament condition. After the selection, subjects were randomly assigned to the piece-rate condition or the tournament condition and completed the test they selected for the assigned condition. At the end, they answered questions designed to assess their attention during the study. Figure 1 summarizes the experimental design.

In the following subsections, we provide details on the assessments, including how we chose which questions to include in our math and verbal assessments. We also describe different treatment arms, participants' incentives, and details about how we measured attention and comprehension. Finally, we provide details about our pre-registration.

Figure 1: Experimental Design Diagram



2.2 Test Questions

We used verbal and math questions from GRE practice tests for our assessment. Initially, we chose 15 verbal questions and 15 math questions from these tests. Then, we conducted a pilot with 100 Prolific users to determine the accuracy rates and time spent in each question as well as the correlations in accuracy across math and verbal questions. Each individual in the pilot survey was provided with 5 randomly chosen math question and 5 randomly chosen verbal question from this initial list of questions. They were paid \$0.20 per correct answer (up to \$2) in addition to their participation payment of \$2.[9]

Using data from this pilot, we selected 4 verbal and 4 math questions for the first part of the main experiment, as well as another 4 verbal and 4 math questions for the second part. We made this selection so that the difficulty of verbal questions and math questions are similar to each other, that the accuracy rates are not different across genders, and that there is strong positive correlation between accuracy in individual math questions and verbal questions. See Appendix Section C for the final list of questions.

In Part 1 of the experiment, all subjects answered 4 math questions and 4 verbal questions. The order of the blocks (math vs verbal) and the order of the questions within a block were randomized. There were two attention checks built in the first part.[10] In Part 2 of the experiment, subjects were given either 4 math or 4 verbal questions based on their choice of test.

To discourage looking up the answers, we took various precautions. The questions included in the experiment were hard to type in a browser because they involved graphics, tables,

[9] To discourage looking up the answers, the copy/paste feature was disabled and we asked subjects to pledge not to look up the answers. At the end of the test, we asked pilot subjects whether they looked up the answers and only 4% admitted that they looked up some answers.

[10] 99% of the subjects passed the verbal attention check and 99.75% of the subjects passed the math attention check.

and/or long text and the copy/paste feature was disabled. Subjects were given one minute per question after which the screen auto advanced. If a subject didn't answer the question within the time limit, we counted that question as incorrect. At the beginning of the study, we told subjects that the research study's validity depends on them answering the questions themselves without getting help. They pledged that they will not get help to answer questions and they will not share the questions and/or answers with anyone. At the end of the experiment, we asked subjects whether they looked up the answers and only 6% admitted that they looked up some of the answers.

2.3 Treatments

After Part 1 of the experiment completed, half of the subjects were randomly assigned to learn their scores (performance feedback treatment) while the remaining half did not (no performance feedback treatment). Subjects in the performance feedback treatment learned how many math questions and how many verbal questions they answered correctly in the first part of the experiment.

Then, subjects were randomly assigned one of the following conditions. 1/6th of the subjects did not receive any advice (no advisor treatment). 1/6th of the subjects received advice from an advisor but they were not provided any information regarding the gender of the advisor (ungendered advisor treatment). 2/6th of the subjects received advice from a female advisor (female advisor treatment) and another 2/6th of the subjects received advice from a male advisor (male advisor treatment). Within each group who received advice, half of the subjects learned that their advisor performed among the top 50% of the advisors in that test and the other half did not receive any information about the knowledge level of their advisors. See Appendix Figure 1 for the visuals of each treatment.

The advice subjects received (math or verbal) depended on their performance in part 1 and randomization. To provide advice without deceiving the subjects, we first ran an experiment with another set of Prolific subjects (the *advisor* experiment). In this experiment, after answering test questions that belonged to part 2 of the main experiment, subjects (*advisors*) were presented with different profiles (in terms of possible number of correct answers in the math and verbal tests of part 1 of the main experiment) and were asked to provide their advice regarding the test (math or verbal) a subject with this profile should choose. To incentivize truthful advice, *advisors* knew that they might get paid based on the part 2 test performance of a subject in the main experiment (*advisee*). Advisors did not know any characteristics (gender, race, age, etc.) of their advisees other than their scores and this was common knowledge to the *advisees*. See Appendix Figure 2 for the relationship between the two experiments. More details about the advisor experiment can be found in Appendix Section D.

Based on the advice provided in the *advisor* experiment, we assigned advice (math or verbal) to the participants in our main experiment. For 10 out 25 score combinations (category 1), we had some *advisors* recommending taking a math test and others recommending a verbal test. Hence, for *advisees* with these score combinations, we were able to randomly assign the advice. For 9 score combinations (category 2), we had all *advisors* recommending verbal test. Hence, *advisees* in this category received Verbal advice. For the remaining 6 score combinations (category 3), we had all *advisors* recommending math test. Consequently, *advisees* in this category received Math advice. The details can be found in Appendix Table 1.

In our setting, message delivered through the performance feedback intervention and through the advice intervention were aligned for most of the subjects. For participants who were in categories 2 and 3, performance feedback and advice were perfectly aligned. For participants who were in category 1, those who were randomly assigned to receive either

Math advice or Verbal advice, advice was either in the same direction with the performance feedback or in the opposite direction. Overall, 79% of individuals who were better in math based on their part 1 scores received math advice and 97% of the individuals who were better in verbal based on their part 1 scores received verbal advice.

Finally, we had one more treatment variation. In part 2, we randomly assigned subjects to either a piece-rate condition or a tournament condition. In the piece-rate condition, subjects were paid for each correct answer on the test they chose to complete in part 2 (math or verbal). In contrast, subjects in the tournament condition were paid only if they outperformed a randomly paired participant in their chosen test, provided both chose the same test for part 2. Before we revealed the assigned treatment condition to the subjects, we asked their choices of test for part 2 (math or verbal) both for piece-rate and tournament conditions (strategy method). In this respect, our paper differs from the literature on competitiveness since we look at the test choices under piece-rate and tournament condition, not the decision to enter competition.[11]

2.4 Payments

Subjects were paid a fixed amount of $2 and were paid a bonus based on the accuracy of their answers and which part was randomly chosen for payment by the computer. If the computer chose part 1 for payment, then subjects received $0.25 per correct answer (up to $2). If the computer chose part 2 and piece-rate for payment, subjects received $0.50 per correct answer (up to $2). If the computer chose part 2 and tournament for payment, subjects were randomly matched with another subject who chose the same assessment with them (math or verbal) and who took it under the tournament condition. If their score was above their

[11] 85.6% of our subjects chose the same test under both the piece-rate and tournament conditions. Furthermore, we do not observe the performance feedback effects or advice effects vary depending on the payment condition. Hence, in the analysis that follows, we pool the choices in the two payment conditions and control for the payment condition in the regressions.

opponent, they received $2 and if their score was below their opponent, they received $0. Ties were resolved randomly by the computer. All the payment procedures were explained to the subjects.

Subjects earned an additional $0.25 bonus if they recalled their advisors' characteristics (gender of the advisor (male, female, or ungendered) and whether they were given information about knowledge level of the advisor), the advice they received (math or verbal), and two true/false questions regarding the details of the experiment[12]. 86% of the subjects correctly recalled the gender of the advisor and 98% of the subjects correctly recalled the advice they received. Only 42% of the subjects correctly recalled whether they were given information about the knowledge level of their advisor, which is less than luck. Due to this issue, we do not conduct any analysis regarding the effects of knowledge level of the advisor. Overall, only 17% of the subjects earned the additional bonus.

2.5 Other details

We recruited 1200 participants (600 female and 600 male) from Prolific (www.prolific.com) with the following criteria: aged between 18-30, a Prolific approval rate is between 95-100% and 10–2,000 prior submissions on the platform. Data from Prolific includes the age, sex, and race of the participants as well as their total approvals on Prolific.[13]

[12]The two true/false questions were the following: "Your advisor knew your performance from the first test when providing their advice." and "Your advisor didn't know your characteristics (gender, race, age, etc.) when providing their advice." 77% of the subjects answered the first question correctly and 75% of the subjects answered the latter correctly.

[13]Prolific researcher guidelines state that any participant who has completed a study and has provided the data should be approved and paid unless they meet any of the following rejection criteria: providing exceptionally fast response, not answering critical questions that were compulsory, failing fair attention checks. Assuming most subjects would be approved, we can view approvals as a proxy for number of submissions on Prolific, which shows us how experienced the subject is with the Prolific platform.

After the instructions of the study provided, we checked for subjects' understanding of the instructions with three survey questions. 82% of the subjects answered all three of these questions correctly. Subjects who did not answer these questions correctly were still able to continue with the experiment and were included in the analysis that follows.

We asked subjects whether they are currently a student and whether their parents have a college degree. We measured subjects' patience and risk tolerance using qualitative survey questions (Falk, Becker, Dohmen, Huffman and Sunde, 2022). We also assessed subjects' self-evaluation of their proficiency in math tasks compared to verbal tasks using a survey question.

At the end of the experiment, in addition to answering the recall questions, subjects in the advice treatments were asked "How likely would you recommend this advisor to another person?" with a 5-item Likert scale (extremely unlikely to extremely likely). Subjects were permitted to use a calculator during the experiment, and 49% reported using one when asked. We also asked an open-ended question about the purpose of the experiment.[14]

2.6 Pre-Registration

We preregistered the study on AEA RCT Registry (https://doi.org/10.1257/rct.11317-2.0). We wrote "this study investigates the impact of advisor-advisee gender match on advisee's propensity to follow advice as well as their test choice (math vs verbal)". We also stated that we will explore heterogeneity by gender of advisee, whether the advisee knows their own score, whether the advice received is stereotypical or non-stereotypical, whether

[14] 572 respondents stated that they do not know or they are not sure. Out of remaining 628 respondents, 50 respondents stated an answer that relates to the effect of performance feedback on test choice and 89 respondents (not mutually exclusive) stated an answer that relates to the effect of advice/advisor on test choice. The most common response was "testing ability or perceived ability in math versus verbal tasks". Other common responses include testing the "effect of incentives on cheating" and "understanding the gap between actual and perceived ability".

the advisee is informed about the knowledge level of advisor, payment scheme (tournament vs piece-rate), risk aversion of the advisee, whether the advisee is a student in real life, parental education level of the advisee, and whether the advisee perceives themselves as better at math. We registered two primary outcome variables: Whether a subject choose the non-stereotypical test (math test being the non-stereotypical test for females and verbal test being the non-stereotypical test for males) and whether a subject follows advice (defined as choice of test by the advisee being the same with the advice of the advisor).

In the analysis that follows (Section 4), we use *choice of math test* as our dependent variable since we think the results are easier to grasp and more objective using this variable rather than using *choice of the non-stereotypical test* as the dependent variable. We also investigate the impact of whether the subject knows their own score (performance feedback) on the choice of math test by itself in addition to conducting the pre-registered analysis of exploring the heterogeneity of the advisor-advisee gender match by performance feedback. Hence, the analysis in the paper diverges from our pre-registration.

3 Data and Descriptive Statistics

Table 1 displays the averages of the observable characteristics both for the overall sample (Column 1) and for different treatment arms (Columns 2-7). 62% of our subjects were White, 12% were Black, 10% were Asian, 9% were Mixed and 6% was in the other category. Mean age was 24.7 years old (standard deviation is 3.32). 40% of our subjects were students at the time of the study and 56% of them had at least one parent with a college degree. Among the participants, 35.5% perceived themselves as better at math tasks, 44.5% as better at verbal tasks, and the remainder as equally skilled in both.

Table 1 Columns 8 and 9 presents the p-values for the equality of means across different

treatments to check the balance across treatments. As shown in Column 8, there are some racial differences and differences in terms parents' education between the individuals who were randomly assigned into performance feedback. There are also some racial differences and differences in Prolific approval ratings among the ones who were randomly assigned to different advisor treatments as seen in Column 9. Hence, we report the results both without and with controls in the analysis that follows.

Table 1: Summary Statistics and Balance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Perf. Feedback:		**Advisor:**				**P-values:**	
	All	**No**	**Yes**	**No**	**Ungendered**	**Male**	**Female**	**2=3**	**5=6=7=8**
Female	0.500	0.489	0.511	0.492	0.544	0.489	0.493	0.453	0.588
Race:									
Asian	0.104	0.119	0.090	0.132	0.093	0.071	0.129	0.096	0.027
Black	0.121	0.121	0.121	0.137	0.113	0.124	0.114	0.981	0.841
Mixed	0.093	0.092	0.095	0.091	0.074	0.094	0.104	0.886	0.684
Other	0.063	0.075	0.051	0.086	0.049	0.043	0.079	0.088	0.075
White	0.618	0.593	0.643	0.553	0.672	0.668	0.574	0.072	0.004
Age	24.69	24.71	24.66	24.68	24.69	24.60	24.77	0.789	0.918
Student	0.396	0.395	0.396	0.396	0.392	0.400	0.394	0.971	0.997
Educated Parent	0.556	0.521	0.590	0.569	0.564	0.572	0.53	0.015	0.633
Risk Preferences	5.398	5.491	5.307	5.310	5.270	5.448	5.458	0.147	0.678
Time Preferences	6.571	6.556	6.585	6.629	6.706	6.395	6.646	0.799	0.189
Perceptions:									
Better at Math	0.355	0.348	0.362	0.350	0.387	0.349	0.347	0.635	0.771
Better at Verbal	0.446	0.439	0.453	0.492	0.407	0.425	0.463	0.629	0.248
Equal	0.199	0.213	0.186	0.157	0.206	0.225	0.191	0.242	0.252
Prolific Approvals	720	741	699	807	687	724	690	0.159	0.054
Subjects	1200	597	603	197	204	395	404	1200	1200

Column 1 shows the means of the observable characteristics for the overall sample. Columns 2 and 3 present these means for subjects who were randomly assigned to no performance feedback and performance feedback, respectively. Columns 4-7 display the averages for subjects in different advisor treatments: no, ungendered, male, and female, respectively. Column 8 tests the equality of means across subjects of performance feedback treatments and Column 9 tests the equality of means across subjects of advisor treatments. Student indicates whether the subjects are currently students and Educated Parent is equal to 1 if at least one of the subjects' parents has a college degree. We measured subjects' risk and time preferences using qualitative survey questions (Falk, Becker, Dohmen, Huffman and Sunde, 2022). The scale ranges from 0 to 10 where the higher number indicates more preference for risk and more patience, respectively. Perceptions about math versus verbal ability measures subjects' self evaluation about how good they are at math tasks compared to verbal tasks. Prolific approvals are subjects' total number of approvals on Prolific.

