Gender, Information Exchange and Choice over Co-workers: Experimental Evidence*

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

August 31, 2021

Abstract

We study whether choice over co-workers matters for information exchange and performance in gender diverse teams. We carried out a lab-in-the-field experiment where 'test-takers' were randomly assigned 'helpers'. Helper allocation was randomised on two dimensions: (1) gender and (2) test-taker preference for that helper at baseline. To directly investigate frictions in information exchange, we cross-randomized the availability of hints to help with the task. We find that random helpers are significantly worse for the performance of female test-takers than preferred helpers and that these differences are entirely driven by male helpers. Results show that the gender stereotypical nature of the task matters for female test-takers. In some categories, the negative effect of random helpers can be explained by the lower use of alternative source of information by the test-taker (hints) in the presence of these helpers. This suggests that working with a partner of the opposite gender creates frictions for female workers, but that these frictions can be mitigated by allowing choice over co-workers. Organizations can benefit from encouraging women to select their co-workers in team formation.

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

Keywords: gender, diversity, team, performance, stereotype, information

*We are grateful to Joydeep Bhattacharya, Leora Friedberg, Pamela Jakiela, Subha Mani, Amalia Miller, 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 and Elizabeth Claire Schroppe 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 (Nova SBE) NOVAFRICA, and BELAB. Email:brais.pereira@novasbe.pt

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

1 Introduction

Collaborative activities and teamwork have gone up by 50% over 20 years (Cross et al., 2016; Lazear and Shaw, 2007). Alongside this increase in teamwork in organizations, increasing gender diversity in teams 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 specialization and combining more skills and knowledge, but a more diverse team will 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 who share affinities.

In this paper we investigate the effect of choosing a co-worker on information exchange and performance and test whether this effect differs across genders. We find that allocating a randomly-chosen male co-worker to a female worker significantly reduces the performance of the female worker, relative to allocating a male co-worker of her choice. This occurs even though the two types of male co-workers have the same ability. This effect disappears when pairing two female co-workers. If workers perform better when paired with diverse co-workers of their choice, rather than those they are randomly allocated to, allowing choice over team formation can 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 improve performance relative to pairing co-workers with higher ability but lower affinity.

We designed a lab-in-the-field experiment conducted with undergraduate students in Pakistan in which we randomised both the gender and the affinity of co-workers. To do that, we created teams of two students, each assigned a different role: 'test-taker' or 'helper'.[1] Test-takers were asked to complete a series of six multiple-choice tests and were the main participants of interest. Helpers were assigned to support test-takers. Our experiment used a within-person design and 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 with a list of helpers from their degree-year and asked them to rank these as helpers. We elicited these preferences before the start of the test and avoided making gender salient at any stage of the experiment.[2] Preference for helpers corresponded to affinities with this helper. The most-cited reason for ranking a helper as more preferred was being close friend with them (71% of female test takers and 47% of male test-takers). We then allocated preferred helpers to test-takers using a random serial dictatorship (Abdulkadiroğlu and Sönmez, 1998; Breza and Chandrasekhar, 2019). This resulted in 91% test-takers working with their top 5 most preferred helpers whenever they were allocated a preferred helper.

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

[2] Studies have shown that making gender salient itslef can affect behaviours (cf. Hoff and Pandey, 2006; Benjamin et al., 2010; Boschini et al., 2012; Bordalo et al., 2012; Bertrand and Duflo, 2017).

This design resulted in four main treatments for test-takers: being allocated a random female helper, a preferred female helper, a random male helper, or a preferred male helper. Each treatment was randomly assigned across tests, with each test consisting of 30 questions. In the control test, test-takers were not assigned any helper. Our main outcome of interest was the performance of a test-taker, measured as a dummy that turns on 1 whenever an answer is correct. 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 piece-rate according to the number of correct answers in a randomly chosen test with no penalty for wrong answers.[3] Incentives were kept fixed throughout the experiment.

In addition to this main treatment, we randomly allocated hints to different questions within a test. Hints are cues to the correct answer. Hints were available on 80% of questions in a test. Hints carried no penalties and were available through scratch sheets. This allowed us to measure for which questions test-takers accessed hints and as a result, whether they used hints at different rates in the presence of different helpers. We use these results to investigate whether there is a change in the use of alternative sources of information by the test-taker (hints) in the presence of these helpers. Finally, the questions were grouped in three categories: economics, sports, and cooking. This allowed us to evaluate whether there is any heterogeneity across categories that are considered more or less gender stereotypical in the context of the experiment.[4]

Our research design relies on recent developments in the empirical estimation of peer effects and departs from previous peer effect studies in a few ways. First, we only study 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 subject and their peers helps overcome mechanical correlations in outcomes, which have 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 allocate helpers with such characteristics to test-takers. This becomes equivalent to an allocation to a treatment arm rather than an allocation to groups (which might have chance variation in characteristics). Finally, we stratified the students on degree-year-gender-GPA and created two groups of participants at baseline. Test-takers and helpers were chosen without replacement from different pools. This helps overcome what has been called 'exclusion bias' in the estimation of peer effects in the literature, a bias that cannot be overcome just through random assignment of peers (Guryan et al., 2009; Caeyers and Fafchamps, 2016; Angrist, 2014).

