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

Fulya Ersoy and Derek Rury

This version: July 18, 2025

Most Recent Version:

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

Abstract

The gender gap in STEM careers is shaped in part by differences in educational investments. This study investigates two interventions—absolute performance feedback and personalized advice—aiming at narrowing the gender disparities in investments in math skills. In an online lab experiment setting, where females are less likely to select a math task over a verbal task, we show that providing performance feedback closes this gap by increasing female participation in math tasks while leaving male participation unchanged. Among the participants eligible for both math and verbal advice, either performance feedback or randomly assigned math advice significantly increases the likelihood of females choosing math, with no measurable effect on males. Notably, the gender of the advisor—randomly assigned in our study—has no significant impact on the effectiveness of the advice. These findings suggest that performance feedback and targeted advice can encourage female participation in math-related tasks, offering insights into strategies for reducing gender disparities in STEM fields.

*Fulya Ersoy, fulya@uchicago.edu, University of Chicago. 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 Elif Demiral, Prachi Jain, Jamin Speer, Srini Vasudevan, and participants of the Loyola Marymount University brownbag seminar and North American ESA for helpful comments. ChatGPT, (https://openai.com) is used for text editing.

1 Introduction

Despite significant progress in gender equality, particularly in education, women remain underrepresented in STEM careers. Educational choices play a critical role in this disparity, as the gender gap in STEM begins early, with girls selecting different subjects in secondary school and boys enrolling in more advanced math and science courses in high school. This pattern persists through college and graduate school (Delaney and Devereux, 2019; Speer, 2023; Ahimbisibwe et al., 2024; Delaney and Devereux, 2024), partly because high-ability women often opt out of math-focused opportunities and into more gender-stereotypical fields, such as the humanities. Research suggests this may stem from uncertainty about their ability in male-typed fields (Beyer, 1990; Niederle and Vesterlund, 2011; Marshman et al., 2018) or from lower preferences for competition (Niederle and Vesterlund, 2007). Beyond individual career implications, this gender disparity imposes substantial economic costs on society (Peri et al., 2015), including exacerbating the gender pay gap. Thus, identifying effective interventions to reduce this gender gap remains a crucial research priority.

This study examines whether absolute performance feedback and personalized advice reduce the gender gap in selecting a math test over a verbal test in an online lab experiment. We test the direct effects of these interventions and their interactions. We also test whether the impact of advice depends on the gender match between the advice giver and the advisee. These interventions are grounded in prior research. First, perceived ability strongly influences educational decisions (Altonji et al., 2016), and existing gender differences in perceived math ability (Exley and Kessler, 2022) suggest that performance feedback may narrow the gender gap by providing objective information about past performance. Second, advice can shape academic choices (Gentry et al., 2023), particularly when advisor and advisee share the same gender (Canaan and Mouganie, 2023), and evidence shows women respond more to encouragement in academic contexts (Unkovic et al., 2016; Wong et al., 2020). Thus, advice—especially from a same-gender advisor—may reduce the gender gap by providing information and/or offering encouragement.

For our experiment, we recruited participants from Prolific, an online platform often used in economic research (Palan and Schitter, 2018; Bordalo et al., 2024). Participants first took a short math and verbal assessment and were then randomized into different groups. A random half of the participants learned both their (absolute) math and verbal scores from the assessment whereas the other half did not receive any feedback on their performance. Participants were also randomly assigned to receive advice - which was generated from an earlier prolific survey - on what type of assessment (Math or Verbal) they should take in the second part of the study.[1] To study the gender-match (e.g. female advisor, female advisee) effects of advice, participants in advice treatments were also randomly assigned a male, female, or ungendered advisor as the source of the advice. In the second part of the study, subjects chose which assessment - math or verbal - to take. To ensure that performance in the first part and assessment selection for the second part were incentivized, subjects were paid based on their performance in either part, which was randomly selected by the computer. The online lab setting allows us to gain control over the environment in ways that are impossible in real-world settings, enabling us to isolate the specific effects of performance feedback and personalized advice from other confounding factors (e.g. role models).

We hypothesize that performance feedback, which provides precise and objective information about prior performance, will reduce the gender gap in math choice. In contrast, advice conveys coarser and potentially noisier information, so its credibility may depend on shared characteristics such as gender match between advisor and advisee. Beyond informa-

[1] Advice was created from an experiment we conducted with another set of Prolific subjects before conducting this experiment (see section 3.3 for more details). The subjects in that experiment were asked to provide advice based on different score profiles; hence, advice was mostly aligned with subjects' scores. For some score profiles, we were able to randomly assign the type of advice (Math or Verbal).

tion, advice can also serve as encouragement or discouragement (e.g., signaling that someone believes the participant is good or bad at math). We therefore expect advice to reduce the gender gap, with stronger effects when the advisor and advisee share the same gender. Finally, if the effects of advice operate purely through informational content, we would expect that combining advice with performance feedback would have no additional effect beyond performance feedback alone.

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—a pattern that mirrors gender disparities observed in real-world STEM participation. Our performance feedback intervention had a notable impact on choice of math, 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 100% of the gender gap in this context. We find no significant effect of being assigned to the advice treatment and no evidence of an interaction between advice and performance feedback.

To examine whether our interventions have heterogeneous effects by initial test performance, we categorize subjects into three bins: better at math, equally skilled in math and verbal, or worse at math. A gender gap in test choice in the control groups exist across all bins, although we are only able to detect statistically significant differences for the individuals who are worse at math. Among those better at math, 16% of females and 32% of males chose the math test, while among those worse at math, only 7% of females did so compared to 42% of males. Our interventions affected males and females differently based on initial test scores. For those better at math, performance feedback (alone or with advice) closed the gender gap by increasing math test selection more among females than males. For those worse at math, advice (alone or with feedback) reduced the gender gap by lowering the proportion of males choosing the math test.

One feature of our study is that, since we aimed to provide truthful advice, we could not offer math or verbal advice to subjects independently of their initial test scores. However, there was a subset of subjects who were eligible to receive both math and verbal advice; hence, we were able to randomize the type of advice they received. For this subset, either receiving performance feedback or being randomly advised to take the math test increased the proportion of females choosing the math test but it did not statistically affect males' choices. Either intervention reduced the gender gap in math choice. Receiving both performance feedback and randomly assigned math advice together increased the proportion of both females and males choosing the math test. We find suggestive evidence that performance feedback and math advice act as substitutes for females but not for males. For females, adding math advice to performance feedback provides no additional benefit beyond feedback alone. For males, however, combining advice with feedback increases math choice, suggesting that the effect of advice for males is not driven solely by its informational content. Taken together, these results imply that if the goal is to reduce the gender gap in math choice, offering either performance feedback or math advice—rather than combining both—may be more effective.

Given the strong effects found in the applied literature on gender-match (Carrell et al., 2010; Ginther et al., 2020; Canaan and Mouganie, 2023; **?**), we randomly assign the gender of the advisor in the advice treatments to test whether gender match between the advisor and advisee is a factor in the effectiveness of advice for the subjects eligible for randomly assigned math or verbal advice. Overall, we do not find any significant differences in subjects' choice of math test across the various advisor treatments (ungendered, female, or male), though we see some suggestive patterns. Verbal advice significantly reduces math test choice for males when given by a gendered advisor. Notably, the gender gap in math choice narrows when verbal advice comes from a female advisor or when math advice is given by an advisor of an unknown gender.

The rest of the paper is structured as follows. Section 2 provides a review of the literature studying the effect of performance feedback and advice and how this paper contributes to these literatures. Section 3 provides details on the experimental design. Section 4 describes the data. Section 5 presents our results and Section 6 contextualizes the results and discusses some policy implications. Section 7 concludes.

2 Literature Review

2.1 Performance Feedback

One fact documented by social science researchers is that learning about your past performance can change your future behavior. This occurs across several domains including the labor market (Buser et al., 2014), health (Kolstad, 2013) and education (Azmat and Iriberri, 2010; Dobrescu et al., 2021; Castagentti and Rury, 2024). One strand of the education literature shows that performance feedback positively affects both effort (Eriksson et al., 2009; Castagentti and Rury, 2024) and performance (Azmat and Iriberri, 2010; Bandiera et al., 2015; Azmat et al., 2019; Dobrescu et al., 2021; Bobba and Frisancho, 2022) at least in the short run, highlighting performance feedback as a useful policy tool for improving educational outcomes. When considering this mechanism, we designed our experiment such that performance feedback cannot impact effort of our subjects. Rather, it affects the decisions of our subjects by resolving uncertainty about their own performance.

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 et al., 2014; Berlin and Dargnies, 2016; Jeworrek, 2019; Coffman et al., 2024).[2] Research also consistently documents that when focusing on the gender differences in reactions to performance feedback, large differences in self-assessed ability across various domains (Beyer, 1990; Niederle and Vesterlund, 2011; Marshman et al., 2018). Even when controlling for actual performance, men tend to be overconfident, while women are underconfident (Exley and Kessler, 2022; Owen, 2023; Demiral and Mollerstrom, 2024). This has implications for education as beliefs about ability are crucial in shaping major choice decisions (Altonji et al., 2016).

Another important fact is that most research on performance feedback in education focuses on *relative* feedback, or comparisons of your performance to how others performed. This research is valuable from a policy-making stand point as most comparisons happen between students and their immediate peers (e.g. classmates) (Murphy and Weinhardt, 2020). In our experiment, we provide *absolute* performance feedback—information about one's own performance—which may help bridge this confidence gap, aligning individuals' perceptions with their actual abilities and influencing their choices independent of one's peers, which the student does not select.