Next, we explore if there are gender differences in part 1 test scores across females and males.[15] Table 2 shows the distribution of math and verbal scores for females and males separately as well as for the overall sample. The average number of correct answers in math test is 1.81 for females and 1.91 for males and we can marginally reject the equality of the means using a t-test (two-sided p-value: 0.097), but we are unable to reject the equality of the distributions (K-Smirnov p-value: 0.485). The average number of correct answers in verbal test is 1.89 for females and 1.76 for males and we can reject the equality of the means using a t-test (two-sided p-value: 0.045) and we are able to reject the equality of the distributions (K-Smirnov p-value: 0.059).[16] 35% of females and 41% of males performed better in the part 1 math test compared to the verbal test. Conversely, 40% of females and 32% of males had more correct answers in the part 1 verbal test than they had in the part 1 math test.

4 Results

In this section, we present how our interventions affect choice of math test and how the effects differ by gender.[17] First, we provide a tabulation of math test choice by treatments, by gender, and by initial test scores. Next, we explore correlates of the math test choice in the control group (no performance feedback or advice received). Then, we look at how performance feedback affects choice of math test. Next, we investigate how advice affects the choice of math test based on the advisor gender and the type of the advice received (math or verbal). Finally, we investigate the interactions between performance feedback and advice on math test choice.

[15] The test questions were selected based on not having accuracy differences across genders in the pilot.

[16] Appendix Table 2 shows that the mean differences between females and males' math scores as well as verbal test scores become insignificant once we control for the observable characteristics.

[17] In our setting, choosing math test is equivalent to not choosing the verbal test since there are only two choices available.

Table 2: Part 1 Test Scores

	(1) All	(2) Female	(3) Male
Math Scores:			
Average (SD)	1.856 (1.059)	1.805 (1.031)	1.907 (1.086)
0	132	69	63
1	317	160	157
2	402	211	191
3	290	139	151
4	59	21	38
Verbal Scores:			
Average (SD)	1.823 (1.165)	1.89 (1.171)	1.755 (1.156)
0	198	97	101
1	274	116	158
2	342	183	159
3	315	164	151
4	71	40	31
Math versus Verbal Scores:			
Math>Verbal	454	209	245
Math=Verbal	314	149	165
Math<Verbal	432	242	190
Subjects	1200	600	600

This table shows the mean, standard deviation (SD), and the distribution of math and verbal scores from part 1 of the experiment for the overall sample in Column 1, for females in Column 2, and for males in Column 3.

4.1 Preliminary Analysis

In this section, we provide a simple tabulation of math test choice by treatments, by gender, and by how part 1 scores in math and verbal tests relate to each other. Table 3 presents the results. Panel A shows the percentage of individuals choosing the math test over verbal test in each treatment for all score combinations. 23% of the individuals who received neither advice nor performance feedback (the control group) chose the math test whereas 27% of the individuals who received advice only chose the math test. About 38% of individuals in the performance feedback treatments (with either no advice or advice) chose the math

test. Panels B, C, and D show the percentage of individuals choosing the math test in each treatment for individuals whose part 1 math score was greater than, equal to, and less than part 1 verbal score, respectively. Comparing the control treatments across panels, percentage of individuals choosing math are similar (between 21% and 24%), which suggests individuals are not aware how well they performed in math versus verbal tests in part 1. This interpretation can be further strengthened comparing the performance feedback only treatments across panels. Once provided with the performance feedback, 74% of individuals whose part 1 math score was greater than part 1 verbal score chose math, compared to 13% of individuals whose part 1 math score was less than part 1 verbal score. We also observe a similar but weaker pattern for the treatments that involve advice, which is expected given the advice is mostly dependent on part 1 test scores.

Comparing males and females, we see patterns suggestive of that females and males react to different treatments differently. Considering all score combinations (panel A), females are less likely to choose math test compared to males, with the exception of the performance feedback only treatment, where we cannot reject the equality of the percentages. Gender gap in math choice in the control group is especially large where only 11% of females choosing math test whereas 35% of males choosing the math test. When part 1 math score was greater than part 1 verbal score (panel B), performance feedback (with either no advice or advice) was effective at eliminating the gender gap in the choice of math test whereas gender gap persists under the advice only treatment. When part 1 math score is less than part 1 verbal score (panel D), only 7% of the females chose the math test in the control group whereas 42% of the males did. For subjects in Panel D, treatments involving performance feedback and/or advice reduced the percentage of males choosing the math test without affecting the proportion of females.

Table 3: Choice of Math Test across Treatments, Genders, and Part 1 Test Scores

	(1) All	(2) Female	(3) Male	(4) p-value (2)=(3)
Panel A: All Scores				
No PF & No Advice	0.226	0.112	0.352	.0001
	(0.419)	(0.317)	(0.480)	
PF & No Advice	0.375	0.365	0.384	.7739
	(0.485)	(0.484)	(0.489)	
No PF & Advice	0.272	0.212	0.328	.0000
	(0.445)	(0.409)	(0.470)	
PF & Advice	0.382	0.340	0.427	.0050
	(0.486)	(0.474)	(0.495)	
Panel B: Part 1 Math Score > Part 1 Verbal Score				
No PF & No Advice	0.243	0.156	0.316	.1210
	(0.432)	(0.369)	(0.471)	
PF & No Advice	0.738	0.844	0.667	.0778
	(0.443)	(0.369)	(0.476)	
No PF & Advice	0.443	0.359	0.520	.0014
	(0.497)	(0.481)	(0.501)	
PF & Advice	0.689	0.706	0.675	.5224
	(0.463)	(0.457)	(0.470)	
Panel C: Part 1 Math Score = Part 1 Verbal Score				
No PF & No Advice	0.229	0.125	0.333	.0860
	(0.425)	(0.338)	(0.482)	
PF & No Advice	0.180	0.227	0.143	.4406
	(0.388)	(0.429)	(0.356)	
No PF & Advice	0.273	0.220	0.315	.0858
	(0.446)	(0.416)	(0.466)	
PF & Advice	0.320	0.276	0.364	.1259
	(0.467)	(0.449)	(0.483)	
Panel D: Part 1 Math Score < Part 1 Verbal Score				
No PF & No Advice	0.206	0.0714	0.423	.0005
	(0.407)	(0.261)	(0.504)	
PF & No Advice	0.128	0.0714	0.194	.1052
	(0.336)	(0.261)	(0.401)	
No PF & Advice	0.0843	0.0598	0.110	.0854
	(0.278)	(0.238)	(0.314)	
PF & Advice	0.113	0.0926	0.144	.1312
	(0.317)	(0.291)	(0.352)	

PF means performance feedback. Standard deviations are reported in parentheses. Two-sided p-value for the equality of proportions across males and females are reported on Column (4).

4.2 Gender Gap in the Choice of Math

Before we evaluate how the interventions affect choice of math test rigorously, we investigate whether there is a gender gap in the choice of math and if so, how much of this gender gap can be explained by the initial test scores and other characteristics.

The analysis in this subsection focuses on individuals who were in the control group, that is they neither received advice nor received performance feedback. Table 4 shows the results. Column 1 reveals that there is a huge gap in math test choice between females and males. Only 11% of females chose the math test whereas 35% of males chose it despite the fact that 35% of females and 41% of males performed better in the part 1 math test compared to the part 1 verbal test. We can reject the equality of the means between females and males in the choice of math test (p-value is 0.002). Gender gap remains similar when we control for part 1 math and verbal scores linearly or by including dummy variables for each score combination as shown in Columns 2 and 3, respectively. In Column 2, we see that Part 1 math score is positively correlated with the choice of math test over verbal test as expected whereas there is no statistically significant correlation between Part 1 verbal score and the choice of the math test.

Table 4 Columns 4 and 5 repeat the analysis in Columns 2 and 3, but include the additional controls. Gender gap drops to 18.6 percentage points (pps) and 16.2 pps (p-values are 0.013 and 0.047) in Columns 4 and 5, respectively, when all the controls are included. In these regressions, we see that risk preferences correlated positively with the choice of math test. In line with the literature, females in our subject pool are more risk averse than males (two-sided t-test p-value: 0.000 and K-Smirnov p-value: 0.000). Hence, it looks like gender differences in risk preferences explain some part of the gender gap in math choice we observe in our data, but a substantial gender gap in math choice remains even after controlling part

1 test scores, demographics, perceptions about being better at math versus verbal tasks, and time and risk preferences.

4.3 Effect of Performance Feedback on the Choice of Math

In this subsection, we look at whether and how performance feedback affects the choice of math test over verbal test.

We restrict our sample to the subjects who did not receive any advice during the experiment and run the regressions of the form:

$$\begin{aligned} ChooseMath_{ij} = \alpha_0 + \alpha_1 PerformanceFeedback_i + \alpha_2 MathScore_i \\ + \alpha_3 VerbalScore_i + \alpha_4 Tournament_{ij} + \Omega X_i + \epsilon_{ij} \end{aligned} \tag{1}$$

where $ChooseMath_{ij}$ is equal to 1 if the individual i chose math test over verbal test under condition $j \in \{piece-rate, tournament\}$, $PerformanceFeedback_i$ is equal to 1 if the individual i learned their scores from part 1 before making their choice for part 2, $MathScore_i$ is math score of individual i from Part 1 (out of 4), $VerbalScore_i$ is verbal score of individual i from Part 1 (out of 4), $Tournament_{ij}$ is equal to 1 if the choice belongs to the tournament condition, and X_i is the set of controls. We estimate coefficient of interest, α_1, using a linear probability model and standard errors are clustered at the individual level.

Table 5 presents the results. Subjects who received performance feedback were 12.8 pps more likely to choose math test compared to the subjects who did not receive performance feedback as shown in column 1 (p-value: 0.020). This effect corresponds to a 56% increase in choosing math since only 22.6% of control subjects chose the math test. Effect size remains similar when we control for observables in column 2. When we repeat the analysis

Table 4: Correlates of Choice of Math Test

	(1)	(2)	(3)	(4)	(5)
Female	-0.240***	-0.226***	-0.208**	-0.186**	-0.162**
	(0.0756)	(0.0755)	(0.0826)	(0.0737)	(0.0803)
Part 1 Math Score		0.0657**		0.0679**	
		(0.0327)		(0.0332)	
Part 1 Verbal Score		0.0478		0.0362	
		(0.0313)		(0.0309)	
Tournament	0.0215	0.0215	0.0215	0.0215	0.0215
	(0.0406)	(0.0408)	(0.0428)	(0.0423)	(0.0444)
Asian				-0.0999	-0.0593
				(0.160)	(0.177)
Black				-0.0141	-0.0184
				(0.140)	(0.170)
Mixed				-0.0153	0.0632
				(0.159)	(0.197)
White				-0.0307	-0.0295
				(0.129)	(0.160)
Age				-0.00245	0.000590
				(0.0118)	(0.0124)
Student				0.111	0.157*
				(0.0821)	(0.0823)
College Educated Parents				-0.0130	-0.0224
				(0.0788)	(0.0918)
Risk Preferences				0.0373**	0.0288*
				(0.0175)	(0.0170)
Time Preferences				-0.0270	-0.0182
				(0.0176)	(0.0191)
Better at Math				0.162	0.187
				(0.114)	(0.144)
Better at Verbal				-0.0887	-0.109
				(0.0959)	(0.113)
Prolific Approvals (in 1000s)				0.0673	-0.00294
				(0.0778)	(0.0824)
Constant	0.342***	0.127	0.197**	0.0977	0.623
	(0.0695)	(0.0920)	(0.0883)	(0.327)	(0.416)
Controls	No	No	No	Yes	Yes
Finer Score Controls	No	No	Yes	No	Yes
Observations	186	186	186	186	186
Subjects	93	93	93	93	93

Sample is restricted to the subjects who did not receive performance feedback or advice. Dependent variable is equal to 1 if the individual i chose math test over verbal. Tournament is equal to 1 if the decision belongs to the tournament condition. Columns (3) and (5) includes dummy variables for each score combination from Part 1 instead of just linearly controlling for math and verbal scores. Columns (4) and (5) include all the controls that are listed in Table 1. Standard Errors are clustered at the individual level and are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

separately for females (columns 3-4) and males (columns 5-6), we see that all of the effect of performance feedback was driven by females. Female subjects who received performance feedback were 24 pps more likely to choose the math test compared to the 11.2% of female subjects in the control group (p-value<0.001). Performance feedback did not significantly affect male subjects' math test choice. We can reject the equality of the effect of performance feedback on choosing math test between females and males (p-value: .038, not reported in the table). We are also unable to reject that the gender gap is equal to zero for subjects who received performance feedback. Hence, receiving performance feedback closed the gender gap in choosing math by making females more likely to choose math.

