Gender and Choice over Co-workers: Experimental Evidence*

Shan Aman-Rana†, Clement Minaudier‡, Brais Álvarez Pereira§ and Shamyla Chaudry¶

November 3, 2021

Abstract

We study whether choice over co-workers matters for performance in gender-diverse teams. We carried out a lab-in-the-field experiment where students were randomly assigned co-workers meant to help them perform on tests. Co-worker allocation was randomized on two dimensions: (1) gender and (2) student preference for that co-worker at baseline. We find that randomly chosen male co-workers reduce the performance of females (12 % of the average score), while preferred male co-workers have a positive yet statistically insignificant effect (6% of the average score). These effects are heterogeneous across the gender stereotype of the questions and materialize even though the two types of male co-workers have the same average ability. To investigate the mechanism behind these effects, we randomly allocated hints as an additional source of information across questions. We find that some (but not all) of these differences are driven by difficulty in accessing those hints in the presence of random male co-workers.

JEL codes: J1, J15, J16, M50, O15

Keywords: Gender, diversity, teams, choice, performance, stereotype, information, communication, advising, help in organizations

*We are grateful to Gaurab Aryal, Joydeep Bhattacharya, Alexia Delfino, Leora Friedberg, Elizabeth Hoffman, Pamela Jakiela, Naureen Karachiwalla, Jessica Leight, Subha Mani, Amalia Miller, John Pepper, Sheetal Seekhri, Sandip Sukhtankar, Basit Zafar, and seminar participants at multiple universities, conferences, and workshops for comments and suggestions. Nida Naz at the Lahore School of Economics, Daniel Kwiatkowski, Ryan Keller, Ashwin Nair, Elizabeth Claire Schroppe and Yanwen Wang at UVA provided excellent research assistance. Financial support from the Quantitative Collaborative is gratefully acknowledged. All mistakes are our own. AEA Trial Registry RCT ID: 0007677. UVA IRB: 3568

†Department of Economics, University of Virginia. Email:sa8ey@virginia.edu

‡Department of Economics, University of Vienna. Email:clement.minaudier@univie.ac.at

§Nova School of Business and Economics NOVAFRICA, and BELAB. Email:brais.pereira@novasbe.pt

¶Lahore School of Economics. Email:shamyla@lahoreschool.edu.pk

1 Introduction

Collaborative activities and teamwork in the workplace have gone up by 50% over 20 years (Lazear and Shaw, 2007; Cross et al., 2016). Alongside this increase in teamwork, increasing gender diversity has become a key objective for organizations (Thakrar, 2017; Pedulla, 2020; Temple-West and Edgecliffe-Johnson, 2020). Improving diversity in organizations comes with trade-offs: a more diverse team increases performance by allowing for specializing and combining more skills and knowledge, but a more diverse team can also have higher costs of communication (Lazear, 1999; Prat, 2002; Hong and Page, 2004; Kahane et al., 2013).

One way communication costs can be mitigated is by pairing workers with co-workers of their choice. Since workers know their own communication costs better than their managers do, allowing workers choice over co-workers can be a cost effective way to achieve the goal of diversity. Existing theoretical and empirical work (e.g. Garicano and Rayo, 2016; Coffman, 2014; Bordalo et al., 2016, 2019; Battiston et al., 2021) has shown that communication problems in organizations can reduce performance. However, whether choosing team members facilitates cooperation and improves performance, especially on gender-stereotypical tasks remains an open question.

In this paper, we investigate the effect of co-worker choice on performance in teams, and test whether this effect differs across genders. We find that choice over co-worker matters for performance, but that the effect of choice differs across genders and across the gender-stereotype of the task. If workers perform better when paired with co-workers of their choice rather than those they are randomly allocated to, allowing workers to choose who they work with could be a low-cost policy for organizations to increase both productivity and gender diversity. Alternatively, this suggests that pairing two co-workers who share affinities can potentially improve performance relative to a pair with marginally higher joint ability but lower affinity.

We designed a lab-in-the-field experiment conducted with undergraduate students in Pakistan. We created teams of two students, with each assigned a different role: 'test-taker' or 'helper'.[1] The test-takers were asked to complete a series of six tests of 30 unique questions each. The helpers were seated next to the test-takers and assigned to support the test-takers during the test.

Our experiment used a within-person design and randomly varied helpers along two dimensions: (1) gender and (2) test-taker preference for the helper. To assign helpers with preferred status, we provided each test-taker at baseline with a list of classmates from their degree year and asked the test-takers to rank them as helpers.[2] Test-takers were then allocated preferred helpers using a random serial dictatorship mechanism (Abdulkadiroğlu and Sönmez, 1998; Breza and Chandrasekhar, 2019). Specifically, test-takers were ordered and then seated based on random numbers. The first seated got the first choice of helpers, the second got their first choice, or their

[1] Using teams of two allows us to minimize the scope for coordination problems that are endemic in large teams (Van Huyck et al., 1990; Weber et al., 2001; Weber, 2006).

[2] The test-takers' preferences for helpers captured their affinities with each helper: The most cited reason for ranking a helper as more preferred was being a close friend to the test-taker (71% of female test-takers and 47% of male test-takers).

second choice if their first choice was already taken by the first test-taker and so on.[3]

This design resulted in four main treatments for test-takers: being allocated either a randomly chosen female helper, a preferred female helper, a randomly chosen male helper, or a preferred male helper. Each treatment was randomly assigned across tests. In the control test, test-takers were not assigned any helper.

Our main outcome of interest was the performance of a test-taker on each question. Our within-person design therefore measures the performance of an individual test-taker in the presence of different types of helpers relative to their performance alone. Test-takers were paid a piece rate according to the number of correct answers in a randomly chosen test, with no penalty for wrong answers.[4] Incentives were kept fixed throughout the experiment. Finally, the questions were grouped into three categories: economics (their subject of study), sports, and cooking. This allowed us to evaluate whether there is any heterogeneity across categories that are considered more or less gender stereotypical by the participants.[5]

Our first result is that random helpers are significantly worse for female test-takers than preferred helpers. We find that being allocated a random helper reduces the score of a female test-taker by 3 percentage points, which corresponds to 6% of their average score. A preferred helper, on the other hand, has a positive yet statistically insignificant effect on the score of female test-takers. The difference between the returns on a preferred vs. random helper are statistically significant at the 5% level. On the other hand, there are no differences across preferred or random helpers for male test-takers.

Our second result is that the difference between random and preferred helpers for female test-takers is entirely driven by male helpers. Random male helpers statistically significantly reduce the performance of females (6 percentage points or 12 % of the average score), while preferred male helpers have a positive yet statistically insignificant effect (3 percentage points or 6% of the average score). This occurs despite the fact that preferred and random helpers have the same average abilities (as measured through a baseline test and their university GPA). There are no differences across preferred or random helpers of either gender for male test-takers. This result suggests that the positive effect of pairing workers based on affinity can be even more important in gender-diverse teams.

Our third result is that the gender-stereotypical nature of the task affects the relationship between the presence of helpers and performance. In particular, in the sports category, random

[3]This resulted in 91% of test-takers working with their top five most preferred helpers whenever they were allocated a preferred helper.

[4]While imposing penalties could have reduced noise by removing the incentives for test-takers to choose an answer at random, existing studies show that men and women tend to react differently when facing penalties for wrong answers (e.g. Baldiga, 2014, Akyol et al., 2020, Coffman and Klinowski, 2020).

[5]A survey of the participants indicated that most participants expected females to have better knowledge in cooking (85% of males and 65% of females), while a majority of males (76%) and a large portion of females (44%) believed that men know more about sports. Around 75% of either gender thought that both genders were as knowledgeable about economics. Testing differences in performance by the gender-stereotypical nature of the task is important, since it has been shown to be an important determinant of behavior differences across men and women (Coffman, 2014; Bordalo et al., 2019; Bordalo et al., 2016).

male helpers have a negative, though statistically insignificant, effect on the score of female test-takers (5 percentage points or 10% of their average score), while preferred male helpers have a positive and statistically significant effect (14 percentage points or 28% of their average score). The difference between preferred and random male helpers is large and statistically different from zero in this category. Since sports is the category with the highest gender gap in knowledge and sports is where females stand to benefit the most from male helpers, these differences are important. This implies that females would have performed significantly better on this task had they always been given the choice of whom to work with. In fact, they would have even performed better alone than with a random male helper. Results are different in the case of a gender-stereotypically female category like cooking. In this category, we find that working with males, irrespective of affinity, reduces performance. Both random and preferred male helpers lower the performance of females (12 percentage points, or 20% of the average score), while female helpers have no effect, irrespective of affinity. We find that the gender stereotype of the tasks also matters for male test-takers. In sports, for example, a random female helper is strictly worse for a male test-taker than a preferred female helper. In cooking, the reverse is true: a random female helper is strictly better for a male test-taker than a preferred female helper. This suggests that choice matters for males too, but only in tasks that are gender-stereotypical, not overall. Like the case of female test-takers, preferred helpers are better in gender-stereotypically male categories like sports.

We then investigate the possible mechanisms behind these results. More specifically, we test whether the presence of certain types of helpers created frictions in accessing useful information for test-takers. To do so we randomly allocated hints to questions within each test taker - helper pair. There was no penalty for taking hints and these provided clues to the correct answer. Hints were available through scratch-off sheets. We use the scratched-off panels from these sheets to measure how often test-takers accessed hints in the presence of different helpers and whether it affected their performance.

To guide our interpretation of this mechanism, we propose a simple theoretical framework. A test-taker begins with a prior belief about which answer is most likely to be correct. If the test-taker does not have access to either a hint or a helper, the test-taker will pick an answer according to this prior belief, and performance alone therefore captures the prior belief or ability of the test-taker. When a helper is available but a hint is not, the test-taker processes the information from the helper as an additional source of information, updates her beliefs, and chooses the most likely answer. Finally, when a hint is available, the test-taker evaluates the expected gains from using the hint and uncovers the hint if the marginal gain of doing so outweighs the marginal cost. Since there is no penalty on accessing hints the marginal costs can be thought of as the direct time costs of uncovering hints, or the psychological costs of accessing help in the presence of helpers.

We find that in the cooking category, the negative effect of random male helpers on female test-takers can be explained by the lower use of alternative sources of information by the test-taker

in their presence. Female test-takers are 7 percentage point (9% of average hints scratched off) less likely to scratch off hints in cooking when they are paired with random male helpers than when they are working alone. This leads them to perform worse with a random male helper than alone in questions where hints are available. Unlike female test-takers, male test-takers never access fewer hints in the presence of helper, regardless of the type of helper, and there is no situation in which their performance suffers.

We conclude the paper with a discussion of an important puzzle for female test-takers that remains unexplained by hint access behavior. Female test-takers perform worse in cooking when paired with preferred male helpers, relative to working alone, but do not access fewer hints. This negative effect is not present in the case of questions without hints, suggesting that this is not the result of sabotage or pernicious helper behavior. We suggest three classes of explanations for this result: social image concerns (Bursztyn and Jensen, 2017; Bursztyn et al., 2017; Chandrasekhar et al., 2018), difficulty in communicating information between helpers and test-takers, and misperceptions of helper's ability.

Our results have important policy implications for workplace diversity initiatives. First, allowing women to select their advisors or teammates of the opposite gender can be a low-cost way for organizations to improve both information flow and productivity. In situations where this is not possible, managers can form teams around workers' affinities, rather than just using skill complementarities. Secondly, cooperative team-building activities can improve performance, as they can effectively turn randomly-assigned co-workers into friendly co-workers (see e.g. Boisjoly et al., 2006; Paluck, 2006, 2016; Lowe, 2021). Finally, the recent push to increase diversity in pink-collared jobs, such as nursing or social care (Delfino, 2021) might need to be carefully implemented along with other measures. For example, if information is lost due to communication frictions, then communication training might mitigate the negative effect of being paired with an unknown co-worker of a different gender.

Our results are particularly relevant to conservative environments where men and women interact less frequently. Beyond Pakistan, our findings have implications for other developing countries like India and Bangladesh and those in the Middle East and a large part of Africa that have similar conservative social norms and a high gender disparity in outcomes (Crotti et al., 2020).

Related literature. Our paper adds to the literature on communication in organizations. Following a large theoretical literature (e.g. Garicano, 2000; Dessein, 2002; Alonso et al., 2008), a recent literature has investigated empirically the importance of communication for performance (e.g. Bloom et al., 2014; Battiston et al., 2021; Menzel, 2021). Specifically, we complement recent studies which show that gender differences can constrain communication or affect the selection of team members, and that these effects can vary by the gender-stereotypical nature of the task (Coffman, 2014; Bordalo et al., 2016, 2019; Coffman et al., 2021). We complement this literature by

showing that the effects of gender differences can be mitigated if workers are allowed choice over co-workers.

Our paper also relates to and complements the organizational economics literature on peer effects on performance at the workplace. Theoretically Kandel and Lazear (1992) and Rotemberg (1994) highlight why peers might matter for performance. Empirical studies show that peers can matter due to productivity spillovers (Mas and Moretti, 2009; Falk and Heckman, 2009; Guryan et al., 2009); or incentives in the presence of social preferences (Hamilton et al., 2003; Bandiera et al., 2005; Blanes i Vidal and Nossol, 2011; Cohn et al., 2014; Breza et al., 2018); or peer pressure or mutual monitoring (Kandel and Lazear, 1992; Jones and Kato, 1995; Knez and Simester, 2001). Ashraf and Bandiera (2018) give an excellent survey of the key theoretical and empirical papers on social incentives at the workplace. Some studies have shown that friendships between co-workers matters for performance in organizations (Bandiera et al., 2010; Park, 2019) and that attending a business training with a friend improves business activity for women who have restricted mobility (Field et al., 2016). A few studies investigate the effects of peer diversity on performance. Rasul and Rogger (2015) find a positive association between increased ethnic diversity and performance in the public sector in a developing country. Alvarez Pereira and Aman-Rana (2021) show that both group size and gender matters for the effects of skills diversity of peers on performance. Hjort (2014), Hedegaard and Tyran (2018), and Dasgupta et al. (2020) show that changing incentives can reduce taste-based discrimination against ethnically diverse co-workers. Using lab experiments in which they vary social identity artifically, Chen and Li (2009) and Chen and Chen (2011) show that diversity in teams reduces performance. Hoogendoorn and Van Praag (2012) and Hoogendoorn et al. (2013) look specifically at the trade-off between increased information sharing and reduced communication costs due to ethnic or gender diversity. They find that the former outweighs the latter and that teams with a better gender balance and more ethnic diversity perform better. Marx et al. (2021) find that ethnic diversity decreases team performance when diversity varies among workers at the same level, but that it increases team performance when diversity varies between workers and their supervisors.

This paper contributes to that literature by investigating gender peer effects, focusing in particular on information exchange and how that can be facilitated by allowing workers to choose their co-workers. Calder-Wang et al. (2021) is the closest related study within this field. Their study exploits two sources of variation. First, in some cohorts, students were randomly assigned to teams based upon a computer algorithm. Second, in other cohorts, students were allowed to choose their teams from among students in their section. They find that exogenously-created diverse teams perform worse than exogenously-created homogeneous teams, but that this negative effect is alleviated in cohorts in which teams are allowed to be endogenously formed. While the results for exogenous teams can be viewed as causal, the authors caution against interpreting the endogenous team results as causal. Our paper complements the Calder-Wang et al. (2021) study by providing

causal evidence using experimental variation in teams, based on the latest developments in the empirical peer effects literature (Angrist, 2014; Caeyers and Fafchamps, 2016). Another closely related paper is Hahn et al. (2019) who study whether primary school students perform better when allocated to groups of friends or groups of randomly assigned peers. They find that working with friends significantly improved the performance of female students but not that of male students, in line with our results. However, in their setting they find that the effect of friendship is not due to the gender of the peers. By contrast, because we randomised both the choice over co-workers and the gender, we can show that the effect of working with preferred peers differs by gender. Our notion of 'preferred' co-worker is also broader than friendship as participants might select workers with whom they work well even if they would not classify them as friends. Our study also differs from these two studies in two important ways. First, by randomly allocating hints across helpers, we are able to directly investigate whether the effect of helpers on performance depends on the availability of alternative sources of information. Second, by varying the gender stereotype of the task, we are also able to test how the nature of the task interacts with gender and preference for helpers.

Since helpers act in an advising capacity, our paper also complements the literature on mentoring. Studies in the psychology and administrative science literature show that mentoring relationships are stronger between members of the same demographic group (Dreher and Cox Jr, 1996). Studies in economics have shown that minority students benefit from minority instructors (Carrell et al., 2010; Fairlie et al., 2014; Hossain, 2021). Matsa and Miller, 2011 show that a higher share of women on boards of directors results in a higher proportion of female top executives the following year. Husain et al., 2018 show that male teachers prefer working with male principals.[6] Our results also show that male helpers can lower performance for female test-takers. However, we contribute to the literature by showing that these problems only exist for those male helpers that are randomly chosen and depends on the gender stereotype of the task.

The paper also relates to the broader literature on how diversity affects economic development.[7] Based in a developing country, our paper provides micro-evidence on how gender interactions take place in organizations and how they can impact information exchange across genders and performance depending on choice.

[6] Athey et al. (2000) and Muller-Itten and Oery (2021) describe the theoretical long-run effects of different affirmative action policies based on the idea that same-type mentorship is more beneficial for minorities. Blau et al. (2010) and Ginther et al. (2020) show that female academics benefit from female mentorship.

[7] Multiple papers consider the effects of ethnic or cultural diversity (Alesina and Ferrara, 2005) and how ethnic diversity increases social anomie and lowers housing quality (Algan et al., 2016); reduces preferences for redistribution (Dahlberg et al., 2012); reduces social solidarity, social capital, or trust (Putnam, 2007, Delavande and Zafar, 2019); negatively impacts choice of public policy and economic growth (Easterly and Levine, 1997); and reduces firm productivity (Parrotta et al., 2014). Chattopadhyay and Duflo (2004) focus on gender diversity and show that reserving a council seat for a woman affects the council's investment in infrastructure that is directly relevant to the needs of women. Cheema et al. (2019) show that physical distance poses a significant hurdle for women to access the opportunities offered to them.

2 Context and research design

2.1 Context

In 2020, we collaborated with a university in Punjab, Pakistan to run the experiment. The experiment was conducted with male and female undergraduates pursuing degrees in Business Administration, Commerce, and Public Administration. All students therefore studied economics, which we could use as a gender-neutral reference category. The experiment took place on campus and was conducted in one session.

2.2 Incentives.

Incentives were kept fixed throughout the experiment. Both test-takers and helpers were given PKR 100 as a participation fee. Test-takers were paid a piece-rate according to the number of correct answers and were therefore incentivized to maximize their performance. For them, one test was randomly selected out of the six and they received PKR 15 for each correct answer in that test. Test-takers were also given an additional PKR 100 for correctly guessing their rank in the group at the end of the test. Overall, test-takers could receive a maximum payment of PKR 650 (approximately equal to 4.22 dollars). Helpers got a flat fee of PKR 500 for helping test-takers perform. Together with the participation fee, helpers received a maximum of PKR 600. In addition, all participants could play a game that could allow them to increase their PKR 100 participation fee by two and a half times. Finally, through a random drawing, six participants were given mobile phone power banks.[8]

2.3 Timing.

The experiment was carried out in one single day and progressed as follows: After consent was obtained, each participant filled out a survey that asked about their preferences for helpers. They were also given a short quiz to test their understanding of questions related to their confidence in their performance. Next, all participants (test-takers and helpers) were given a baseline test to measure their test-specific ability in the absence of hints and helpers. Tests two through six were then given to the test-taker group with assignment into different helper treatments in each test. After the sixth test, a second survey was given on demographics, time use, opinions, and other traits. After the experiment, their scores were calculated and payments were made through a mobile payment app (see Figure 2 for details of the sequence of activities).

[8]The stakes were high enough that we received many requests later to repeat the experiments so that participants could earn more money.

2.4 Differences in research design to estimate peer effects.

Our research design relies on recent developments in the empirical estimation of peer effects and differs from previous peer effect studies in at least two important ways. First, we only study the outcomes of test-takers. Helpers are only used to induce variation in co-worker characteristics, and we do not measure their performance. This bifurcation of the subjects and their peers helps overcome mechanical correlations in outcomes that has been highlighted as a problem in the estimation of peer effects in the literature (Angrist, 2014). Moreover, we ensure strong variation in peer characteristics by identifying the peer characteristics of interest at baseline, i.e., gender and preference over helpers, and then randomly allocating helpers with such characteristics to test-takers. This becomes equivalent to an allocation to a treatment arm rather than an allocation to groups (with chance variation in characteristics). Second, we stratified the students on degree year-gender-GPA and created separate groups of students at baseline from whom test-takers and helpers were chosen. Specifically, for each degree-year we created lists of male and female students. Within each separate list of male and female students we created groups of above and below median students. These were then randomly split into two groups. One was considered the group of test-takers and other the group of helpers. Test-takers and helpers were chosen without replacement from each of these separate groups. This sampling (from separate groups created at baseline) helps overcome what has been called "exclusion bias" in the estimation of peer effects in the literature: a bias that cannot be overcome through random assignment of peers alone (Guryan et al., 2009; Caeyers and Fafchamps, 2016; Angrist, 2014).[9]

2.5 Experimental design

Our experiment aims to assess whether gender and preference over co-worker matter for performance. Our lab-in-the-field approach allows us to control the team composition as well as the nature of the task, while at the same time allowing students to perform on knowledge-intensive tasks with which they are familiar. Indeed, these students regularly complete assignments in groups and frequently take tests.

The experiment was conducted in person. We created teams of two people whom we call "test-takers" and "helpers." As the name suggests, test-takers took tests and were the main participants of interest, while helpers helped test-takers perform on the tests. The research design was a within-person rotation design intended to observe how the same test-taker's behavior and performance changed across treatments with different helpers. This allowed us to control for unobserved ability of the test-takers, including their ability to choose a preferred helper that could maximize their performance. Since we were constrained in the number of female students, rotation design was

[9]Exclusion bias arises from the fact that subject i and i's peers all come from the same pool and, without replacement, "i cannot be his own peer". This creates a small sample negative relationship between subjects' characteristics (Caeyers and Fafchamps, 2016, p.2).

also helpful to increase statistical power.

At baseline, we randomly picked some participants to be test-takers and others to be helpers. This resulted in 85 male and 85 female students (as test-takers) and 69 male and 69 female students (as helpers) that we approached via email to participate in the study (see Appendix Table A1 for the distribution of test-takers and helpers sampled across degrees). The 154 female students we approached constituted the universe of female students across these degree-years. Out of the invited participants, 48 male and 54 female students participated in the experiment as test-takers, while 42 males and 51 females participated as helpers.

We matched helpers and test-takers on the day of the experiment. We began by randomly seating the test-takers. Each test-taker selected a number from a bowl and lined up in ascending order based on their number. Test-takers were then allocated a seat using that random order. Test-takers remained seated for the whole experiment, while helpers moved. The next sections describe how the helpers were randomly allocated to test-takers in different treatments.