We also contribute to this literature by showing how absolute performance feedback impacts the *choice of task* (e.g. choosing a mathematical task over a verbal task), rather than performance (e.g. score on an exam). Papers that focus on the effects of performance feedback on task choice are scarce apart from Baier et al. (2024) and Coffman et al. (2024). Baier et al. (2024) finds that there is no significant gender gap in the choice of male-typed

[2] 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.

task when receiving absolute performance feedback or absolute performance feedback combined with relative performance feedback.[3] Coffman et al. (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.[4] Hence, we view our paper as providing further evidence to inform performance feedback policies, especially those aimed at influencing educational choices.

2.2 Advice

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 et al., 2015; Mulhern, 2023; Gentry et al., 2023). 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). Advice can serve as encouragement, mitigating stereotypes and self-doubt or as a means of providing information. However, data from field studies often entangle advice effects with other factors, such as an advisor's broader investment in a student's success or role modeling (Porter and Serra, 2020; Patnaik et al., 2023). Our paper adds to this literature by investigating whether advice alone, delivered online in a one-shot setting, meaningfully influences decision-making.

[3] Baier et al. (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.

[4] This outcome variable is not the main outcome variable studied in the paper and it is only presented in the supplementary appendix. Coffman et al. (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.

Research studying advice shows that not only advice influences individuals' behavior, but also that it increases the efficiency of individuals' decisions, even when the advice is naive (i.e. provided by agents with little or no more knowledge than you) (Schotter, 2003; Çelen et al., 2010; Chaudhuri et al., 2009). There is a large literature studying how advice influences individual's decisions in a lab setting, particularly on how advice influences competitive preferences. Brandts et al. (2015) design a lab experiment to study how advice influences women and men's propensity to enter a tournament. They find that naive advice shrinks the gender gap in tournament entry and that this intervention is efficiency improving.

Our advice intervention also builds on research highlighting the importance of gender match between the advisee/student and the advisor/instructor (Carrell et al., 2010; **?**). However, in a field setting, students choose whether to seek advice and from whom to seek advice (Heikensten and Isaksson, 2019; Gallen and Wasserman, 2023). Additionally, propensity to provide advice differs by advisor gender and students receive different advice based on their own gender (Lordan and Lekfuangfu, 2023; Gallen and Wasserman, 2024; Coutts et al., 2024; Osun, 2024). Hence, it is unclear whether gender match will play a role in more controlled settings. For example, in a laboratory setting, Brandts and Rott (2021) finds that gender matching does not affect advice giving or advice following in tournament entry decisions, suggesting that gender match matters less when what the better advice or decision is clear.

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 advisor. Our *null* results on the importance of gender match between the advisor and advisee suggest that the observed propensity for girls to follow the advice of female advisors in field settings may not simply be due to sharing the same gender identity. Instead, these findings point to alternative explanations, such as a role model effect (Porter and Serra, 2020; Patnaik et al., 2023) or importance of continuous, high-touch interactions, for the documented importance of gender match in prior studies (Carrell et al., 2010; ?; Ginther et al., 2020; Canaan and Mouganie, 2023).

3 Experimental Design

3.1 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. Their payment was based on their part 1 or part 2 performance (randomly chosen at the end of the experiment). Figure 1 summarizes the experimental design.

Figure 1: Experimental Design Diagram

Instructions, Understanding Checks, Initial Survey Questions

Part 1: Subjects answer 4 Math Questions **and** 4 Verbal Questions

Random Assignment to Treatments		**Advisor**			
		No	**Ungendered**	**Female**	**Male**
Performance Feedback	**No**	93	99	184	221
	Yes	104	105	220	174

Number of subjects in each treatment are reported in each cell. Each subject is randomly assigned to only one treatment. Subjects in advisor treatments receive Math **or** Verbal Advice based on their Part 1 scores. A subset of subjects receive randomly assigned Math or Verbal Advice. See Table 1 for details.

Choice: Subjects choose Test (Math **or** Verbal) under Tournament **and** Piece-rate Condition

Part 2: Random Assignment to Tournament **or** Piece-rate Condition. Subjects learn their condition and answer 4 Math Questions **or** 4 Verbal Questions (based on their choice in the assigned condition).

Final Survey Questions

3.2 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.[5]

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, the accuracy rates are not different across genders, and there is a 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.[6] In Part 2 of the experiment, subjects answered either 4 math questions or 4 verbal questions based on their choice.

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,

[5]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.

[6]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.

3.3 Treatments

After Part 1 of the experiment completed, half of the subjects were randomly assigned to learn their (absolute) math and verbal scores (i.e., performance feedback) from part 1 while the remaining half did not (i.e., no performance feedback).

Then, subjects were randomly assigned one of the following conditions. 1/6th of the subjects did not receive any advice (i.e., no advisor). 1/6th of the subjects received advice from an advisor but they were not provided any information regarding the gender of the advisor (i.e., ungendered advisor). 2/6th of the subjects received advice from a female advisor (i.e., female advisor) and another 2/6th of the subjects received advice from a male advisor (i.e., male advisor).[7] See Appendix Figure 1 for the visuals of advice treatments.

[7] To test whether subjects' beliefs about their advisors' competence influence their decisions, half of the subjects in each of these groups were randomly assigned to learn whether their advisor performed among the top 50% of the advisors. At the end of the experiment, we asked subjects whether they were given information about knowledge level of their advisor (incentivized). 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 and we pooled these treatments.

The advice subjects received (math or verbal) depended on their performance in part 1 and was randomized when possible. 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 the 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 for part 2. 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. This figure was also provided to the advisors and advisees. Further 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 some score combinations, our advisors never recommended math or verbal (e.g. advisors never recommended math for a score of 0 correct in math and 5 correct in verbal). For clarity, we can break down the advisors' recommendations into three different groups, or categories which we call R (for random), V (for verbal) and M (for math). For 10 out 25 score combinations (category R in Table 1), some advisors recommended taking the math test whereas others recommended taking the verbal test. Hence, *advisees* with these score combinations were eligible to receive randomly assigned Math or Verbal advice (we focus on this subsample in Section 5.3). For 9 score combinations (category V in Table 1), all *advisors* recommended taking the verbal test. Hence, *advisees* in this category were eligible to receive only Verbal advice. For the remaining 6 score combinations (category M in Table 1), all *advisors* recommended taking the math test. Consequently, *advisees* in this category were eligible to receive only Math advice. Table 1 shows the details.

Table 1: Advice Eligibility Categories based on Part 1 Scores

Math Score \ Verbal Score	0	1	2	3	4
0	R	R	V	V	V
1	R	R	V	V	V
2	M	R	R	V	V
3	M	M	M	R	V
4	R	M	M	R	R

The letters in the table represent which types of advice subjects are eligible to receive based on their part 1 math and verbal test scores. R = Randomly Assigned Math or Verbal Advice, V = Verbal Advice, M = Math Advice.

In our setting, messages delivered through the performance feedback intervention and through the advice intervention were aligned for most of the subjects. 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 (See Appendix Table 1). For participants who were in categories V and M, performance feedback and advice were perfectly aligned. For participants in Category R who were randomly assigned to receive Math or Verbal advice, the advice was consistent with performance feedback, conflicting with performance feedback, or neither consistent nor conflicting (when math and verbal scores were equal) (See Appendix Table 2).

Finally, we asked subjects' choices of part 2 test (math or verbal) under both piece-rate and tournament conditions (strategy method). We then randomly assigned subjects to either a piece-rate condition or a tournament condition and informed them about their condition before the part 2 test starts. 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 (math or verbal), provided both chose the same test for part 2 under tournament condition. 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.[8]

3.4 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 other subjects who chose the same test (math or verbal) and who were randomly assigned to take it under the tournament condition. If their scores were above their opponents, they received $2 and if their scores were below their opponent, they received $0. Ties were resolved randomly by the computer. All the payment procedures were explained to the subjects.

To measure participant attention, subjects earned an additional $0.25 bonus if they recalled their advisors' characteristics, the advice they received (math or verbal), and two true/false questions regarding the details of the experiment.[9] 86% of the subjects correctly

[8] 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 for the two conditions and control for the condition in the regressions.

[9] 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,

recalled the gender of the advisor and 98% of the subjects correctly recalled the advice they received.

3.5 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 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 chooses 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 5), 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, impact of advice and advice type - Math or Verbal -, and the interactions between the two interventions on the choice of math test in addition to conducting the pre-registered analysis of the effects of the advisor-advisee gender match. Hence, the analysis in the paper diverges from our pre-registration. We present the pre-registered analysis in Appendix Tables 3 (main analysis)

etc.) when providing their advice." 77% of the subjects answered the first question correctly and 75% of the subjects answered the latter correctly.

and 4 (heterogeneity analysis).

4 Data and Descriptive Statistics

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.[10]

Table 2 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 2 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 of 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 will report the results both without and with controls in the tables of Section 5. Throughout the text, we will only

[10]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.

discuss the results based on the regression specifications that include the full set of controls.