Table 5: Performance Feedback and Choice of Math Test

Sample	All		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)
Performance Feedback	0.128**	0.137***	0.241***	0.226***	0.0187	0.0402
	(0.0549)	(0.0515)	(0.0641)	(0.0623)	(0.0877)	(0.0695)
Part 1 Math Score	0.138***	0.131***	0.136***	0.138***	0.123***	0.109***
	(0.0248)	(0.0249)	(0.0248)	(0.0279)	(0.0450)	(0.0335)
Part 1 Verbal Score	-0.0919***	-0.0791***	-0.121***	-0.124***	-0.0614*	-0.00806
	(0.0237)	(0.0229)	(0.0295)	(0.0295)	(0.0368)	(0.0266)
Tournament	0.0102	0.0102	0.0412	0.0412	-0.0200	-0.0200
	(0.0250)	(0.0255)	(0.0360)	(0.0372)	(0.0351)	(0.0362)
Constant	0.141**	0.155	0.0985	0.587*	0.227*	0.290
	(0.0696)	(0.263)	(0.0789)	(0.309)	(0.123)	(0.369)
Controls	No	Yes	No	Yes	No	Yes
Observations	394	394	194	194	200	200
Subjects	197	197	97	97	100	100

Sample is restricted to the subjects who did not receive advice. Dependent variable is equal to 1 if the individual i chose math test over verbal. Performance Feedback is equal to 1 if the individual learned their scores from part 1. Tournament is equal to 1 if the decision belongs to the tournament condition. Columns 1-2 include all genders and Columns 3-4 (5-6) include female (male) subjects. Odd columns do not include any additional controls and the even columns include the controls that are listed in Table 1. Standard Errors are clustered at the individual level and are in parentheses. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Previous work examining how absolute performance feedback influences academic performance does not find significant differences in responses by females and males (Azmat and

Iriberri, 2010). In our setting, we find absolute performance feedback increases females' probability of choosing math test without influencing males' test choice. One potential mechanism, which we are unable to test directly, is that performance feedback affects beliefs about one's own ability differently for males and females.[18] Table 3 Panels B-D provide some suggestive evidence in line with this mechanism. Looking at individuals who performed better at math than verbal in part 1, we see that only 16% of the females in the control group chose the math test whereas 85% of the females in the performance feedback only treatment did so and we see that 32% of the males in the control group chose the math test and 67% of the males did so in the performance feedback only treatment. Conversely, among the individuals who performed worse at math than verbal in part 1, 42% of the males in the control group chose the math test and 19% of the males did so in the performance feedback only treatment.

4.4 Effect of Advisors and Advice on the Choice of Math

In this subsection, we look at how receiving advice affected subjects' test choice for Part 2 across different advisor treatments.

We restrict our sample to the subjects who did not receive any performance feedback during the experiment and run the regressions of the form:

[18] Coffman, Collis and Kulkarni (2023) shows that individuals update their beliefs about own ability in a specific domain in response to noisy but informative feedback about one's own score in that domain. They also find men are significantly more responsive to information in male-typed domains, while women are significantly more responsive in female-typed domains. They focus on beliefs across tasks in different domains as their outcome variable and they do not collect data on choice of the task. Their subjects were randomly assigned to complete tasks in three different domains out of eight domains and their beliefs about own ability in that domain were elicited after each domain.

$$ChooseMath_{ij} = \alpha_0 + \alpha_1 Ungendered_i + \alpha_2 Male_i + \alpha_3 Female_i + \alpha_4 MathScore_i + \alpha_5 VerbalScore_i + \alpha_6 Tournament_{ij} + \Omega X_i + \epsilon_{ij} \quad (2)$$

where $ChooseMath_{ij}$ is equal to 1 if the individual i chose math test over verbal test under condition $j \in \{piece-rate, tournament\}$, $Ungendered_i$ is equal to 1 if the subject was assigned an ungendered advisor, $Male_i$ is equal to 1 if the subject was assigned a male advisor, $Female_i$ is equal to 1 if the subject was assigned a female advisor, $MathScore_i$ is math score of individual i from Part 1 (out of 4), $VerbalScore_i$ is verbal score of individual i from Part 1 (out of 4), $Tournament_{ij}$ is equal to 1 if the decision belongs to the tournament condition, and X_i is the set of controls. The baseline category is the subjects who did not receive any advice. We estimate coefficients of interest using a linear probability model and standard errors are clustered at the subject level.

Table 6 presents the results. There are a couple of things worth mentioning in this table. First, there is no statistically significant effect of advice treatments[19] on the math test choice, neither for the overall sample nor for the male and female subsamples. Second, treatment effects do not differ significantly between males and females; that is, gender gap in math test choice is not affected by advice. Third, treatment effects do not differ significantly based on advisor gender or the gender match between the advisee and advisor.

Since individuals' response to the advice treatments is likely to vary based on the type of advice they receive (math or verbal), we analyze the treatment effects by the type of advice in the analysis that follows. Crucially, the type of advice was orthogonal to the advisor gender.

[19]Overall, 55% of the students in the advice treatments were told to take the verbal test and the remaining were told to take the math test.

Table 6: Advice and Choice of Math Test

Sample	All		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)
Advice from:						
Ungendered Advisor	0.0499	0.0304	0.0719	0.0670	0.0271	-0.0135
	(0.0554)	(0.0499)	(0.0613)	(0.0568)	(0.0902)	(0.0822)
Male Advisor	0.0550	0.0340	0.0904*	0.0697	0.0157	-0.00852
	(0.0469)	(0.0437)	(0.0499)	(0.0498)	(0.0773)	(0.0715)
Female Advisor	0.00686	-0.00885	0.0615	0.0390	-0.0559	-0.0743
	(0.0475)	(0.0434)	(0.0507)	(0.0493)	(0.0786)	(0.0721)
Part 1 Math Score	0.118***	0.0946***	0.0710***	0.0534***	0.149***	0.126***
	(0.0148)	(0.0140)	(0.0172)	(0.0169)	(0.0216)	(0.0209)
Part 1 Verbal Score	-0.0838***	-0.0830***	-0.0959***	-0.0883***	-0.0690***	-0.0801***
	(0.0136)	(0.0129)	(0.0171)	(0.0173)	(0.0211)	(0.0200)
Tournament	-0.0101	-0.0101	-0.00685	-0.00685	-0.0131	-0.0131
	(0.0150)	(0.0151)	(0.0223)	(0.0226)	(0.0203)	(0.0205)
Constant	0.173***	0.446***	0.182***	0.655***	0.186**	0.234
	(0.0505)	(0.171)	(0.0559)	(0.213)	(0.0846)	(0.248)
P-values:						
ungendered=male	.914	.932	.755	.96	.87	.938
ungendered=female	.364	.353	.863	.6	.244	.35
male=female	.193	.211	.551	.487	.186	.206
ungendered=male=female	.392	.411	.834	.756	.328	.408
ungendered=male=female=0	.465	.585	.325	.506	.522	.554
Controls	No	Yes	No	Yes	No	Yes
Observations	1194	1194	584	584	610	610
Subjects	597	597	292	292	305	305

*Sample is restricted to the subjects who did not receive performance feedback. Dependent variable is equal to 1 if the individual i chose math test over verbal. Advice from Ungendered/Male/Female Advisor is equal to 1 if the advisor gender was unknown/male/female. Tournament is equal to 1 if the decision belongs to the tournament condition. Columns 1-2 include all genders and Columns 3-4 (5-6) include female (male) subjects. Odd columns do not include any additional controls and the even columns include the controls that are listed in Table 1. Standard Errors are clustered at the individual level and are in parentheses.** $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

As explained in Section 2, type of advice was only random for Category 1 subjects.[20] Hence, we now restrict our analysis to category 1 subjects. Once again, we also restrict our sample to the subjects who did not receive any performance feedback during the experiment. We

[20]Details can be found in Appendix Table 1.

run the regressions of the form:

$$\begin{aligned} ChooseMath_{ij} = \alpha_0 &+ \alpha_1 UngenderedVerbal_i + \alpha_2 MaleVerbal_i + \alpha_3 FemaleVerbal_i \\ &+ \alpha_4 UngenderedMath_i + \alpha_5 MaleMath_i + \alpha_6 FemaleMath_i \\ &+ \alpha_7 MathScore_i + \alpha_8 VerbalScore_i + \alpha_9 Tournament_{ij} + \Omega X_i + \epsilon_{ij} \end{aligned} \tag{3}$$

where $ChooseMath_{ij}$ is equal to 1 if the individual i chose math test over verbal test under condition $j \in \{piece{-}rate, tournament\}$, $UngenderedVerbal_i/MaleVerbal_i/FemaleVerbal_i$ is equal to 1 if an ungendered/male/female advisor recommended the subject to take the verbal test, respectively, $UngenderedMath_i/MaleMath_i/FemaleMath_i$ is equal to 1 if an ungendered/male/female advisor recommended the subject to take the math test, respectively, $MathScore_i$ is math score of individual i from Part 1 (out of 4), $VerbalScore_i$ is verbal score of individual i from Part 1 (out of 4), $Tournament_{ij}$ is equal to 1 if the decision belongs to the tournament condition, and X_i is the set of controls. The baseline category is the subjects who did not receive any advice. We estimate coefficients of interest using a linear probability model and standard errors are clustered at the subject level.

Table 7 presents the results. Being randomly advised to take the verbal test does not significantly reduce the likelihood of choosing the math test for the overall sample or the females compared to not receiving any advice, but it does for male subjects if the advice comes from a gendered advisor (p-values are less than 0.10 only in the regressions with controls). Treatment effects are significantly different between males and females if they are assigned a female advisor (p-value is .036 in the regression with controls); that is, gender gap in math test choices is reduced by receiving verbal advice if the advice comes from a female advisor. There is also some suggestive evidence of differences across advice treatments' effectiveness for the overall sample and females. For the overall sample, those who received

the random Verbal advice from a male or a female advisor are marginally less likely to choose math test compared to those who received the random Verbal advice from an ungendered advisor (p-values are less than 0.10 only in the regressions with controls). Furthermore, females who received the Verbal advice from a male advisor are less likely to choose math test compared to those who received Verbal advice from an ungendered advisor (p-value$<$0.10 only in the regression with controls).

Next, we look at the effect of being randomly advice to take the math test. Looking at the overall sample, receiving math advice in any advice treatment makes individuals more likely to choose math test compared to receiving no advice for the overall sample. The effects are between 21 pps and 23 pps in the regressions without controls and statistically significant at least at the 5% level whereas the effects are between 14 pps and 20 pps in the regressions with controls and statistically significant at the 5% level only for gendered advisor treatments. Effects are stronger for female subjects. They range between 25 pps and 40 pps in the regressions without controls and between 23 pps and 29 pps in the regressions with controls, all statistically significant at least at the 5% level. Effects for males subjects are smaller, sometimes negative, and never statistically significant. There is suggestive evidence that gender gap in math test choices is reduced by receiving math advice (p-values are .082, .109, and .155 in ungendered, male, and female advisor treatments, respectively). Finally, the advice treatments are not differentiable from each other neither for the overall sample nor for the male or female samples; but they are jointly differentiable from zero for the overall sample and female sample.

The effects discussed above are causal since both advice treatments and advice type were randomly assigned but these effects are only relevant for a selected group of individuals since most subjects in category 1 have similar math and verbal scores (see Appendix Table

Table 7: Type of Advice and Choice of Math Test-Category 1 Subjects

Sample	All		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)
Verbal Advice from:						
Ungendered Advisor	0.0514	0.0606	0.0671	0.131	0.146	-0.0353
	(0.0971)	(0.0940)	(0.0984)	(0.102)	(0.223)	(0.193)
Male Advisor	-0.0383	-0.0951	-0.0566	-0.0438	-0.0672	-0.206*
	(0.0725)	(0.0734)	(0.0770)	(0.0860)	(0.113)	(0.112)
Female Advisor	-0.0680	-0.0996	0.00931	0.0234	-0.159	-0.260**
	(0.0712)	(0.0717)	(0.0980)	(0.0903)	(0.107)	(0.108)
Math Advice From:						
Ungendered Advisor	0.234**	0.143	0.402**	0.286**	0.0859	-0.0284
	(0.115)	(0.0929)	(0.190)	(0.132)	(0.153)	(0.134)
Male Advisor	0.211***	0.162**	0.251**	0.230**	0.140	-0.00865
	(0.0795)	(0.0774)	(0.103)	(0.105)	(0.120)	(0.115)
Female Advisor	0.209**	0.201**	0.249**	0.276***	0.154	0.0484
	(0.0934)	(0.0824)	(0.119)	(0.0994)	(0.145)	(0.134)
Part 1 Math Score	0.116***	0.0720*	0.111**	0.0344	0.103*	0.0793
	(0.0406)	(0.0366)	(0.0510)	(0.0481)	(0.0602)	(0.0506)
Part 1 Verbal Score	-0.0918**	-0.0767**	-0.127**	-0.0841	-0.0448	-0.0575
	(0.0410)	(0.0386)	(0.0544)	(0.0518)	(0.0615)	(0.0587)
Tournament	-0.0146	-0.0146	0.0394	0.0394	-0.0612*	-0.0612*
	(0.0233)	(0.0236)	(0.0330)	(0.0338)	(0.0329)	(0.0336)
Constant	0.155**	0.270	0.102	0.933***	0.234**	-0.187
	(0.0676)	(0.230)	(0.0792)	(0.334)	(0.108)	(0.305)
P-values:						
Verbal Advice:						
ungendered=male	.338	.079	.197	.082	.33	.368
ungendered=female	.199	.071	.608	.302	.151	.225
male=female	.659	.946	.495	.446	.302	.564
ungendered=male=female	.437	.159	.415	.218	.249	.453
ungendered=male=female=0	.563	.174	.613	.372	.268	.091
Math Advice:						
ungendered=male	.848	.841	.451	.705	.715	.874
ungendered=female	.844	.562	.466	.939	.684	.588
male=female	.98	.636	.988	.68	.919	.644
ungendered=male=female	.978	.824	.733	.897	.911	.848
ungendered=male=female=0	.021	.065	.016	.017	.636	.954
Controls	No	Yes	No	Yes	No	Yes
Observations	548	548	254	254	294	294
Subjects	274	274	127	127	147	147