For logistical ease, as well as to ensure that test-takers were allocated their preferred helpers, we had to bifurcate random and preferred helpers across rounds. In the second and sixth tests, half of the test-takers were randomly selected to work with preferred helpers, while the other half did their tests alone (see figure 3 for an example of the allocation across rounds 2-6). While the random helper treatment was in the third, fourth, and fifth round of tests (see figure 3 for an example of the allocation across rounds 2-6).

2.5.1 Random Helpers

For the random helper treatment, we asked the students in the helper group to select papers from a bowl with numbers on them. We then asked them to stand in ascending order based on their number, and used this order to allocate them to test-takers.

Gender was never mentioned to the participants throughout the experiment to avoid making gender salient and avoid any potential change in behavior (see e.g. Hoff and Pandey, 2006; Benjamin et al., 2010; Boschini et al., 2012; Bordalo et al., 2012; Bertrand and Duflo, 2017). However, the enumerators were trained to alternate between allocating a random female helper, no helper, and a random male helper within each row.

The pattern was switched across rounds. After each test round, helpers were asked to come to the front of the room while test-takers remained seated. Helpers were asked to choose papers from the bowl again, and the allocation exercise was repeated but with a different type of helper for each test-taker across rounds. This generated random variation in the types of helpers that test-takers worked with across tests.

2.5.2 Preferred Helpers

After consent and before the start of the experiment, we asked test-takers to rank their preferred helpers for the study, from the sample of helpers. The gender of the helpers was not indicated in the list, and the ranking of preferences was not separated by gender. Preferred helpers were allocated across test-takers using a random serial dictatorship allocation mechanism. The first test-taker to be allocated a seat was matched with the most preferred helper on their list, the second one was matched with her most preferred helper unless that helper was already allocated to the first test-taker. This process then continued until all test-takers who were due to be allocated a helper were paired with one. This allocation mechanism has been shown to have several desirable properties that prevent test-taker from strategically manipulating the allocation.[10]. It is also easy to implement in practice and easy to analyze given the randomization of the choice order. Table 3 describes whether test-takers worked with their most preferred helpers in the preferred treatment. 57% of female and 50% of male test-takers worked with their most preferred helpers, while 87% of female and 78% of male test-takers worked with one of their top three most preferred helpers. 91% of test-takers (male and female) worked with one of their top five most preferred helpers.

Table 2 describes the proportion of random and preferred helpers of each gender, by the gender of test-takers. Among those allocated as preferred helpers to female test-takers, 75% were female and 25% were male. This suggests that we have low power in the case of preferred male helpers matched with female test-takers. The distribution is similar but reversed and less skewed for male test-takers. 39% of male test-takers worked with preferred helpers that were female, while 61% worked with preferred helpers that were male. A stronger preference for female (male) helpers by female (male) test-takers is in line with the literature on homophily.[11]

2.5.3 Hint availability

To investigate whether there is a change in the use of alternative sources of information by the test-taker in the presence of these helpers, we randomized the availability of hints for questions within each test. Hints were randomly available for 80% of questions in any test. Hints were prepared by the research team.[12] These were simple cues to help answer the corresponding question. We made hints available through sheets which could be scratched off to uncover hints for any question. Providing hints in such a way allowed us to observe the exact questions for which information was accessed by test-takers. There was no penalty for taking hints.

The table below summarizes our treatments. Across test rounds, each test-taker was randomly

[10]Random serial dictatorship is Pareto efficient and strategy-proof (i.e., it cannot be manipulated by misrepresenting preferences) (Abdulkadiroğlu and Sönmez, 1998; Breza and Chandrasekhar, 2019)

[11]Homophily is the tendency of various types of individuals to associate with others who are similar to themselves (Lazarsfeld et al., 1954). There is a vast literature that documents the presence of homophily across many characteristics like age, religion, gender, or race (see McPherson et al., 2001). In the case of gender, Ibarra (1992) shows that the structure of social networks depends on gender.

[12]A pilot test was able to simplify hints so that students found them useful.

matched with either a random or preferred helper of either gender. The counterfactual was test-takers performing alone without any helpers.

Table 1: Treatments for test-takers

Helpers	Hint available	
Random female helper	yes	no
Preferred female helper	yes	no
Random male helper	yes	no
Preferred male helper	yes	no
No helper (control)	yes	no

Figure 4 presents details of the distribution of treatments across test-takers for each round of the experiment.[13] Appendix tables A2 and A3 describe the proportion of test-takers with helpers of different types across different test rounds. Each test-taker was supposed to get a treatment once. However, there were a few instances when a few test-takers either got a treatment more than once or did not get either the random or preferred helper treatment. Appendix Table A6 to A15 presents balance tables showing that these test-takers were not systematically different across rounds. Finally, in 75 test-taker-test rounds, some test-taker-helper matches were not from within the same degree-year. This was the result of a misunderstanding by one enumerator. All questions in these 75 test rounds were dropped from the analysis.

3 Theoretical framework

We conceptualize helpers and hints as two independent sources of information that provide the test-taker with a signal correlated with the correct answer. To fix ideas, we consider a simplified model where test-takers face two possible choices per question: A or B. Let $X \in \{A, B\}$ denote the correct answer, and let $x \in \{A, B\}$ denote the answer chosen by the test-taker. The test-taker gets a payoff of 1 if she picks the right answer: $x = X$ and a payoff of 0 otherwise.

The test-test-taker has some prior $\mu = \Pr(X = A)$ that the correct answer is A. Without loss of generality we let A be the answer that the test-taker would choose by default, that is: $\mu > \frac{1}{2}$. Without any additional information, the test-taker's expected score would be: $S_T = \mathbb{P}(X = A) \times 1 + \mathbb{P}(X = B) \times 0 = \mu$.

The hint provides an informative signal $h \in \{h_A, h_B\}$ according to $\mathbb{P}(h_X|X) = p_h > \frac{1}{2}$. The hint therefore indicates the correct answer with probability p_h, but can also suggest the wrong answer with probability $1 - p_h$ (it could be misinterpreted for example). Similarly, the helper

[13]Since this is a within-person design, conditional on test-taker fixed effect, the variation exploited in the study is within person across multiple test rounds. Therefore, non-random attrition does not create systematic differences across the treatment and control groups in the usual sense. Balance tables that tested whether people who left were systematically different from those who stayed can be found in appendix tables A5 and A11. Appendix Table A5 shows that for female test-takers those who left were not systematically different from those who stayed. However, for male test-takers Appendix Table A11 shows that those who were younger and had a higher GPA were more likely to leave the test.

shares a message $m \in \{m_A, m_B\}$ correlated with the correct answer, but uncorrelated with the hint conditional on the correct answer. We assume that that message is informative and that its precision is $\mathbb{P}(m_X|X) = p_m > \frac{1}{2}$. Given that helpers have no incentives to compete with the test-takers in the experiment, we assume that the helper truthfully discloses his signal and does not attempt to mislead the test-taker. Following the helper's advice increases the chances of getting the correct answer, but it does not guarantee the right answer.

Suppose first that the test-taker only has access to the hint. The test-taker reads the hint and updates her belief about the right answer according to Bayes rule: $\Pr(X|h_X) = \frac{p_h \mu}{p_h \mu + (1-p_h)(1-\mu)}$. If that posterior belief remains above $\frac{1}{2}$ the test-taker chooses answer $x(h) = A$, otherwise she chooses $x(h) = B$. Similarly, when the test-taker has access to the helper only, she updates her belief given the message shared by the helper and chooses the answer $x(m)$ that she believes is most likely to be correct. Finally, when the test-taker has access to both the hint and the helper she updates her beliefs given both signals and chooses the most likely answer $x(h, m)$.

Since hints are hidden, there is a (potentially very small) cost of uncovering them. Test-takers therefore strategically choose whether to uncover the hint or not by comparing their expected score with the hint to their expected score without it. Let κ denote the cost of uncovering a hint. The cost of taking the hint includes both the direct effort cost and the opportunity cost of spending more time on future questions. To summarize:

1. If a hint is available but a helper is not, the test-taker evaluates her expected score with and without the hint and decides whether to take the hint. If she does, she updates her beliefs and chooses an answer.

2. If a helper is available but a hint is not, the test-taker listens to the helper's advice, updates her beliefs and chooses an answer.

3. If both a helper and a hint are available, and if the hint is uncovered, the test-taker receives the two independent signals, updates her beliefs using both signals and chooses an answer. If the hint is not uncovered, she only processes the helper's advice and chooses an answer.

Let S_T denote the expected score of the test-taker when she has access to neither hint nor helper, S_H the expected score when the test-taker has access to the hint only (and takes it), and S_M the expected score when she has access to the helper only. Finally, let S_{MH} denote the score when she has access to a helper and a hint (and chooses to use the hint).

The test-taker's utility is her expected score given the information available net of the cost of taking the hint, should she decide to do so. Let $k \in \{0, 1\}$ denote her decision to scratch off the hint

($k = 1$) or not ($k = 0$). If no hint is available, set $k = 0$. The test-taker's expected utility is then:

$$U(k) = \begin{cases} k(S_H - \kappa) + (1-k)S_T & \text{with no helper,} \\ k(S_{MH} - \kappa) + (1-k)S_M & \text{with a helper.} \end{cases}$$

3.1 Expected value of hints, hint uncovering, and expected score

To compute the expected score of the test-taker in the different treatments (hint available, helper available, both available), we proceed as follows. We first compute the expected score with and without a hint to obtain the test-taker's expected gain from scratching off a hint. Comparing this expected gain to the cost of uncovering the hint (κ) determines whether the test-taker chooses to scratch off the hint ($k = 1$) or not ($k = 0$). Given this choice, we can derive predictions about the score we should observe in different treatments.

The answer chosen by the test-taker depends on the precision of the helper and the hint, relative to the strength of the test-taker's prior belief. For instance, if the helper's signal is not sufficiently precise, and the test-taker is confident that the correct answer is A, then the test-taker might choose answer $x = A$ even when the helper suggests that the answer is B ($m = m_B$). The relative strength of the signals also determines whether the test-taker chooses to follow the hint or the helper when the two provide contradictory signals. Since the hints were designed by the research team who knew the correct answers, we assume throughout the paper that the hint is more informative than the helper. We focus on the case where the helper is more knowledgeable than the test-taker, as this is the case where the helper has the most influence over the test-taker. These assumptions imply that $p_H > p_M > \mu$.

The expected scores given different sources of information are:

- Test-taker alone: $S_T = \mathbb{P}(X = A) = \mu$
- Test-taker with hint: $S_H = \sum_{X \in \{A,B\}} \mathbb{P}(X) \sum_{h \in \{h_A, h_B\}} \mathbb{P}(h|X) \mathbb{1}\{x(h) = X\}$
- Test-taker and helper, $S_M = \sum_{X \in \{A,B\}} \mathbb{P}(X) \sum_{m \in \{m_A, m_B\}} \mathbb{P}(m|X) \mathbb{1}\{x(m) = X\}$
- Test-taker with hint and helper:

$$S_{MH} = \sum_{X \in \{A,B\}} \mathbb{P}(X) \sum_{m \in \{m_A, m_B\}} \sum_{h \in \{h_A, h_B\}} \mathbb{P}(h|X) \mathbb{P}(m|X) \mathbb{1}\{x(m,h) = X\}$$

where $\mathbb{1}\{x = y\}$ denotes the indicator function and takes value 1 if $x = y$.

Given these expected scores, we can obtain the expected gain from scratching off the hint when alone as $S_H - S_T$ and its expected gain in the presence of the helper as $S_{MH} - S_M$, where S_H and S_{MH} are the expected scores if the test-taker *were to uncover the hint*. The test-taker uncovers the hint when alone if $S_H - S_T > \kappa$ and does so in the presence of the helper if $S_{MH} - S_M > \kappa$.

Finally, let S^*_H and S^*_{MH} be the expected scores of the test-taker (with and without the helper respectively) when the hint is available, taking into account the choice of whether to uncover the hint. If the hint is uncovered in treatments without helpers, then $S^*_H = S_H$, while if the hint is not uncovered in those treatments, $S^*_H = S_T$. Similarly, in the presence of a helper, $S^*_{MH} = S_{MH}$ if the hint is uncovered and $S^*_{MH} = S_M$ otherwise.

In the appendix, we list all possible outcomes (expected score and hint taking behavior) for all possible combinations of parameters.

3.2 Testable predictions

We can use the stylized model above to derive some hypotheses. Our main question is whether different types of helpers (preferred or random, male or female) have different effects on performance. The model suggests that when the helpers have the same ability, there should be no difference in performance across types of helpers. We therefore have the following testable result:

Result 1. *If preferred and random helpers have the same ability, then the gain from adding a random helper should be the same as the gain from adding a preferred helper.*

The statement of result 1 holds both when adding a helper in questions with hints and in questions without hints, and therefore holds when averaging across questions with and without hints.

In addition, the model generates some predictions about the rate at which test-takers uncover hints when facing different types of helpers, and the effect of these hints on the score of the test-takers. We first note that since the test-taker uncovers hints based on the expected marginal gain of the hint, the rate of uncovering hints should only depend on the precision of the hint and the helper relative to the test-taker's prior knowledge.

Result 2. *If the preferred and random helpers have the same ability, then the test-taker should be equally likely to uncover hints in the presence of both types of helpers.*

It is worth emphasizing that what matters for this statement is whether the test-taker perceives these abilities to be the same. In the model, we assume that test-takers have correct perceptions. We discuss the possibility that they misperceive these abilities in Section 6.

Second, an interesting feature of the model is that making a helper available can reduce the test-taker's score when hints are available. The reason this can happen is that the presence of the helper can make the test-taker less likely to uncover the hint. This occurs when the marginal gain from having both helper and hint relative to having the helper alone is smaller than the marginal gain of having the hint relative to having neither. In this case, it is possible for the cost of uncovering the hint to be larger than the marginal gain from having the hint in the presence of the helper, while being smaller than this marginal gain in the absence of the helper. To illustrate

this logic, suppose that the expected scores are as follows: when the test-taker is alone $S_T = 0.2$, with just a hint $S_H = 0.6$, with the helper $S_M = 0.5$, and with both helper and hint it is $S_{MH} = 0.7$.[14] Consider a cost of uncovering hints of $\kappa = 0.3$. When the test-taker considers taking the hint alone, she finds it valuable to uncover the hint since her expected utility if she scratches off the hint is $U(k = 1) = S_H - \kappa = 0.3$ while her expected utility if she does not is $U(k = 0) = S_T = 0.2 < 0.3$. When the test-taker considers taking the hint in the presence of the helper, she no longer uncovers the hint since her expected utility when doing so is $U(k = 1) = S_{MH} - \kappa = 0.4$, while her utility from not scratching off the hint is $U(k = 0) = S_M = 0.5 > 0.4$. In other words, the marginal gain of scratching off the hint when alone ($S_H - S_T = 0.4$) is larger than the marginal cost, but the marginal gain of scratching it off in the presence of the helper ($S_{MH} - S_M = 0.2$) is lower than the marginal cost. As a result, her score with the helper and the hint is $S^*_{MH} = S_M = 0.5$, which is less than her score with just the hint of $S^*_H = S_H = 0.6$.

Since the two sources of information are valuable, the only way that the score can be reduced in the presence of the helper is through this mechanism. This implies that a lower score in the presence of a helper should be accompanied by a lower rate of hint taking in the presence of that helper.

Result 3. *The test-taker's score is strictly lower when both the hint and the helper are present than when only the hint is present if, and only if, the test-taker is strictly less likely to take the hint in the presence of the helper. The test-taker's score can never be lower in the presence of the helper in the questions with no hints.*

Our main result focuses on the value of different types of helpers to the test-taker (Result 1). In the mechanism section, we investigate how the rate of hint-taking can explain the difference in the returns to helpers (Results 2 and 3).

4 Data and descriptive statistics

4.1 Measuring performance

Performance was measured on six rounds of twenty-five minute long tests. A baseline test was administered to both test-takers and helpers. Only test-takers took tests two through six. The questions were multiple choice questions, with four choices and only one correct choice per question. There was no penalty for wrong answers. Each test had three subjects: economics, cooking, and sports. The test had thirty questions in total, with ten questions in each of the three categories. The order of the sections for each test was random. This was done to ensure that there were no section order effects in performance. The questions were randomly chosen for each test from a pool of sixty questions per category. This helped keep the difficulty level similar across tests.

[14] The combination of hint and helper can be less than the sum of the two if, for example, part of the information provided by the helper is redundant when the hint is uncovered.

Questions were not repeated across tests. The cooking and sports category questions were sourced from the internet, while economics questions were taken from economics textbooks (see appendix subsection 10.2 for details).

4.2 Measuring hint uncovered

Hints were available on a question-by-question basis. 80% of the questions had hints available. Test-takers were given a separate sheet that had scratch-off panels where hints for each question were hidden. An example of one of these sheets is shown in Figure 1. The number of panels scratched off could then be counted to determine whether or how many hints the test-takers used and for which questions. Hints did not give direct answers. They only provided some information that could help test-takers arrive at the right answer.

4.3 Gender gap in confidence, baseline test scores and hints uncovered

We measured self-confidence in an incentive-compatible way, using two measures. First, we employed a mechanism suggested by Karni (2009) and implemented by Mobius et al. (2014) and Coffman (2014), in which participants are offered the option to replace their answer by that of a computer with a predetermined accuracy. The lowest accuracy of the computer's answer that the participants are willing to substitute for their own answer measures their confidence.[15]

A second way we elicited self-confidence was by asking test-takers at the end of each test to guess whether they were above the median in the group that took the test with them. If they guessed correctly, they could earn PKR 100.

The top panel of Table 4 presents the test-takers' belief that the answer is correct in the baseline test (without helpers or hint). This is presented both for male and female test-takers. Overall, male test-takers are more confident than female test-takers that their answer is correct. The gender gap in confidence arises in the economics and sports categories but not in cooking. Sports is the category with the largest gender gap in confidence. On the other hand, there is almost no gender gap in confidence in cooking. The bottom panel describes whether the test-takers believed that they performed above average in the group of test-takers present in baseline test. Here, females more than males believe that they performed better than the class in economics and cooking, however the gap is not very large. The gender gap in self-confidence is quite large in sports, consistent with the other measure.

Figure 5 shows the gender gap (female minus male test-takers) in baseline test score by subjects (economics, cooking, and sports) as well as hints uncovered in test rounds when the test-takers were working alone without helpers. There are two main takeaways. First, the overall gender gap in scores is negligible (-0.1); however, the gap in hints taken is much larger (1.2 more hints by

[15]Specifically, the participants were asked: "I think my answer has a % chance of being right. The randomly-drawn computer should answer for me if its accuracy is greater than that." See Coffman (2014), p.1633 for details.

females than males). This remains the case by subjects as well. Irrespective of the gender gap in baseline test scores, the gender gap in hints is positive. Second, females have the biggest knowledge gap in sports. Females score 1 point out of 10 lower than their male colleagues in sports. In line with the knowledge gap, female test-takers take 1.7 more hints in sports than males. Sports is the category in which female test-takers seek the most new information. Therefore, working with a male helper can potentially benefit female test-takers the most in the sports category.

Finally, we asked participants how they perceived the ability of different genders to answer questions in each category. Most participants expected females to have better knowledge in cooking (85% of males and 65% of females), while a majority of males (76%) and a large portion of females (44%) believed that men know more about sports. Around 75% of either gender thought that both genders were as knowledgeable about economics. The results of this survey are presented in Appendix figures A2, A3, and A4. We use these perceptions to categorize cooking as a female-stereotypical category, sports as a male-stereotypical category, and economics as a gender-neutral one.

4.4 Ability of preferred versus random helpers

To investigate whether preferred helpers are systematically selected by test-takers based on their ability, we use two measures.

We first look at survey data to analyze the reasons stated by test-takers for ranking a given helper as their most-preferred helper. Table 5 describes these reasons for the sample of matched test-takers and helpers. 71% of female test-takers stated that being a close friend was the reason they ranked a helper first, while 21% stated that their choice was based on the helper's GPA. Female test-takers also stated being "female" (4%) and "nice" (4%) as reasons for ranking a helper at the top. Male test-takers were similar. 47% gave a helper a top rank because the helper was a close friend, while 13% specified the helper's high GPA as a reason. Male test-takers also stated their reasons for choosing test-takers as: "female" (13%), "nice" (13%), class fellow (7%), caste (7%), and "younger than me" (7%).

Secondly, we run a regression to test whether the allocated preferred and random helpers had systematically different observed ability. This allows us to understand whether a difference in the performance of test-takers across preferred and random helpers is due to a difference in ability of these helpers. We measure the ability of helpers using two sources of data. The first is their GPA, obtained from the administrative data of the university. The second source is the baseline test. This second measure allows us to observe the helpers' ability in all three categories (economics, cooking, and sports). Using this data we estimate the following regressions for a helper (j) matched

with test-taker (i) in a test round (t):[16]

$$Female\ Helper\ Ability_{jit} = \alpha + \beta Preferred\ female\ helper_{jit} + \epsilon_{jit}, \quad (1)$$

$$Male\ Helper\ Ability_{jit} = \delta + \mu Preferred\ male\ helper_{jit} + \varepsilon_{jit}, \quad (2)$$

where $Female\ Helper\ Ability_{jit}$ is measured as either the female helper's GPA in the university or her performance in the baseline test. $Male\ Helper\ Ability_{jit}$ is measured similarly for male helpers. $Preferred\ female\ helper_{iqst}$ is a dummy variable that takes a value of 1 if a test-taker is allocated a female helper that they ranked among their preferred helpers. $Preferred\ male\ helper_{iqst}$ are similarly defined for male helpers. ϵ_{jit} and ε_{jit} are the error terms clustered at the test-taker level. The reference category in equation (1) is random female helpers, while the reference category in equation (2) is random male helpers. Estimating equations (1) and (2) are estimated separately both for male and female test-takers.

Table 6 presents the results. It shows that preferred helpers for both female and male test-takers are not systematically selected on ability. This is the case irrespective of how ability is measured. Following Result 1 from the theoretical model, Table 6 suggests that the gain from adding a random helper should be the same as the gain from adding a preferred helper, since their observed ability is the same.

5 Results: What is the effect of gender and choice over helpers on performance?

We now proceed to test our main research questions: Does performance differ in the presence of preferred rather than random helpers, and do these differences depend on the gender of the helper and the gender-stereotype of the task?

5.1 Estimation

We estimate the following regression for test-taker i in question q, subject s, and test round t:

$$\begin{aligned} Correct\ ans_{iqst} = {} & \pi\ Random\ female\ helper_{iqst} + \phi\ Preferred\ female\ helper_{iqst} \\ & + \omega\ Random\ male\ helper_{iqst} + \theta\ Preferred\ male\ helper_{iqst} + \alpha_i + u_{iqst}, \end{aligned} \quad (3)$$

where $Correct\ ans_{iqst}$ is a dummy variable taking value 1 if a question was correctly answered and 0 otherwise. $Random\ female\ helper_{iqst}$ is a dummy variable that takes a value of 1 if a test-taker is allocated a randomly chosen female helper, while $Preferred\ female\ helper_{iqst}$ is a dummy variable that takes a value of 1 if a test-taker is allocated a female helper that they ranked among

[16] We collapse the data at test-round level as there is no variation in helper types across questions within test-rounds.

their preferred helpers. $Random\ male\ helper_{iqst}$ and $Preferred\ male\ helper_{iqst}$ are similarly defined. All specifications include test-taker fixed effects (α_i), so the variation exploited is within person across treatments. The reference category is test-takers without helpers. u_{iqst} are standard errors that are clustered at a test-taker level, as that is the level at which the treatments are allocated (Abadie et al., 2017). The comparison of interest is between π and ϕ, and ω and θ.