Table 2: Summary Statistics and Balance

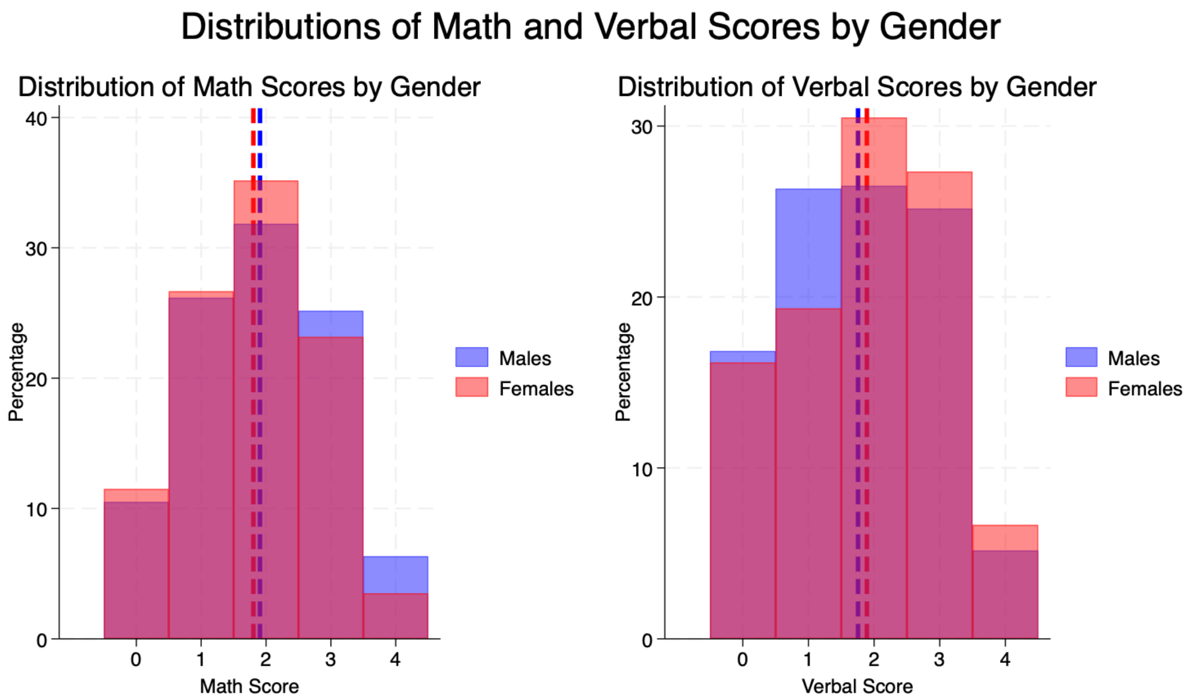
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Perf. Feedback:		**Advisor:**				**P-values:**	
	All	**No**	**Yes**	**No**	**Ungendered**	**Male**	**Female**	**2=3**	**5=6=7=8**
Female	0.500	0.489	0.511	0.492	0.544	0.489	0.493	0.453	0.588
Race:									
Asian	0.104	0.119	0.090	0.132	0.093	0.071	0.129	0.096	0.027
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 et al., 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.[11] 35% of females and 41% of males performed better in the part 1 math test compared

[11] These test questions were selected based on absence of accuracy differences across genders in the pilot we conducted with Prolific subjects.

Figure 2: Part 1 Test Scores



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. Figure 2 shows the distribution of math and verbal scores for females and males. 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). Appendix Table 5 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.

5 Results

In this section, we first show the overall effects of the interventions. Then, we describe how these effects differ based on the initial test scores of the subjects. Next, we depict the effects of the interventions for category R subjects, subjects who were eligible to receive randomly assigned math or verbal advice. Finally, we look at whether advisor gender impacts the effectiveness of the advice intervention for the category R subjects.

5.1 How do interventions affect the choice of math test?

First, we look at whether and how performance feedback and advice affect the choice of math test over verbal test. To study these effects, we estimate the following model:

$$\begin{aligned} ChooseMath_{ij} = \alpha_0 + \alpha_1 PerformanceFeedback_i + \alpha_2 Advice_i + \alpha_3 Advice_i * Feedback_i \\ + \alpha_4 MathScore_i + \alpha_5 VerbalScore_i + \alpha_6 Tournament_{ij} + \Omega X_i + \epsilon_{ij} \end{aligned} \quad (1)$$

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, $Advice_i$ is equal to 1 if the individual i received advice about which test to take in part 2, $Advice_i * Feedback_i$ is equal to 1 if the individual i both received performance feedback and advice, $MathScore_i$ 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 to 0 if it belongs to the piece-rate 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 interventions are neither complements nor substitutes, α_3 should be not distinguishable from zero. To investigate what happens to the gender gaps, we estimate the equations for males and females using seemingly unrelated regression and look at whether the coefficients of interest are different across these two groups.

Table 3 presents the results. Among all subjects in the control group, 24.7% chose the math test. There is also a significant gender gap; while 35.1% of males chose math, only 15% of the females did so, indicating a 20 percentage-point gender gap in the choice of math test after controlling for various factors.

Performance feedback intervention increased the percentage of individuals choosing the math test in the overall sample by 11.5 percentage points (statistically significant at the 5% level), corresponding an effect size of 46.5%. The effect is driven by female subjects (23 percentage points, statistically significant at the 1% level) and no effect is detected for the proportion of males choosing the math test. The difference between the male and female coefficients are statistically significant at the 5% level. Receiving performance feedback closes the gender gap in the choice of math observed in the control group. We will investigate this result more in Subsection 5.2 where we look at the heterogeneous effects of the interventions based on initial test scores.

Advice intervention, on the other hand, did not have a significant effect on choice of math test for either the overall sample or the male and female subsamples. This non-existence of an effect might be due to the fact that 45% of subjects receiving advice (47% of males and 42% of females) received Math advice while the remaining received Verbal advice. Since, the advice is endogenous, we choose not to analyze the effect of type of advice (Math vs Verbal) here.[12] We will investigate the effect of type of advice in Subsection 5.3 for the subjects who were eligible to receive randomly assigned Math vs Verbal advice.

[12]Although this is not our preferred specification due to the endogeneity issue, Appendix Table 6 reports the results of this analysis for completeness.

Table 3: Performance Feedback, Advice, and Choice of Math Test

Sample	All		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)
Performance Feedback (α_1)	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 (α_2)	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 (α_3)	-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 (α_4)	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 (α_5)	-0.128***	-0.119***	-0.139***	-0.136***	-0.114***	-0.104***†
	(0.00901)	(0.00900)	(0.0116)	(0.0120)	(0.0138)	(0.0133)
Constant (α_0)	0.189***	0.298**	0.150***	0.413***	0.243***	0.186
	(0.0466)	(0.116)	(0.0510)	(0.152)	(0.0788)	(0.169)
Control Mean	0.237	0.247	0.137	0.15	0.343	0.351
	(.0413)	(.0376)	(.0412)	(.0401)	(.0709)	(.0628)
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 included in all regressions and equal to 1 if the choice belongs to the tournament condition. Control Mean is the predicted level of choice of math in No Performance Feedback and No Advice Condition, with the relevant controls, averaged over all subjects in all treatments. Standard errors of the prediction are reported in the parentheses. 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 even columns include the controls that are listed in Table 2. 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. * represents whether the coefficients are statistically significantly different from zero and † represents statistically significant differences between coefficients of males and females estimated through seemingly unrelated regressions. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. †$p < 0.10$, ††$p < 0.05$, †††$p < 0.01$.*

Furthermore, there is no detectable interaction effect between performance feedback and advice interventions, which suggests that performance feedback and advice interventions are neither complements nor substitutes. We find this result interesting given that although performance feedback and advice are targeting different mechanisms, they are conveying similar information in most cases in our setting.

The p-values reported in the table allow for comparisons across different treatments. For example, there is no statistically significant difference between the effect of the performance feedback intervention alone and the intervention that combines performance feedback and advice. However, the effects differ significantly between the advice intervention alone and the combined intervention. Finally, the combined intervention significantly increases the probability of choosing the math test compared to the control group, both in the overall sample and the female subsample.

5.2 Are the effects of interventions different based on Part 1 performance?

The observed effect of interventions on math test choice may depend on Part 1 performance, since performance feedback intervention is solely based on Part 1 performance and advice intervention is primarily based on Part 1 performance. To explore, we provide a simple tabulation of math test choice by treatments, gender, and part 1 scores in math and verbal test scores. Table 4 presents the results. Panels A, B, and C show the percentage of individuals choosing the math test over verbal 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 were not aware how well they performed in math versus verbal tests in part 1 without performance feedback and advice.

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 primarily, but not solely dependent on part 1 test scores.

Comparing males and females, we see patterns suggestive of that females and males react to different interventions differently based on initial test scores. When the Part 1 math score exceeded the Part 1 verbal score (Panel A), performance feedback—whether alone or combined with advice—effectively eliminated the gender gap in math test choice. In contrast, the gender gap persisted under the advice-only intervention. When the Part 1 math score was lower than the Part 1 verbal score (Panel C), only 7% of females in the control group chose the math test, compared to 42% of males (statistically significant at the 5% level), suggesting a significant gender gap in overconfidence. Among subjects in Panel C, treatments involving advice-whether alone or combined with performance feedback-eliminated the gender gap in math test choice by reducing the percentage of males choosing math while not impacting females' choices.

For robustness, Appendix Table 7 repeats the regression analysis in Table 3, separately for individuals whose part 1 math score was greater than, equal to, and less than part 1 verbal score. This analysis confirms the findings from the simple tabulation exercise: Among the subjects whose part 1 math scores were greater than verbal scores, Performance Feedback intervention was more effective at increasing the choice of math for females than males. Among the subjects whose part 1 math scores were less than verbal scores, Advice intervention was more effective at decreasing the choice of math for males than females.