Sample is restricted to the subjects in Category 1 (see Appendix Table 1) who did not receive performance feedback. Dependent variable is equal to 1 if the individual i chose math test over verbal. Verbal/Math Advice from Ungendered/Male/Female Advisor is equal to 1 if the advisor gender was unknown/male/female and the advice received was Verbal/Math, respectively. Tournament is equal to 1 if the decision belongs to the tournament condition. Columns 1-2 include all genders and Columns 3-4 (5-6) include female (male) subjects. Odd columns do not include any additional controls and the even columns include the controls that are listed in Table 1. Standard Errors are clustered at the individual level and are in parentheses. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

1).[21] For the sake of completeness, Appendix Table 4 repeats the results for subjects in categories 2 and 3. Category 2 subjects (Columns 1-6) in advice treatments always received the verbal advice. For this group, we see that the advice, compared to not receiving any advice, statistically decreases the overall probability of choosing the math test (effects vary between 11pps and 19pps across different advisor treatments). The effects for males are stronger than females; hence, gender gap in choosing math is reduced upon receiving verbal advice (statistically significant at the 5% for all advisor treatments.) Category 3 subjects (Columns 7-12) in advice treatments always received the math advice. For this group, the effects are generally positive, indicating an increase in the probability of choosing the math test, but not statistically significant, potentially due to the smaller sample sizes.

4.5 Interactions between Performance Feedback and Advice Treatments

As the previous subsection documented, the effect of advisor treatments are generally not differentiable from each other. Hence, in this section, we combine the three advisor treatments into one and analyze how receiving performance feedback and advice affected the choice of math test and whether the two interventions acted as substitutes, complements, or neither.

We first run the regressions of the form:

[21] Appendix Table 3 conducts the analysis for subjects in all categories combined. That specification is not our preferred one since (i) the control group includes some subjects who would have received Math advice and some subjects who would have received Verbal advice if they were in advice treatments so it might not be an adequate comparison for the people who received Verbal advice or for the people who received Math advice and (ii) since advice is not random for Category 2 and 3 subjects, it wouldn't be appropriate to compare the coefficients for subjects who received Verbal advice versus Math advice.

$$ChooseMath_{ij} = \alpha_0 + \alpha_1 PerformanceFeedback_i + \alpha_2 Advice_i + \alpha_3 Advice_i * Feedback_i + \alpha_4 MathScore_i + \alpha_5 VerbalScore_i + \alpha_6 Tournament_{ij} + \Omega X_i + \epsilon_{ij} \quad (4)$$

where $ChooseMath_{ij}$ is equal to 1 if the individual i chose math test over verbal test under condition $j \in \{piece - rate, tournament\}$, $PerformanceFeedback_i$ is equal to 1 if the individual i learned their scores from part 1 before making their choice for part 2, $Advice_i$ is equal to 1 if the individual was in one of the advisor treatments, $Advice_i * Feedback_i$ is equal to 1 if the individual received performance feedback and was in one of the advisor treatments, $MathScore_i$ is math score of individual i from Part 1 (out of 4), $VerbalScore_i$ is verbal score of individual i from Part 1 (out of 4), $Tournament_{ij}$ is equal to 1 if the decision belongs to the tournament condition, and X_i is the set of controls. The baseline category is subjects who did not receive any advice or performance feedback. We estimate coefficients of interest using a linear probability model and standard errors are clustered at the subject level. If the performance feedback and advice are neither complements nor substitutes, we will expect α_3 to be not distinguishable from zero.

Table 8 presents the results. This table confirms the results discussed previously. Receiving performance feedback increased the probability of choosing the math test in the overall sample. The effect is stronger for female subjects and no effect is detected for male subjects. Receiving performance feedback closes the gender gap in the choice of math. Neither the overall sample nor the male and female subsamples experience a statistically significant effect of receiving advice on their choice of math test. We are also unable to reject α_3 is equal to zero which suggests that performance feedback and advice treatments are neither complements nor substitutes. We find this result interesting given that in our setting performance feedback

and advice are conveying similar information in most cases. Finally, receiving performance feedback with or without advice increases the probability of choosing the math test compared to being in the control group or just receiving advice for the overall sample and the female subsample as evidenced by the p-values in the table.

Table 8: Performance Feedback, Advice, and Choice of Math Test

Sample	All		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)
Performance Feedback	0.124**	0.115**	0.242***	0.228***	0.00674	0.00904
	(0.0553)	(0.0507)	(0.0632)	(0.0600)	(0.0896)	(0.0790)
Advice	0.0324	0.0185	0.0594	0.0453	-0.00691	-0.0217
	(0.0445)	(0.0407)	(0.0464)	(0.0453)	(0.0751)	(0.0670)
Feedback*Advice	-0.00699	0.00728	-0.0909	-0.0763	0.0837	0.0924
	(0.0600)	(0.0554)	(0.0702)	(0.0670)	(0.0963)	(0.0861)
Part 1 Math Score	0.150***	0.130***	0.139***	0.125***	0.156***	0.135***
	(0.00999)	(0.00958)	(0.0128)	(0.0126)	(0.0147)	(0.0140)
Part 1 Verbal Score	-0.128***	-0.119***	-0.139***	-0.136***	-0.114***	-0.104***
	(0.00901)	(0.00900)	(0.0116)	(0.0120)	(0.0138)	(0.0133)
Tournament	0.00250	0.00250	-3.70e-19	-3.70e-19	0.00500	0.00500
	(0.0110)	(0.0110)	(0.0159)	(0.0159)	(0.0152)	(0.0153)
Constant	0.189***	0.298**	0.150***	0.413***	0.243***	0.186
	(0.0466)	(0.116)	(0.0510)	(0.152)	(0.0788)	(0.169)
P-values:						
$\alpha_2 + \alpha_3 = 0$	.529	.5	.549	.538	.205	.198
$\alpha_1 + \alpha_3 = 0$	0	0	0	0	.011	.002
$\alpha_1 + \alpha_2 + \alpha_3 = 0$	.001	.001	0	0	.268	.236
Controls	No	Yes	No	Yes	No	Yes
Observations	2400	2400	1200	1200	1200	1200
Subjects	1200	1200	600	600	600	600

*Sample includes all subjects. Dependent variable is equal to 1 if the individual i chose math test over verbal. Performance Feedback is equal to 1 if the individual learned their scores from part 1. Advice is equal to 1 if the individual was assigned to one of the advice treatments. Feedback*Advice is equal to 1 if the individual received performance feedback and was assigned one of the advice treatments. Tournament is equal to 1 if the decision belongs to the tournament condition. Columns 1-2 include all genders and Columns 3-4 (5-6) include female (male) subjects. Odd columns do not include any additional controls and the even columns include the controls that are listed in Table 1. Standard Errors are clustered at the individual level and are in parentheses.* $\alpha_2 + \alpha_3 = 0$ *compares the effect of Performance Feedback treatment to the Performance Feedback and Advice treatment.* $\alpha_1 + \alpha_3 = 0$ *compares the effect of Advice treatment to the Performance Feedback and Advice treatment.* $\alpha_1 + \alpha_2 + \alpha_3 = 0$ *compares the effect of the Performance Feedback and Advice treatment to the control treatment.* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Since individuals' response to the advice treatments is likely to vary based on the type of advice they receive (math or verbal), we analyze the treatment effects by the type of advice in the analysis that follows. We restrict our analysis to category 1 subjects since type of advice was only random for these subjects. We run the regressions of the form:

$$\begin{aligned} ChooseMath_{ij} = \alpha_0 &+ \alpha_1 PerformanceFeedback_i + \alpha_2 AdviceVerbal_i + \alpha_3 AdviceMath_i \\ &+ \alpha_4 AdviceVerbal_i * Feedback_i + \alpha_5 AdviceMath_i * Feedback_i \\ &+ \alpha_6 MathScore_i + \alpha_7 VerbalScore_i + \alpha_8 Tournament_{ij} + \Omega X_i + \epsilon_{ij} \end{aligned} \tag{5}$$

where $ChooseMath_{ij}$ is equal to 1 if the individual i chose math test over verbal test under condition $j \in \{piece-rate, tournament\}$, $PerformanceFeedback_i$ is equal to 1 if the individual i learned their scores from part 1 before making their choice for part 2, $AdviceVerbal_i$ is equal to 1 if the advisor recommended the individual to take the verbal test, $AdviceMath_i$ is equal to 1 if the advisor recommended the individual to take the math test, $AdviceVerbal_i * Feedback_i$ is equal to 1 if the individual received performance feedback and was also recommended to take the verbal test, $AdviceMath_i * Feedback_i$ is equal to 1 if the individual received performance feedback and was also recommended to take the math test, $MathScore_i$ is math score of individual i from Part 1 (out of 4), $VerbalScore_i$ is verbal score of individual i from Part 1 (out of 4), $Tournament_{ij}$ is equal to 1 if the decision belongs to the tournament condition, and X_i is the set of controls. The baseline category is subjects who did not receive any advice or performance feedback. We estimate coefficients of interest using a linear probability model and standard errors are clustered at the subject level. If the performance feedback and advice are neither complements nor substitutes, we will expect α_4 and α_5 to be not distinguishable from zero.

Table 9 Columns 1 and 2 present the results for all subjects in Category 1 without and with controls, respectively. Receiving performance feedback does not statistically significantly affect individuals' probability of choosing math compared to the control group. This result is different than what we found for the all subjects in Table 8; but it is expected given that most Category 1 individuals have similar math and verbal scores in Part 1 test, hence performance feedback might not convey a clear message in these cases. Receiving random math advice only or in conjunction with the performance feedback result in individuals being more likely to choose the math test compared to the control group. We do not see any significant effects of receiving random verbal advice alone or with performance feedback on choosing math compared to the control group. We also do not detect any interactions between performance feedback and receiving Verbal advice (i.e., α_4 is not statistically significantly different from 0) or between performance feedback or receiving Math advice (i.e., α_5 is not statistically significantly different from 0).

Next, we compare the probability of choosing math across our treatments. Figure 2a presents the result for all subjects in Category 1, controlling for observables. Individuals who were randomly assigned to receive both Math Advice and performance feedback were *more* likely to choose math test compared to individuals who were randomly assigned to receive only performance feedback, who in turn were *more* likely to choose math test compared to individuals who were randomly assigned to receive only Verbal advice. Individuals who were randomly assigned to receive only Math advice were *more* likely to choose math test compared to individuals who were randomly assigned to receive verbal advice with performance feedback, who in turn were *more* likely to choose the math test compared to individuals who were randomly assigned to receive only Verbal advice (marginally significant).