5.2 Results for female test-takers

Table 7, columns (1)-(4) show that a preferred helper has a marginally positive effect on performance while a random helper significantly decreases a female test-taker's performance. To obtain these coefficients, we collapse the gender of the helper and focus on the difference between random and preferred helpers. Recall that Table 6 showed that preferred and random helpers for female test-takers were not systematically selected on ability. Despite that, working with a preferred helper versus a random one generates different effects on performance. Column (1) shows that, overall, working with a random helper reduces the probability of a correct answer by 3 percentage points, relative to working alone. The effect is statistically significant and is 6% of the mean of 0.51. This overall negative effect is significantly different from the effect of a preferred helper. An F-test at the bottom of the table that tests whether the two effects are equal has a p-value of 0.02. This difference remains in all categories except cooking.

In the sports category, shown in Column (4), female test-takers have a 6 percentage points higher probability of a correct answer when they work with their preferred helper. The effect is also economically significant and 12% of the control mean of outcome. This is in contrast to their performance in a stereotypically-female category like cooking. In cooking, female test-takers perform worse when paired with helpers of either type relative to when they are alone. An F-test at the bottom of the table that tests the similarity of the effect has a p-value of 0.60, so we cannot reject the hypothesis that the two types of helpers have different effects in this category. These results suggest that working with preferred helpers has benefits for females when the category is not gender-stereotypically female. Female perform better in stereotypically-female questions when they are alone, than when paired with a male.

To investigate the source of the differences between preferred and random helpers, we break down the helpers by gender in table 7, columns (5)-(8). The results suggest that the entire negative effect of random helpers seems to be driven by male helpers. Overall, random and preferred male helpers have opposite effects on performance. Working with a random male helper reduces the probability of a correct answer by 6 percentage points (12% of the control mean) relative to when female test-takers work alone. The effect is positive for preferred male helpers, although not significant. An F-test that tests the similarity of the effect across random and preferred male helpers has a p-value of 0.15.

The results are similar when we look at a gender-stereotypically male category like sports. In

sports, working with a preferred male helper increases the probability of a correct answer by 14 percentage points relative to working alone. The effect is statistically significant, and the magnitude of the effect is large (28% of the control mean). This is not the case for random male helpers. When paired with random male helpers, the probability of a correct answer decreases by 5 percentage points relative to being alone. Although the effect is not statistically significant, an *F*-test at the bottom of the table testing the similarity of the effects across random and preferred male helpers rejects the null hypothesis (p-value=0.03). These results suggest that in a gender-stereotypically male category like sports female test-takers can benefit from working with males of their choice rather than just any male. Note that in this context, sports is also the category with the greatest gender gap in knowledge, and so the questions in which female test-takers could gain most from males' knowledge.

The results for a gender-stereotypically female category like cooking show a different picture. In such categories, pairing female test-takers with *any* type of male helper reduces their performance. Although the effect is negative and statistically significant only in the case of random male helpers, the magnitude of the effect is similar across random and preferred male helpers. An *F*-test at the bottom of the table that tests the similarity of the effect across random and preferred male helpers fails to reject the null hypothesis (p-value=0.96).

These effects disappear when looking at female helpers. The choice of female helpers makes no difference to the performance of female test-takers. While the magnitude of the positive effect on performance in categories other than cooking is larger for preferred female helpers, the effects are not statistically significant. An *F*-test that tests the similarity of the effect across preferred and random female helpers fails to reject the null. This is true overall, as well as by subjects.

These results suggest two main ideas. First, choice over *male* co-workers is an important determinant of performance for female workers in mixed-gender teams. This is particularly the case in fields that are gender-stereotypically male, which in this context is also the category with the greatest gap in knowledge. This is not the case in gender-stereotypically female fields. In that case, pairing female workers with male workers (irrespective of how they are chosen) results in a decrease in the performance of female workers relative to being alone.

5.3 Results for male test-takers

Table 8 presents the results for male test-takers. Columns (1)-(4) show the effect of a random or preferred helper of any gender on the performance of male test-takers. While Columns (5)-(8) further split these by the gender of the helper.

The lack of overall difference across preferred and random helpers in Column (1) masks considerable heterogeneity across subjects, particularly in gender-stereotypical categories like cooking and sports.

In a stereotypically-female category like cooking, working with a random helper increases the

probability of a correct answer by 6 percentage points (12% of the control mean), relative to being alone. The effect is marginally significant. The effect is opposite with a preferred helper. Working with a preferred helper lowers the probability of a correct answer by 3 percentage points (6% of control mean), however, the effect is not statistically significant. An *F*-test at the bottom of the table that tests the similarity of the effect across random and preferred helpers rejects the null (*p*-value=0.02). This suggests that for males in a gender-stereotypically female category, it is better to work with random helpers than people they prefer.

This difference across random and preferred helpers is reversed when we consider a gender-stereotypically male category like sports. Working with a preferred helper increases the probability of a correct answer by 7 percentage points (13% of the control mean). The effect is negative, though not statistically significant in the case of random helpers. An *F*-test at the bottom of the table that tests whether the effect of random and preferred helpers are equal rejects the null hypothesis (*p*-value=0.03). This suggests that, as with female test-takers, the gender stereotype of the task is important for males. While the results in sports go in the same direction as female test-takers, the effects are very different in the case of cooking.

Table 8, columns (5)-(8) present the results for helpers broken down by gender as well as choice. Column (5) shows that overall neither gender nor choice over helpers is important for male test-takers. However, as before, this average effect masks heterogeneity by category of questions.

The results in Column (8) show that the positive effect of a preferred helper was driven by female helpers. Working with a preferred female helper increases the probability of a correct answer for male test-takers in the sports category by 7 percentage points (13% of the control mean). This is despite the fact that baseline test results show that females know less than males in sports. On the other hand, the effect of a random female helper is negative and an *F*-test at the bottom of the table that tests the similarity of the effect for random and preferred female helpers rejects the null hypothesis (*p*-value=0.01).

This difference across preferred and randomly chosen helpers disappears for male helpers in sports. Neither type of effects are significant and the *p*-value of an *F*-test at the bottom of the table that tests the similarity of the effect for random and preferred male helpers is 0.49. Both these results are similar to the results for female test-takers. Working with the preferred helper of the opposite gender matters more in a gender stereotypically male category of questions.

In the case of cooking things are different. Table 8, Column (7) (as in Column (3)) shows that in this case, random helpers of either gender are better than preferred helpers. An *F*-test at the bottom of the table that tests the similarity of the effect for random and preferred female (male) helpers almost rejects the null (*p*-value=0.10 (0.12)). This is different from the results for female test-takers.

It is possible that the pressure to be perceived as high ability works in the opposite direction in a gender-stereotypically female category. While females do worse with any male, males on the

other hand work better with random helpers of any gender in this category.

Taken together, the results in Table 7 and Table 8 show that affinity with helpers matter and they matter more for gender-stereotypical category of tasks. While in gender-stereotypically male category of questions, preferred helpers of the opposite gender are most helpful for both; in gender-stereotypically female tasks choice matters for males, while gender alone matters for females.

6 Mechanism

The opposite effects of preferred and random helpers on performance are puzzling in light of Result 1 of the model since preferred and random helpers have the same average abilities. When test-takers have access to additional sources of information to help them answer questions, the presence of different types of helpers can affect whether test-takers choose to use these sources of information. There can be several reasons for these differences even when the test-taker expects the different types of helpers to have the same ability. For instance, the decision to access additional information can signal to the helper that the test-taker is not very confident about an answer. A test-taker who is concerned about the helper's perception of her her ability might therefore refrain from accessing this additional information even when it would have been valuable.[17] Alternatively, accessing additional sources of information could signal that the test-taker has doubts about the value of the helper's advice, which could offend the helper. This effect could be differential across preferred and random helpers leading to the opposite effects on performance that we observe.

Our research design allows us to investigate whether additional information was accessed differently in the presence of different types of helpers. We exploit the random allocation of hints across questions as variation in the availability of additional sources of information. The data on hint uncovering allows to evaluate whether test-takers accessed hints differently in the present of different types of helpers and whether the presence of different helpers affected the performance of test-takers in questions where hints were available.

6.1 Hints uncovered and performance

To evaluate whether the rate of uncovering hints varies across types of helpers, we estimate the following regression for a test-taker (i), for question (q), subject (s), and test round (t):

$$\begin{aligned} Hint\ uncovered_{iqst} = \tau\ Random\ female\ helper_{iqst} + \alpha\ Preferred\ female\ helper_{iqst} \\ + \mu\ Random\ male\ helper_{iqst} + \chi\ Preferred\ male\ helper_{iqst} + \beta_i + \epsilon_{iqst}, \end{aligned} \quad (4)$$

where $Hint\ uncovered_{iqst}$ is a dummy variable equal to 1 if a hint was scratched for a question,

[17] Studies have shown that social signalling concerns can lead to frictions in the flow of information and affect outcomes (Bursztyn and Jensen, 2017; Bursztyn et al., 2017; Chandrasekhar et al., 2018)

and the helper dummy variables are defined as in previous specifications. The regression includes test-taker fixed effects (β_i) and exploits within person across helper variations. ϵ_{iqst} are standard errors clustered at a test-taker level.

The counterfactual is test-takers uncovering hints on their own without any helpers. This allows to us to see the effects relative to a situation where the demand for information is uninhibited by the presence of a helper yet depends on each test-taker's ability.

Along with these results we also present results on performance in questions in which hints were available. To test whether changes in hints uncovering behavior had any effect on performance, we use the estimation in equation (3), but restrict the data to the set of questions in which hints were available. The reference category in this case is a test-taker with hints available.

6.1.1 Results for female test-takers

Table 9, Columns (1) - (4), show the effects on the probability of uncovering hints, and Columns (5) - (8) show the results for performance. The data is restricted to questions for which hints are available. Column (1) shows the results for hints uncovered overall, while Columns (2), (3) and (4) show results for economics, cooking and sports, respectively. Similarly, Column (5) shows the effect on performance overall, while Columns (2), (3) and (4) show results for economics, cooking and sports, respectively.

The first thing to note is that when female test-takers work alone, the probability of uncovering hints for any question is 76%. It is the highest in sports (84%) and the lowest in economics (70%) (bottom row in table 9, columns (1), (4) and (2)). We also find that hints had a large positive effect on performance of test-takers (highest in sports and lowest in economics) relative to when they work alone without hints. This suggests that female test-takers found hints to be a useful alternative source of information.

Table 9 confirms that in cooking (Column (3)), female test-takers uncovered significantly fewer hints in the presence of a random male helper. The probability of uncovering hints is 7 percentage points (9% of the mean) lower when test-takers are accessing hints in the presence of random male helpers than when they access hints alone. This effect is marginally significant at the 10% level. Column (7) shows that this is also the category in which test-takers performed 15 percentage points (23% of the control mean) lower in the presence of a helper relative to when they worked alone on questions with hints. This suggests that the mechanism behind a lower score in cooking with random male helpers is a reduction in accessing other sources of information in their presence.

These results are consistent with Result 3: the test-taker's score is strictly lower when both the hint and the helper are present than when only the hint is present if, and only if, the test-taker is strictly less likely to take the hint in the presence of the helper.

We can see a similar, though weaker, negative effect in the sports category. The probability of uncovering hints is 6 percentage points (7% of the mean) lower when test-takers are accessing

hints through random male helpers than when they access hints alone. This effect is marginally significant at the 10% level. Column (8) shows that test-takers performed 6 percentage points (11% of the control mean) lower relative to when they worked alone in questions with hints. However, the effect is not significant.

Some remaining puzzles. Some of the effects of different types of helpers on performance that we observe cannot be rationalised by the hint taking behavior of the test-takers.

Puzzle #1. First, the overall reduction in performance with a random male helper across all categories is much larger than what can be explained by a reduction in hints uncovered. In fact, in Table 9 Column (1) the probability of uncovering hints is just 2 percentage points (3% of control mean) lower with random male helpers and the effect is not significant. On the other hand, in Table 9 Column (5) working with a random male helper lowers the overall probability of a correct answer by 9 percentage points (16% of control mean). Other mechanisms must also be at work.

One possible explanation is that the random male helpers could have actively discriminated against or sabotaged (Lazear, 1989) the female test-takers that they were meant to help. They could do this by wilfully giving them wrong answers or by putting pressure on the test-takers to accept the answer suggested by the helper. Such negative helper behavior could result in a reduction in scores, relative to when the test-taker is working alone. To explore this we look at performance in questions without hints.

Result 3 from the theoretical model suggests that in questions without hints, the test-taker's score can never be lower in the presence of the helper than alone. If the information from the helper is not helpful, the test-taker could always ignore it. However, if the presence of the helper reduces the performance of the test-taker even in questions without hints then this could suggest sabotage by the helper.

Table 10 presents results on performance of female test-takers in questions without hints. Column (1) show the results for correct answer overall, while Columns (2), (3) and (4) show results by economics, cooking and sports, respectively. The reference category is test-taker alone, without a helper or hint. Table 10, Column (1), shows that in the presence of a random male helper, the probability of a correct answer *increases* by 2 percentage points, relative to working alone. However, the effect is not statistically significant. These results rule out negative helper behavior as a main determinant of the fall in performance of test-takers in questions with hints available.

Puzzle #2. It is not clear why working with a preferred male helper in questions with hints in the cooking category lowers the score by 17 percentage points (27% of the control mean), without a corresponding sizeable reduction in hint access. Table 10 shows that there is no negative effect of preferred male helpers in questions without hints either, suggesting that negative helper behavior cannot explain the reduction in performance. Since the rate of hint uncovering is the same with or without this type of helper, the test-taker must be processing the information from the helper and

the hint incorrectly in the presence of the helper.

We consider the following explanations for this puzzle. The unexplained reduction in performance could be due to *(1) social image concerns and (2) difficulty in communicating information between helpers and test-takers*.

Social image concerns can lead the test-taker to choose an incorrect answer in the presence of a preferred male helper because choosing a different answer than the one suggested by the helper signals a lower ability of the test-taker. This can lead the helper to form a negative opinion of the test-taker's ability. If both the test-taker and the helper are of high ability, then the test-taker should be more likely to pick a given answer when the helper also believes that answer to be correct. This would lead the test-taker to pick answers that agree with the helper, even when she does not think this is the most likely answer, and therefore reduce her score. This is particularly plausible in a gender-stereotypically female category where the female test-taker might have additional pressure to appear knowledgeable (particularly when paired with a helper that she knows personally). Similarly, the test-taker might rush into answering questions in this category in order to impress the helper.[18]

Communication issues can also explain why the presence of a helper would reduce the score of the test-taker. If communication issues between the helper and the test-taker distract the test-taker then the test-taker could perform better by herself. This could be particularly true in questions with hints if the presence of a hint leads the helper and test-taker to debate longer about the correct answer. This could explain the lower score in cooking in the presence of a preferred helper that is not accompanied by a lower rate of hint uncovering. It is also plausible that these types of debates are more likely when the test-taker and the helper know each other.

There is a third class of issues that could affect information exchange and performance: misperception of the helper's ability by test-takers. If female test-takers misperceived that male helpers had a high ability in cooking, they could have relied more heavily on helpers than hints, so their scores with a helper could be lower relative to when they worked alone in questions with hints. We think this explanation is unlikely for two reasons. First, female test-takers misperceive male ability in cooking in the opposite direction: females (65%) expected males to have a *lower* knowledge of cooking, even though there was no difference in performance across genders at baseline. Second, we would expect misperceptions to be stronger with a random helper as the test-taker would be less likely to know their ability than that of a preferred helper.

[18]The students in this context are of marriageable age and knowledge in cooking can signal desirable characteristics for the marriage market. It has been shown in the literature that actions that signal desirable characteristics for the marriage market are important for women. Bursztyn et al., 2017 show that female MBA students avoid career-enhancing actions because these actions signal undesirable traits, like ambition, for the marriage market.

6.1.2 Results for male test-takers

Table 11, Columns (1) - (4) show the effects on probability of uncovering hints, and Columns (5) - (8) show results for performance. The data is restricted to questions for which hints were available. Column (1) show the results for hints uncovered overall, while Columns (2), (3) and (4) show results for economics, cooking and sports, respectively. Similarly, Column (5) shows the effect on performance overall, while Columns (2), (3) and (4) show results for economics, cooking and sports, respectively.

For male test-takers the probability of uncovering hints for any question is 61% when they work alone. It is the highest in sports (65%) and the lowest in economics (56%). We also find (results not reported here) that hints had a large positive effect on performance of test-takers (highest in sports and lowest in economics) relative to when they work alone without hints. This suggests that male test-takers also found hints to be a useful alternative source of information.

There are three big takeaways. First, unlike female test-takers, male test-takers never access fewer hints with any type of helper, and there is no situation in which their performance suffers. Second, male test-takers access even more hints in the presence of helpers than they do alone. Third, this extra information in the presence of helpers does not necessarily result in a higher performance. This suggests that the information of helpers and hints potentially overlap, since test-takers perform as well as they would with just hints, in the absence of helpers. The fact that test-takers uncover more hints with these helpers but do not perform better suggests that the test-takers may overestimate the benefits that the hint will bring at the point where they decide to uncover it. This is true in all categories other than sports.

Table 8 showed that, for male test-takers, random helpers were more helpful in cooking, while preferred helpers were more helpful in sports, irrespective of gender.

Table 11 shows that the positive effect of preferred helpers in sports comes in part from the fact that, in questions with hints, preferred female helpers lead the test-taker to take significantly more hints and to perform significantly better than random female helpers. In this category, male test-takers have a 12 percentage points (18% of the control mean) higher probability of accessing hints in the presence of preferred female helpers and this is accompanied by a 10 percentage points (17% of the control mean) increase in the probability of a correct answer. In cooking, the positive effect of random helpers is driven in part by random male helpers improving performance significantly more than preferred male helpers in questions where hints are available. This is despite there being no difference in the rate of hint taking between the two types of male helpers.

Table 12 shows the results in questions without hints. Column (4) of that table shows that the value of a preferred helpers in the sports category relative to a random helper is also present in questions without hints, but that in these questions it is preferred male helpers rather than preferred female helpers that drive the result.

Conclusion

Meaningful gender diversity at work remains an important goal of many organizations. This study highlights that the question of choice over co-workers is central to this problem.

In particular, we find that female test-takers benefit significantly more from being paired with helpers that they chose, rather than randomly assigned ones. We show that this only occurs when the helper is a male, and that the effect is particularly strong in categories where participants have strong gender stereotypes about ability. We show that part of these results can be explained by the fact that test-takers are less likely to access independent information in the presence of these helpers.

These results have implications for team formation in organizations. Organizations can pursue diversity goals by understanding affinities between workers and the gender stereotypical nature of the task. In categories that are not perceived by workers as gender-stereotypically female, allowing female workers to select co-workers of the opposite gender will improve performance in mixed-gender teams.

This study also opens up avenues for further research to explain the puzzles we observe. What can explain the poor performance of female test-takers with random male helpers in a gender-stereotypically female category like cooking? Why do they take fewer hints with random male helpers in sports? Most importantly, understanding the specific mechanism behind the results we observe can determine what specific set of policies are best suited to address the potential productivity losses we identify. Finally, in this study we kept incentives fixed and investigated the role of choice as a policy lever. Future work will need to investigate whether team incentives or alternative reward schemes can help reduce frictions in mixed gender teams.

References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2017). When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.

Abdulkadiroğlu, A. and T. Sönmez (1998). Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica 66*(3), 689–701.

Akyol, Ş. P., J. Key, and K. Krishna (2020). Hit or miss? test taking behavior in multiple choice exams.

Alesina, A. and E. L. Ferrara (2005). Ethnic diversity and economic performance. *Journal of Economic Literature 43*(3), 762–800.

Algan, Y., C. Hémet, and D. D. Laitin (2016). The social effects of ethnic diversity at the local level: A natural experiment with exogenous residential allocation. *Journal of Political Economy 124*(3), 696–733.

Alonso, R., W. Dessein, and N. Matouschek (2008). When does coordination require centralization? *American Economic Review 98*(1), 145–79.

Alvarez Pereira, B. and S. Aman-Rana (2021). Diversity in small or large groups? Experimental evidence. *Working Paper*.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics 30*, 98–108.

Ashraf, N. and O. Bandiera (2018). Social incentives in organizations. *Annual Review of Economics 10*, 439–463.

Athey, S., C. Avery, and P. Zemsky (2000). Mentoring and diversity. *American Economic Review 90*(4), 765–786.

Baldiga, K. (2014). Gender differences in willingness to guess. *Management Science 60*(2), 434–448.

Bandiera, O., I. Barankay, and I. Rasul (2005). Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics*, 917–962.

Bandiera, O., I. Barankay, and I. Rasul (2010). Social incentives in the workplace. *The Review of Economic Studies 77*(2), 417–458.

Battiston, D., J. Blanes i Vidal, and T. Kirchmaier (2021). Face-to-face communication in organizations. *The Review of Economic Studies 88*(2), 574–609.

Benjamin, D. J., J. J. Choi, and A. J. Strickland (2010). Social identity and preferences. *American Economic Review 100*(4), 1913–28.

Bertrand, M. and E. Duflo (2017). Field experiments on discrimination. *Handbook of economic field experiments 1*, 309–393.

Blanes i Vidal, J. and M. Nossol (2011). Tournaments without prizes: Evidence from personnel records. *Management Science 57*(10), 1721–1736.

Blau, F. D., J. M. Currie, R. T. Croson, and D. K. Ginther (2010). Can mentoring help female assistant professors? interim results from a randomized trial. *American Economic Review 100*(2), 348–52.

Bloom, N., L. Garicano, R. Sadun, and J. Van Reenen (2014). The distinct effects of information technology and communication technology on firm organization. *Management Science 60*(12), 2859–2885.

Boisjoly, J., G. J. Duncan, M. Kremer, D. M. Levy, and J. Eccles (2006). Empathy or antipathy? the impact of diversity. *American Economic Review 96*(5), 1890–1905.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016). Stereotypes. *The Quarterly Journal of Economics 131*(4), 1753–1794.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2019). Beliefs about gender. *American Economic Review 109*(3), 739–73.

Bordalo, P., N. Gennaioli, and A. Shleifer (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics 127*(3), 1243–1285.

Boschini, A., A. Muren, and M. Persson (2012). Constructing gender differences in the economics lab. *Journal of Economic Behavior & Organization 84*(3), 741–752.

Breza, E. and A. G. Chandrasekhar (2019). Social networks, reputation, and commitment: evidence from a savings monitors experiment. *Econometrica 87*(1), 175–216.

Breza, E., S. Kaur, and Y. Shamdasani (2018). The morale effects of pay inequality. *The Quarterly Journal of Economics 133*(2), 611–663.

Bursztyn, L., T. Fujiwara, and A. Pallais (2017). 'acting wife': Marriage market incentives and labor market investments. *American Economic Review 107*(11), 3288–3319.