Table 4: Choice of Math Test across Treatments, Genders, and Part 1 Test Scores

	(1) All	(2) Female	(3) Male
Panel A: Math Score > Verbal Score			
No PF & No Advice	0.243	0.156	0.316
	(0.432)	(0.369)	(0.471)
	[70]	[32]	[38]
PF & No Advice	0.738***	0.844***	0.667**
	(0.443)	(0.369)	(0.476)
	[80]	[32]	[48]
No PF & Advice	0.443**	0.359	0.520††
	(0.497)	(0.481)	(0.501)
	[388]	[184]	[204]
PF & Advice	0.689***	0.706***	0.675***
	(0.463)	(0.457)	(0.470)
	[370]	[170]	[200]
Panel B: Math Score = Verbal Score			
No PF & No Advice	0.229	0.125	0.333
	(0.425)	(0.338)	(0.482)
	[48]	[24]	[24]
PF & No Advice	0.180	0.227	0.143
	(0.388)	(0.429)	(0.356)
	[50]	[22]	[28]
No PF & Advice	0.273	0.220	0.315
	(0.446)	(0.416)	(0.466)
	[264]	[118]	[146]
PF & Advice	0.320	0.276	0.364
	(0.467)	(0.449)	(0.483)
	[266]	[134]	[132]
Panel C: Math Score < Verbal Score			
No PF & No Advice	0.206	0.0714	0.423††
	(0.407)	(0.261)	(0.504)
	[68]	[42]	[26]
PF & No Advice	0.128	0.0714	0.194
	(0.336)	(0.261)	(0.401)
	[78]	[42]	[36]
No PF & Advice	0.0843**	0.0598	0.110***
	(0.278)	(0.238)	(0.314)
	[356]	[184]	[172]
PF & Advice	0.113	0.0926	0.144**
	(0.317)	(0.291)	(0.352)
	[362]	[216]	[146]

PF means performance feedback. Standard deviations are reported in parentheses. Number of observations is reported in brackets. Number of subjects is number of observations divided by two. * *represents whether the coefficients of No PF & No Advice treatment are statistically significantly different from coefficients of other treatments based on the two-sample test of proportions clustered at the subject level with an intra-class correlation of 1.* † *represents statistically significant differences between coefficients of males and females based on the two-sample test of proportions clustered at the subject level with an intra-class correlation of 1. Note that using a different intra-class correlation would result in more statistically significant estimates.* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. † $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$.

5.3 How do interventions affect the choice of math test among subjects who are eligible for randomly assigned advice?

A limitation of our experiment is that not every subject was eligible to receive both types of advice (math and verbal). For example, a subject who answered no math questions correctly but all verbal questions correctly was not eligible to receive advice to choose math test since no advisor recommended math test in this scenario.[13] However, there exists a subset of subjects (Category R) who was eligible to receive randomly assigned Math or Verbal advice. Most of these subjects had similar math and verbal scores in part 1 tests (see Table 1). In the following analysis, we restrict our sample to subjects in Category R.[14] Since effects of the advice intervention likely depend on whether math or verbal advice was received, we analyze the treatment effects separately by randomly assigned advice type in the following analysis. We run the regressions of the form:

$$\begin{aligned} ChooseMath_{ij} = \alpha_0 + \alpha_1 PerformanceFeedback_i + \alpha_2 AdviceVerbal_i + \alpha_3 AdviceMath_i \\ + \alpha_4 AdviceVerbal_i * Feedback_i + \alpha_5 AdviceMath_i * Feedback_i \\ + \alpha_6 MathScore_i + \alpha_7 VerbalScore_i + \alpha_8 Tournament_{ij} + \Omega X_i + \epsilon_{ij} \end{aligned} \quad (2)$$

where $AdviceVerbal_i$ 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

[13]Advisors get paid either based on their own performance or based on the performance of advisee who received their advice; hence, they were incentivized to give reasonable advice. Details on how the advisor experiment was conducted are located in Appendix Section D.

[14]For the sake of completeness, Appendix Table 8 repeats the results for subjects in Category V (those who were only eligible to receive Verbal Advice) and Category M (those who were only eligible to receive Math Advice).

to 1 if the individual received performance feedback and was also recommended to take the math test. The baseline category is Category R subjects who did not receive any advice or performance feedback. If the performance feedback and advice are neither complements nor substitutes, α_4 and α_5 should be not distinguishable from zero. To investigate what happens to the gender gaps, we estimate the equations for males and females using seemingly unrelated regression and look at whether the coefficients of interest are different across these two groups.

Table 5 presents the results. Among all subjects in Category R, receiving performance feedback does not statistically significantly affect their probability of choosing math compared to the control group. This result is different than the result of Table 3; but it is expected given that most Category R individuals have similar math and verbal scores in Part 1 test -indeed, 58% of them had equal math and verbal scores-, hence performance feedback might not be as effective for these individuals. Receiving randomly assigned math advice -alone 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 randomly assigned verbal advice -alone or combined 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).

For female subjects in Category R, receiving performance feedback alone and receiving randomly assigned math advice alone were both highly effective at increasing their math choice. Receiving performance feedback alone increased the math choice by 25 percentage points and receiving randomly assigned math advice alone increased math choice by 29.5 percentage points, both statistically significant at the 1% level. Receiving verbal advice alone didn't result in significant changes. We also find suggestive evidence of that performance feedback and math advice were substitutes to each other for Category R females, since receiving both interventions together was less effective than the sum of the effects of the individual interventions (only statistically significant at the 10% level). Combining performance feedback with randomly assigned verbal or math advice has no additional effect beyond performance feedback alone (the p-value for $\alpha_2 + \alpha_4 = 0$ is 0.415 (verbal) and for $\alpha_3 + \alpha_5 = 0$ is 0.42 (math)), suggesting that the impact of advice for females may operate primarily through its informational content.

For male subjects in Category R, we find no statistically significant effects of any of the interventions alone. Interactions between performance feedback and receiving Verbal advice or between performance feedback and receiving Math advice were also not statistically significant. However, we see that receiving the performance feedback and the randomly assigned math advice together increased the proportion of Category R males choosing the math test (the p-value for $\alpha_1 + \alpha_3 + \alpha_5 = 0$ is 0.049). Furthermore, combining performance feedback with randomly assigned math advice has an additional positive impact beyond performance feedback alone (p-value for $\alpha_3 + \alpha_5 = 0$ is 0.05), suggesting that the effect of advice for males is not driven solely by its informational content.

For Category R subjects, interventions administered alone significantly reduce the gender gap relative to the control group. The reduction is statistically significant for performance feedback (p = 0.029) and marginally significant for randomly assigned verbal advice (p = 0.075) and randomly assigned math advice (p = 0.078). When performance feedback is combined with random math advice, we do not find a significant difference in the gender gap compared to control (p = 0.305). However, combining performance feedback with random verbal advice does reduce the gender gap relative to control (p = 0.037).

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

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 (α_6)	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 (α_7)	-0.149***	-0.126***	-0.176***	-0.156***	-0.125***	-0.0979***
	(0.0251)	(0.0240)	(0.0359)	(0.0337)	(0.0363)	(0.0352)
Constant (α_0)	0.148**	0.0971	0.0594	0.342	0.238**	-0.0629
	(0.0615)	(0.171)	(0.0733)	(0.261)	(0.0988)	(0.228)
Control Mean	.216	.226	.126	.098	.315	.365
	(.0544)	(.052)	(.0569)	(.0566)	(.091)	(.0848)
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 who are eligible to receive randomly assigned Math or Verbal Advice (see Category R in 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 included in all regressions and equal to 1 if the decision belongs to the tournament condition. Control Mean is the predicted level of choice of math in No Performance Feedback and No Advice Condition, with the relevant controls, averaged over Condition 1 subjects in all treatments. Standard errors of the prediction are reported in the parentheses. 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 even columns include the controls that are listed in Table 2. Standard Errors are clustered at the individual level and are in parentheses.* represents whether the coefficients are statistically significantly different from zero and † represents statistically significant differences between coefficients of males and females estimated through seemingly unrelated regressions . * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. † $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$.*

5.4 Does Gender Match between Advisor and Advisee Matter in the Role of Advice on Choice of Math Test?

Given the strong effects of randomly assigned math advice in our study and the applied findings on gender match between advisors and advisees—particularly the impact of female advisors on female students' persistence in quantitative or STEM fields (Carrell et al., 2010; Porter and Serra, 2020; Canaan and Mouganie, 2023)—we examine whether the advisor's gender is a mechanism contributing to our results. Crucially, in our experiment, advisor's gender was randomly assigned and the type of advice was independent of the advisor's gender, which allow us cleanly test this mechanism. If the effects of advice stem primarily from advice coming from a same-gender advisor rather than personalized interactions with a same-gender advisor, advice can be delivered at scale through structured interventions like this one. These interventions will be cost effective compared to personalized interactions with an advisor since the latter requires more sustained, resource-intensive relationships.

To explore this mechanism, we restrict our sample to Category R subjects (those who were eligible to receive randomly assigned Math and Verbal advice) who did not receive performance feedback.[15] We 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 $UngenderedVerbal_i$ / $MaleVerbal_i$ / $FemaleVerbal_i$ is equal to 1 if an ungendered

[15] For the sake of completeness, Appendix Table 9 repeats the results for subjects in Category V (those who were only eligible to receive Verbal Advice) and Category M (those who were only eligible to receive Math Advice).

/ 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. The baseline category is the Category R subjects who did not receive performance feedback or advice. To investigate what happens to the gender gaps, we estimate the equations for males and females using seemingly unrelated regression and look at whether the coefficients of interest are different across these two groups.