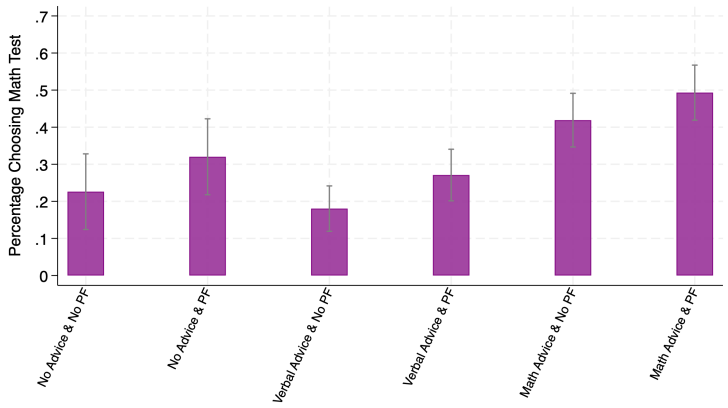
Table 9: Performance Feedback, Type of Advice, and Choice of Math Test-Category 1 Subjects

Sample	All		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)
Performance Feedback (α_1)	0.0991	0.0941	0.206*	0.248***	-0.00602	-0.0611
	(0.0835)	(0.0727)	(0.111)	(0.0947)	(0.126)	(0.109)
Verbal Advice (α_2)	-0.0279	-0.0456	0.00947	0.0527	-0.0817	-0.157
	(0.0633)	(0.0616)	(0.0707)	(0.0714)	(0.103)	(0.0966)
Math Advice (α_3)	0.217***	0.193***	0.277***	0.295***	0.139	0.0721
	(0.0688)	(0.0639)	(0.0845)	(0.0808)	(0.108)	(0.101)
Feedback*Verbal Advice (α_4)	-0.0131	-0.00363	-0.0557	-0.134	0.0356	0.128
	(0.0959)	(0.0871)	(0.130)	(0.119)	(0.143)	(0.130)
Feedback*Math Advice (α_5)	-0.0573	-0.0199	-0.202	-0.216*	0.0982	0.188
	(0.102)	(0.0907)	(0.138)	(0.120)	(0.150)	(0.136)
Part 1 Math Score	0.159***	0.123***	0.174***	0.136***	0.150***	0.110***
	(0.0244)	(0.0227)	(0.0343)	(0.0325)	(0.0359)	(0.0328)
Part 1 Verbal Score	-0.149***	-0.126***	-0.176***	-0.156***	-0.125***	-0.0979***
	(0.0251)	(0.0240)	(0.0359)	(0.0337)	(0.0363)	(0.0352)
Tournament	0.0129	0.0129	0.0428*	0.0428*	-0.0140	-0.0140
	(0.0174)	(0.0176)	(0.0250)	(0.0253)	(0.0244)	(0.0247)
Constant	0.148**	0.0971	0.0594	0.342	0.238**	-0.0629
	(0.0615)	(0.171)	(0.0733)	(0.261)	(0.0988)	(0.228)
P-values:						
Verbal Advice:						
$\alpha_2 + \alpha_4 = 0$	.572	.435	.676	.415	.643	.745
$\alpha_1 + \alpha_4 = 0$	.071	.062	.025	.103	.667	.347
$\alpha_1 + \alpha_2 + \alpha_4 = 0$	.371	.477	.041	.036	.615	.357
Math Advice:						
$\alpha_3 + \alpha_5 = 0$	.034	.008	.496	.42	.024	.005
$\alpha_1 + \alpha_5 = 0$	.471	.162	.959	.67	.257	.099
$\alpha_1 + \alpha_3 + \alpha_5 = 0$	0	0	0	0	.032	.049
Controls	No	Yes	No	Yes	No	Yes
Observations	1086	1086	514	514	572	572
Subjects	543	543	257	257	286	286

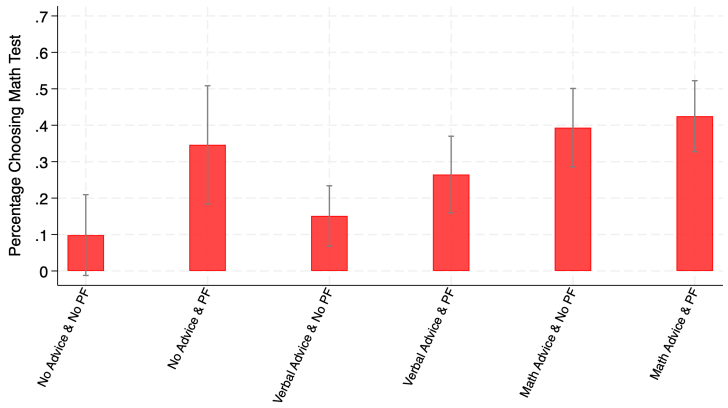
*Sample is restricted to subjects in Category 1 (see Appendix Table 1). Dependent variable is equal to 1 if the individual i chose math test over verbal. Performance Feedback is equal to 1 if the individual learned their scores from part 1. Verbal/Math Advice is equal to 1 if the advisor recommended the individual to take the Verbal/Math test. Feedback*Verbal/Math Advice is equal to 1 if the individual received performance feedback and was recommended to take the Verbal/Math test. Tournament is equal to 1 if the decision belongs to the tournament condition. Columns 1-2 include all genders and Columns 3-4 (5-6) include female (male) subjects. Odd columns do not include any additional controls and the even columns include the controls that are listed in Table 1. Standard Errors are clustered at the individual level and are in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Figure 2: Choice of Math across Treatments-Category 1 Subjects

(a) *All*



(b) *Female*



(c) *Male*

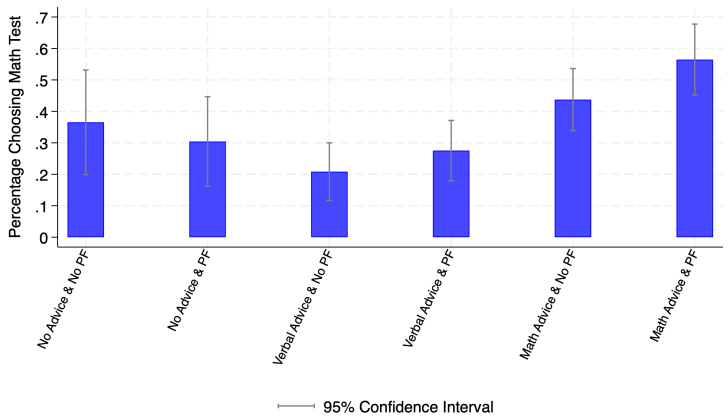


Figure depicts the percentage of individuals in each treatment choosing the math test over verbal test after controlling for observables. Sample is restricted to Category 1 subjects. Panel A includes both females and males, Panel B has females only, and Panel C has males only. PF stands for Performance Feedback. 95% Confidence Intervals are depicted in the figure.

Table 9 Columns 3 and 4 present the results for Category 1 female subjects without and with controls, respectively. Receiving performance feedback alone and receiving random math advice alone result in females being more likely to choose the math test compared to the control group. We do not see any significant effect of receiving random verbal advice alone on choosing math compared to the control group. Different than the results for the overall sample (Columns 1-2), we see that there is a potential interaction between receiving random Math advice and receiving performance feedback. The interaction coefficient is negative, large, and marginally significant in the specification with controls, suggesting that Performance Feedback and Math advice are substitutes to each other for females.

Next, we compare females' probability of choosing the math test across the treatments. Figure 2b presents the result. Females who received both random Math Advice and performance feedback were not distinguishable from females who received only performance feedback or who received only random Math Advice in terms of their math choice, but they were all *more* likely to choose math compared to the control group. Females who received both random Verbal Advice and performance feedback were *more* likely to choose math test compared to females who received only random Verbal advice or who were in the control group.

Table 9 Columns 5 and 6 present the results for Category 1 male subjects without and with controls, respectively. We do not detect any significant effects of receiving performance feedback alone or receiving random math advice alone or random verbal advice alone for male subjects. The interaction coefficients are positive but not statistically significant. Figure 2c compares males' probability of choosing the math test across the treatments. Males who received both random Math Advice and performance feedback were *more* likely to choose math test than the males who received performance feedback alone or the males in the control group. There is also suggestive evidence of that males who received both random

Math Advice and performance feedback were *more* likely to choose math test compared to males who received random Math advice alone.

Comparing the effects for males and females in a seemingly unrelated regression framework and using the specifications with controls, we can reject the equality of the effect for males and females of performance feedback alone (p-value<0.05), of Verbal advice alone or Math advice alone (p-values<0.10), and of performance feedback and Verbal advice combined (p-value<0.05) compared to the control group, but we cannot reject the equality of the effect of performance feedback and Math advice combined for males and females (p-value=0.13) These findings suggest that performance feedback and advice affect males and females differently.

The effects discussed above are causal since advice type was randomly assigned for category 1 subjects, but these effects are only relevant for category 1 subjects, a select group of individuals who mostly have similar math and verbal scores (see Appendix Table 1).[22] For the sake of completeness, Appendix Table 6 repeats the results for subjects in categories 2 and 3. Category 2 subjects (Columns 1-6) in advice treatments always received the verbal advice. For this group, we see that the advice statistically decreases the overall probability of choosing the math test by 15 pps compared to the control group. This effect is solely driven by males; hence, gender gap in choosing math is reduced upon receiving verbal advice (p-value<0.05). The interaction coefficient is positive (only marginally significant for overall sample and males in the specification with controls) which suggests performance feedback and verbal advice act as substitutes since the coefficients are negative for performance feedback alone and advice alone. Category 3 subjects (Columns 7-12) in advice treatments always received the math advice. For this group, the effect of performance feedback is positive, large, and statistically significant at the 1% level and the effect of advice is positive, smaller,

[22]Appendix Table 5 conducts the analysis for subjects in all categories combined. For the reasons discussed in Footnote 21, this is not our preferred specification.

and only marginally significant in some specifications. The interaction coefficient is negative (but mostly not significant) which suggests performance feedback and math advice might act as substitutes since the coefficients are positive for feedback alone and math advice alone.

5 Conclusion

In this paper, we study how performance feedback and advice influence the choice of math over verbal tasks in an online experiment. To accomplish this, we conducted two sets of randomizations in the experiment. First, participants completed a math and verbal assessment, after which they were randomly assigned to either learn their math and verbal performance or not. In a second randomization, students either did not received advice or received advice from a randomly assigned advisor (male, female, or ungendered) about which assessment to choose next. We then observe how these treatments influence subjects' choice of math or verbal assessment in the second part of the experiment.

Our findings reveal a substantial gender gap in the choice of math task which persists even after controlling for various observables. Receiving absolute performance feedback increases the likelihood of choosing the math task over verbal task and closes the gender gap in the choice of math task. Receiving advice does not significantly affect the math test choice or the gender gap in the math test choice. The effects do not differ based on advisor gender either. However, type of advice matters. Receiving random math advice, compared to no advice, significantly increases the propensity to choose math task, particularly for female students. Receiving random verbal advice, compared to no advice, has no detectable influence on the choice of math task for the overall sample, though there is suggestive evidence that Verbal advice reduces males' likelihood of choosing the math task, hence reducing the gender gap in the choice of math test. Our setting allows us to assess how performance feedback and

advice interventions interact to influence the choice of math, specifically assessing whether they act as substitutes or complements. Given that the advice is highly correlated with test scores in our setting, these two convey similar information. Overall, we do not find strong evidence of the interaction between the two interventions.

Our results contribute to the discussion on the gender gap in quantitative fields and STEM careers. Our experiment isolates the effect of advisor gender from other factors, showing that randomly assigned Math advice increases the selection of math task, *independent of the gender of the advisor*. This finding suggests that other mechanisms, such as role model effects or endogeneity of advice based on advisor and advisee gender, may play a significant role in field settings.

Our findings on performance feedback have implications for students' field choices, especially for the ones who are uncertain about their skills. The effectiveness of providing performance feedback in both math and verbal assessments underscores the value of comprehensive evaluation across multiple skill areas. Clear information about absolute performance could encourage students, particularly females, to pursue math-focused fields by reducing uncertainty about their abilities (Owen, 2023; Rury, 2022). If the policy maker's aim is to reduce gender gap in quantitative fields, our results suggest that providing performance feedback may be more effective than providing advice.

We expect our findings to generalize to some settings but not all settings. In this study, we examine the effects of immediate absolute performance feedback on choice of math and find strong effects. We do not know whether the results hold in settings when there is a delay between feedback and task choice (as in Coffman, Ugalde Araya and Zafar (2024)) or when relative performance feedback is provided instead of or in addition to absolute performance feedback. In this study, we investigate the role of advice and advisee-advisor gender in the

choice of math in a setting where there is no interaction between the advisor and the advisee. Hence, we do not expect the effects to be similar in environments where interactions are in-person or extensive. Finally, information provided through advice and performance feedback were mostly aligned in our setting. The effects may not generalize to the settings where they are complementary to each other.

References

Ahimbisibwe, Isaac, Adam Altjmed, Georgy Artemov, Andrés Barrios-Fernández, Aspasia Bizopoulou, Martti Kaila, Jin-Tan Liu, Rigissa Megalokonomou, José Montalbán, Christopher Neilson, Jintao Sun, Sebastián Otero, and Xiaoyang Ye, "The STEM Major Gender Gap: Evidence from Coordinated College Application Platforms Across Five Continents," 2024. Working Paper.

Azmat, Ghazala and Nagore Iriberri, "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment Using High School Students," *Journal of Public Economics*, 2010, *94* (7-8), 435–452.

_ , Manuel Bagues, Antonio Cabrales, and Nagore Iriberri, "What You Don't Know...Can't Hurt You? A Natural Field Experiment on Relative Performance Feedback in Higher Education," *Management Science*, 2019, *64* (8), 3714–3736.

Baier, Alexandra, Brent Davis, and Tarek Jaber-Lopez, "Gender, choice of task, and the effect of feedback on competition: An experiment," *Journal of Economic Psychology*, 2024, *103*.

Bandiera, Oriana, Valentino Larcinese, and Imran Rasul, "Blissful Ignorace? A Natural Experience on the Effect of Feedback on Students' Performance," *Labour Economics*, 2015, *34* (1), 13–25.

Berlin, Noémi and Marie-Pierre Dargnies, "Gender differences in reactions to feedback and willingness to compete," *Journal of Economic Behavior Organization*, 2016, *130*, 320–336.

Bobba, Matteo and Veronica Frisancho, "Self-perceptions about academic achievement: Evidence from Mexico City," *Journal of Econometrics*, 2022, *231* (1), 58–73. Annals Issue: Subjective Expectations Probabilities in Economics.

Bonaccio, Silvia and Reeshad S Dalal, "Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences," *Organizational Behavior and Human Decision Processes*, 2006, *101* (2), 127–151.

Borghans, Lex, Bart HH Golsteyn, and Anders Stenberg, "Does Expert Advice Improve Educational Choice?," *PLoS ONE*, 2015, *10* (12), e0145378.

Brandts, Jordi, Victoria Groenert, and Christoph Rott, "The impact of advice on women's and men's selection into competition," *Management Science*, 2015, *61* (5), 1018–1035.

Canaan, Serena and Pierre Mouganie, "The Impact of Advisor Gender on Female Students' STEM Enrollment and Persistence," *Journal of Human Resources*, 2023, *58* (2).