Bursztyn, L. and R. Jensen (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics 9*, 131–153.

Caeyers, B. and M. Fafchamps (2016). Exclusion bias in the estimation of peer effects. Technical report, National Bureau of Economic Research.

Calder-Wang, S., P. A. Gompers, and K. Huang (2021). Diversity and performance in entrepreneurial teams. Technical report, National Bureau of Economic Research.

Carrell, S. E., M. E. Page, and J. E. West (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics 125*(3), 1101–1144.

Chandrasekhar, A. G., B. Golub, and H. Yang (2018). Signaling, shame, and silence in social learning. Technical report, National Bureau of Economic Research.

Chattopadhyay, R. and E. Duflo (2004). Women as policy makers: Evidence from a randomized policy experiment in india. *Econometrica 72*(5), 1409–1443.

Cheema, A., A. I. Khwaja, F. Naseer, and J. N. Shapiro (2019). Glass walls: Experimental evidence on access constraints faced by women. Technical report, Mimeo, Harvard University.

Chen, R. and Y. Chen (2011). The potential of social identity for equilibrium selection. *American Economic Review 101*(6), 2562–89.

Chen, Y. and S. X. Li (2009). Group identity and social preferences. *American Economic Review 99*(1), 431–57.

Coffman, K., C. B. Flikkema, and O. Shurchkov (2021). Gender stereotypes in deliberation and team decisions. *Games and Economic Behavior*.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics 129*(4), 1625–1660.

Coffman, K. B. and D. Klinowski (2020). The impact of penalties for wrong answers on the gender gap in test scores. *Proceedings of the National Academy of Sciences 117*(16), 8794–8803.

Cohn, A., E. Fehr, B. Herrmann, and F. Schneider (2014). Social comparison and effort provision: Evidence from a field experiment. *Journal of the European Economic Association 12*(4), 877–898.

Cross, R., R. Rebele, and A. Grant (Jan-Feb, 2016). Collaborative overload. *Harvard Business Review*, 74–79.

Crotti, R., T. Geiger, V. Ratcheva, and S. Zahidi (2020). Global gender gap report 2020. In *World Economic Forum. http://www3. weforum. org/docs/WEF_GGGR_2020. pdf*.

Dahlberg, M., K. Edmark, and H. Lundqvist (2012). Ethnic diversity and preferences for redistribution. *Journal of Political Economy 120*(1), 41–76.

Dasgupta, U., S. Mani, J. Vecci, and T. Želinskỳ (2020). Game of prejudice: Experiments at the extensive and intensive margin.

Delavande, A. and B. Zafar (2019). Gender discrimination and social identity: Evidence from urban pakistan. *Economic Development and Cultural Change 68*(1), 1–40.

Delfino, A. (2021). Breaking gender barriers: Experimental evidence on men in pink-collar jobs. *IZA Discussion Paper 14083*.

Dessein, W. (2002). Authority and communication in organizations. *The Review of Economic Studies 69*(4), 811–838.

Dreher, G. F. and T. H. Cox Jr (1996). Race, gender, and opportunity: a study of compensation attainment and the establishment of mentoring relationships. *Journal of applied psychology 81*(3), 297.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: policies and ethnic divisions. *The Quarterly Journal of Economics 112*(4), 1203–1250.

Fairlie, R. W., F. Hoffmann, and P. Oreopoulos (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review 104*(8), 2567–91.

Falk, A. and J. J. Heckman (2009). Lab experiments are a major source of knowledge in the social sciences. *Science 326*(5952), 535–538.

Field, E., S. Jayachandran, R. Pande, and N. Rigol (2016). Friendship at work: Can peer effects catalyze female entrepreneurship? *American Economic Journal: Economic Policy 8*(2), 125–53.

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *Journal of Political Economy 108*(5), 874–904.

Garicano, L. and L. Rayo (2016). Why organizations fail: Models and cases. *Journal of Economic Literature 54*(1), 137–92.

Ginther, D. K., J. M. Currie, F. D. Blau, and R. T. Croson (2020). Can mentoring help female assistant professors in economics? an evaluation by randomized trial. In *AEA Papers and Proceedings*, Volume 110, pp. 205–09.

Guryan, J., K. Kroft, and M. J. Notowidigdo (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics 1*(4), 34–68.

Hahn, Y., A. Islam, E. Patacchini, and Y. Zenou (2019, 11). Friendship and Female Education: Evidence from a Field Experiment in Bangladeshi Primary Schools. *The Economic Journal 130*(627), 740–764.

Hamilton, B. H., J. A. Nickerson, and H. Owan (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy 111*(3), 465–497.

Hedegaard, M. S. and J.-R. Tyran (2018). The price of prejudice. *American Economic Journal: Applied Economics 10*(1), 40–63.

Hjort, J. (2014). Ethnic divisions and production in firms. *The Quarterly Journal of Economics 129*(4), 1899–1946.

Hoff, K. and P. Pandey (2006). Discrimination, social identity, and durable inequalities. *American Economic Review 96*(2), 206–211.

Hong, L. and S. E. Page (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences 101*(46), 16385–16389.

Hoogendoorn, S., H. Oosterbeek, and M. Van Praag (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science 59*(7), 1514–1528.

Hoogendoorn, S. M. and M. Van Praag (2012). Ethnic diversity and team performance: A field experiment. Technical report, IZA Discussion Papers.

Hossain, M. A. (2021). Performance gains from gender match in higher education: Evidence from a setting with entrenched gender stereotypes. *Working Paper*.

Husain, A. N., D. A. Matsa, and A. R. Miller (2018). Do male workers prefer male leaders? an analysis of principals' effects on teacher retention. Technical report, National Bureau of Economic Research.

Ibarra, H. (1992). Homophily and differential returns: Sex differences in network structure and access in an advertising firm. *Administrative science quarterly*, 422–447.

Jones, D. C. and T. Kato (1995). The productivity effects of employee stock-ownership plans and bonuses: evidence from japanese panel data. *The American Economic Review*, 391–414.

Kahane, L., N. Longley, and R. Simmons (2013). The effects of coworker heterogeneity on firm-level output: assessing the impacts of cultural and language diversity in the national hockey league. *Review of Economics and Statistics 95*(1), 302–314.

Kandel, E. and E. P. Lazear (1992). Peer pressure and partnerships. *Journal of Political Economy 100*(4), 801–817.

Karni, E. (2009). A mechanism for eliciting probabilities. *Econometrica 77*(2), 603–606.

Knez, M. and D. Simester (2001). Firm-wide incentives and mutual monitoring at continental airlines. *Journal of Labor Economics 19*(4), 743–772.

Lazarsfeld, P. F., R. K. Merton, et al. (1954). Friendship as a social process: A substantive and methodological analysis. *Freedom and Control in Modern Society 18*(1), 18–66.

Lazear, E. P. (1989). Pay equality and industrial politics. *Journal of Political Economy 97*(3), 561–580.

Lazear, E. P. (1999). Globalisation and the market for team-mates. *The Economic Journal 109*(454), 15–40.

Lazear, E. P. and K. L. Shaw (2007). Personnel economics: The economist's view of human resources. *Journal of Economic Perspectives 21*(4), 91–114.

Lowe, M. (2021, June). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review 111*(6), 1807–44.

Marx, B., V. Pons, and T. Suri (2021). Diversity and team performance in a kenyan organization. *Journal of Public Economics 197*, 104332.

Mas, A. and E. Moretti (2009). Peers at work. *The American Economic Review 99*(1), 112–145.

Matsa, D. A. and A. R. Miller (2011). Chipping away at the glass ceiling: Gender spillovers in corporate leadership. *American Economic Review 101*(3), 635–39.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology 27*(1), 415–444.

Menzel, A. (2021). Knowledge exchange and productivity spill-overs in bangladeshi garment factories. *Journal of Economic Behavior & Organization 185*, 721–746.

Mobius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2014). Managing self-confidence.

Muller-Itten, M. and A. Oery (2021). Mentoring and the dynamics of affirmative action. *Economic Policy* (forthcoming).

Paluck, E. L. (2006). Diversity training and intergroup contact: A call to action research. *Journal of Social Issues 62*(3), 577–595.

Paluck, E. L. (2016). How to overcome prejudice. *Science 352*(6282), 147–147.

Park, S. (2019). Socializing at work: Evidence from a field experiment with manufacturing workers. *American Economic Journal: Applied Economics 11*(3), 424–55.

Parrotta, P., D. Pozzoli, and M. Pytlikova (2014). Labor diversity and firm productivity. *European Economic Review 66*, 144–179.

Pedulla, D. (2020, May). Diversity and inclusion efforts that really work. *Harvard Business Review*.

Prat, A. (2002). Should a team be homogeneous? *European Economic Review 46*(7), 1187–1207.

Putnam, R. D. (2007). E pluribus unum: Diversity and community in the twenty-first century the 2006 johan skytte prize lecture. *Scandinavian Political Studies 30*(2), 137–174.

Rasul, I. and D. Rogger (2015). The impact of ethnic diversity in bureaucracies: evidence from the nigerian civil service. *American Economic Review 105*(5), 457–61.

Rotemberg, J. J. (1994). Human relations in the workplace. *Journal of Political Economy 102*(4), 684–717.

Temple-West, P. and A. Edgecliffe-Johnson (2020, Dec). Push for diversity brings rush of business for executive headhunters. *Financial Times*.

Thakrar, M. (2017, Jan). How to lead the push for diversity in the workplace. *Forbes*.

Van Huyck, J. B., R. C. Battalio, and R. O. Beil (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *The American Economic Review 80*(1), 234–248.

Weber, R., C. Camerer, Y. Rottenstreich, and M. Knez (2001). The illusion of leadership: Misattribution of cause in coordination games. *Organization Science 12*(5), 582–598.

Weber, R. A. (2006). Managing growth to achieve efficient coordination in large groups. *American Economic Review 96*(1), 114–126.

7 Figures

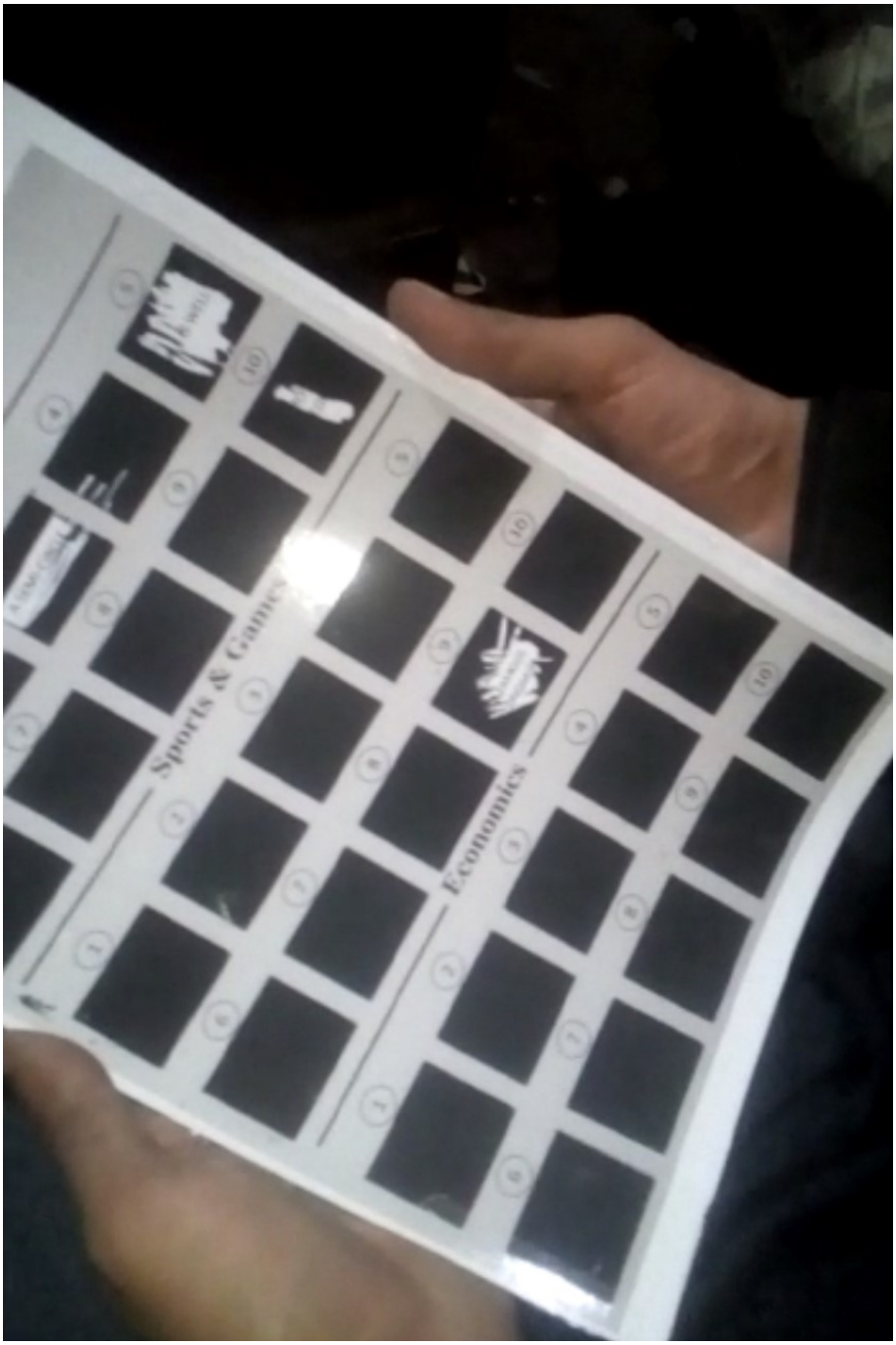

Figure 1: Hint sheet used by test-takers.

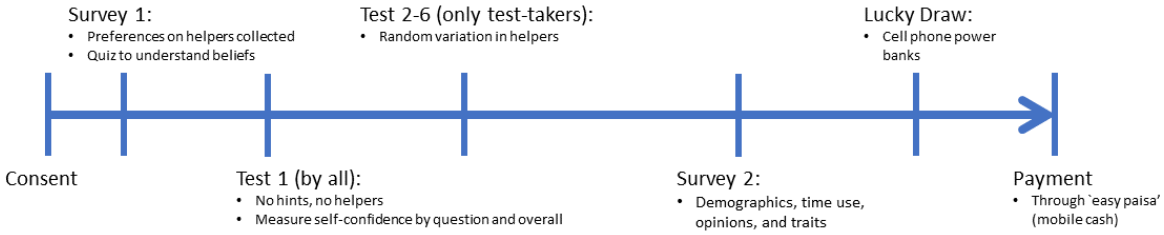

Figure 2: Sequence of activities

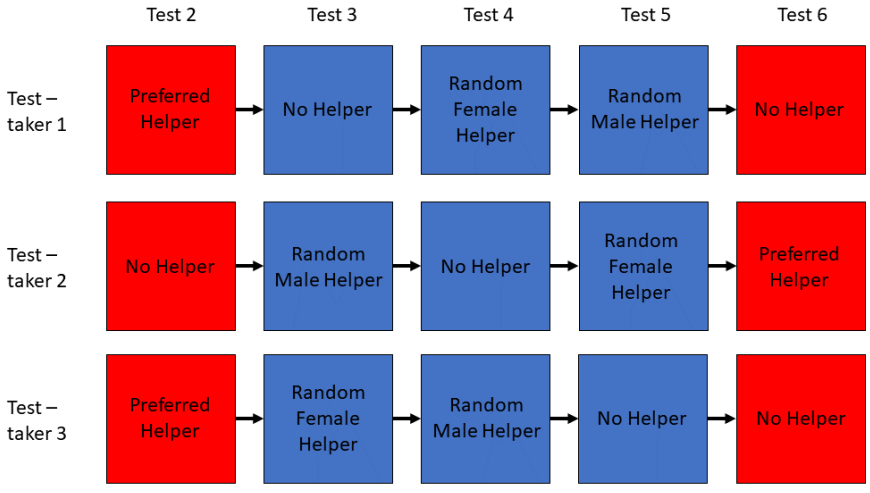

Figure 3: Example of random allocation of treatments per test-taker

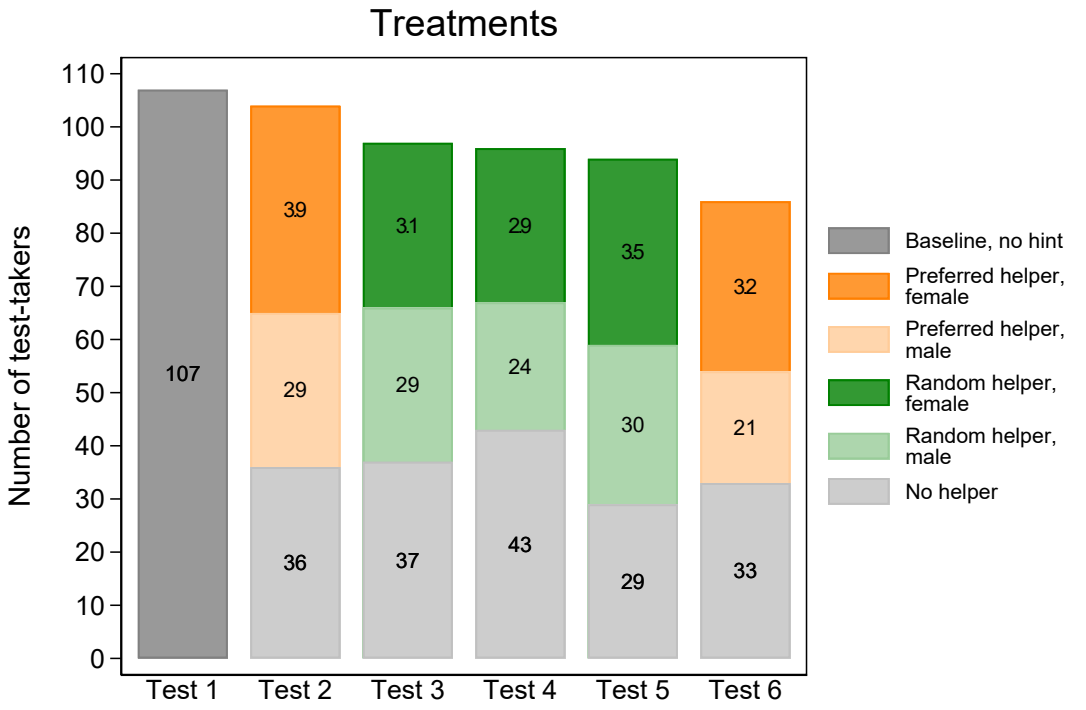

Figure 4: Sample of test-takers with different treatments

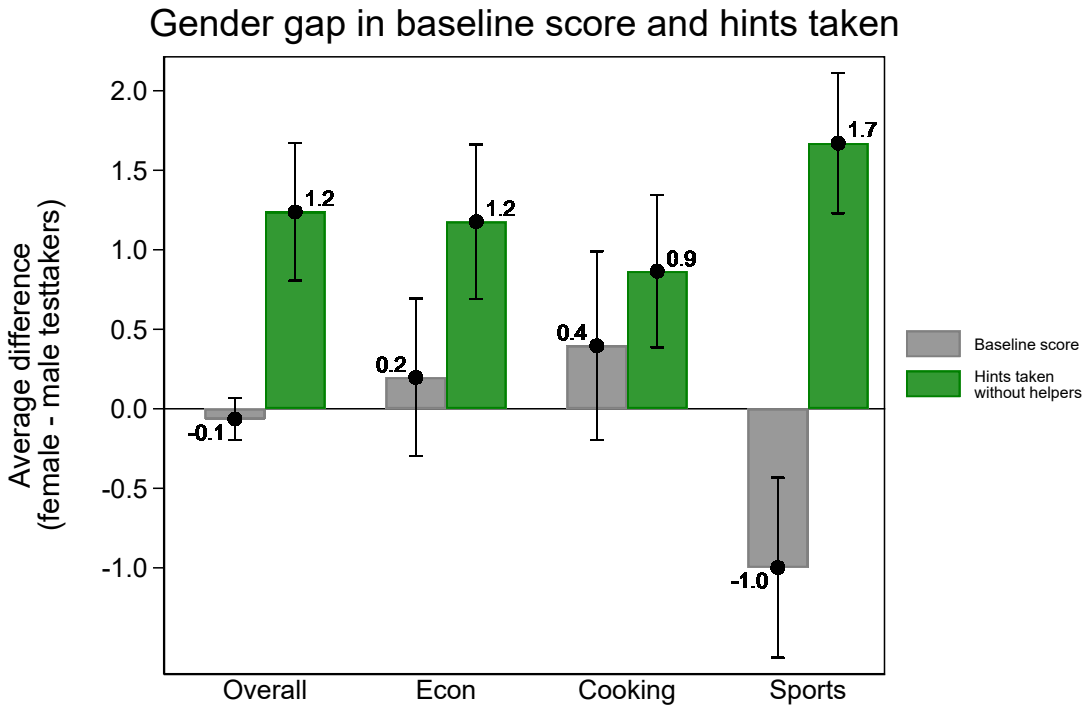

Figure 5: Gender gap in performance and hints taken by test-takers.

8 Tables

Table 2: Proportion of male and female helpers with test-takers

	Female test-taker	Male test-taker
P(female | helper type = preferred)	0.75	0.39
P(female | helper type = random)	0.59	0.56
P(male | helper type = preferred)	0.25	0.61
P(male | helper type = random)	0.41	0.44

Notes: P(female | helper type = preferred) - Proportion of female helpers given that the allocated helper is of a preferred type. P(female | helper type = random) - Proportion of female helpers given that the helper is randomly selected. P(male | helper type = preferred) - Proportion of male helpers given that the allocated helper is of a preferred type. P(male | helper type = random) - Proportion of male helpers given that the helper is randomly selected.

Table 3: In the preferred helper treatment did test-takers work with their preferred helpers?

	Test-taker worked with helper that was:				
	Most preferred	Top three	Top five	Top ten	# of test-takers
Female test-taker	0.57	0.87	0.91	0.96	47
Male test-taker	0.50	0.78	0.91	0.94	32

Table 4: Baseline Descriptive Statistics

	Female	Male	Total
Belief that answer is correct			
Overall	68.96	73.06	70.80
Economics	68.89	72.03	70.28
Cooking	73.73	71.95	72.93
Sports	63.82	74.78	68.79
Proportion that believe they are above the median			
Overall	0.74	0.75	0.74
Economics	0.72	0.69	0.71
Cooking	0.82	0.81	0.82
Sports	0.68	0.82	0.74

Notes: Beliefs that the answer is correct were solicited from test-takers in the baseline test (without helpers or hints). Proportion that believe they are above the median describes the proportion of test-takers that believe that they performed above average in the group of test-takers present in the baseline test.