Table 6 presents the results. Crucially, both the advisor gender and advice type (math versus verbal) were randomly assigned for the subjects included in this table. In general, we fail to reject the effects of the verbal advice or math advice treatments are the same regardless of the gender of the advisor (supported by the p-values reported in the table).

Although we cannot differentiate between different advisor treatments, a couple of findings are worth highlighting. Being randomly advised to take the verbal test reduces the likelihood of choosing the math test for male subjects if the advice comes from a gendered advisor. Furthermore, effects of randomly assigned verbal advice are significantly different between males and females if the advice comes from a female advisor; that is, gender gap in math test choice is reduced by receiving verbal advice from a female advisor. Gender gap in math test choice is also marginally reduced when math advice is given by an ungendered advisor.

6 Discussion

Across many settings, it is shown that women are less likely to choose STEM paths compared to men. According to the National Center for Education Statistics, while women in the U.S. earn 57% of all bachelors degrees, they account for only 21% of degrees in engineering and 19% of degrees in computer science. Ahimbisibwe et al. (2024) shows a large gender gap

Table 6: Type of Advice, Gender of Advisor, and Choice of Math Test-Category R

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)
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)
Control Mean	.215	.247	.131	.124	.315	.425
	(.0536)	(.0543)	(.0567)	(.062)	(.0892)	(.0872)
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
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
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 subjects who are eligible to receive randomly assigned Math or Verbal Advice (see Category R in 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 included in all regressions and equal to 1 if the decision belongs to the tournament condition. Control Mean is the predicted level of choice of math in No Performance Feedback and No Advice Condition, with the relevant controls, averaged over Condition 1 subjects in No Performance Feedback treatments Standard errors of the prediction are reported in the parentheses. 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 even columns include the controls that are listed in Table 2. Standard Errors are clustered at the individual level and are in parentheses. * represents whether the coefficients are statistically significantly different from zero and † represents statistically significant differences between coefficients of males and females estimated through seemingly unrelated regressions . * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. † $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$.*

in applications to STEM majors across ten countries even among talented students. The setting of our online experiment matches these patterns. Even in this simplified setting and after controlling various observables including test scores, we find that men are 20 percentage points more likely to choose math path over the verbal path than women.

Many explanations are proposed for this gap including differences in self-confidence about STEM ability (Owen, 2023), differences in tastes and preferences (Zafar, 2013; Niederle and Vesterlund, 2007), differences in belief updating (Coffman et al., 2023), lack of female role models (Porter and Serra, 2020), and gender stereotypes (Carlana, 2019). Inspired by these explanations, we designed two interventions: absolute performance feedback intervention and personalized advice intervention.

In our setting, we find absolute performance feedback increases females' probability of choosing the math test without influencing proportion of males choosing math. One potential mechanism is that performance feedback affects beliefs about one's own ability (self-confidence) differently for males and females, which is supported by our heterogeneity analysis. Looking at individuals who performed better at math than verbal on the initial test, we see that receiving performance feedback increases both males and females' math choices. Since increases are larger for females, this intervention closes the gender gap in math choice. Conversely, among the individuals who performed worse at math than verbal on the initial test, performance feedback reduced the proportion of males choosing math, without affecting the math choices of females.

Compared to the literature, this is an interesting finding since literature suggests females are less responsive to feedback in male-typed domains and/or more discouraged by it. Owen (2023) finds that men are overconfident and women are underconfident about their relative performance in STEM courses and men, but not women, correct their beliefs about their own relative course rank when provided with relative performance feedback. Goulas and Megalokonomou (2021) shows females in all parts of the ability distribution are more discouraged by relative performance feedback. Coffman et al. (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 and men are significantly more responsive to information in male-typed domains, while women are significantly more responsive in female-typed domains.

One implication of this finding is that we should provide absolute performance feedback to students before their educational choices. Although students receive absolute performance feedback during their studies, this feedback does not always come right before their track choices. Reminding females their prior performances before important educational decisions such as track choices and math course choices might increase proportion of female students choosing STEM fields. One can also consider selective provision of performance feedback and/or advice based on initial test score distributions.

In our setting, most of the individuals who were eligible to receive randomly assigned advice performed similarly in their math and verbal assessments. For females in this subgroup, receiving performance feedback only, receiving math advice only, and receiving both the performance feedback and math advice are equally effective at increasing the proportion of females choosing the math test. For males, receiving performance feedback only or receiving math advice only does not increase the proportion of males choosing the math test significantly; however receiving both at the same time does.

Consequently, although a combined intervention leads to the highest percentage of individuals choosing the math test, the performance feedback intervention is the most effective at decreasing the gender gap. Hence, if the policy makers' aim is to encourage individuals who are equally talented at math and verbal tasks to choose math, performance feedback combined with math advice appears to be the optimal strategy whereas if the aim is to reduce gender gap in math choice for these individuals, providing performance feedback alone might be the best solution.

Our study contributes to the discussion on the gender gap in quantitative fields and STEM careers by disentangling the effect of advisor gender from other confounding factors. In applied settings, it is challenging to separate the impact of role models from the advice itself, as advisors and advisees often interact in ways influenced by their gender. Prior research suggests that exposure to same-gender role models may influence educational and career choices, but it remains unclear whether the advice given by these role models has an independent effect (Carrell et al., 2010; Canaan and Mouganie, 2023; Porter and Serra, 2020). Our experiment isolates these factors and finds that randomly assigned math advice increases the likelihood of selecting the math task, independent of the gender of the advisor. This finding highlights that the gender-match effects found in the applied literature are not purely due to receiving advice from a same gender advisor.

7 Conclusion

In this paper, we study how absolute performance feedback and personalized advice influence the choice of math test over a verbal test in an online experiment. To accomplish this, we conducted two sets of randomization in the experiment. First, participants were randomly assigned to either learn their math and verbal performance from the initial assessment or not. In a second randomization, participants were randomly assigned to either receive advice or not about which assessment to choose next. We then observe how these interventions influence subjects' choice of math assessment.

Our findings reveal a substantial gender gap in the choice of math test which persists even after controlling for various observables. Receiving absolute performance feedback increases the likelihood of choosing the math test over verbal test and closes the gender gap in the choice of math test. Receiving advice does not significantly affect the math test choice or the gender gap in the math test choice, nor does the interaction of performance feedback and advice. However, type of advice matters. Among the subset of subjects who were eligible to receive randomly assigned Math or Verbal advice, receiving math advice, compared to no advice, significantly increases the propensity to choose math task, particularly for female participants. Our results suggest that performance feedback and advice are substitutes to each other for females in this subset. Receiving randomly assigned 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 it reduces the gender gap in the choice of math test by reducing males' likelihood of choosing the math test. In general, the effects do not differ based on randomly assigned advisor gender.

Our findings have implications for students' educational investment decisions. 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 can encourage students, particularly females, to pursue math-focused fields by reducing uncertainty about their abilities (Owen, 2023; Rury, 2022). Providing advice on pursuing math-focused fields to the students who are equally talented at math and verbal tasks can also be an important catalyzer in encouraging students, particularly females, to pursue quantitative fields.

We expect our findings to generalize to the settings that share important features with our setting. In this study, we examine the effects of *immediate absolute performance* feedback on choice of math test 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 et al. (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 test 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 conflicting with 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.

Altonji, Joseph G, Peter Arcidiacono, and Arnaud Maurel, "The analysis of field choice in college and graduate school: Determinants and wage effects," in "Handbook of the Economics of Education," Vol. 5, Elsevier, 2016, pp. 305–396.

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.

Beyer, Sylvia, "Gender differences in the accuracy of self-evaluations of performance.," *Journal of personality and social psychology*, 1990, *59* (5), 960.

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.

Bordalo, Pedro, Giovanni Burro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, "Imagining the future: memory, simulation, and beliefs," *Review of Economic Studies*, 2024, p. rdae070.

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

Brandts, Jordi and Christina Rott, "Advice from women and men and selection into competition," *Journal of Economic Psychology*, 2021, *82*, 102333.

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

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek, "Gender, competitiveness, and career choices," *The quarterly journal of economics*, 2014, *129* (3), 1409–1447.

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

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.

Çelen, Boğaçhan, Shachar Kariv, and Andrew Schotter, "An experimental test of advice and social learning," *Management Science*, 2010, *56* (10), 1687–1701.

Chaudhuri, Ananish, Andrew Schotter, and Barry Sopher, "Talking ourselves to efficiency: Coordination in inter-generational minimum effort games with private, almost common and common knowledge of advice," *The Economic Journal*, 2009, *119* (534), 91–122.

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.

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.

Demiral, Elif E. and Johanna Mollerstrom, "Signaling confidence," *Journal of Economic Behavior Organization*, 2024, *226*, 106691.

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.

Exley, Christine L. and Judd B. Kessler, "The Gender Gap in Self-Promotion," *The Quarterly Journal of Economics*, August 2022, *137* (3), 1345–1381.

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.

Ginther, Donna K, Janet M Currie, Francine D Blau, and Rachel TA Croson, "Can mentoring help female assistant professors in economics? An evaluation by randomized trial," *AEA Papers and Proceedings*, 2020, *110*, 205–209.

Goulas, Sofoklis and Rigissa Megalokonomou, "Knowing who you actually are: The effect of feedback on short- and longer-term outcomes," *Journal of Economic Behavior Organization*, 2021, *183*, 589–615.

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.