_ , Antoine Deeb, and Pierre Mouganie, "Adviser Value Added and Student Outcomes: Evidence from Randomly Assigned College Advisers," *American Economic Journal: Economic Policy*, 2022, *14* (4).

Carlana, Michela, "Implicit Stereotypes: Evidence from Teachers' Gender Bias," *The Quarterly Journal of Economics*, August 2019, *134* (3), 1163–1224.

Carrell, Scott E., Marianne Page, and Jim West, "Sex and Science: How Professor Gender Perpetuates the Gender Gap," *Quarterly Journal of Economics*, 2010, *125* (3).

Castagentti, Alessandro and Derek Rury, "Ego-Relevance, Information Avoidance and Performance Feedback: Evidence from a Field Experiment," *Working Paper*, 2024.

Coffman, Katherin, Manuela Collis, and Leena Kulkarni, "Stereotypes and Belief Updating," *Journal of the Economic Economic Association*, 2023.

Coffman, Katherine, M. Paz Ugalde Araya, and Basit Zafar, "A (dynamic) investigation of stereotypes, belief-updating, and behavior," *Economic Inquiry*, 2024, *62* (3), 957–983.

Coutts, Alexander, Boon Han Koh, and Zahra Murad, "The Signals We Give: Performance Feedback, Gender, and Competition," 2024. SSRN Working Paper 4635599.

Croson, Rachel and Uri Gneezy, "Gender Differences in Preferences," *Journal of Economic Perspectives*, 2009, *47*.

Delaney, Judith M. and Paul J. Devereux, "Understanding gender differences in STEM:Evidence from college applications," *Economics of Education Review*, 2019, *72*, 219–238.

_ **and** _ , "Gender Differences in Graduate Degree Choices," 2024. IZA Working Paper No.16918.

Dobrescu, L I, M Faravelli, R Megalokonomou, and A Motta, "Relative Performance Feedback in Education: Evidence from a Randomised Controlled Trial," *The Economic Journal*, 05 2021, *131* (640), 3145–3181.

Eriksson, Tor, Anders Poulsen, and Marie C. Villeval, "Feedback and Incentives: Experimental Evidence," *Labour Economics*, 2009, *16*, 679–688.

Ertac, Seda and Balázs Szentes, "The effect of information on gender differences in competitiveness: Experimental evidence," 2011. Working Paper.

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde, "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences," *Management Science*, 2022, *69* (4), 1935–1950.

Gallen, Yana and Melanie Wasserman, "Does information affect homophily?," *Journal of Public Economics*, 2023, *222*, 104876.

_ **and** _ , "Informed Choices: Gender Gaps in Career Advice," *Working Paper*, 2024.

Gentry, Melissa, Jonathan Meer, and Danila Serra, "Can High School Counselors Help the Economics Pipeline?," *AEA Papers and Proceedings*, 2023, *113*, 462–466.

Heikensten, Emma and Siri Isaksson, "Simon Says: Examining Gender Differences in Advice Seeking and Influence in the Lab," *SSRN Electronic Journal*, 2019.

Jeworrek, Sabrina, "Gender stereotypes still in MIND: Information on relative performance and competition entry," *Journal of Behavioral and Experimental Economics*, 2019, *82*, 101448.

Lordan, Grace and Warn Lekfuangfu, "Stephen versus Stephanie? Does Gender Matter for Peer-to-Peer Career Advice," IZA Discussion Paper 16161, Institute of Labor Economics (IZA) May 2023.

Mulhern, Christine, "Beyond Teachers: Estimating Individual School Counselor's Effects on Educational Attainment," *American Economic Review*, 2023, *113* (11), 2846–93.

Owen, Stephanie, "College field specialization and beliefs about relative performance: An experimental intervention to understand gender gaps in STEM," *Economics of Education Review*, 2023, *97*, 102479.

Patnaik, Arpita, Gwyn C. Pauley, Joanna Venator, and Matthew J. Wiswall, "The Impacts of Same and Opposite Gender Alumni Speakers on Interest in Economics," *NBER Working Paper*, 2023.

Peri, Giovanni, Kevin Shih, and Chad Sparber, "STEM Workers, H-1B Visas, and Productivity in US Cities," *Journal of Labor Economics*, 2015, *33* (S1), S225–S255.

Porter, Catherine and Danila Serra, "Gender Differences in the Choice of Major: The Importance of Role Models," *American Economic Journal: Applied Economics*, 2020, *12* (3).

Rury, Derek, "Tightening the Leaky Pipeline(s): The Role of Beliefs About Ability in STEM Major Choice," *Working Paper*, 2022.

Schotter, Andrew, "Decision Making with Naive Advice," *American Economic Review*, May 2003, *93* (2), 196–201.

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

Welsch, David M. and Matthew Winden, "Student gender, counselor gender, and college advice," *Education Economics*, 2018, *27* (2), 112–131.

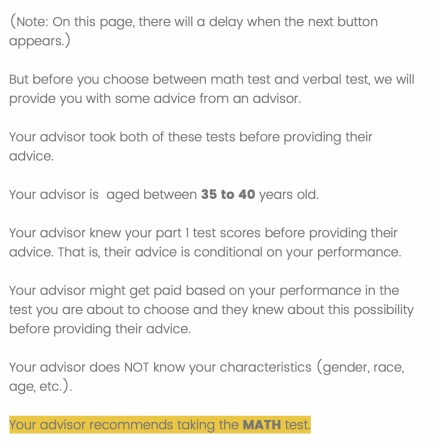
Wozniak, David, William T. Harbaugh, and Ulrich Mayr, "The menstrual cycle and performance feedback alter gender differences in competitive choices," *Journal of Labor Economics*, 2014, *32* (1), 161–198.

_ , _ , and _ , "The effect of feedback on gender differences in competitive choices," 2016. SSRN Working Paper.

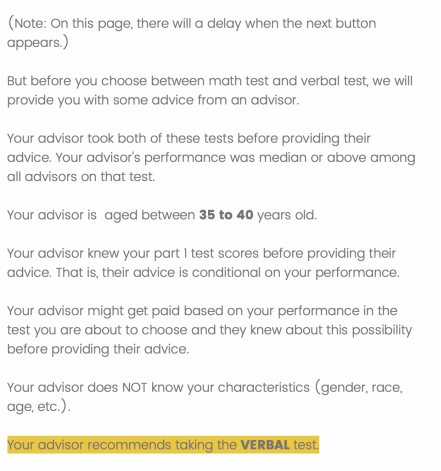
Appendix

A Figures

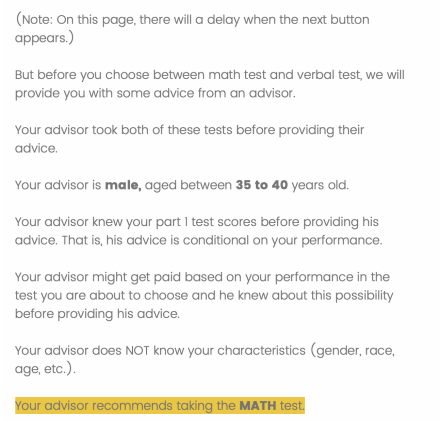
Figure 1: Treatment Variations: Advisor's Sex x Advisor's Knowledge



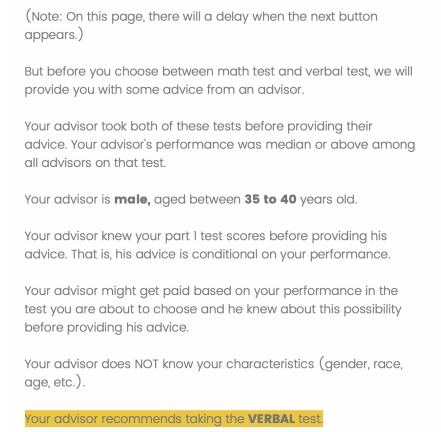
(a) *Advisor's Sex is unknown, Advisor's Score: Unknown*



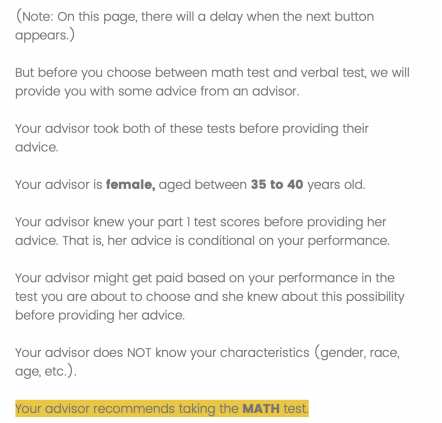
(b) *Advisor's Sex: Unknown, Advisor's Score: Above Median*



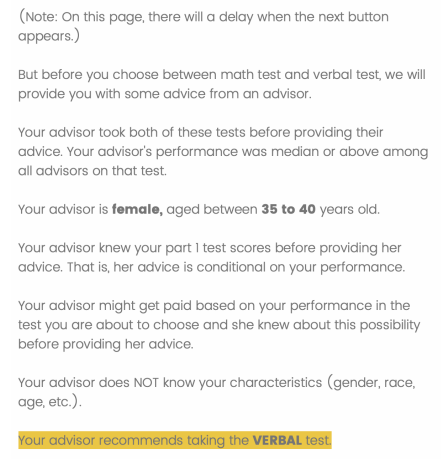
(c) *Advisor's Sex: Male, Advisor's Score: Unknown*



(d) *Advisor's Sex: Male, Advisor's Score: Above Median*

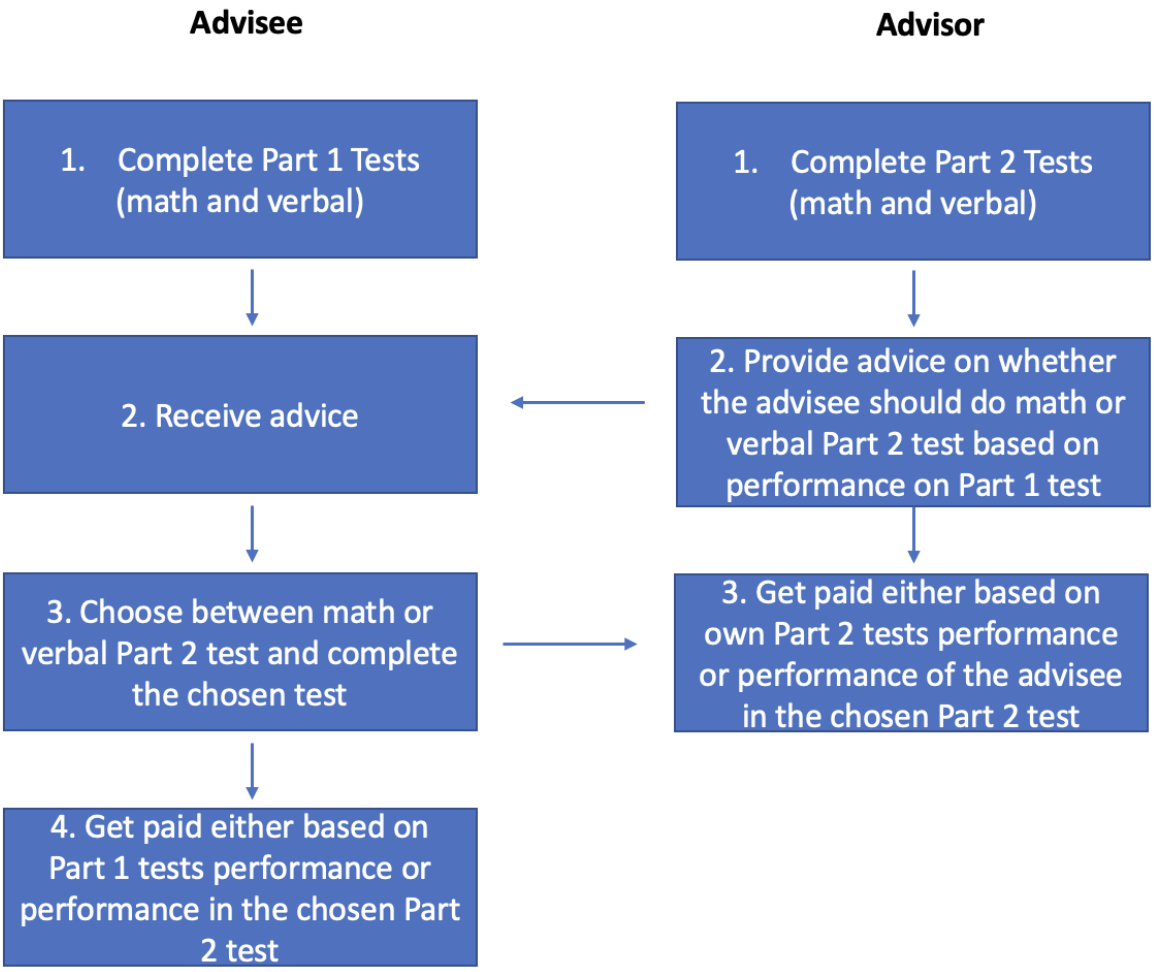


(e) *Advisor's Sex: Female, Advisor's Score: Unknown*



(f) *Advisor's Sex: Female, Advisor's Score: Above Median*

Figure 2: Relationship between Advisor and Advisee Experiment



B Tables

Table 1: Advice Categories

Math Score	Verbal Score	Advice Provided	Category
0	0	Math or Verbal (Randomly Assigned)	1
0	1	Math or Verbal (Randomly Assigned)	1
0	2	Verbal	2
0	3	Verbal	2
0	4	Verbal	2
1	0	Math or Verbal (Randomly Assigned)	1
1	1	Math or Verbal (Randomly Assigned)	1
1	2	Verbal	2
1	3	Verbal	2
1	4	Verbal	2
2	0	Math	3
2	1	Math or Verbal (Randomly Assigned)	1
2	2	Math or Verbal (Randomly Assigned)	1
2	3	Verbal	2
2	4	Verbal	2
3	0	Math	3
3	1	Math	3
3	2	Math	3
3	3	Math or Verbal (Randomly Assigned)	1
3	4	Verbal	2
4	0	Math or Verbal (Randomly Assigned)	1
4	1	Math	3
4	2	Math	3
4	3	Math or Verbal (Randomly Assigned)	1
4	4	Math or Verbal (Randomly Assigned)	1

This table shows the advice provided to the subjects based on their part 1 math and verbal scores. Subjects to whom the advice was randomly assigned are called Category 1 subjects, subjects who received Verbal advice are called Category 2 subjects, and subjects who received Math advice are called Category 3 subjects.