Table 5: Stated reasons for preference of helpers (sample of helpers matched with test-takers)

	Female test-taker				Male test-taker			
	Most preferred	Top three	Top five	Top ten	Most preferred	Top three	Top five	Top ten
Close friend	0.71	0.55	0.52	0.50	0.47	0.36	0.33	0.32
GPA	0.21	0.17	0.18	0.20	0.13	0.20	0.20	0.19
Female	0.04	0.14	0.14	0.13	0.13	0.08	0.10	0.10
Nice	0.04	0.10	0.11	0.11	0.13	0.12	0.10	0.10
Class fellow	0.00	0.14	0.16	0.17	0.07	0.04	0.10	0.13
From Zaat	0.00	0.00	0.00	0.00	0.07	0.08	0.07	0.06
Not from Zaat	0.00	0.00	0.00	0.02	0.00	0.04	0.03	0.03
Older than me	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Younger than me	0.00	0.00	0.00	0.00	0.07	0.04	0.03	0.03
Male	0.00	0.02	0.02	0.02	0.00	0.28	0.23	0.23
Other	0.00	0.00	0.02	0.02	0.00	0.00	0.03	0.03

Notes: Test-takers could state multiple reasons for preferring a helper. There were seven test-takers who did not state a reason for their partner and were not included. Three of these were female and four were male. Most preferred - The stated reason for selecting the helper that was the most preferred. Top three - The stated reason for selecting the helper that was the among the top three preferred. Top five - The stated reason for selecting the helper that was the among the top five preferred. Top ten - The stated reason for selecting the helper that was the among the top ten preferred.

Table 6: **Are preferred helpers of higher ability than random helpers?**

	Reference category: Random helpers of each category									
Dependent Variables:	Helper	**Helper Baseline Scores**				Helper	**Helper Baseline Scores**			
	GPA	Overall	By Subjects			GPA	Overall	By Subjects		
			Econ	Cooking	Sports			Econ	Cooking	Sports
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Helpers with female test-takers										
Preferred Female Helper	0.06	0.08	-0.14	0.20	0.02					
	(0.12)	(0.62)	(0.33)	(0.39)	(0.29)					
Preferred Male Helper						0.10	0.24	0.42	0.07	-0.25
						(0.18)	(0.77)	(0.40)	(0.51)	(0.51)
Person FE	No	No	No	No	No	No	No	No	No	No
Observations	84	83	83	83	83	42	41	41	41	41
Control mean of outcome	3.30	10.49	2.62	4.67	3.21	3.09	10.33	2.37	4.00	3.96
Panel B: Helpers with male test-takers										
Preferred Female Helper	-0.07	0.45	-0.03	0.46	0.02					
	(0.16)	(0.83)	(0.37)	(0.49)	(0.38)					
Preferred Male Helper						-0.12	-0.44	0.16	-0.20	-0.40
						(0.13)	(0.69)	(0.43)	(0.43)	(0.56)
Person FE	No	No	No	No	No	No	No	No	No	No
Observations	48	47	47	47	47	50	49	49	49	49
Control mean of outcome	3.30	10.48	2.97	4.23	3.29	3.36	11.48	2.76	4.28	4.44

* p<0.1, ** p<0.05, *** p<0.01. Standard errors in parenthesis.
Notes: The unit of observation is at test taker-test round level. Only observations with a helper in any test round have been used. Preferred female helper is a dummy variable that turns on 1 if a test-taker is allocated to a female helper stated to be preferred by the test-takers at baseline. Preferred male helper is a dummy variable that turns on 1 if a test-taker is allocated to a male helper stated to be preferred by the test-takers at baseline. Helper GPA is the GPA of the allocated helper obtained from the university's administrative data. Helper baseline scores are the allocated helper's test-scores at baseline.

Table 7: **What is the effect of gender and choice over helpers on performance?**

Dependent Variables:	*Reference category: Test-takers without helpers*							
	Correct answer (dummy)							
	Overall	By Subjects			Overall	By Subjects		
		Econ	Cooking	Sports		Econ	Cooking	Sports
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample: Female Test-takers								
Random Helper	-0.03**	-0.00	-0.06**	-0.02				
	(0.01)	(0.03)	(0.03)	(0.03)				
Preferred Helper	0.02	0.05	-0.04	0.06*				
	(0.02)	(0.03)	(0.03)	(0.03)				
Random female helper					-0.01	0.01	-0.03	0.00
					(0.02)	(0.03)	(0.02)	(0.03)
Preferred female helper					0.02	0.05	-0.02	0.04
					(0.02)	(0.03)	(0.03)	(0.03)
Random male helper					-0.06**	-0.02	-0.12**	-0.05
					(0.03)	(0.04)	(0.05)	(0.05)
Preferred male helper					0.03	0.07	-0.12	0.14**
					(0.06)	(0.07)	(0.08)	(0.06)
P val: Random = Preferred Helper	0.02	0.10	0.60	0.02	-	-	-	-
P val: Random Female = Preferred Female Helper	-	-	-	-	0.23	0.32	0.88	0.30
P val: Random Male = Preferred Male Helper	-	-	-	-	0.15	0.20	0.96	0.03
Person FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6270	2090	2090	2090	6270	2090	2090	2090
Clusters	54	54	54	54	54	54	54	54
Control mean of outcome	0.51	0.44	0.59	0.50	0.51	0.44	0.59	0.50

* p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered at test-taker level in parenthesis.

Notes: The unit of observation is at test taker-question level. Random helper is a dummy variable that turns on 1 if a test-taker is allocated a helper chosen randomly by the research team. Preferred helper on the other hand is a dummy variable that turns on 1 if a test-taker is allocated a helper stated to be preferred by the test-takers at baseline. Random female helper is a dummy variable that turns on 1 if a test-taker is allocated a female helper chosen randomly by the research team. Preferred female helper is a dummy variable that turns on 1 if a test-taker is allocated to a female helper stated to be preferred by the test-takers at baseline. Random male helper is a dummy variable that turns on 1 if a test-taker is allocated a male helper chosen randomly by the research team. Preferred male helper is a dummy variable that turns on 1 if a test-taker is allocated to a male helper stated to be preferred by the test-takers at baseline. Correct answer is a dummy variable that turns on 1 if the question is answered correctly.

Table 8: **What is the effect of gender and choice over helpers on performance?**

Dependent Variables:	*Reference category: Test-takers without helpers*							
	Correct answer (dummy)							
	Overall	By Subjects			Overall	By Subjects		
		Econ	Cooking	Sports		Econ	Cooking	Sports
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample: Male Test-takers								
Random Helper	0.04*	0.07***	0.06*	-0.01				
	(0.02)	(0.02)	(0.03)	(0.04)				
Preferred Helper	0.04*	0.08***	-0.03	0.07*				
	(0.02)	(0.03)	(0.04)	(0.04)				
Random female helper					0.03	0.07**	0.06	-0.04
					(0.03)	(0.03)	(0.04)	(0.05)
Preferred female helper					0.03	0.04	-0.03	0.07*
					(0.02)	(0.04)	(0.06)	(0.04)
Random male helper					0.05	0.06	0.05	0.03
					(0.03)	(0.04)	(0.05)	(0.05)
Preferred male helper					0.05	0.10***	-0.04	0.07
					(0.03)	(0.03)	(0.06)	(0.05)
P val: Random = Preferred Helper	0.97	0.69	0.02	0.03	-	-	-	-
P val: Random Female = Preferred Female Helper	-	-	-	-	0.88	0.49	0.10	0.01
P val: Random Male = Preferred Male Helper	-	-	-	-	0.91	0.44	0.12	0.49
Person FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5790	1930	1930	1930	5790	1930	1930	1930
Clusters	48	48	48	48	48	48	48	48
Control mean of outcome	0.48	0.40	0.50	0.54	0.48	0.40	0.50	0.54

* p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered at test-taker level in parenthesis.

Notes: The unit of observation is at test taker-question level. Random helper is a dummy variable that turns on 1 if a test-taker is allocated a helper chosen randomly by the research team. Preferred helper on the other hand is a dummy variable that turns on 1 if a test-taker is allocated a helper stated to be preferred by the test-takers at baseline. Random female helper is a dummy variable that turns on 1 if a test-taker is allocated a female helper chosen randomly by the research team. Preferred female helper is a dummy variable that turns on 1 if a test-taker is allocated to a female helper stated to be preferred by the test-takers at baseline. Random male helper is a dummy variable that turns on 1 if a test-taker is allocated a male helper chosen randomly by the research team. Preferred male helper is a dummy variable that turns on 1 if a test-taker is allocated to a male helper stated to be preferred by the test-takers at baseline. Correct answer is a dummy variable that turns on 1 if the question is answered correctly.

Table 9: **Mechanism: Does choice of co-worker matter for uncovering hints?**

Sample: Female Test-takers	Questions with hints available							
	Reference category: Test-takers alone, with hints							
Dependent Variables:	**Hint uncovered (dummy)**				**Correct answer (dummy)**			
	Overall	By Subjects			Overall	By Subjects		
		Econ	Cooking	Sports		Econ	Cooking	Sports
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Random Female Helper	-0.02	-0.02	-0.02	-0.03	-0.02	-0.00	-0.05	-0.00
	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.04)	(0.03)	(0.03)
Preferred Female Helper	-0.00	-0.01	-0.00	0.01	0.02	0.06*	-0.03	0.02
	(0.03)	(0.06)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)
Random Male Helper	-0.02	0.06	-0.07*	-0.06*	-0.09***	-0.05	-0.15***	-0.06
	(0.03)	(0.06)	(0.04)	(0.03)	(0.03)	(0.04)	(0.05)	(0.05)
Preferred Male Helper	-0.02	-0.04	-0.01	-0.01	0.01	0.11	-0.17*	0.11
	(0.09)	(0.17)	(0.09)	(0.07)	(0.08)	(0.08)	(0.09)	(0.10)
P val: Random female = preferred female helper	0.54	0.82	0.77	0.30	0.26	0.19	0.75	0.64
P val: Random male = preferred male helper	0.95	0.59	0.55	0.47	0.24	0.06	0.79	0.15
Person FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5016	1672	1672	1672	5016	1672	1672	1672
Clusters	54	54	54	54	54	54	54	54
Control mean of outcome	0.76	0.70	0.75	0.84	0.56	0.46	0.64	0.57

* p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered at test-taker level in parenthesis.

Notes: The unit of observation is at test taker-question level. Random female helper is a dummy variable that turns on 1 if a test-taker is allocated a female helper chosen randomly by the research team. Preferred female helper is a dummy variable that turns on 1 if a test-taker is allocated to a female helper stated to be preferred by the test-takers at baseline. Random male helper is a dummy variable that turns on 1 if a test-taker is allocated a male helper chosen randomly by the research team. Preferred male helper is a dummy variable that turns on 1 if a test-taker is allocated to a male helper stated to be preferred by the test-takers at baseline. Hint uncovered is a dummy variable that turns on 1 if a hint was available and uncovered to help answer the question. Columns (5)-(8) results report performance on questions where hints were available, irrespective of whether they were scratched.

Table 10: **Mechanism: Does choice over co-worker matter in questions without hints?**

Sample: Female Test-takers	Questions without hints			
	Reference category: Test-takers alone, no hints			
Dependent Variables:	**Correct answer (dummy)**			
	Overall	By Subjects		
		Econ	Cooking	Sports
	(1)	(2)	(3)	(4)
Random female helper	0.04	0.03	0.06	0.02
	(0.03)	(0.06)	(0.07)	(0.06)
Preferred female helper	0.04	-0.02	0.03	0.13*
	(0.03)	(0.06)	(0.08)	(0.07)
Random male helper	0.02	0.13	-0.01	-0.04
	(0.04)	(0.10)	(0.08)	(0.07)
Preferred male helper	0.10	-0.07	0.12	0.25**
	(0.08)	(0.12)	(0.14)	(0.11)
P val: Random female = preferred female helper	0.86	0.42	0.70	0.20
P val: Random male = preferred male helper	0.43	0.15	0.43	0.02
Person FE	Yes	Yes	Yes	Yes
Observations	1254	418	418	418
Clusters	54	54	54	54
Control mean of outcome	0.32	0.33	0.40	0.22

* p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered at test-taker level in parenthesis.

Notes: The unit of observation is at test taker-question level. Random female helper is a dummy variable that turns on 1 if a test-taker is allocated a female helper chosen randomly by the research team. Preferred female helper is a dummy variable that turns on 1 if a test-taker is allocated to a female helper stated to be preferred by the test-takers at baseline. Random male helper is a dummy variable that turns on 1 if a test-taker is allocated a male helper chosen randomly by the research team. Preferred male helper is a dummy variable that turns on 1 if a test-taker is allocated to a male helper stated to be preferred by the test-takers at baseline. Correct answer is a dummy variable that turns on 1 if the question is answered correctly.

Table 11: **Mechanism: Does choice of co-worker matter for uncovering hints?**

Sample: Male Test-takers	Questions with hints available							
	Reference category: Test-takers alone, with hints							
Dependent Variables:	**Hint uncovered (dummy)**				**Correct answer (dummy)**			
	Overall	By Subjects			Overall	By Subjects		
		Econ	Cooking	Sports		Econ	Cooking	Sports
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Random Female Helper	0.07*	0.10**	0.01	0.11**	0.04	0.06	0.07	-0.03
	(0.04)	(0.05)	(0.05)	(0.04)	(0.03)	(0.04)	(0.05)	(0.05)
Preferred Female Helper	0.10*	0.02	0.17**	0.12**	0.04	0.02	-0.00	0.10**
	(0.05)	(0.08)	(0.08)	(0.05)	(0.03)	(0.04)	(0.07)	(0.05)
Random Male Helper	0.03	0.10*	0.03	-0.04	0.05	0.04	0.08	0.04
	(0.04)	(0.06)	(0.05)	(0.05)	(0.03)	(0.04)	(0.06)	(0.05)
Preferred Male Helper	0.10**	0.09	0.05	0.16***	0.03	0.12***	-0.06	0.03
	(0.05)	(0.07)	(0.06)	(0.05)	(0.03)	(0.04)	(0.06)	(0.06)
P val: Random female = preferred female helper	0.59	0.28	0.10	0.75	0.96	0.36	0.30	0.03
P val: Random male = preferred male helper	0.10	0.90	0.63	0.00	0.52	0.12	0.03	0.87
Person FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4632	1544	1544	1544	4632	1544	1544	1544
Clusters	48	48	48	48	48	48	48	48
Control mean of outcome	0.61	0.56	0.62	0.65	0.52	0.42	0.54	0.60

* p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered at test-taker level in parenthesis.

Notes: The unit of observation is at test taker-question level. Random female helper is a dummy variable that turns on 1 if a test-taker is allocated a female helper chosen randomly by the research team. Preferred female helper is a dummy variable that turns on 1 if a test-taker is allocated to a female helper stated to be preferred by the test-takers at baseline. Random male helper is a dummy variable that turns on 1 if a test-taker is allocated a male helper chosen randomly by the research team. Preferred male helper is a dummy variable that turns on 1 if a test-taker is allocated to a male helper stated to be preferred by the test-takers at baseline. Hint uncovered is a dummy variable that turns on 1 if a hint was available and uncovered to help answer the question. Columns (5)-(8) results report performance on questions where hints were available, irrespective of whether they were scratched.

Table 12: **Mechanism: Does choice over co-worker matter in questions without hints?**

Sample: Male Test-takers	Questions without hints			
	Reference category: Test-takers alone, no hints			
Dependent Variables:	**Correct answer (dummy)**			
	Overall	By Subjects		
		Econ	Cooking	Sports
	(1)	(2)	(3)	(4)
Random female helper	0.01	0.10	0.03	-0.11
	(0.04)	(0.08)	(0.08)	(0.08)
Preferred female helper	-0.02	0.13	-0.14	-0.06
	(0.06)	(0.11)	(0.11)	(0.11)
Random male helper	0.03	0.16*	-0.06	0.00
	(0.05)	(0.09)	(0.08)	(0.10)
Preferred male helper	0.11**	0.04	0.07	0.23*
	(0.06)	(0.08)	(0.08)	(0.12)
P val: Random female = preferred female helper	0.70	0.81	0.23	0.66
P val: Random male = preferred male helper	0.16	0.20	0.22	0.02
Person FE	Yes	Yes	Yes	Yes
Observations	1158	386	386	386
Clusters	48	48	48	48
Control mean of outcome	0.32	0.31	0.36	0.31

* p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered at test-taker level in parenthesis.

Notes: The unit of observation is at test taker-question level. Random female helper is a dummy variable that turns on 1 if a test-taker is allocated a female helper chosen randomly by the research team. Preferred female helper is a dummy variable that turns on 1 if a test-taker is allocated to a female helper stated to be preferred by the test-takers at baseline. Random male helper is a dummy variable that turns on 1 if a test-taker is allocated a male helper chosen randomly by the research team. Preferred male helper is a dummy variable that turns on 1 if a test-taker is allocated to a male helper stated to be preferred by the test-takers at baseline. Correct answer is a dummy variable that turns on 1 if the question is answered correctly.

9 Technical appendix

In this section we derive the expected score for all possible combinations of parameters μ (prior), p_H (precision of hint) and p_M (precision of helper), given that hints are more precise than helpers ($p_H > p_M$) and that obtaining advice from the helper is costless.

9.1 Expected score and hint uncovering behavior

1. **Very low confidence**

   - **Conditions:** both hint alone and helper alone can change answer. If helper says A and hint says B, following hint. $\frac{\mu}{1-\mu} < \frac{p_H}{1-p_H}\frac{1-p_M}{p_M}$ and $\frac{\mu}{1-\mu} < \frac{p_M}{1-p_M} < \frac{p_H}{1-p_H}$
   - **Expected score:** Alone: $S_T = \mu$, Hint: $S_H = p_H$, Helper: $S_M = p_M$, Helper and hint: $S_{MH} = p_H$.
   - **Use of hint:** take hint at cost κ if
     - Alone: $S_H - S_T = p_H - \mu > \kappa$
     - With helper: $S_{MH} - S_M = p_H - p_M > \kappa$
   - **Expected scores given hint available:**
     (a) If $\kappa < p_H - p_M < p_H - \mu$: $S_H = S_{MH} = p_H$,
     So $S_H - S_T = p_H - \mu > 0$ and $S_{MH} - S_H = p_H - p_H = 0$.
     (b) If $p_H - p_M < \kappa < p_H - \mu$: $S_H = p_H$ but $S_{MH} = p_M$.
     So $S_H - S_T = p_H - \mu > 0$ and $S_{MH} - S_H = p_M - p_H < 0$.
     (c) If $\kappa > p_H - \mu > p_H - p_M$: $S_H = \mu$, $S_{MH} = p_M$.
     So $S_H - S_T = \mu - \mu = 0$ and $S_{MH} - S_H = p_M - \mu > 0$.

2. **Low confidence**

   - **Conditions:** both hint alone and helper alone can change answer. If helper says A and hint says B, follow helper. $\frac{p_H}{1-p_H}\frac{1-p_M}{p_M} < \frac{\mu}{1-\mu}$ and $\frac{\mu}{1-\mu} < \frac{p_M}{1-p_M} < \frac{p_H}{1-p_H}$
   - **Expected score:** Alone: $S_T = \mu$, Hint: $S_H = p_H$, Helper: $S_M = p_M$, Helper and hint: $S_{MH} = p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)]$.
   - **Use of hint:** take hint at cost κ if
     - Alone: $S_H - S_T = p_H - \mu > \kappa$
     - With helper: $S_{MH} - S_M = p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)] - p_M = \mu p_H(1-p_M) - (1-\mu)p_M(1-p_H) > \kappa$
       Note $LHS > 0$ since $\frac{\mu}{1-\mu} > \frac{p_H}{1-p_H}\frac{1-p_M}{p_M} > \frac{p_M}{1-p_M}\frac{1-p_H}{p_H}$), while $(p_H - \mu) - (\mu p_H(1-p_M) - (1-\mu)p_M(1-p_H))$ can be either positive or negative.
   - **Expected scores given hint available:**
     (a) If $\kappa < \min\{p_H - \mu, \mu p_H(1-p_M)) - (1-\mu)p_M(1-p_H))\}$: $S_H = p_H$, $S_{MH} = p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)]$.
     So $S_H - S_T = p_H - \mu > 0$ and $S_{MH} - S_H = \mu p_M(1-p_H) - (1-\mu)p_H(1-p_M) > 0$.
     (b) If $\mu p_H(1-p_M) - (1-\mu)p_M(1-p_H)) < \kappa < p_H - \mu$: $S_H = p_H$, $S_{MH} = p_M$.
     So $S_H - S_T = p_H - \mu > 0$, $S_{MH} - S_H = p_M - p_H < 0$.
     (c) If $p_H - \mu < \kappa < \mu p_H(1-p_M)) - (1-\mu)p_M(1-p_H)$: $S_H = \mu$, and $S_{MH} = p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)]$.
     So $S_H - S_T = \mu - \mu = 0$ and $S_{MH} - S_H = \mu(p_H p_M + p_M(1-p_H) + p_H(1-p_M)) + (1-\mu)p_H p_M - \mu = \mu(1-(1-p_H)(1-p_M)-1) + (1-\mu)p_H p_M = (1-\mu)p_H p_M - \mu(1-p_H)(1-p_M) > 0$ since $\frac{\mu}{1-\mu} < \frac{p_M}{1-p_M} < \frac{p_M}{1-p_M}\frac{p_H}{1-p_H}$.
     (d) If $\kappa > \max\{p_H - \mu, \mu p_H(1-p_M)) - (1-\mu)p_M(1-p_H))\}$.
     Then: $S_H = S_{MH} = \mu$ so $S_H - S_T = 0$ and $S_{MH} - S_H = 0$.

3. **Moderate-low confidence**

- **Conditions:** hint alone can change answer, but not helper alone. If helper says A and hint says B, follow hint. $\frac{\mu}{1-\mu} < \frac{p_H}{1-p_H}\frac{1-p_M}{p_M}$ and $\frac{p_M}{1-p_M} < \frac{\mu}{1-\mu} < \frac{p_H}{1-p_H}$
- **Expected score:** Alone: $S_T = \mu$, Hint: $S_H = p_H$, Helper: $S_M = \mu$, Helper and hint: $S_{MH} = p_H$.
- **Use of hint:** take hint at cost κ if
  - Alone: $S_H - S_T = p_H - \mu > \kappa$
  - With helper: $S_{MH} - S_M = p_H - \mu > \kappa$.
- **Expected scores given hint available:**
  (a) If $\kappa < p_H - \mu$: $S_H = S_{MH} = p_H$, so $S_H - S_T = p_H - \mu > 0$, $S_{HM} - S_H = 0$.
  (b) If $\kappa > p_H - \mu$: $S_H = S_{MH} = \mu$ so $S_H - S_T = 0$ and $S_{HM} - S_H = 0$.