Kolstad, Jonathan T, "Information and quality when motivation is intrinsic: Evidence from surgeon report cards," *American Economic Review*, 2013, *103* (7), 2875–2910.

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.

Marshman, Emily M, Z Yasemin Kalender, Timothy Nokes-Malach, Christian Schunn, and Chandralekha Singh, "Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm?," *Physical review physics education research*, 2018, *14* (2), 020123.

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

Murphy, Richard and Felix Weinhardt, "Top of the class: The importance of ordinal rank," *The Review of Economic Studies*, 2020, *87* (6), 2777–2826.

Niederle, Muriel and Lise Vesterlund, "Do Women Shy Away from Compeition? Do Men Compete too Much?," *Quarterly Journal of Economics*, 2007.

_ **and** _ , "Gender and competition," *Annu. Rev. Econ.*, 2011, *3* (1), 601–630.

Osun, Elif B, "Gender differences in advice giving," *Experimental Economics*, 2024, pp. 1–38.

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

Palan, Stefan and Christian Schitter, "Prolific. ac—A subject pool for online experiments," *Journal of behavioral and experimental finance*, 2018, *17*, 22–27.

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.

Unkovic, Cait, Maya Sen, and Kevin M. Quinn, "Does Encouragement Matter in Improving Gender Imbalances in Technical Fields? Evidence from a Randomized Controlled Trial," *PLOS ONE*, 04 2016, *11* (4), 1–15.

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

Wong, Y. Joel, Nelson O. O. Zounlome, Nancy Goodrich Mitts, and Emily Murphy, "You can do it! An experimental evaluation of an encouragement intervention for female students," *The Journal of Positive Psychology*, 2020, *15* (4), 427–437.

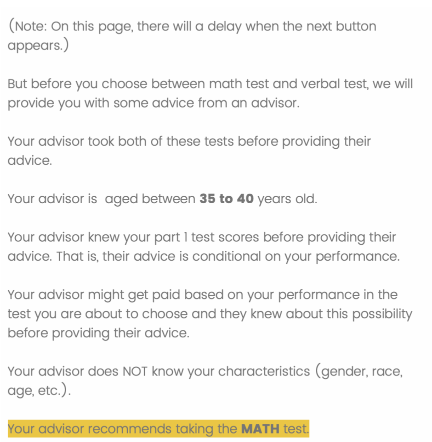
Wozniak, David, William T. Harbaugh, and Ulrich Mayr, "The menstrual cycle and performance feedback alter gender differences in competitive choices," *Journal of Labor Economics*, 2014, *32* (1), 161–198.

Zafar, Basit, "College major choice and the gender gap," *Journal of Human Resources*, 2013, *48* (3), 545 – 595. Cited by: 270; All Open Access, Green Open Access.

Appendix

A Figures

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



(Note: On this page, there will a delay when the next button appears.)

But before you choose between math test and verbal test, we will provide you with some advice from an advisor.

Your advisor took both of these tests before providing their advice.

Your advisor is aged between **35 to 40** years old.

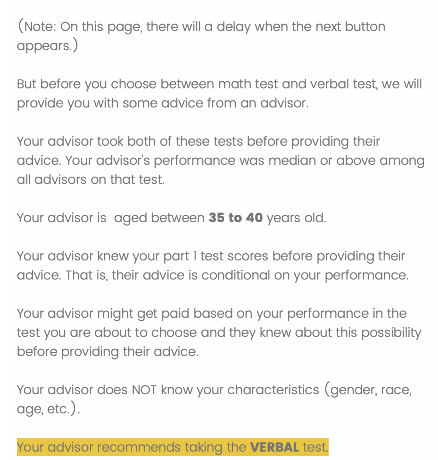
Your advisor knew your part 1 test scores before providing their advice. That is, their advice is conditional on your performance.

Your advisor might get paid based on your performance in the test you are about to choose and they knew about this possibility before providing their advice.

Your advisor does NOT know your characteristics (gender, race, age, etc.).

Your advisor recommends taking the **MATH** test.

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



(Note: On this page, there will a delay when the next button appears.)

But before you choose between math test and verbal test, we will provide you with some advice from an advisor.

Your advisor took both of these tests before providing their advice. Your advisor's performance was median or above among all advisors on that test.

Your advisor is aged between **35 to 40** years old.

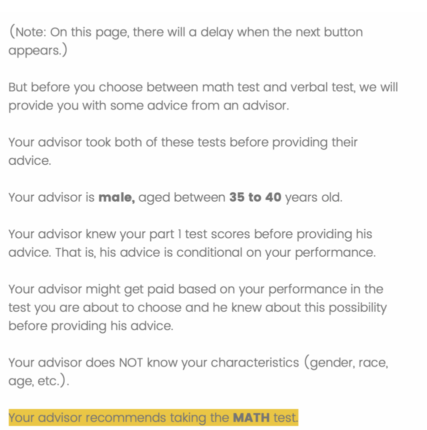
Your advisor knew your part 1 test scores before providing their advice. That is, their advice is conditional on your performance.

Your advisor might get paid based on your performance in the test you are about to choose and they knew about this possibility before providing their advice.

Your advisor does NOT know your characteristics (gender, race, age, etc.).

Your advisor recommends taking the **VERBAL** test.

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



(Note: On this page, there will a delay when the next button appears.)

But before you choose between math test and verbal test, we will provide you with some advice from an advisor.

Your advisor took both of these tests before providing their advice.

Your advisor is **male,** aged between **35 to 40** years old.

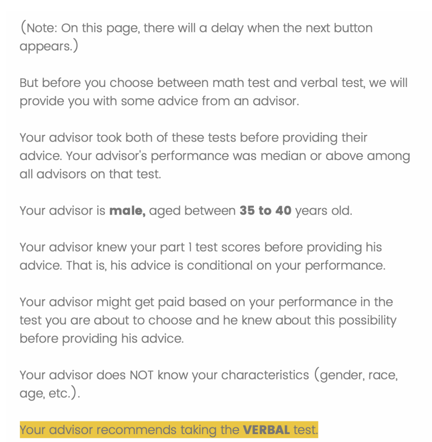
Your advisor knew your part 1 test scores before providing his advice. That is, his advice is conditional on your performance.

Your advisor might get paid based on your performance in the test you are about to choose and he knew about this possibility before providing his advice.

Your advisor does NOT know your characteristics (gender, race, age, etc.).

Your advisor recommends taking the **MATH** test.

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



(Note: On this page, there will a delay when the next button appears.)

But before you choose between math test and verbal test, we will provide you with some advice from an advisor.

Your advisor took both of these tests before providing their advice. Your advisor's performance was median or above among all advisors on that test.

Your advisor is **male,** aged between **35 to 40** years old.

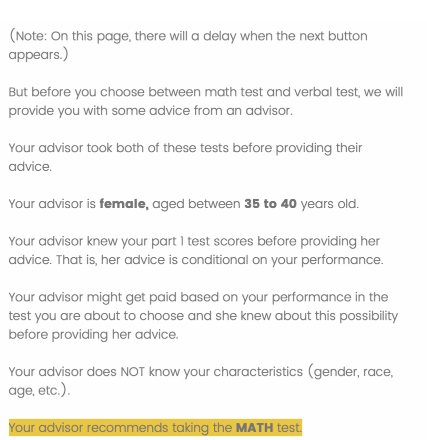
Your advisor knew your part 1 test scores before providing his advice. That is, his advice is conditional on your performance.

Your advisor might get paid based on your performance in the test you are about to choose and he knew about this possibility before providing his advice.

Your advisor does NOT know your characteristics (gender, race, age, etc.).

Your advisor recommends taking the **VERBAL** test.

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



(Note: On this page, there will a delay when the next button appears.)

But before you choose between math test and verbal test, we will provide you with some advice from an advisor.

Your advisor took both of these tests before providing their advice.

Your advisor is **female,** aged between **35 to 40** years old.

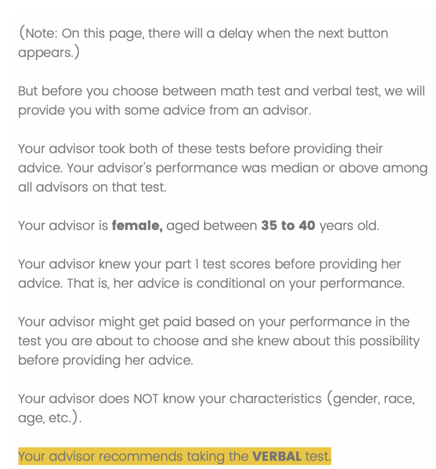
Your advisor knew your part 1 test scores before providing her advice. That is, her advice is conditional on your performance.

Your advisor might get paid based on your performance in the test you are about to choose and she knew about this possibility before providing her advice.

Your advisor does NOT know your characteristics (gender, race, age, etc.).

Your advisor recommends taking the **MATH** test.

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



(Note: On this page, there will a delay when the next button appears.)

But before you choose between math test and verbal test, we will provide you with some advice from an advisor.

Your advisor took both of these tests before providing their advice. Your advisor's performance was median or above among all advisors on that test.

Your advisor is **female,** aged between **35 to 40** years old.

Your advisor knew your part 1 test scores before providing her advice. That is, her advice is conditional on your performance.