Table 2: Correlates of Part 1 Test Scores

	Part 1 Math Score		Part 1 Verbal Score	
	(1)	(2)	(3)	(4)
Female	-0.102*	-0.0939	0.135**	0.0569
	(0.0611)	(0.0627)	(0.0672)	(0.0690)
Asian		0.176		0.0934
		(0.147)		(0.163)
Black		-0.303**		-0.484***
		(0.143)		(0.157)
Mixed		-0.0544		-0.0458
		(0.158)		(0.166)
White		-0.0688		-0.0549
		(0.123)		(0.134)
Age		0.00471		0.0222**
		(0.0100)		(0.0105)
Student		0.118*		0.0805
		(0.0683)		(0.0735)
College Educated Parents		0.189***		0.299***
		(0.0632)		(0.0677)
Risk Preferences		-0.0526***		-0.0882***
		(0.0153)		(0.0171)
Time Preferences		0.0404**		0.0573***
		(0.0160)		(0.0172)
Better at Math		0.100		-0.00132
		(0.0864)		(0.0928)
Better at Verbal		-0.259***		0.0992
		(0.0687)		(0.0763)
Prolific Approvals (in 1000s)		0.0948*		-0.0995
		(0.0572)		(0.0656)
Constant	1.907***	1.786***	1.755***	1.260***
	(0.0443)	(0.308)	(0.0472)	(0.336)
Subjects	1200	1200	1200	1200

Dependent variable is the part 1 math test score in columns 1 and 2 and part 1 verbal test score in columns 3 and 4. Other is the omitted race category and "Equal at Math and Verbal" is the omitted perception category. Robust standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Type of Advice and Choice of Math Test

Sample	All		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)
Verbal Advice from:						
Ungendered Advisor	-0.0495	-0.0571	0.0142	0.0268	-0.110	-0.195*
	(0.0571)	(0.0544)	(0.0592)	(0.0587)	(0.106)	(0.100)
Male Advisor	-0.0521	-0.0711	-0.00544	-0.0228	-0.110	-0.117
	(0.0478)	(0.0449)	(0.0502)	(0.0503)	(0.0795)	(0.0749)
Female Advisor	-0.0914*	-0.110**	-0.00762	-0.0451	-0.183**	-0.180**
	(0.0469)	(0.0438)	(0.0508)	(0.0497)	(0.0778)	(0.0745)
Math Advice From:						
Ungendered Advisor	0.188**	0.147**	0.230*	0.172*	0.123	0.117
	(0.0820)	(0.0713)	(0.121)	(0.104)	(0.111)	(0.0971)
Male Advisor	0.174***	0.148***	0.209***	0.181***	0.131	0.100
	(0.0588)	(0.0546)	(0.0699)	(0.0685)	(0.0913)	(0.0848)
Female Advisor	0.160**	0.141**	0.184**	0.180**	0.118	0.0656
	(0.0676)	(0.0592)	(0.0847)	(0.0744)	(0.103)	(0.0946)
Part 1 Math Score	0.0678***	0.0472***	0.0319	0.0154	0.0901***	0.0728***
	(0.0172)	(0.0158)	(0.0198)	(0.0187)	(0.0266)	(0.0246)
Part 1 Verbal Score	-0.0384**	-0.0403***	-0.0562***	-0.0508***	-0.0200	-0.0305
	(0.0149)	(0.0141)	(0.0199)	(0.0192)	(0.0234)	(0.0225)
Tournament	-0.0101	-0.0101	-0.00685	-0.00685	-0.0131	-0.0131
	(0.0150)	(0.0151)	(0.0224)	(0.0226)	(0.0204)	(0.0206)
Constant	0.179***	0.444***	0.171***	0.617***	0.216***	0.245
	(0.0486)	(0.167)	(0.0539)	(0.214)	(0.0821)	(0.240)
P-values:						
Verbal Advice:						
ungendered=male	.958	.774	.734	.392	1	.348
ungendered=female	.378	.266	.706	.196	.39	.861
male=female	.285	.281	.962	.619	.185	.281
ungendered=male=female	.487	.413	.925	.433	.368	.461
ungendered=male=female=0	.26	.092	.984	.595	.113	.093
Math Advice:						
ungendered=male	.868	.992	.864	.933	.938	.86
ungendered=female	.745	.935	.722	.942	.965	.611
male=female	.826	.908	.777	.991	.889	.696
ungendered=male=female	.946	.993	.928	.996	.99	.868
ungendered=male=female=0	.01	.023	.01	.023	.482	.584
Controls	No	Yes	No	Yes	No	Yes
Observations	1194	1194	584	584	610	610
Subjects	597	597	292	292	305	305

Sample is restricted to all subjects who didn't receive performance feedback. Dependent variable is equal to 1 if the individual i chose math test over verbal. Verbal/Math Advice from Ungendered/Male/Female Advisor is equal to 1 if the advisor gender was unknown/male/female and the advice received was Verbal/Math. Tournament is equal to 1 if the decision belongs to the tournament condition. Columns 1-2 include all genders and Columns 3-4 (5-6) include female (male) subjects. Odd columns do not include any additional controls and the even columns include the controls that are listed in Table 1. Standard Errors are clustered at the individual level and are in parentheses. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Type of Advice and Choice of Math Test-Category 2 and 3 subjects

	Category 2: Verbal Advice						Category 3: Math Advice					
Sample	All		Female		Male		All		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Advice from:												
Ungendered Advisor	-0.184***	-0.188***	-0.184***	-0.0503	-0.369***	-0.432***	0.220	0.142	0.154	-0.106	0.273	0.272
	(0.0686)	(0.0636)	(0.0686)	(0.0703)	(0.124)	(0.105)	(0.138)	(0.141)	(0.148)	(0.139)	(0.192)	(0.195)
Male Advisor	-0.119	-0.115*	-0.119	-0.0128	-0.279**	-0.312***	0.231**	0.173	0.253**	0.0580	0.245	0.163
	(0.0727)	(0.0646)	(0.0727)	(0.0700)	(0.130)	(0.106)	(0.115)	(0.120)	(0.101)	(0.116)	(0.179)	(0.170)
Female Advisor	-0.165**	-0.176***	-0.165**	-0.0751	-0.331**	-0.363***	0.187	0.0900	0.121	-0.104	0.206	0.0433
	(0.0693)	(0.0611)	(0.0693)	(0.0670)	(0.130)	(0.103)	(0.120)	(0.129)	(0.125)	(0.154)	(0.180)	(0.179)
Part 1 Math Score	0.0653**	0.0621**	0.0653**	0.0564**	0.0336	0.0733	0.175**	0.149*	0.0878	0.0431	0.165	0.141
	(0.0255)	(0.0256)	(0.0255)	(0.0257)	(0.0420)	(0.0454)	(0.0841)	(0.0771)	(0.114)	(0.115)	(0.121)	(0.104)
Part 1 Verbal Score	-0.0129	-0.00569	-0.0129	-0.00799	0.0122	-0.0184	-0.0855	-0.113**	-0.159**	-0.184***	-0.000519	-0.00497
	(0.0253)	(0.0251)	(0.0253)	(0.0201)	(0.0522)	(0.0507)	(0.0567)	(0.0485)	(0.0656)	(0.0672)	(0.0871)	(0.0674)
Tournament	0.0312	0.0312	0.0312	0	0.0674*	0.0674*	-0.0611	-0.0611	-0.113	-0.113	-0.0145	-0.0145
	(0.0209)	(0.0213)	(0.0209)	(0.0250)	(0.0356)	(0.0369)	(0.0376)	(0.0386)	(0.0711)	(0.0751)	(0.0333)	(0.0350)
Constant	0.162*	0.270	0.162*	-0.0507	0.304	0.434	-0.122	0.635	0.130	1.002	-0.138	0.421
	(0.0957)	(0.214)	(0.0957)	(0.177)	(0.196)	(0.336)	(0.222)	(0.517)	(0.294)	(0.720)	(0.307)	(0.652)
P-values:												
ungendered=male	.124	.109	.124	.496	.163	.101	.924	.754	.528	.158	.864	.426
ungendered=female	.596	.756	.596	.547	.494	.25	.787	.62	.837	.991	.694	.075
male=female	.279	.15	.279	.145	.447	.504	.652	.348	.316	.255	.783	.243
ungendered=male=female	.302	.233	.302	.318	.367	.215	.901	.642	.582	.285	.919	.165
ungendered=male=female=0	.038	.015	.038	.422	.027	.001	.234	.485	.106	.469	.497	.22
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	384	384	384	206	178	178	262	262	124	124	138	138
Subjects	192	192	192	103	89	89	131	131	62	62	69	69

Sample is restricted to individuals who did not receive performance feedback. Sample in Columns 1-6 is Category 2 subjects for whom the advice received was verbal. Sample in Columns 7-12 is Category 3 subjects for whom the advice received was math. Dependent variable is equal to 1 if the individual i chose math test over verbal. Advice from Ungendered/Male/Female Advisor is equal to 1 if the advisor gender was unknown/male/female. Tournament is equal to 1 if the decision belongs to the tournament condition. Columns 1-2 and 7-8 include all genders in their respective categories. Columns 3-4 and 9-10 include female subjects in their respective categories. Columns 5-6 and 11-12 include male subjects in their respective categories. Odd columns do not include any additional controls and the even columns include the controls that are listed in Table 1. Standard Errors are clustered at the individual level and are in parentheses. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Performance Feedback, Type of Advice, and Choice of Math Test

Sample	All		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)
Performance Feedback	0.131**	0.122**	0.244***	0.231***	0.0180	0.0195
	(0.0549)	(0.0505)	(0.0636)	(0.0609)	(0.0878)	(0.0778)
Verbal Advice	-0.0348	-0.0512	0.0288	0.0170	-0.113	-0.132**
	(0.0434)	(0.0401)	(0.0444)	(0.0437)	(0.0738)	(0.0670)
Math Advice	0.123**	0.110**	0.121**	0.108*	0.0999	0.0883
	(0.0508)	(0.0458)	(0.0591)	(0.0557)	(0.0804)	(0.0717)
Feedback*Verbal Advice (α_4)	-0.0725	-0.0567	-0.159**	-0.150**	0.0172	0.0339
	(0.0602)	(0.0564)	(0.0709)	(0.0685)	(0.0965)	(0.0881)
Feedback*Math Advice (α_5)	0.0550	0.0699	-0.0101	0.00902	0.137	0.140
	(0.0680)	(0.0624)	(0.0846)	(0.0801)	(0.104)	(0.0930)
Part 1 Math Score	0.109***	0.0875***	0.106***	0.0911***	0.106***	0.0845***
	(0.0115)	(0.0108)	(0.0149)	(0.0144)	(0.0172)	(0.0158)
Part 1 Verbal Score	-0.0869***	-0.0790***	-0.109***	-0.105***	-0.0635***	-0.0547***
	(0.0103)	(0.0102)	(0.0136)	(0.0139)	(0.0156)	(0.0147)
Tournament	0.00250	0.00250	0.00000	0.00000	0.00500	0.00500
	(0.0110)	(0.0110)	(0.0159)	(0.0160)	(0.0152)	(0.0153)
Constant	0.189***	0.279**	0.146***	0.380**	0.253***	0.191
	(0.0446)	(0.113)	(0.0484)	(0.150)	(0.0751)	(0.160)
P-values:						
Verbal Advice:						
Advice+Feedback*Advice=0	.012	.009	.021	.016	.139	.104
Feedback+Feedback*Advice=0	.02	.009	.007	.01	.38	.186
Feedback+Advice+Feedback*Advice=0	.59	.73	.013	.03	.3	.247
Math Advice:						
Advice+Feedback*Advice=0	0	0	.099	.071	.001	0
Feedback+Feedback*Advice=0	0	0	0	0	.006	.002
Feedback+Advice+Feedback*Advice=0	0	0	0	0	.001	.001
Controls	No	Yes	No	Yes	No	Yes
Observations	2400	2400	1200	1200	1200	1200
Subjects	1200	1200	600	600	600	600

*Sample includes all subjects. Dependent variable is equal to 1 if the individual i chose math test over verbal. Performance Feedback is equal to 1 if the individual learned their scores from part 1. Verbal/Math Advice is equal to 1 if the advisor recommended the individual to take the Verbal/Math test. Feedback*Verbal/Math Advice is equal to 1 if the individual received performance feedback and was recommended to take the Verbal/Math test. Tournament is equal to 1 if the decision belongs to the tournament condition. Columns 1-2 include all genders and Columns 3-4 (5-6) include female (male) subjects. Odd columns do not include any additional controls and the even columns include the controls that are listed in Table 1. Standard Errors are clustered at the individual level and are in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