4. **Moderate-high confidence**

- **Conditions:** hint alone can change answer, but not helper alone. If helper says A and hint says B, follow helper. $\frac{p_H}{1-p_H}\frac{1-p_M}{p_M} < \frac{\mu}{1-\mu}$ and $\frac{p_M}{1-p_M} < \frac{\mu}{1-\mu} < \frac{p_H}{1-p_H}$
- **Expected score:** Alone: $S_T = \mu$, Hint: $S_H = p_H$, Helper: $S_M = \mu$, Helper and hint: $S_{MH} = p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)]$.
- **Use of hint:** take hint at cost κ if
  - Alone: $S_H - S_T = p_H - \mu > \kappa$
  - With helper: $S_{MH} - S_M = p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)] - \mu > \kappa$
    Note: $LHS > 0$ and $[p_H - \mu] - [p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)] - \mu] = p_H(1-p_M) - \mu[p_H(1-p_M) + p_M(1-p_H)] = p_H(1-p_M)(1-\mu) - \mu p_M(1-p_H) < 0$ given conditions for moderate-high confidence.
- **Expected scores given hint available:**
  (a) If $\kappa < \min\{p_H - \mu, p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)] - \mu\}$: $S_H = p_H$ and $S_{MH} = p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)]$.
  So $S_H - S_T = p_H - \mu > 0$ and $S_{MH} - S_H = p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)] - p_H = \mu p_M(1-p_H) - (1-\mu)p_H(1-p_M) > 0$ given the conditions for moderate-high confidence.
  (b) If $p_H - \mu < \kappa < p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)] - \mu$: $S_H = \mu$, $S_{MH} = p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)]$
  So $S_H - S_T = \mu - \mu = 0$ and $S_{MH} - S_H = p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)] - \mu = \mu(1-(1-p_H)(1-p_M)-1) + (1-\mu)p_H p_M = (1-\mu)p_H p_M - \mu(1-p_H)(1-p_M) > 0$.
  (c) $\kappa > \max\{p_H - \mu, p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)] - \mu\}$: $S_H = S_{MH} = \mu$.
  So $S_H - S_T = 0$ and $S_H - S_{MH} = 0$.
  (d) Finally, note that $p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)] - \mu < \kappa < p_H - \mu$ is not possible as $p_H - \mu < p_H p_M + \mu[p_H(1-p_M) + p_M(1-p_H)] - \mu$.

5. **High confidence**

- **Conditions:** neither hint alone nor helper alone can change answer, but if both helper and hint say B then choose B. $\frac{\mu}{1-\mu} < \frac{p_H}{1-p_H}\frac{p_M}{1-p_M}$ and $\frac{p_M}{1-p_M} < \frac{p_H}{1-p_H} < \frac{\mu}{1-\mu}$
- **Expected score:** Alone: $S_T = \mu$, Hint: $S_H = \mu$, Helper: $S_M = \mu$, Helper and hint: $S_{MH} = \mu[1-(1-p_H)(1-p_M)] + (1-\mu)p_H p_M$.
- **Use of hint:** take hint at cost κ if
  - Alone: $S_H - S_T = \mu - \mu > \kappa \Rightarrow$ never.
  - With helper: $S_{MH} - S_M = \mu[1-(1-p_H)(1-p_M)] + (1-\mu)p_H p_M - \mu > \kappa$ (note $LHS > 0$ given condition).
- **Expected scores given hint available:**

(a) If $\kappa < \mu[1-(1-p_H)(1-p_M)] + (1-\mu)p_H p_M - \mu$: $S_H = \mu$ and $S_{MH} = \mu[1-(1-p_H)(1-p_M)] + (1-\mu)p_H p_M$.
So $S_H - S_T = \mu - \mu = 0$, $S_{MH} - S_H = \mu[1-(1-p_H)(1-p_M)] + (1-\mu)p_H p_M - \mu = (1-\mu)p_H p_M - \mu(1-p_H)(1-p_M) > 0$ given condition for high confidence.

(b) If $\kappa > \mu[1-(1-p_H)(1-p_M)] + (1-\mu)p_H p_M - \mu$: $S_H - S_T = \mu - \mu = 0$ and $S_H = S_{MH} = \mu - \mu = 0$.

6. **Very high confidence**

- **Conditions:** neither hint alone nor helper alone can change answer, test-taker always chooses A. $\frac{p_H}{1-p_H}\frac{p_M}{1-p_M} < \frac{\mu}{1-\mu}$ and $\frac{p_M}{1-p_M} < \frac{p_H}{1-p_H} < \frac{\mu}{1-\mu}$
- **Expected score:** Alone: $S_T = \mu$, Hint: $S_H = \mu$, Helper: $S_M = \mu$, Helper and hint: $S_{MH} = \mu$.
- **Use of hint:** Alone or with helper: $\mu - \mu > \kappa \Rightarrow$ never.
- **Expected scores given hint available:** $S_H - S_T = \mu - \mu = 0$ and $S_H = S_{MH} = \mu - \mu = 0$.

9.2 Proof of results

Proof of Result 1. It is straightforward to see that this statement holds from the derivations above. The difference between the test-taker's score with a helper vs. alone is given by $S_M - S_T$. This depends on p_M and μ only. Let p_{MP} and p_{MR} the precision of the preferred and random helpers respectively. If $p_{MP} = p_{MR}$, then $S_M - S_T$ is the same for random and preferred helpers.

The same logic applies when looking at the score of the test-taker in questions with hints, with and without a helper: $S_{MH} - S_H$. This difference depends on p_M, p_H and μ. If $p_{MP} = p_{MR}$, then $S_{MH} - S_H$ is the same for random and preferred helpers.

□

Proof of Result 2. Result 2 follows directly from Result 1. Since the test-taker's decision to uncover the hint only depends on the difference in expected score with and without the hint, and since by Result 1 the expected score is the same across different types of helpers if the helpers have the same ability, then the decision to uncover hint should be the same across types of helpers.

□

Proof of Result 3. Note that the only cases where the rate of hint taking decreases when adding a helper (i.e. $S_{MH} - S_H < 0$) are the very low confidence and low confidence cases with an intermediate value of κ. The intermediate value of κ in turns implies that the test-taker takes a hint when alone, but not in the presence of the helper. Therefore, lower hint uncovering in the presence of the helper is a necessary condition for the score to decrease in the presence of a helper.

In addition, these two cases are the only cases where the test-taker takes fewer hints in the presence of a helper than alone.

Therefore taking fewer hints in the presence of a helper is a necessary and sufficient condition for the test-taker's score to be lower in the presence of the helper.

Finally, note that the expected score in the presence of a helper but no hint is either p_M or μ while the expected score alone is μ, so given $p_M \geq \mu$, the helper always weakly improves the performance of the test-taker.

□

10 Appendix: For Online Publication

10.1 Appendix Tables

Table A1: Distribution of male and female students across degrees

	Overall		Test-taker Sample		Helper Sample	
Semesters	# of females	# of males	# of females	# of males	# of females	# of males
BBA						
BBA 1st	24	65	14	14	10	10
BBA 3rd	24	71	14	14	10	10
BBA 5th	28	136	16	16	12	12
BBA 7th	6	37	3	3	3	3
BBA IM 7th	8	23	4	4	4	4
BSPA						
BSPA 1st	12	49	7	7	5	5
BSPA 3rd	7	41	4	4	3	3
BSPA 5th	11	50	6	6	5	5
BSPA 7th	13	24	7	7	6	6
BS Comm						
BS Comm 1st	13	44	7	7	6	6
BS Comm 3rd	3	33	1	1	2	2
BS Comm 5th	3	31	1	1	2	2
BS Comm 7th	2	25	1	1	1	1

Notes: BBA - Bachelor's in business administration. BSPA - Bachelor's of science in public administration. BS Comm - Bachelor's of science in commerce.

Table A2: Proportion of test takers with helpers of different types

	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
Sample: Female test takers	(1)	(2)	(3)	(4)	(5)	(6)
Random female helper	-	-	0.29	0.24	0.34	-
Random male helper	-	-	0.31	0.19	0.31	-
No helper	1.00	0.25	0.27	0.44	0.19	0.27
Preferred female helper	-	0.53	-	-	-	0.34
Preferred male helper	-	0.19	-	-	-	0.15
Observations	59	59	59	59	59	59

Notes: Random female helper is a dummy variable that turns on 1 if a test-taker is allocated a female helper chosen randomly by the research team. Random male helper is a dummy variable that turns on 1 if a test-taker is allocated a male helper chosen randomly by the research team. No helper is a dummy variable that turns on 1 if a test-taker is not allocated a helper. Preferred female helper is a dummy variable that turns on 1 if a test-taker is allocated to a female helper stated to be preferred by the test-takers at baseline. Preferred male helper is a dummy variable that turns on 1 if a test-taker is allocated to a male helper stated to be preferred by the test-takers at baseline.

Table A3: Proportion of test takers with helpers of different types

	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
Sample: Male test takers	(1)	(2)	(3)	(4)	(5)	(6)
Random female helper	-	-	0.29	0.31	0.31	-
Random male helper	-	-	0.22	0.27	0.24	-
No helper	1.00	0.45	0.43	0.35	0.37	0.35
Preferred female helper	-	0.16	-	-	-	0.24
Preferred male helper	-	0.37	-	-	-	0.24
Observations	49	49	49	49	49	49

Notes: Random female helper is a dummy variable that turns on 1 if a test-taker is allocated a female helper chosen randomly by the research team. Random male helper is a dummy variable that turns on 1 if a test-taker is allocated a male helper chosen randomly by the research team. No helper is a dummy variable that turns on 1 if a test-taker is not allocated a helper. Preferred female helper is a dummy variable that turns on 1 if a test-taker is allocated to a female helper stated to be preferred by the test-takers at baseline. Preferred male helper is a dummy variable that turns on 1 if a test-taker is allocated to a male helper stated to be preferred by the test-takers at baseline.

Table A4: Deviations from Experimental Protocol

	Random male helper twice	Random female helper twice	No random helper	Preferred male helper twice	Preferred female helper twice	No preferred helper
Sample: female test-takers						
Proportion of test-takers	0.06	0.08	0.00	0.09	0.24	0.04
Sample: male test-takers						
Proportion of test-takers	0.04	0.13	0.04	0.12	0.10	0.07

Notes: Random male helper twice - a test-taker was assigned a random male helper twice in the six tests. Random female helper twice - a test-taker was assigned a random female helper twice in the six tests. No random helper - a test-taker was not assigned a random helper in any of the six tests. Preferred male helper twice - a test-taker was assigned a preferred male helper twice in the six tests. Preferred female helper twice- a test-taker was assigned a preferred female helper twice in the six tests. No preferred helper - a test-taker was not assigned a preferred helper in any of the six tests.

Table A5: Balance on Baseline Characteristics of Test-Takers

	(1)	(2)	(3)
	Female test-takers who		Difference
	Completed tests 1 - 6	Did not complete tests 1 - 6	(1) - (2)
GPA	3.33	3.32	0.01
	(0.59)	(0.64)	(0.18)
Age	20.51	19.50	1.01
	(1.43)	(1.29)	(0.75)
Gender of test-taker (male = 1)	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Number of children	0.02	0.00	0.02
	(0.15)	(0.00)	(0.04)
Marital status (married = 1)	0.05	0.00	0.05
	(0.21)	(0.00)	(0.05)
Language: Urdu	0.43	0.27	0.17
	(0.50)	(0.46)	(0.15)
Language: English	0.02	0.00	0.02
	(0.15)	(0.00)	(0.04)
Father's occupation: private business	0.18	0.13	0.05
	(0.39)	(0.35)	(0.11)
Father's occupation: kissan	0.11	0.00	0.11
	(0.32)	(0.00)	(0.08)
Father's occupation: government job	0.11	0.00	0.11
	(0.32)	(0.00)	(0.08)
Father's occupation: private job	0.09	0.00	0.09
	(0.29)	(0.00)	(0.08)
Mother's occupation: housewife	0.59	0.13	0.46***
	(0.50)	(0.35)	(0.14)
Mother's occupation: government job	0.11	0.07	0.05
	(0.32)	(0.26)	(0.09)
Family's monthly income: less than Rs. 10,000	0.05	0.00	0.05
	(0.21)	(0.00)	(0.05)
Family's monthly income: Rs. 10,000 - Rs. 100,000	0.57	0.27	0.30**
	(0.50)	(0.46)	(0.15)
Family's monthly income: greater than Rs. 100,000	0.05	0.00	0.05
	(0.21)	(0.00)	(0.05)
Household size: 1-5	0.23	0.00	0.23**
	(0.42)	(0.00)	(0.11)
Household size: 6-10	0.32	0.20	0.12
	(0.47)	(0.41)	(0.14)
Household size: greater than 10	0.45	0.80	-0.35**
	(0.50)	(0.41)	(0.14)
Observations	44	15	59

*p<0.1, **p<0.05, ***p<0.01. Standard errors in parentheses.

Table A6: Balance on Baseline Characteristics of Test-Takers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Female test-takers with female helpers in				Difference					
	Test 3	Test 4	Test 5	>1 test	(1) - (2)	(1) - (3)	(2) - (3)	(1) - (4)	(2) - (4)	(3) - (4)
GPA	3.65	3.45	3.21	3.28	0.20	0.44*	0.24	0.37*	0.17	-0.07
	(0.32)	(0.54)	(0.71)	(0.45)	(0.19)	(0.23)	(0.25)	(0.19)	(0.26)	(0.32)
Age	20.25	20.27	20.08	21.20	-0.02	0.17	0.20	-0.95	-0.93	-1.12**
	(1.16)	(2.05)	(1.04)	(0.84)	(0.81)	(0.49)	(0.65)	(0.60)	(0.97)	(0.52)
Gender of test-taker (male = 1)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of children	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Marital status (married = 1)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Language: Urdu	0.45	0.33	0.86	0.33	0.12	-0.40**	-0.52***	0.12	0.00	0.52**
	(0.52)	(0.49)	(0.36)	(0.52)	(0.21)	(0.18)	(0.17)	(0.26)	(0.25)	(0.20)
Language: English	0.00	0.00	0.07	0.00	0.00	-0.07	-0.07	0.00	0.00	0.07
	(0.00)	(0.00)	(0.27)	(0.00)	(0.00)	(0.08)	(0.08)	(0.00)	(0.00)	(0.11)
Father's occupation: private business	0.18	0.17	0.21	0.17	0.02	-0.03	-0.05	0.02	0.00	0.05
	(0.40)	(0.39)	(0.43)	(0.41)	(0.17)	(0.17)	(0.16)	(0.21)	(0.20)	(0.21)
Father's occupation: kissan	0.00	0.17	0.07	0.00	-0.17	-0.07	0.10	0.00	0.17	0.07
	(0.00)	(0.39)	(0.27)	(0.00)	(0.12)	(0.08)	(0.13)	(0.00)	(0.16)	(0.11)
Father's occupation: government job	0.00	0.08	0.14	0.33	-0.08	-0.14	-0.06	-0.33**	-0.25	-0.19
	(0.00)	(0.29)	(0.36)	(0.52)	(0.09)	(0.11)	(0.13)	(0.15)	(0.19)	(0.20)
Father's occupation: private job	0.09	0.08	0.14	0.00	0.01	-0.05	-0.06	0.09	0.08	0.14
	(0.30)	(0.29)	(0.36)	(0.00)	(0.12)	(0.14)	(0.13)	(0.12)	(0.12)	(0.15)
Mother's occupation: housewife	0.45	0.58	0.71	0.50	-0.13	-0.26	-0.13	-0.05	0.08	0.21
	(0.52)	(0.51)	(0.47)	(0.55)	(0.22)	(0.20)	(0.19)	(0.27)	(0.26)	(0.24)
Mother's occupation: government job	0.18	0.08	0.14	0.00	0.10	0.04	-0.06	0.18	0.08	0.14
	(0.40)	(0.29)	(0.36)	(0.00)	(0.15)	(0.15)	(0.13)	(0.17)	(0.12)	(0.15)
Family's monthly income: less than Rs. 10,000	0.00	0.08	0.07	0.00	-0.08	-0.07	0.01	0.00	0.08	0.07
	(0.00)	(0.29)	(0.27)	(0.00)	(0.09)	(0.08)	(0.11)	(0.00)	(0.12)	(0.11)
Family's monthly income: Rs. 10,000 - Rs. 100,000	0.64	0.58	0.64	0.50	0.05	-0.01	-0.06	0.14	0.08	0.14
	(0.50)	(0.51)	(0.50)	(0.55)	(0.21)	(0.20)	(0.20)	(0.26)	(0.26)	(0.25)
Family's monthly income: greater than Rs. 100,000	0.00	0.00	0.14	0.00	0.00	-0.14	-0.14	0.00	0.00	0.14
	(0.00)	(0.00)	(0.36)	(0.00)	(0.00)	(0.11)	(0.11)	(0.00)	(0.00)	(0.15)
Household size: 1-5	0.09	0.25	0.29	0.33	-0.16	-0.19	-0.04	-0.24	-0.08	-0.05
	(0.30)	(0.45)	(0.47)	(0.52)	(0.16)	(0.16)	(0.18)	(0.20)	(0.24)	(0.24)
Household size: 6-10	0.45	0.42	0.29	0.17	0.04	0.17	0.13	0.29	0.25	0.12
	(0.52)	(0.51)	(0.47)	(0.41)	(0.22)	(0.20)	(0.19)	(0.25)	(0.24)	(0.22)
Household size: greater than 10	0.45	0.33	0.43	0.50	0.12	0.03	-0.10	-0.05	-0.17	-0.07
	(0.52)	(0.49)	(0.51)	(0.55)	(0.21)	(0.21)	(0.20)	(0.27)	(0.26)	(0.26)
Observations	11	12	14	6	23	25	26	17	18	20

*p<0.1, **p<0.05, ***p<0.01. Standard errors in parentheses.

Table A7: Balance on Baseline Characteristics of Test-Takers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Female test-takers with male helpers in				Difference					
	Test 3	Test 4	Test 5	>1 test	(1) - (2)	(1) - (3)	(2) - (3)	(1) - (4)	(2) - (4)	(3) - (4)
GPA	3.45	3.31	3.39	3.38	0.14	0.06	-0.08	0.07	-0.07	0.01
	(0.46)	(0.82)	(0.59)	(0.63)	(0.27)	(0.20)	(0.29)	(0.28)	(0.46)	(0.34)
Age	20.00	19.86	20.93	20.75	0.14	-0.93	-1.07	-0.75	-0.89	0.18
	(1.21)	(1.35)	(1.54)	(2.22)	(0.60)	(0.55)	(0.69)	(0.86)	(1.06)	(0.96)
Gender of test-taker (male = 1)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of children	0.07	0.00	0.00	0.00	0.07	0.07	0.00	0.07	0.00	0.00
	(0.27)	(0.00)	(0.00)	(0.00)	(0.09)	(0.07)	(0.00)	(0.14)	(0.00)	(0.00)
Marital status (married = 1)	0.14	0.00	0.00	0.00	0.14	0.14	0.00	0.14	0.00	0.00
	(0.36)	(0.00)	(0.00)	(0.00)	(0.12)	(0.09)	(0.00)	(0.19)	(0.00)	(0.00)
Language: Urdu	0.64	0.67	0.20	0.25	-0.02	0.44**	0.47**	0.39	0.42	-0.05
	(0.50)	(0.50)	(0.41)	(0.50)	(0.21)	(0.17)	(0.19)	(0.28)	(0.30)	(0.24)
Language: English	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Father's occupation: private business	0.43	0.00	0.13	0.00	0.43**	0.30*	-0.13	0.43	0.00	0.13
	(0.51)	(0.00)	(0.35)	(0.00)	(0.17)	(0.16)	(0.12)	(0.26)	(0.00)	(0.18)
Father's occupation: kissan	0.00	0.11	0.27	0.00	-0.11	-0.27**	-0.16	0.00	0.11	0.27
	(0.00)	(0.33)	(0.46)	(0.00)	(0.09)	(0.12)	(0.18)	(0.00)	(0.17)	(0.23)
Father's occupation: government job	0.07	0.00	0.00	0.25	0.07	0.07	0.00	-0.18	-0.25	-0.25**
	(0.27)	(0.00)	(0.00)	(0.50)	(0.09)	(0.07)	(0.00)	(0.18)	(0.16)	(0.12)
Father's occupation: private job	0.21	0.00	0.07	0.00	0.21	0.15	-0.07	0.21	0.00	0.07
	(0.43)	(0.00)	(0.26)	(0.00)	(0.14)	(0.13)	(0.09)	(0.22)	(0.00)	(0.13)
Mother's occupation: housewife	0.71	0.33	0.67	0.25	0.38*	0.05	-0.33	0.46	0.08	0.42
	(0.47)	(0.50)	(0.49)	(0.50)	(0.21)	(0.18)	(0.21)	(0.27)	(0.30)	(0.28)
Mother's occupation: government job	0.07	0.33	0.07	0.00	-0.26	0.00	0.27*	0.07	0.33	0.07
	(0.27)	(0.50)	(0.26)	(0.00)	(0.16)	(0.10)	(0.15)	(0.14)	(0.26)	(0.13)
Family's monthly income: less than Rs. 10,000	0.07	0.11	0.00	0.00	-0.04	0.07	0.11	0.07	0.11	0.00
	(0.27)	(0.33)	(0.00)	(0.00)	(0.13)	(0.07)	(0.08)	(0.14)	(0.17)	(0.00)
Family's monthly income: Rs. 10,000 - Rs. 100,000	0.71	0.33	0.60	0.25	0.38*	0.11	-0.27	0.46	0.08	0.35
	(0.47)	(0.50)	(0.51)	(0.50)	(0.21)	(0.18)	(0.21)	(0.27)	(0.30)	(0.28)
Family's monthly income: greater than Rs. 100,000	0.07	0.11	0.00	0.00	-0.04	0.07	0.11	0.07	0.11	0.00
	(0.27)	(0.33)	(0.00)	(0.00)	(0.13)	(0.07)	(0.08)	(0.14)	(0.17)	(0.00)
Household size: 1-5	0.29	0.11	0.07	0.25	0.17	0.22	0.04	0.04	-0.14	-0.18
	(0.47)	(0.33)	(0.26)	(0.50)	(0.18)	(0.14)	(0.12)	(0.27)	(0.23)	(0.18)
Household size: 6-10	0.29	0.44	0.47	0.00	-0.16	-0.18	-0.02	0.29	0.44	0.47*
	(0.47)	(0.53)	(0.52)	(0.00)	(0.21)	(0.18)	(0.22)	(0.24)	(0.27)	(0.26)
Household size: greater than 10	0.43	0.44	0.47	0.75	-0.02	-0.04	-0.02	-0.32	-0.31	-0.28
	(0.51)	(0.53)	(0.52)	(0.50)	(0.22)	(0.19)	(0.22)	(0.29)	(0.31)	(0.29)
Observations	14	9	15	4	23	29	24	18	13	19

*p<0.1, **p<0.05, ***p<0.01. Standard errors in parentheses.