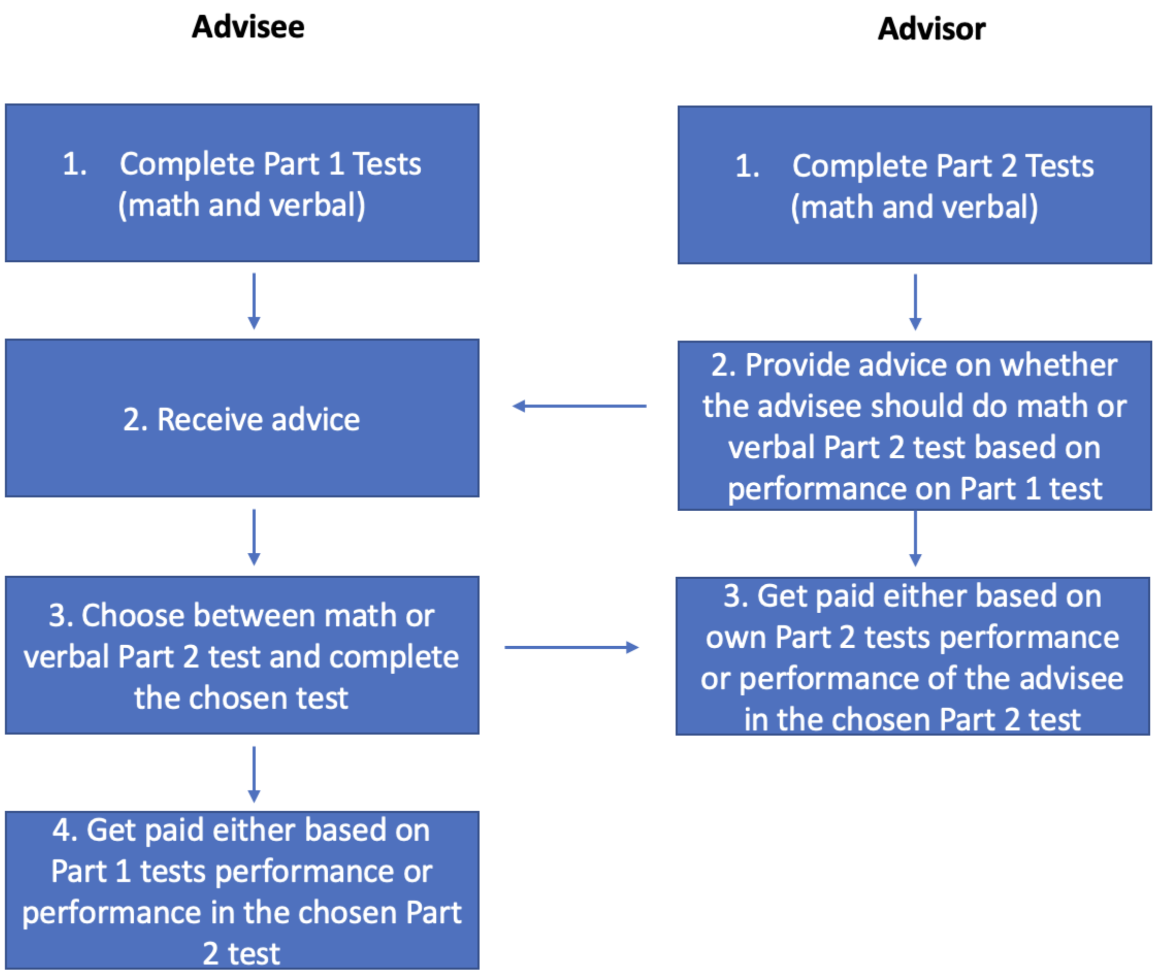
Your advisor might get paid based on your performance in the test you are about to choose and she knew about this possibility before providing her advice.

Your advisor does NOT know your characteristics (gender, race, age, etc.).

Your advisor recommends taking the **VERBAL** test.

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

Figure 2: Relationship between Advisor and Advisee Experiment



B Tables

Table 1: Relationship between Performance Feedback and Advice

	All (1)	Female (2)	Male (3)
Panel A: Math Advice			
Math Score > Verbal Score	0.667	0.657	0.675
Math Score = Verbal Score	0.302	0.310	0.295
Math Score < Verbal Score	0.031	0.033	0.030
Panel B: Verbal Advice			
Math Score > Verbal Score	0.143	0.128	0.160
Math Score = Verbal Score	0.233	0.207	0.262
Math Score < Verbal Score	0.624	0.666	0.578

Rows show the share of subjects whose Part 1 Math score is (i) greater than, (ii) equal to, or (iii) less than their Part 1 Verbal score among those who received Math advice (Panel A) and who received Verbal Advice (Panel B). Column (1) pools all genders; Columns (2) and (3) restrict to female and male subjects, respectively.

Table 2: Relationship between Performance Feedback and Advice-Category R Subjects

	All (1)	Female (2)	Male (3)
Panel A: Math Advice (randomly assigned)			
Math Score > Verbal Score	0.351	0.336	0.364
Math Score = Verbal Score	0.589	0.600	0.579
Math Score < Verbal Score	0.061	0.064	0.058
Panel B: Verbal Advice (randomly assigned)			
Math Score > Verbal Score	0.346	0.346	0.347
Math Score = Verbal Score	0.566	0.561	0.570
Math Score < Verbal Score	0.088	0.093	0.083

Sample is restricted to Category R subjects. Rows show the share of subjects whose Part 1 Math score is (i) greater than, (ii) equal to, or (iii) less than their Part 1 Verbal score among those who received randomly assigned Math advice (Panel A) and who received randomly assigned Verbal Advice (Panel B). Column (1) pools all genders; Columns (2) and (3) restrict to female and male subjects, respectively.

Table 3: Pre-Registration Main Analysis

	Follow Advice		Non-Stereotypical Choice	
	(1)	(2)	(3)	(4)
Gender Match	-0.0173	-0.0193	-0.0267	-0.0237
	(0.0280)	(0.0278)	(0.0299)	(0.0300)
Part 1 Verbal Score	0.0670	0.0609	-0.0118	-0.00766
	(0.0112)	(0.0117)	(0.0135)	(0.0140)
Part 1 Math Score	-0.0731	-0.0849	-0.0263	-0.0195
	(0.0131)	(0.0133)	(0.0148)	(0.0148)
Female Advisee	-0.0355	-0.0239	-0.351	-0.353
	(0.0280)	(0.0289)	(0.0301)	(0.0317)
Tournament	-0.0338	-0.0338	-0.0138	-0.0138
	(0.0139)	(0.0139)	(0.0139)	(0.0140)
Constant	0.773	0.658	0.714	0.646
	(0.0392)	(0.156)	(0.0414)	(0.161)
Observations	1598	1598	1598	1598
Subjects	799	799	799	799
Controls	No	Yes	No	Yes

Sample is restricted to the subjects who received gendered advisor treatments. Dependent variable in Columns 1-2 is equal to 1 if individual i's choice in condition j (tournament or piece-rate) is the same as the advice received. Dependent variable in Columns 3-4 is equal to 1 if a female advisee chooses the math test or a male advisee chooses the verbal test in condition j. Gender Match is equal to 1 if individual i's sex is the same as their advisors. Tournament is equal to 1 if the decision belongs to the tournament condition. Odd columns do not include any additional controls, and even columns include the controls that are listed in Table 2. Standard Errors are clustered at the individual level and are in parentheses. $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Table 4: Pre-Registration Heterogeneity Analysis

	Follow Advice		Non-Stereotypical Choice	
	(1)	(2)	(3)	(4)
Gender Match	-0.0323	-0.0403	-0.0208	-0.0175
	(0.0396)	(0.0395)	(0.0438)	(0.0435)
*Female Advisee	0.0305	0.0427	-0.0120	-0.0125
	(0.0559)	(0.0557)	(0.0597)	(0.0602)
Gender Match	-0.0140	-0.0156	-0.0371	-0.0300
	(0.0307)	(0.0306)	(0.0353)	(0.0331)
*Tournament	-0.00726	-0.00726	0.0126	0.0126
	(0.0277)	(0.0278)	(0.0278)	(0.0279)
Gender Match	0.00952	-0.00826	-0.0649	-0.0586
	(0.0752)	(0.0739)	(0.0878)	(0.0803)
*Risk Pref	-0.00498	-0.00202	0.00672	0.00641
	(0.0130)	(0.0127)	(0.0151)	(0.0140)
Gender Match	-0.0381	-0.0338	0.0258	0.0406
	(0.0375)	(0.0375)	(0.0385)	(0.0359)
*Non-Stereotypical Advice	0.0266	0.0152	-0.0424	-0.0632
	(0.0559)	(0.0553)	(0.0573)	(0.0530)
Gender Match	-0.0323	-0.0357	-0.0103	-0.00563
	(0.0392)	(0.0390)	(0.0451)	(0.0416)
*Knowledgeable Advisor	0.0318	0.0355	-0.0437	-0.0358
	(0.0561)	(0.0554)	(0.0649)	(0.0598)
Gender Match	-0.0120	-0.0125	-0.00856	-0.0410
	(0.0401)	(0.0388)	(0.0459)	(0.0398)
*Performance Feedback	-0.00826	-0.0111	-0.0460	0.0350
	(0.0558)	(0.0554)	(0.0646)	(0.0602)
Gender Match	-0.0122	-0.00912	-0.0322	-0.0306
	(0.0364)	(0.0359)	(0.0422)	(0.0383)
*Student	-0.0136	-0.0256	0.00441	0.0175
	(0.0570)	(0.0570)	(0.0657)	(0.0627)
Gender Match	-0.0740	-0.0735	-0.0988	-0.0797
	(0.0429)	(0.0431)	(0.0476)	(0.0442)
*College Educated Parents	0.103	0.0985	0.123	0.102
	(0.0563)	(0.0565)	(0.0648)	(0.0600)
Gender Match	-0.0453	-0.0485	-0.0310	-0.0228
	(0.0344)	(0.0340)	(0.0405)	(0.0343)
*Better at Math	0.0797	0.0842	0.000554	-0.00250
	(0.0587)	(0.0583)	(0.0672)	(0.0669)
Observations	1598	1598	1598	1598
Subjects	799	799	799	799