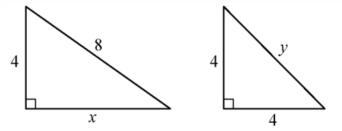
Table 6: Performance Feedback, Type of Advice, and Choice of Math Test-Category 2 and 3 Subjects

	Category 2: Verbal Advice						Category 3: Math Advice					
Sample	All		Female		Male		All		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Feedback	-0.0880	-0.0899	-0.0173	-0.0412	-0.211	-0.178	0.524***	0.431***	0.736***	0.575***	0.381**	0.276
	(0.0787)	(0.0723)	(0.0729)	(0.0749)	(0.144)	(0.117)	(0.122)	(0.121)	(0.0947)	(0.121)	(0.179)	(0.174)
Advice	-0.149**	-0.152***	-0.0493	-0.0483	-0.304**	-0.324***	0.229**	0.144	0.198**	0.0504	0.266*	0.205
	(0.0662)	(0.0571)	(0.0600)	(0.0588)	(0.122)	(0.0971)	(0.104)	(0.101)	(0.0889)	(0.107)	(0.153)	(0.147)
Feedback*Advice	0.120	0.136*	0.0516	0.0799	0.250	0.251*	-0.185	-0.102	-0.252**	-0.0878	-0.161	-0.0878
	(0.0830)	(0.0779)	(0.0793)	(0.0795)	(0.152)	(0.128)	(0.133)	(0.130)	(0.111)	(0.130)	(0.198)	(0.187)
Part 1 Math Score	0.0645***	0.0634***	0.0601***	0.0506***	0.0586**	0.0858***	0.0980*	0.0829	0.0507	0.0174	0.0846	0.0724
	(0.0161)	(0.0157)	(0.0172)	(0.0158)	(0.0291)	(0.0318)	(0.0576)	(0.0511)	(0.0689)	(0.0630)	(0.0853)	(0.0648)
Part 1 Verbal Score	-0.0201	-0.0167	-0.00547	-0.000141	-0.0411	-0.0385	-0.0501	-0.0649**	-0.116***	-0.121***	0.0326	-0.00575
	(0.0200)	(0.0191)	(0.0182)	(0.0175)	(0.0404)	(0.0395)	(0.0358)	(0.0320)	(0.0400)	(0.0389)	(0.0575)	(0.0426)
Tournament	0.0254	0.0254	0.00450	0.00450	0.0523*	0.0523*	-0.0532**	-0.0532**	-0.0992**	-0.0992**	-0.0141	-0.0141
	(0.0169)	(0.0170)	(0.0208)	(0.0211)	(0.0279)	(0.0284)	(0.0240)	(0.0243)	(0.0435)	(0.0447)	(0.0247)	(0.0253)
Constant	0.185**	0.407***	0.0370	0.102	0.421**	0.661***	0.0469	0.343	0.180	0.777*	0.0426	0.127
	(0.0821)	(0.147)	(0.0661)	(0.143)	(0.162)	(0.237)	(0.169)	(0.323)	(0.189)	(0.436)	(0.237)	(0.382)
P-values:												
$\alpha_2 + \alpha_3 = 0$	.571	.756	.967	.593	.541	.399	.622	.612	.422	.607	.431	.316
$\alpha_1 + \alpha_3 = 0$	.229	.076	.249	.165	.416	.148	0	0	0	0	.005	.002
$\alpha_1 + \alpha_2 + \alpha_3 = 0$	.082	.06	.806	.865	.035	.012	0	0	0	0	.001	.006
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	788	788	444	444	344	344	526	526	242	242	284	284
Subjects	394	394	222	222	172	172	263	263	121	121	142	142

*Sample in Columns 1-6 is Category 2 subjects for whom the advice received was verbal. Sample in Columns 7-12 is Category 3 subjects for whom the advice received was math. Dependent variable is equal to 1 if the individual i chose math test over verbal. Feedback is equal to 1 if the individual learned their scores from part 1. Advice is equal to 1 if the individual was assigned to one of the advice treatments. Feedback*Advice is equal to 1 if the individual received performance feedback and was assigned one of the advice treatments. Tournament is equal to 1 if the decision belongs to the tournament condition. Columns 1-2 and 7-8 include all genders in their respective categories. Columns 3-4 and 9-10 include female subjects in their respective categories. Columns 5-6 and 11-12 include male subjects in their respective categories. Odd columns do not include any additional controls and the even columns include the controls that are listed in Table 1. Standard Errors are clustered at the individual level and are in parentheses.* $\alpha_2 + \alpha_3 = 0$ *compares the effect of Performance Feedback treatment to the Performance Feedback and Advice treatment.* $\alpha_1 + \alpha_3 = 0$ *compares the effect of Advice treatment to the Performance Feedback and Advice treatment.* $\alpha_1 + \alpha_2 + \alpha_3 = 0$ *compares the effect of control treatment to the Performance Feedback and Advice treatment.* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Test Questions

Figure 3: Part 1 Math Questions



Which one of the following is true?

- ○ x is larger
- ○ y is larger
- ○ x and y are equal
- ○ the relationship between x and y cannot be determined.

In triangle ABC, the measure of angle A is 25° and the measure of angle B is greater than 90° but not greater than 100. Which of the following could be the measure of angle C?

- ○ 50
- ○ 60
- ○ 70
- ○ 80
- ○ 90

Distribution of Enrollment by Class and Gender
Total Enrollment: 1,400

Class	Males	Females
Freshmen	303	259
Sophomores	215	109
Juniors	182	88
Seniors	160	84
Total	860	540

Given the data, how many students are either juniors or males or both?

- ○ 678
- ○ 766
- ○ 948
- ○ 1,130
- ○ 1,312

The system of equations 7x + 3y = 12, and 3x + 7y = 6 is given. If x and y satisfy the system of equations given, what is the value of x-y?

- ○ 0.15
- ○ 0.67
- ○ 1
- ○ 1.5
- ○ 1.65

Figure 4: Part 1 Verbal Questions

Select the two answer choices that, when used to complete the sentence, fit the meaning of the sentence as a whole and produce completed sentences that are alike in meaning.

In medieval philosophy every physical phenomenon is presumed to have some determinate cause, leaving no place for _____ in the explanation of particular events.

- ☐ happenstance
- ☐ chance
- ☐ error
- ☐ experience
- ☐ context
- ☐ miscalculation

Please fill in the blank in the following sentence:

This filmmaker is not outspoken on political matters: her films are known for their aesthetic qualities rather than for their _____ ones.

- ○ dramatic
- ○ polemical
- ○ narrative
- ○ commercial
- ○ cinematic

Extensive housing construction is underway in Pataska Forest, the habitat of a large population of deer. Because deer feed at the edges of forests, these deer will be attracted to the spaces alongside the new roads being cut through Pataska Forest to serve the new residential areas. Consequently, once the housing is occupied, the annual number of the forest's deer hit by cars will be much higher than before construction started.

Which of the following is an assumption on which the argument depends?

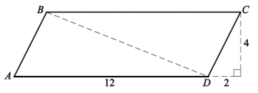
- ○ The development will leave sufficient forest to sustain a significant population of deer.
- ○ The number of deer hit by commercial vehicles will not increase significantly when the housing is occupied.
- ○ Deer will be as attracted to the forest edge around new houses as to the forest edge alongside roads.
- ○ No deer hunting will be allowed in Pataska Forest when the housing is occupied.
- ○ In years past, the annual number of deer that have been hit by cars on existing roads through Pataska Forest has been very low.

Please fill in the blank in the following sentence:

In the 1950's, the country's inhabitants were_____: most of them knew very little about foreign countries.

- ○ imperturbable
- ○ insular
- ○ erudite
- ○ cosmopolitan
- ○ partisan

Figure 5: Part 2 Math Questions



What is the area of parallelogram ABCD?

- ○ 32
- ○ 40
- ○ 48
- ○ 56
- ○ 64

The average (arithmetic mean) of the 11 numbers in a list is 11. If the average of 9 of the numbers in the list is 9, what is the average of the other 2 numbers?

- ○ 9
- ○ 12
- ○ 15
- ○ 18
- ○ 21

It is given that a is between 6 and 7, and b is equal to 9.
Quantity x: a/b
Quantity y: 0.85
From the answer choices given, select and indicate the one that describes the relationship between Quantity x and Quantity y.

- ○ x is larger
- ○ y is larger
- ○ x and y are equal
- ○ the relationship between x and y cannot be determined.

The system of equations $3x - y = -5$, and $x + 2y = 3$ is given. If x and y satisfy the system of equations given, what is the value of x+y?

- ○ -2
- ○ -1
- ○ 0
- ○ 1
- ○ 2

Figure 6: Part 2 Verbal Questions

Please fill in the blank in the following sentence:
Most spacecraft are still at little risk of collision with space debris during their operational lifetimes, but given the numbers of new satellites launched each year, the orbital environment in the future is likely to be less _____ .

- ○ protected
- ○ polluted
- ○ benign
- ○ invulnerable
- ○ crowded

Select the two answer choices that, when used to complete the sentence, fit the meaning of the sentence as a whole and produce completed sentences that are alike in meaning.
The detective's conviction that there were few inept crimes in her district led her to impute some degree of _____ to every suspect she studied.

- ☐ deceit
- ☐ acume
- ☐ duplicity
- ☐ shrewdness
- ☐ evasiveness
- ☐ equivocation

Computers cannot accurately predict climate change unless the mathematical equations fed into them adequately capture the natural meteorological processes they are intended to simulate. Moreover, there are processes that influence climate, such as modifications in land use, that scientists do not know how to simulate. The failure to incorporate such a process into a computer climate model can lead the model astray because a small initial effect can initiate a feedback cycle: a perturbation in one variable modifies a second variable, which in turn amplifies the original disturbance. An increase in temperature, for example, can boost the moisture content of the atmosphere, which then causes further warming because water vapor is a greenhouse gas.
In the context in which it appears, "amplifies" most nearly means

- ○ explicates
- ○ expatiates
- ○ adds detail to
- ○ exacerbates
- ○ makes louder

Please select the word that best completes the following sentence.
James Boswell's "Life of Samuel Johnson" is generally thought to have established Boswell as the first great modern biographer; yet the claim of _____ could be made for Johnson himself as author of a life of Richard Savage.

- ○ opportunism
- ○ perseverance
- ○ partisanship
- ○ precedence
- ○ omniscience

D Details about the Advisor Experiment

To create the advice that will be provided in the main experiment, we conducted another study first. We recruited 200 participants (100 male and 100 female) from Prolific with the following criteria: aged between 35-40, approval rate is between 95-100% and the number of previous submissions on prolific is between 50 to 2000.

First, advisors completed 4 math questions and 4 verbal questions which belonged to the Part 2 Test for the advisees in the main experiment. Then, they were shown five randomly chosen scenarios out of 25 possible score combinations (combinations of possible correct answers in math and verbal on a test that is similar to the one they took) in a random order and were asked to provide advice to other participants on whether they should choose the math test or the verbal test for Part 2. Advisors did not know the gender of the advisee when providing this advice. Advisors know that this is not a hypothetical situation and their advice will be passed on a real participant. They also knew that the person that they are advising will be randomly assigned to a piece-rate condition or a tournament condition for their second test, with equal chance.

Advisors received $2 for completing the study and a bonus payment between $0 and $2 depending on either (i) the advisor's number of correct answers in their own test or (ii) the number of correct answers that the advisee has on the part 2 test. Computer randomly picks whether their bonus payment was calculated based on (i) or (ii).

Our pre-analysis plan states that the main goal of this experiment is to create the advice that will provided in the advisee experiment and we will also explore whether the gender of the advisor and whether the advisor knows about their own performance affects the advice they provide. Appendix Table 7 presents these results. Female advisors are more likely to advice math whereas there is no detectable effect of neither learning own scores

before providing advice nor the interaction between the advisor gender and receiving the performance feedback.

Table 7: Correlates of Math Advice

	(1)	(2)	(3)	(4)
Female Advisor	0.0382	0.0532**	0.0652*	0.0697*
	(0.0235)	(0.0262)	(0.0344)	(0.0400)
Performance Feedback	0.0214	0.0344	0.0460	0.0484
	(0.0241)	(0.0229)	(0.0366)	(0.0346)
Female Advisor*Performance Feedback			-0.0495	-0.0293
			(0.0484)	(0.0514)
Advisee Math Score	0.0314**	0.0315**	0.0313*	0.0316**
	(0.0158)	(0.0156)	(0.0160)	(0.0158)
Advisee Verbal Score	-0.0158	-0.0168	-0.0147	-0.0159
	(0.0108)	(0.0108)	(0.0110)	(0.0110)
Constant	0.202***	0.471	0.191***	0.459
	(0.0640)	(0.295)	(0.0644)	(0.303)
Controls	No	Yes	No	Yes
Observations	1000	1000	1000	1000
Subjects	200	200	200	200

Dependent variable is equal to 1 if the advisor recommends choosing math test and 0 if the advisor recommends choosing verbal test. Performance Feedback is equal to 1 if the advisor learned their scores before giving advice. Female Advisor is equal to 1 if the gender of the advisor is female. All columns control for score combination fixed effects (i.e., for which score combination the advice was provided) and advisor's own math and verbal scores. Columns 2 and 4 also control for advisor's race, age, employment status, highest level of education, and total approvals on Prolific. Standard Errors are clustered at the individual level and are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*