Table A8: Balance on Baseline Characteristics of Test-Takers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Female test-takers with no helper in				Difference					
	Test 3	Test 4	Test 5	>1 test	(1) - (2)	(1) - (3)	(2) - (3)	(1) - (4)	(2) - (4)	(3) - (4)
GPA	3.07	3.41	3.64	3.29	-0.34	-0.57	-0.23	-0.23	0.12	0.34
	(0.81)	(0.46)	(0.33)	(0.68)	(0.24)	(0.33)	(0.19)	(0.35)	(0.22)	(0.28)
Age	21.11	20.59	19.67	20.11	0.52	1.44	0.92	1.00	0.48	-0.44
	(1.83)	(1.06)	(1.86)	(1.27)	(0.56)	(0.97)	(0.62)	(0.74)	(0.47)	(0.80)
Gender of test-taker (male = 1)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of children	0.00	0.00	0.00	0.11	0.00	0.00	0.00	-0.11	-0.11	-0.11
	(0.00)	(0.00)	(0.00)	(0.33)	(0.00)	(0.00)	(0.00)	(0.11)	(0.07)	(0.13)
Marital status (married = 1)	0.00	0.00	0.00	0.22	0.00	0.00	0.00	-0.22	-0.22**	-0.22
	(0.00)	(0.00)	(0.00)	(0.44)	(0.00)	(0.00)	(0.00)	(0.15)	(0.10)	(0.17)
Language: Urdu	0.44	0.58	0.57	0.22	-0.13	-0.13	0.01	0.22	0.36*	0.35
	(0.53)	(0.51)	(0.53)	(0.44)	(0.21)	(0.27)	(0.23)	(0.23)	(0.20)	(0.24)
Language: English	0.00	0.00	0.00	0.11	0.00	0.00	0.00	-0.11	-0.11	-0.11
	(0.00)	(0.00)	(0.00)	(0.33)	(0.00)	(0.00)	(0.00)	(0.11)	(0.07)	(0.13)
Father's occupation: private business	0.00	0.21	0.29	0.33	-0.21	-0.29*	-0.08	-0.33*	-0.12	-0.05
	(0.00)	(0.42)	(0.49)	(0.50)	(0.14)	(0.16)	(0.19)	(0.17)	(0.18)	(0.25)
Father's occupation: kissan	0.33	0.00	0.00	0.22	0.33***	0.33	0.00	0.11	-0.22**	-0.22
	(0.50)	(0.00)	(0.00)	(0.44)	(0.11)	(0.19)	(0.00)	(0.22)	(0.10)	(0.17)
Father's occupation: government job	0.00	0.16	0.00	0.11	-0.16	0.00	0.16	-0.11	0.05	-0.11
	(0.00)	(0.37)	(0.00)	(0.33)	(0.13)	(0.00)	(0.14)	(0.11)	(0.15)	(0.13)
Father's occupation: private job	0.00	0.16	0.14	0.00	-0.16	-0.14	0.02	0.00	0.16	0.14
	(0.00)	(0.37)	(0.38)	(0.00)	(0.13)	(0.12)	(0.17)	(0.00)	(0.13)	(0.12)
Mother's occupation: housewife	0.67	0.68	0.43	0.44	-0.02	0.24	0.26	0.22	0.24	-0.02
	(0.50)	(0.48)	(0.53)	(0.53)	(0.20)	(0.26)	(0.22)	(0.24)	(0.20)	(0.27)
Mother's occupation: government job	0.22	0.00	0.29	0.22	0.22**	-0.06	-0.29**	0.00	-0.22**	0.06
	(0.44)	(0.00)	(0.49)	(0.44)	(0.10)	(0.23)	(0.11)	(0.21)	(0.10)	(0.23)
Family's monthly income: less than Rs. 10,000	0.11	0.00	0.14	0.00	0.11	-0.03	-0.14	0.11	0.00	0.14
	(0.33)	(0.00)	(0.38)	(0.00)	(0.07)	(0.18)	(0.08)	(0.11)	(0.00)	(0.12)
Family's monthly income: Rs. 10,000 - Rs. 100,000	0.44	0.68	0.57	0.67	-0.24	-0.13	0.11	-0.22	0.02	-0.10
	(0.53)	(0.48)	(0.53)	(0.50)	(0.20)	(0.27)	(0.22)	(0.24)	(0.20)	(0.26)
Family's monthly income: greater than Rs. 100,000	0.11	0.05	0.00	0.00	0.06	0.11	0.05	0.11	0.05	0.00
	(0.33)	(0.23)	(0.00)	(0.00)	(0.11)	(0.13)	(0.09)	(0.11)	(0.08)	(0.00)
Household size: 1-5	0.11	0.26	0.29	0.11	-0.15	-0.17	-0.02	-0.00	0.15	0.17
	(0.33)	(0.45)	(0.49)	(0.33)	(0.17)	(0.21)	(0.20)	(0.16)	(0.17)	(0.21)
Household size: 6-10	0.56	0.26	0.43	0.33	0.29	0.13	-0.17	0.22	-0.07	0.10
	(0.53)	(0.45)	(0.53)	(0.50)	(0.19)	(0.27)	(0.21)	(0.24)	(0.19)	(0.26)
Household size: greater than 10	0.33	0.47	0.29	0.56	-0.14	0.05	0.19	-0.22	-0.08	-0.27
	(0.50)	(0.51)	(0.49)	(0.53)	(0.21)	(0.25)	(0.22)	(0.24)	(0.21)	(0.26)
Observations	9	19	7	9	28	16	26	18	28	16

*p<0.1, **p<0.05, ***p<0.01. Standard errors in parentheses.

Table A9: Balance on Baseline Characteristics of Test-Takers

	(1)	(2)	(3)	(4)	(5)	(6)
	Female test-takers with preferred helper in			Difference		
	Test 2	Test 6	Both tests	(1) - (2)	(1) - (3)	(2) - (3)
GPA	3.39	3.13	3.46	0.26	-0.07	-0.33
	(0.36)	(0.78)	(0.57)	(0.24)	(0.18)	(0.25)
Age	20.33	20.80	20.29	-0.47	0.04	0.51
	(1.15)	(0.92)	(1.79)	(0.45)	(0.59)	(0.61)
Gender of test-taker (male = 1)	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of children	0.00	0.00	0.05	0.00	-0.05	-0.05
	(0.00)	(0.00)	(0.23)	(0.00)	(0.06)	(0.07)
Marital status (married = 1)	0.08	0.00	0.05	0.08	0.02	-0.05
	(0.28)	(0.00)	(0.23)	(0.09)	(0.09)	(0.07)
Language: Urdu	0.62	0.50	0.26	0.12	0.35**	0.24
	(0.51)	(0.53)	(0.45)	(0.22)	(0.17)	(0.19)
Language: English	0.00	0.10	0.00	-0.10	0.00	0.10
	(0.00)	(0.32)	(0.00)	(0.09)	(0.00)	(0.07)
Father's occupation: private business	0.38	0.10	0.11	0.28	0.28*	-0.01
	(0.51)	(0.32)	(0.32)	(0.18)	(0.14)	(0.12)
Father's occupation: kissan	0.00	0.10	0.11	-0.10	-0.11	-0.01
	(0.00)	(0.32)	(0.32)	(0.09)	(0.09)	(0.12)
Father's occupation: government job	0.00	0.30	0.11	-0.30**	-0.11	0.19
	(0.00)	(0.48)	(0.32)	(0.13)	(0.09)	(0.15)
Father's occupation: private job	0.15	0.00	0.11	0.15	0.05	-0.11
	(0.38)	(0.00)	(0.32)	(0.12)	(0.12)	(0.10)
Mother's occupation: housewife	0.69	0.70	0.42	-0.01	0.27	0.28
	(0.48)	(0.48)	(0.51)	(0.20)	(0.18)	(0.20)
Mother's occupation: government job	0.08	0.20	0.11	-0.12	-0.03	0.09
	(0.28)	(0.42)	(0.32)	(0.15)	(0.11)	(0.14)
Family's monthly income: less than Rs. 10,000	0.15	0.00	0.00	0.15	0.15*	0.00
	(0.38)	(0.00)	(0.00)	(0.12)	(0.09)	(0.00)
Family's monthly income: Rs. 10,000 - Rs. 100,000	0.62	0.80	0.42	-0.18	0.19	0.38*
	(0.51)	(0.42)	(0.51)	(0.20)	(0.18)	(0.19)
Family's monthly income: greater than Rs. 100,000	0.00	0.00	0.05	0.00	-0.05	-0.05
	(0.00)	(0.00)	(0.23)	(0.00)	(0.06)	(0.07)
Household size: 1-5	0.23	0.40	0.16	-0.17	0.07	0.24
	(0.44)	(0.52)	(0.37)	(0.20)	(0.14)	(0.17)
Household size: 6-10	0.46	0.40	0.16	0.06	0.30*	0.24
	(0.52)	(0.52)	(0.37)	(0.22)	(0.16)	(0.17)
Household size: greater than 10	0.31	0.20	0.68	0.11	-0.38**	-0.48**
	(0.48)	(0.42)	(0.48)	(0.19)	(0.17)	(0.18)
Observations	13	10	19	23	32	29

*p<0.1, **p<0.05, ***p<0.01. Standard errors in parentheses.

Table A10: Balance on Baseline Characteristics of Test-Takers

	(1)	(2)	(3)	(4)	(5)	(6)
	Female test-takers with no helper in			Difference		
	Test 2	Test 6	Both tests	(1) - (2)	(1) - (3)	(2) - (3)
GPA	3.13	3.39	2.64	-0.26	0.49	0.75**
	(0.78)	(0.36)	(0.59)	(0.24)	(0.59)	(0.29)
Age	20.80	20.33	22.00	0.47	-1.20	-1.67*
	(0.92)	(1.15)	(1.41)	(0.45)	(0.76)	(0.90)
Gender of test-taker (male = 1)	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of children	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Marital status (married = 1)	0.00	0.08	0.00	-0.08	0.00	0.08
	(0.00)	(0.28)	(0.00)	(0.09)	(0.00)	(0.20)
Language: Urdu	0.50	0.62	0.50	-0.12	0.00	0.12
	(0.53)	(0.51)	(0.71)	(0.22)	(0.42)	(0.40)
Language: English	0.10	0.00	0.00	0.10	0.10	0.00
	(0.32)	(0.00)	(0.00)	(0.09)	(0.23)	(0.00)
Father's occupation: private business	0.10	0.38	0.00	-0.28	0.10	0.38
	(0.32)	(0.51)	(0.00)	(0.18)	(0.23)	(0.37)
Father's occupation: kissan	0.10	0.00	1.00	0.10	-0.90***	-1.00
	(0.32)	(0.00)	(0.00)	(0.09)	(0.23)	(0.00)
Father's occupation: government job	0.30	0.00	0.00	0.30**	0.30	0.00
	(0.48)	(0.00)	(0.00)	(0.13)	(0.35)	(0.00)
Father's occupation: private job	0.00	0.15	0.00	-0.15	0.00	0.15
	(0.00)	(0.38)	(0.00)	(0.12)	(0.00)	(0.27)
Mother's occupation: housewife	0.70	0.69	1.00	0.01	-0.30	-0.31
	(0.48)	(0.48)	(0.00)	(0.20)	(0.35)	(0.35)
Mother's occupation: government job	0.20	0.08	0.00	0.12	0.20	0.08
	(0.42)	(0.28)	(0.00)	(0.15)	(0.31)	(0.20)
Family's monthly income: less than Rs. 10,000	0.00	0.15	0.00	-0.15	0.00	0.15
	(0.00)	(0.38)	(0.00)	(0.12)	(0.00)	(0.27)
Family's monthly income: Rs. 10,000 - Rs. 100,000	0.80	0.62	0.50	0.18	0.30	0.12
	(0.42)	(0.51)	(0.71)	(0.20)	(0.35)	(0.40)
Family's monthly income: greater than Rs. 100,000	0.00	0.00	0.50	0.00	-0.50**	-0.50***
	(0.00)	(0.00)	(0.71)	(0.00)	(0.17)	(0.15)
Household size: 1-5	0.40	0.23	0.00	0.17	0.40	0.23
	(0.52)	(0.44)	(0.00)	(0.20)	(0.38)	(0.32)
Household size: 6-10	0.40	0.46	0.50	-0.06	-0.10	-0.04
	(0.52)	(0.52)	(0.71)	(0.22)	(0.42)	(0.41)
Household size: greater than 10	0.20	0.31	0.50	-0.11	-0.30	-0.19
	(0.42)	(0.48)	(0.71)	(0.19)	(0.35)	(0.38)
Observations	10	13	2	23	12	15

*p<0.1, **p<0.05, ***p<0.01. Standard errors in parentheses.

Table A11: Balance on Baseline Characteristics of Test-Takers

	(1)	(2)	(3)
	Male test-takers who		Difference
	Completed tests 1 - 6	Did not complete tests 1 - 6	(1) - (2)
GPA	3.13	3.51	-0.37*
	(0.57)	(0.33)	(0.19)
Age	21.12	19.20	1.92*
	(2.09)	(1.10)	(0.96)
Gender of test-taker (male = 1)	1.00	1.00	0.00
	(0.00)	(0.00)	(0.00)
Number of children	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Marital status (married = 1)	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Language: Urdu	0.38	0.20	0.18
	(0.49)	(0.42)	(0.17)
Language: English	0.03	0.00	0.03
	(0.16)	(0.00)	(0.05)
Father's occupation: private business	0.18	0.20	-0.02
	(0.39)	(0.42)	(0.14)
Father's occupation: kissan	0.15	0.10	0.05
	(0.37)	(0.32)	(0.13)
Father's occupation: government job	0.13	0.00	0.13
	(0.34)	(0.00)	(0.11)
Father's occupation: private job	0.03	0.00	0.03
	(0.16)	(0.00)	(0.05)
Mother's occupation: housewife	0.54	0.50	0.04
	(0.51)	(0.53)	(0.18)
Mother's occupation: government job	0.08	0.00	0.08
	(0.27)	(0.00)	(0.09)
Family's monthly income: less than Rs. 10,000	0.05	0.00	0.05
	(0.22)	(0.00)	(0.07)
Family's monthly income: Rs. 10,000 - Rs. 100,000	0.51	0.30	0.21
	(0.51)	(0.48)	(0.18)
Family's monthly income: greater than Rs. 100,000	0.08	0.20	-0.12
	(0.27)	(0.42)	(0.11)
Household size: 1-5	0.21	0.20	0.01
	(0.41)	(0.42)	(0.15)
Household size: 6-10	0.18	0.30	-0.12
	(0.39)	(0.48)	(0.14)
Household size: greater than 10	0.62	0.50	0.12
	(0.49)	(0.53)	(0.18)
Observations	39	10	49

*p<0.1, **p<0.05, ***p<0.01. Standard errors in parentheses.

Table A12: Balance on Baseline Characteristics of Test-Takers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Male test-takers with female helpers in				Difference					
	Test 3	Test 4	Test 5	>1 test	(1) - (2)	(1) - (3)	(2) - (3)	(1) - (4)	(2) - (4)	(3) - (4)
GPA	3.04	3.31	3.31	3.01	-0.27	-0.27	0.00	0.03	0.30	0.30
	(0.57)	(0.51)	(0.53)	(0.74)	(0.26)	(0.25)	(0.23)	(0.33)	(0.31)	(0.30)
Age	20.50	20.50	20.40	21.17	-0.00	0.10	0.10	-0.67	-0.67	-0.77
	(2.62)	(1.87)	(2.01)	(2.14)	(1.26)	(1.09)	(1.01)	(1.31)	(1.16)	(1.06)
Gender of test-taker (male = 1)	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of children	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Marital status (married = 1)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Language: Urdu	0.56	0.33	0.36	0.14	0.22	0.19	-0.03	0.41	0.19	0.22
	(0.53)	(0.50)	(0.50)	(0.38)	(0.24)	(0.23)	(0.23)	(0.24)	(0.23)	(0.22)
Language: English	0.00	0.11	0.00	0.00	-0.11	0.00	0.11	0.00	0.11	0.00
	(0.00)	(0.33)	(0.00)	(0.00)	(0.11)	(0.00)	(0.10)	(0.00)	(0.13)	(0.00)
Father's occupation: private business	0.22	0.22	0.27	0.14	0.00	-0.05	-0.05	0.08	0.08	0.13
	(0.44)	(0.44)	(0.47)	(0.38)	(0.21)	(0.20)	(0.20)	(0.21)	(0.21)	(0.21)
Father's occupation: kissan	0.00	0.22	0.18	0.14	-0.22	-0.18	0.04	-0.14	0.08	0.04
	(0.00)	(0.44)	(0.40)	(0.38)	(0.15)	(0.14)	(0.19)	(0.12)	(0.21)	(0.19)
Father's occupation: government job	0.11	0.00	0.18	0.00	0.11	-0.07	-0.18	0.11	0.00	0.18
	(0.33)	(0.00)	(0.40)	(0.00)	(0.11)	(0.17)	(0.14)	(0.13)	(0.00)	(0.15)
Father's occupation: private job	0.00	0.11	0.00	0.00	-0.11	0.00	0.11	0.00	0.11	0.00
	(0.00)	(0.33)	(0.00)	(0.00)	(0.11)	(0.00)	(0.10)	(0.00)	(0.13)	(0.00)
Mother's occupation: housewife	0.67	0.67	0.73	0.14	0.00	-0.06	-0.06	0.52**	0.52**	0.58**
	(0.50)	(0.50)	(0.47)	(0.38)	(0.24)	(0.22)	(0.22)	(0.23)	(0.23)	(0.21)
Mother's occupation: government job	0.11	0.00	0.09	0.00	0.11	0.02	-0.09	0.11	0.00	0.09
	(0.33)	(0.00)	(0.30)	(0.00)	(0.11)	(0.14)	(0.10)	(0.13)	(0.00)	(0.12)
Family's monthly income: less than Rs. 10,000	0.00	0.11	0.09	0.00	-0.11	-0.09	0.02	0.00	0.11	0.09
	(0.00)	(0.33)	(0.30)	(0.00)	(0.11)	(0.10)	(0.14)	(0.00)	(0.13)	(0.12)
Family's monthly income: Rs. 10,000 - Rs. 100,000	0.78	0.56	0.45	0.14	0.22	0.32	0.10	0.63***	0.41	0.31
	(0.44)	(0.53)	(0.52)	(0.38)	(0.23)	(0.22)	(0.24)	(0.21)	(0.24)	(0.23)
Family's monthly income: greater than Rs. 100,000	0.00	0.00	0.27	0.14	0.00	-0.27*	-0.27*	-0.14	-0.14	0.13
	(0.00)	(0.00)	(0.47)	(0.38)	(0.00)	(0.16)	(0.16)	(0.12)	(0.12)	(0.21)
Household size: 1-5	0.11	0.11	0.36	0.29	0.00	-0.25	-0.25	-0.17	-0.17	0.08
	(0.33)	(0.33)	(0.50)	(0.49)	(0.16)	(0.20)	(0.20)	(0.21)	(0.21)	(0.24)
Household size: 6-10	0.33	0.33	0.27	0.00	0.00	0.06	0.06	0.33	0.33	0.27
	(0.50)	(0.50)	(0.47)	(0.00)	(0.24)	(0.22)	(0.22)	(0.19)	(0.19)	(0.18)
Household size: greater than 10	0.56	0.56	0.36	0.71	0.00	0.19	0.19	-0.16	-0.16	-0.35
	(0.53)	(0.53)	(0.50)	(0.49)	(0.25)	(0.23)	(0.23)	(0.26)	(0.26)	(0.24)
Observations	9	9	11	7	18	20	20	16	16	18

*p<0.1, **p<0.05, ***p<0.01. Standard errors in parentheses.

Table A13: Balance on Baseline Characteristics of Test-Takers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Male test-takers with male helpers in				Difference					
	Test 3	Test 4	Test 5	>1 test	(1) - (2)	(1) - (3)	(2) - (3)	(1) - (4)	(2) - (4)	(3) - (4)
GPA	3.17	3.14	3.37	3.66	0.03	-0.20	-0.23	-0.50*	-0.53	-0.30
	(0.32)	(0.73)	(0.36)	(0.26)	(0.25)	(0.15)	(0.25)	(0.24)	(0.54)	(0.27)
Age	21.86	19.73	20.50	20.00	2.13**	1.36	-0.77	1.86	-0.27	0.50
	(1.95)	(1.95)	(1.69)	(1.41)	(0.94)	(0.94)	(0.86)	(1.51)	(1.47)	(1.31)
Gender of test-taker (male = 1)	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of children	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Marital status (married = 1)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Language: Urdu	0.30	0.55	0.36	0.50	-0.25	-0.06	0.18	-0.20	0.05	-0.14
	(0.48)	(0.52)	(0.50)	(0.71)	(0.22)	(0.22)	(0.22)	(0.39)	(0.42)	(0.40)
Language: English	0.00	0.00	0.09	0.00	0.00	-0.09	-0.09	0.00	0.00	0.09
	(0.00)	(0.00)	(0.30)	(0.00)	(0.00)	(0.10)	(0.09)	(0.00)	(0.00)	(0.22)
Father's occupation: private business	0.30	0.36	0.00	0.00	-0.06	0.30*	0.36**	0.30	0.36	0.00
	(0.48)	(0.50)	(0.00)	(0.00)	(0.22)	(0.15)	(0.15)	(0.35)	(0.37)	(0.00)
Father's occupation: kissan	0.10	0.27	0.27	0.00	-0.17	-0.17	-0.00	0.10	0.27	0.27
	(0.32)	(0.47)	(0.47)	(0.00)	(0.18)	(0.18)	(0.20)	(0.23)	(0.34)	(0.34)
Father's occupation: government job	0.20	0.00	0.09	0.00	0.20	0.11	-0.09	0.20	0.00	0.09
	(0.42)	(0.00)	(0.30)	(0.00)	(0.13)	(0.16)	(0.09)	(0.31)	(0.00)	(0.22)
Father's occupation: private job	0.00	0.00	0.09	0.00	0.00	-0.09	-0.09	0.00	0.00	0.09
	(0.00)	(0.00)	(0.30)	(0.00)	(0.00)	(0.10)	(0.09)	(0.00)	(0.00)	(0.22)
Mother's occupation: housewife	0.60	0.82	0.73	0.50	-0.22	-0.13	0.09	0.10	0.32	0.23
	(0.52)	(0.40)	(0.47)	(0.71)	(0.20)	(0.21)	(0.19)	(0.42)	(0.34)	(0.38)
Mother's occupation: government job	0.00	0.18	0.00	0.00	-0.18	0.00	0.18	0.00	0.18	0.00
	(0.00)	(0.40)	(0.00)	(0.00)	(0.13)	(0.00)	(0.12)	(0.00)	(0.30)	(0.00)
Family's monthly income: less than Rs. 10,000	0.10	0.09	0.00	0.00	0.01	0.10	0.09	0.10	0.09	0.00
	(0.32)	(0.30)	(0.00)	(0.00)	(0.13)	(0.10)	(0.09)	(0.23)	(0.22)	(0.00)
Family's monthly income: Rs. 10,000 - Rs. 100,000	0.30	0.73	0.82	0.00	-0.43*	-0.52**	-0.09	0.30	0.73*	0.82**
	(0.48)	(0.47)	(0.40)	(0.00)	(0.21)	(0.19)	(0.19)	(0.35)	(0.34)	(0.30)
Family's monthly income: greater than Rs. 100,000	0.10	0.18	0.00	0.50	-0.08	0.10	0.18	-0.40	-0.32	-0.50**
	(0.32)	(0.40)	(0.00)	(0.71)	(0.16)	(0.10)	(0.12)	(0.29)	(0.34)	(0.16)
Household size: 1-5	0.20	0.36	0.18	0.50	-0.16	0.02	0.18	-0.30	-0.14	-0.32
	(0.42)	(0.50)	(0.40)	(0.71)	(0.20)	(0.18)	(0.19)	(0.35)	(0.40)	(0.34)
Household size: 6-10	0.20	0.36	0.18	0.00	-0.16	0.02	0.18	0.20	0.36	0.18
	(0.42)	(0.50)	(0.40)	(0.00)	(0.20)	(0.18)	(0.19)	(0.31)	(0.37)	(0.30)
Household size: greater than 10	0.60	0.27	0.64	0.50	0.33	-0.04	-0.36*	0.10	-0.23	0.14
	(0.52)	(0.47)	(0.50)	(0.71)	(0.21)	(0.22)	(0.21)	(0.42)	(0.38)	(0.40)
Observations	10	11	11	2	21	21	22	12	13	13

*p<0.1, **p<0.05, ***p<0.01. Standard errors in parentheses.