Sample is restricted to the subjects who received gendered advisor treatments. Dependent variable in Columns 1-2 is equal to 1 if individual i's choice in condition j (tournament or piece-rate) is the same as the advice received. Dependent variable in Columns 3-4 is equal to 1 if a female advisee chooses the math test or a male advisee chooses the verbal test in condition j. Each panel reports coefficients from a different heterogeneity analysis. Gender Match is equal to 1 if individual i's sex is the same as their advisors. The second row in each panel is the interaction between gender match and the heterogeneity explored in that panel. All columns control for the non-interacted version of the heterogeneity explored, tournament dummy, and part 1 math and verbal scores. Odd columns do not include any additional controls, and even columns include the controls that are listed in Table 2. Standard Errors are clustered at the individual level and are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: 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 6: Performance Feedback, Type of Advice, and Choice of Math Test

Sample	All		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)
Performance Feedback (PF)	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)
PF*Verbal Advice	-0.0725	-0.0567	-0.159**	-0.150**	0.0172	0.0339†
	(0.0602)	(0.0564)	(0.0709)	(0.0685)	(0.0965)	(0.0881)
PF*Math Advice	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)
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+PF*Advice=0	.012	.009	.021	.016	.139	.104
PF+PF*Advice=0	.02	.009	.007	.01	.38	.186
PF+Advice+PF*Advice=0	.59	.73	.013	.03	.3	.247
Math Advice:						
Advice+PF*Advice=0	0	0	.099	.071	.001	0
PF+PF*Advice=0	0	0	0	0	.006	.002
PF+Advice+PF*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 included in all regressions and 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 even columns include the controls that are listed in Table 2. Standard Errors are clustered at the individual level and are in parentheses. * represents whether the coefficients are statistically significantly different from zero and † represents statistically significant differences between coefficients of males and females estimated through seemingly unrelated regressions . * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. † $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$.*

Table 7: Performance Feedback, Advice, and Choice of Math Test by Part 1 Test Scores

Sample	All		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Math Score > Verbal Score						
Performance Feedback (α_1)	0.437***	0.363***	0.628***	0.579***	0.304**†	0.182††
	(0.0945)	(0.0885)	(0.109)	(0.102)	(0.141)	(0.132)
Advice (α_2)	0.169**	0.120*	0.179**	0.135	0.158	0.0742
	(0.0743)	(0.0707)	(0.0829)	(0.0832)	(0.112)	(0.106)
Feedback*Advice (α_3)	-0.192*	-0.128	-0.283**	-0.237**	-0.138	-0.0373
	(0.103)	(0.0957)	(0.121)	(0.114)	(0.154)	(0.142)
Constant (α_0)	0.00242	0.101	-0.0224	0.294	0.0475	0.0512
	(0.0846)	(0.209)	(0.107)	(0.331)	(0.123)	(0.276)
Observations	908	908	418	418	490	490
Subjects	454	454	209	209	245	245
Panel B: Math Score = Verbal Score						
Performance Feedback (α_1)	-0.0489	-0.0292	0.104	0.150*	-0.187	-0.188†
	(0.0998)	(0.0921)	(0.119)	(0.0873)	(0.154)	(0.154)
Advice (α_2)	0.0462	0.0485	0.0904	0.103	-0.00963	-0.0245
	(0.0801)	(0.0747)	(0.0826)	(0.0769)	(0.130)	(0.123)
Feedback*Advice (α_3)	0.0953	0.109	-0.0478	-0.109	0.233	0.319*††
	(0.112)	(0.105)	(0.138)	(0.111)	(0.170)	(0.173)
Constant (α_0)	0.217**	0.217	0.134	0.725**	0.299**	-0.170††
	(0.0870)	(0.238)	(0.110)	(0.336)	(0.133)	(0.316)
Observations	628	628	298	298	330	330
Subjects	314	314	149	149	165	165
Panel C: Math Score < Verbal Score						
Performance Feedback (α_1)	-0.0775	-0.0764	-0.00662	-0.0256	-0.222	-0.186
	(0.0757)	(0.0696)	(0.0693)	(0.0720)	(0.143)	(0.118)
Advice (α_2)	-0.123*	-0.118**	-0.0295	-0.0142	-0.295**††	-0.288***††
	(0.0637)	(0.0556)	(0.0561)	(0.0565)	(0.123)	(0.101)
Feedback*Advice (α_3)	0.107	0.112	0.0392	0.0520	0.260*	0.240*
	(0.0801)	(0.0753)	(0.0760)	(0.0772)	(0.150)	(0.128)
Constant (α_0)	0.229***	0.345**	0.0819	0.0318	0.473***††	0.585**††
	(0.0744)	(0.136)	(0.0609)	(0.128)	(0.149)	(0.229)
Observations	864	864	484	484	380	380
Subjects	432	432	242	242	190	190
Controls	No	Yes	No	Yes	No	Yes

*Sample includes all subjects. Panels A, B, and C includes individuals whose part 1 math score was greater than, equal to, and less than part 1 verbal score, respectively. 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. Columns 1-2 include all genders and Columns 3-4 (5-6) include female (male) subjects. All columns include Part 1 Math Score, Part 1 Verbal Score, and Tournament dummy as controls. Additionally, even columns include the controls that are listed in Table 2. Standard Errors are clustered at the individual level and are in parentheses. * represents whether the coefficients are statistically significantly different from zero and † represents statistically significant differences between coefficients of males and females estimated through seemingly unrelated regressions . * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. † $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$.*

Table 8: Performance Feedback, Type of Advice, and Choice of Math Test-Category V and M Subjects

	Category V: Verbal Advice Eligible						**Category 3: Math Advice Eligible**					
Sample	All		Female		Male		All		Female		Male	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PF (α_1)	-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 (α_2)	-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)
PF*Advice (α_3)	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)
Part1 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)
Part1 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)
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 V 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. PF is equal to 1 if the individual learned their scores from part 1 (i.e., received performance feedback). 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 included in all regressions and 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 even columns include the controls that are listed in Table 2. 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. * represents whether the coefficients are statistically significantly different from zero and † represents statistically significant differences between coefficients of males and females estimated through seemingly unrelated regressions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. † $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$.*

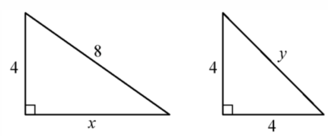
Table 9: Type of Advice, Gender of Advisor, and Choice of Math Test-Category V and M Subjects

	Category V: Verbal Advice Eligible						**Category 3: Math Advice Eligible**					
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***	-0.0731	-0.0503	-0.369***††	-0.432***†††	0.220	0.142	0.154	-0.106	0.273	0.272†
	(0.0686)	(0.0636)	(0.0695)	(0.0703)	(0.124)	(0.105)	(0.138)	(0.141)	(0.148)	(0.139)	(0.192)	(0.195)
Male Advisor	-0.119	-0.115*	-0.0260	-0.0128	-0.279**†	-0.312***††	0.231**	0.173	0.253**	0.0580	0.245	0.163
	(0.0727)	(0.0646)	(0.0726)	(0.0700)	(0.130)	(0.106)	(0.115)	(0.120)	(0.101)	(0.116)	(0.179)	(0.170)
Female Advisor	-0.165**	-0.176***	-0.0583	-0.0751	-0.331**†	-0.363***††	0.187	0.0900	0.121	-0.104	0.206	0.0433
	(0.0693)	(0.0611)	(0.0670)	(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**	0.0673**	0.0564**	0.0336	0.0733	0.175**	0.149*	0.0878	0.0431	0.165	0.141
	(0.0255)	(0.0256)	(0.0275)	(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	-0.0234	-0.00799	0.0122	-0.0184	-0.0855	-0.113**	-0.159**	-0.184***	-0.000519	-0.00497††
	(0.0253)	(0.0251)	(0.0212)	(0.0201)	(0.0522)	(0.0507)	(0.0567)	(0.0485)	(0.0656)	(0.0672)	(0.0871)	(0.0674)
Constant	0.162*	0.270	0.0790	-0.0507	0.304	0.434	-0.122	0.635	0.130	1.002	-0.138	0.421
	(0.0957)	(0.214)	(0.0659)	(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
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 V 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 included in all regressions and 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 even columns include the controls that are listed in Table 2. Standard Errors are clustered at the individual level and are in parentheses.* represents whether the coefficients are statistically significantly different from zero and † represents statistically significant differences between coefficients of males and females estimated through seemingly unrelated regressions . * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. † $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$.

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

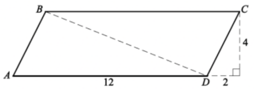
Extensive housing construction is underway in Pataska Forest, the habitat of a large population of deer. Because deer feed at the edges of forests, these deer will be attracted to the spaces alongside the new roads being cut through Pataska Forest to serve the new residential areas. Consequently, once 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 <u>two answer choices</u> 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 <u>amplifies</u> 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 10 presents these results. Female advisors are more likely to advice math whereas there is no detectable effect of either learning own scores (receiving performance feedback) before providing advice or the interaction between the advisor gender and receiving the performance feedback.

Table 10: 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.