Table A14: Balance on Baseline Characteristics of Test-Takers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Male test-takers with no helper in				Difference					
	Test 3	Test 4	Test 5	>1 test	(1) - (2)	(1) - (3)	(2) - (3)	(1) - (4)	(2) - (4)	(3) - (4)
GPA	3.47	3.17	2.67	3.28	0.30	0.80***	0.50**	0.19	-0.11	-0.61**
	(0.53)	(0.35)	(0.53)	(0.55)	(0.20)	(0.23)	(0.21)	(0.22)	(0.21)	(0.24)
Age	20.18	21.17	21.00	21.80	-0.98	-0.82	0.17	-1.62**	-0.63	-0.80
	(1.33)	(2.40)	(2.93)	(1.87)	(0.89)	(0.99)	(1.47)	(0.70)	(1.07)	(1.13)
Gender of test-taker (male = 1)	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of children	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Marital status (married = 1)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Language: Urdu	0.21	0.22	0.67	0.45	-0.01	-0.45**	-0.44*	-0.24	-0.23	0.21
	(0.43)	(0.44)	(0.50)	(0.52)	(0.18)	(0.19)	(0.22)	(0.19)	(0.22)	(0.23)
Language: English	0.07	0.00	0.00	0.00	0.07	0.07	0.00	0.07	0.00	0.00
	(0.27)	(0.00)	(0.00)	(0.00)	(0.09)	(0.09)	(0.00)	(0.08)	(0.00)	(0.00)
Father's occupation: private business	0.14	0.11	0.44	0.18	0.03	-0.30	-0.33	-0.04	-0.07	0.26
	(0.36)	(0.33)	(0.53)	(0.40)	(0.15)	(0.18)	(0.21)	(0.15)	(0.17)	(0.21)
Father's occupation: kissan	0.21	0.00	0.11	0.18	0.21	0.10	-0.11	0.03	-0.18	-0.07
	(0.43)	(0.00)	(0.33)	(0.40)	(0.14)	(0.17)	(0.11)	(0.17)	(0.14)	(0.17)
Father's occupation: government job	0.00	0.22	0.00	0.27	-0.22*	0.00	0.22	-0.27**	-0.05	-0.27*
	(0.00)	(0.44)	(0.00)	(0.47)	(0.12)	(0.00)	(0.15)	(0.12)	(0.20)	(0.16)
Father's occupation: private job	0.07	0.00	0.00	0.00	0.07	0.07	0.00	0.07	0.00	0.00
	(0.27)	(0.00)	(0.00)	(0.00)	(0.09)	(0.09)	(0.00)	(0.08)	(0.00)	(0.00)
Mother's occupation: housewife	0.57	0.67	0.56	0.55	-0.10	0.02	0.11	0.03	0.12	0.01
	(0.51)	(0.50)	(0.53)	(0.52)	(0.22)	(0.22)	(0.24)	(0.21)	(0.23)	(0.24)
Mother's occupation: government job	0.07	0.00	0.11	0.09	0.07	-0.04	-0.11	-0.02	-0.09	0.02
	(0.27)	(0.00)	(0.33)	(0.30)	(0.09)	(0.13)	(0.11)	(0.11)	(0.10)	(0.14)
Family's monthly income: less than Rs. 10,000	0.07	0.00	0.11	0.00	0.07	-0.04	-0.11	0.07	0.00	0.11
	(0.27)	(0.00)	(0.33)	(0.00)	(0.09)	(0.13)	(0.11)	(0.08)	(0.00)	(0.10)
Family's monthly income: Rs. 10,000 - Rs. 100,000	0.43	0.56	0.44	0.64	-0.13	-0.02	0.11	-0.21	-0.08	-0.19
	(0.51)	(0.53)	(0.53)	(0.50)	(0.22)	(0.22)	(0.25)	(0.21)	(0.23)	(0.23)
Family's monthly income: greater than Rs. 100,000	0.14	0.00	0.11	0.09	0.14	0.03	-0.11	0.05	-0.09	0.02
	(0.36)	(0.00)	(0.33)	(0.30)	(0.12)	(0.15)	(0.11)	(0.14)	(0.10)	(0.14)
Household size: 1-5	0.21	0.11	0.22	0.18	0.10	-0.01	-0.11	0.03	-0.07	0.04
	(0.43)	(0.33)	(0.44)	(0.40)	(0.17)	(0.18)	(0.18)	(0.17)	(0.17)	(0.19)
Household size: 6-10	0.21	0.11	0.33	0.27	0.10	-0.12	-0.22	-0.06	-0.16	0.06
	(0.43)	(0.33)	(0.50)	(0.47)	(0.17)	(0.19)	(0.20)	(0.18)	(0.19)	(0.22)
Household size: greater than 10	0.57	0.78	0.44	0.55	-0.21	0.13	0.33	0.03	0.23	-0.10
	(0.51)	(0.44)	(0.53)	(0.52)	(0.21)	(0.22)	(0.23)	(0.21)	(0.22)	(0.24)
Observations	14	9	9	11	23	23	18	25	20	20

*p<0.1, **p<0.05, ***p<0.01. Standard errors in parentheses.

Table A15: Balance on Baseline Characteristics of Test-Takers

	(1)	(2)	(3)	(4)	(5)	(6)
	Male test-takers with preferred helper in			Difference		
	Test 2	Test 6	Both tests	(1) - (2)	(1) - (3)	(2) - (3)
GPA	3.09	3.40	3.00	-0.31	0.09	0.40*
	(0.63)	(0.55)	(0.50)	(0.24)	(0.23)	(0.21)
Age	21.18	20.60	21.78	0.58	-0.60	-1.18
	(2.64)	(1.17)	(2.28)	(0.91)	(1.12)	(0.82)
Gender of test-taker (male = 1)	1.00	1.00	1.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of children	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Marital status (married = 1)	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Language: Urdu	0.42	0.42	0.25	-0.00	0.17	0.17
	(0.51)	(0.51)	(0.45)	(0.21)	(0.20)	(0.20)
Language: English	0.00	0.08	0.00	-0.08	0.00	0.08
	(0.00)	(0.29)	(0.00)	(0.08)	(0.00)	(0.08)
Father's occupation: private business	0.42	0.17	0.00	0.25	0.42**	0.17
	(0.51)	(0.39)	(0.00)	(0.19)	(0.15)	(0.11)
Father's occupation: kissan	0.17	0.08	0.17	0.08	-0.00	-0.08
	(0.39)	(0.29)	(0.39)	(0.14)	(0.16)	(0.14)
Father's occupation: government job	0.00	0.25	0.08	-0.25*	-0.08	0.17
	(0.00)	(0.45)	(0.29)	(0.13)	(0.08)	(0.15)
Father's occupation: private job	0.00	0.08	0.00	-0.08	0.00	0.08
	(0.00)	(0.29)	(0.00)	(0.08)	(0.00)	(0.08)
Mother's occupation: housewife	0.67	0.67	0.33	0.00	0.33	0.33
	(0.49)	(0.49)	(0.49)	(0.20)	(0.20)	(0.20)
Mother's occupation: government job	0.08	0.08	0.00	-0.00	0.08	0.08
	(0.29)	(0.29)	(0.00)	(0.12)	(0.08)	(0.08)
Family's monthly income: less than Rs. 10,000	0.08	0.00	0.08	0.08	-0.00	-0.08
	(0.29)	(0.00)	(0.29)	(0.08)	(0.12)	(0.08)
Family's monthly income: Rs. 10,000 - Rs. 100,000	0.67	0.42	0.42	0.25	0.25	0.00
	(0.49)	(0.51)	(0.51)	(0.21)	(0.21)	(0.21)
Family's monthly income: greater than Rs. 100,000	0.00	0.25	0.00	-0.25*	0.00	0.25*
	(0.00)	(0.45)	(0.00)	(0.13)	(0.00)	(0.13)
Household size: 1-5	0.33	0.33	0.00	-0.00	0.33**	0.33**
	(0.49)	(0.49)	(0.00)	(0.20)	(0.14)	(0.14)
Household size: 6-10	0.25	0.17	0.08	0.08	0.17	0.08
	(0.45)	(0.39)	(0.29)	(0.17)	(0.15)	(0.14)
Household size: greater than 10	0.42	0.50	0.92	-0.08	-0.50***	-0.42**
	(0.51)	(0.52)	(0.29)	(0.21)	(0.17)	(0.17)
Observations	12	12	12	24	24	24

*p<0.1, **p<0.05, ***p<0.01. Standard errors in parentheses.

Table A16: Balance on Baseline Characteristics of Test-Takers

	(1)	(2)	(3)	(4)	(5)	(6)
	Male test-takers with no helper in			Difference		
	Test 2	Test 6	Both tests	(1) - (2)	(1) - (3)	(2) - (3)
GPA	3.40	3.09	2.85	0.31	0.56*	0.25
	(0.55)	(0.63)	(0.32)	(0.24)	(0.29)	(0.33)
Age	20.60	21.18	20.00	-0.58	0.60	1.18
	(1.17)	(2.64)	(2.16)	(0.91)	(0.88)	(1.48)
Gender of test-taker (male = 1)	1.00	1.00	1.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of children	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Marital status (married = 1)	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Language: Urdu	0.42	0.42	0.50	0.00	-0.08	-0.08
	(0.51)	(0.51)	(0.58)	(0.21)	(0.31)	(0.31)
Language: English	0.08	0.00	0.00	0.08	0.08	0.00
	(0.29)	(0.00)	(0.00)	(0.08)	(0.15)	(0.00)
Father's occupation: private business	0.17	0.42	0.25	-0.25	-0.08	0.17
	(0.39)	(0.51)	(0.50)	(0.19)	(0.24)	(0.30)
Father's occupation: kissan	0.08	0.17	0.25	-0.08	-0.17	-0.08
	(0.29)	(0.39)	(0.50)	(0.14)	(0.20)	(0.24)
Father's occupation: government job	0.25	0.00	0.25	0.25*	0.00	-0.25*
	(0.45)	(0.00)	(0.50)	(0.13)	(0.27)	(0.13)
Father's occupation: private job	0.08	0.00	0.00	0.08	0.08	0.00
	(0.29)	(0.00)	(0.00)	(0.08)	(0.15)	(0.00)
Mother's occupation: housewife	0.67	0.67	0.50	0.00	0.17	0.17
	(0.49)	(0.49)	(0.58)	(0.20)	(0.30)	(0.30)
Mother's occupation: government job	0.08	0.08	0.25	0.00	-0.17	-0.17
	(0.29)	(0.29)	(0.50)	(0.12)	(0.20)	(0.20)
Family's monthly income: less than Rs. 10,000	0.00	0.08	0.00	-0.08	0.00	0.08
	(0.00)	(0.29)	(0.00)	(0.08)	(0.00)	(0.15)
Family's monthly income: Rs. 10,000 - Rs. 100,000	0.42	0.67	0.75	-0.25	-0.33	-0.08
	(0.51)	(0.49)	(0.50)	(0.21)	(0.30)	(0.29)
Family's monthly income: greater than Rs. 100,000	0.25	0.00	0.00	0.25*	0.25	0.00
	(0.45)	(0.00)	(0.00)	(0.13)	(0.23)	(0.00)
Household size: 1-5	0.33	0.33	0.00	0.00	0.33	0.33
	(0.49)	(0.49)	(0.00)	(0.20)	(0.25)	(0.25)
Household size: 6-10	0.17	0.25	0.50	-0.08	-0.33	-0.25
	(0.39)	(0.45)	(0.58)	(0.17)	(0.25)	(0.28)
Household size: greater than 10	0.50	0.42	0.50	0.08	0.00	-0.08
	(0.52)	(0.51)	(0.58)	(0.21)	(0.31)	(0.31)
Observations	12	12	4	24	16	16

*p<0.1, **p<0.05, ***p<0.01. Standard errors in parentheses.

10.2 Source of questions

Economics questions sourced from Macroeconomics, 9th Edition, Michael Parkin, University of Western Ontario and Microeconomics, 8th Edition, Pindyck & Rubinfeld. Sports questions sourced from online sources.[19] Cooking questions were also sourced from online sources.[20] All questions were tested in a pilot so ensure that they were easy to understand. There was only 1 correct answer per question.

10.3 Appendix figures

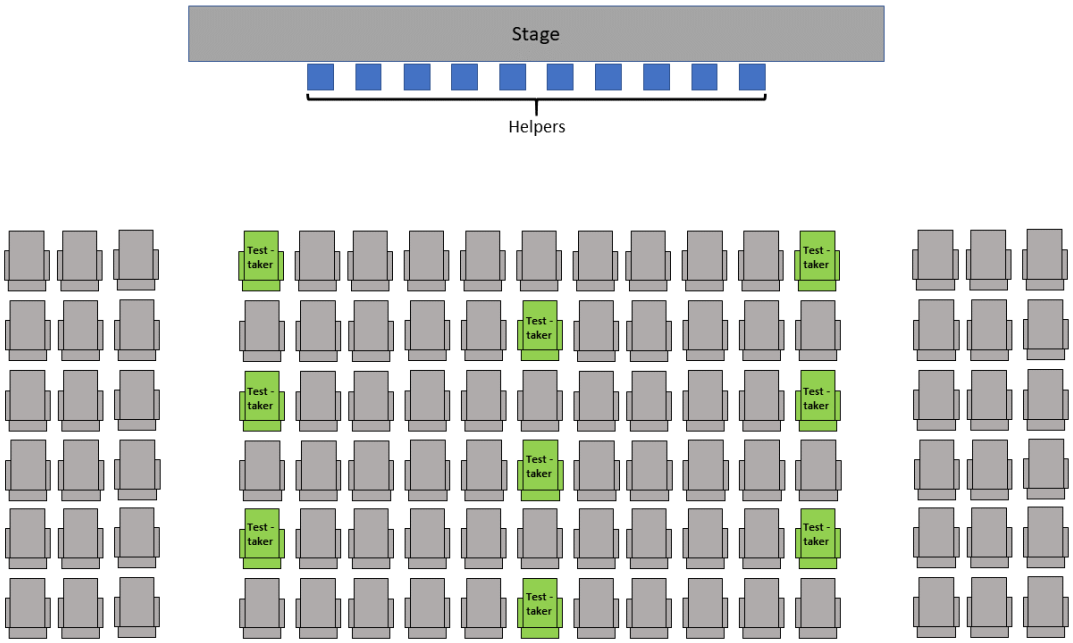

Figure A1: Seating arrangement

[19](1) https://www.quiz-questions.net/sports.php; (2) http://www.pakpassion.net/pakpassion-quiz/411-a-quiz-about-pakistan-cricket.html; and https://www.britannica.com/quiz/cricket-quiz.

[20](1) http://whyfund.net; (2) https://www.dayjob.com/chef-quiz-3042/; (3) https://www.britannica.com/quiz/cuisine-of-india; (4) https://howtocookgreatfood.com/food-quiz-indian-curry/; (5) https://cooking.stackexchange.com/questions/87659/can-sour-cream-be-whipped?rq=1; and (6) https://www.dayjob.com/food-hygiene-quiz-829/.

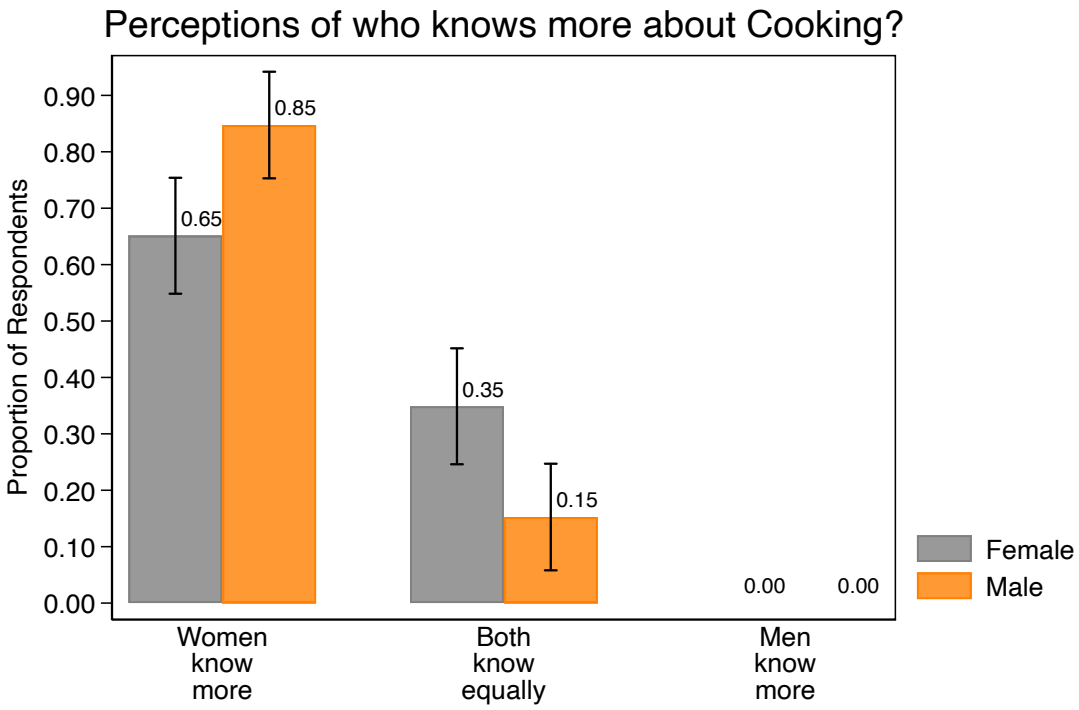

Figure A2: Perceptions of gender stereotypical nature of the different subjects

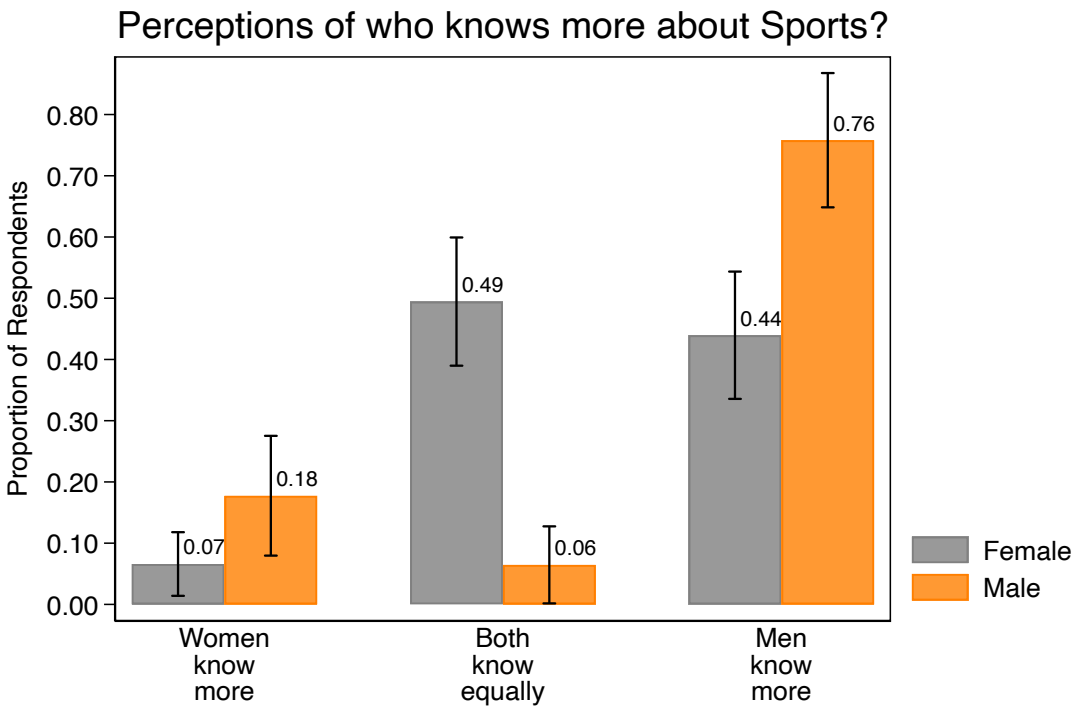

Figure A3: Perceptions of gender stereotypical nature of the different subjects

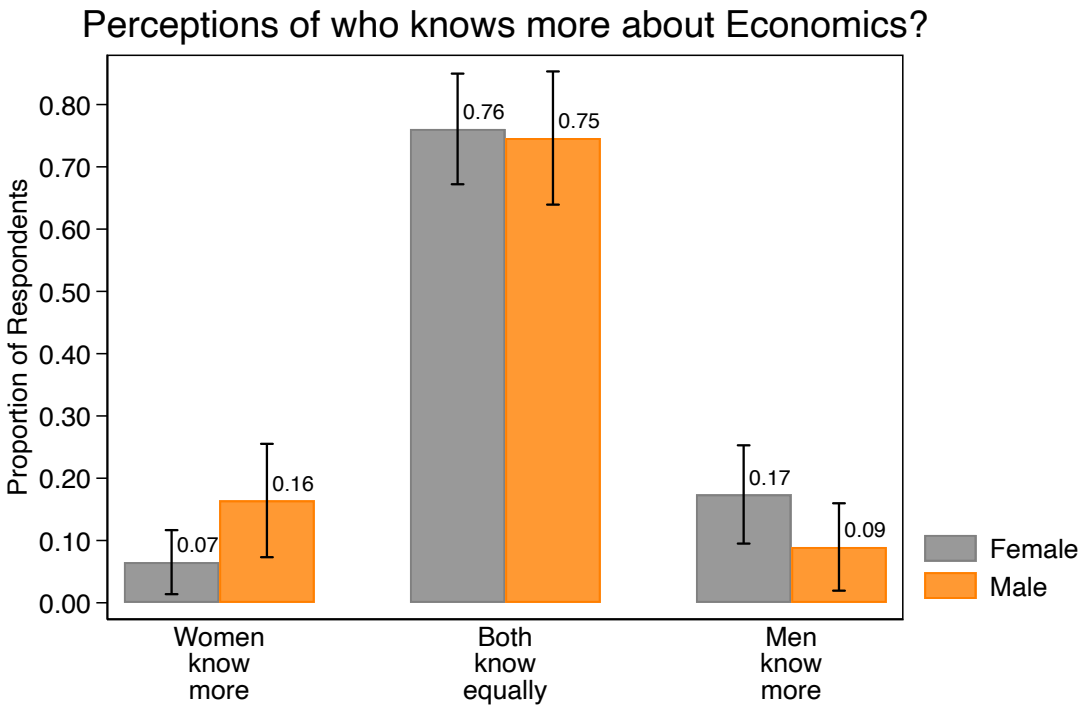

Figure A4: Perceptions of gender stereotypical nature of the different subjects