[3] 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).

[4] 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 female (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 gender stereotypical nature of the task is important, since it has been shown to be an important determinant of behaviour differences across men and women (Coffman, 2014; Bordalo et al., 2019; Bordalo et al., 2016).

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%, 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 to a preferred vs. random helper are statistically significant at the 5% level.

Our second result is that the differences between random and preferred helpers for female test-takers is entirely driven by male helpers. Random male helpers statistically significantly reduce performance of women (6 percentage points, 12 % of average score), while preferred male helpers have a positive yet statistically insignificant effect (3 percentage points, 6% of average score). This arises despite the fact that preferred and random helpers have the same abilities (measured through a baseline test and their university GPA). This result suggest that the positive effect of pairing workers based on affinity can be even more important in gender-diverse teams. An *F*-test testing similarity of the effect for random and preferred male helpers is almost significant with a p-value of 0.15.

Our third results is that gender stereotypical nature of the task is important for performance with helpers. In particular, in the sports category, random male helpers have a negative though statistically insignificant effect on the score of female test-takers (5 percentage points, 10% of their average score), while preferred male helpers have a positive and statistically significant effect (14 percentage points, 28% of their average score). The difference between preferred and random male helpers is large and statistically different than zero in this category. It is important to note that sports is the category with the highest gender gap in knowledge and the category where women stand to benefit the most from male helpers. This implies that females would have performed significantly better in this task had they always been given the choice of whom to work with. In fact, the latter would have even performed better alone. Since in principle, test-takers could ignore the information provided by the helper if it is not helpful, this suggests that the presence of random male helpers prevents test-takers from making the optimal use of the information available. This first result supports our claim that pairing workers who share affinities can be a low cost strategy to improve performance. The case of gender stereotypically female category like cooking is different. We find that working with men, irrespective of affinity, reduces performance. Both random and preferred male helpers lower the performance of women (12 percentage points, 20% of average score). While, female helpers, irrespective of affinity, have no effect.

In cooking, the negative effect of random helpers can be explained by the lower use of alternative source of information by the test-taker (hints) in the presence of these helpers. Female test-takers have a 7 percentage points (9% of average hints scratched) lower probability of scratching hints in cooking when they are paired with random male helpers than working alone. To assess this mechanism and guide our interpretation of these results, 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 has access to neither hint nor helper, the test-taker picks an answer according to this prior and performance alone therefore captures the prior or ability of the test-taker. When a helper is present but not a hint, the test-taker processes the information from the helper as an additional source of information, updates her beliefs and chooses the most likely answer. The increase in performance in the presence of the helper therefore depends on the relative ability of the helper and test-taker. 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 the expected gain in the presence of the helper depends on the test-taker's perception of the helper's ability, the rate of hint uncovering across different treatments provides us with some information about the test-taker's expected gains from different types of helpers. Since test-takers face piece-rate incentive to perform well, we should expect them to optimally use all sources of information available to them (hints and helpers).

The theory predicts that if there is no difference in ability between two types of helpers we should see no difference in the test-taker's score between these two types of helpers. Our results are consistent with this prediction for female helpers paired with female test-takers. However, they are inconsistent with the difference between random and preferred male helpers that we observe for female test-takers. In some categories, these differences can be explained by a lower rate of hint-taking in the presence of random helpers, suggesting that test-takers face a higher cost of uncovering hints with these helpers. In other categories, changes in performance cannot be explained by hint-taking behaviour.

We conclude the paper by proposing three different mechanisms that can explain the difference in costs of uncovering hints across helper types and the remaining differences in performance that are not driven by differences in accessing of hints. First, test-taker may face social image concerns or peer pressure. These can lead them to take fewer hints either because they want to signal to a helper that they have high ability, or because they worry about offending the helper by questioning their advice. It can also generate lower performance even without differences in hint uncovering, if it leads test-takers to follow the advice of helpers rather than that of the hint for fear of offending helpers. Second, different types of helpers might affect both the performance and the hint-uncovering behaviour of test-takers if test-takers misjudge the ability of different types of helpers. This is consistent with our result that a majority of female expected males to have lower knowledge of cooking even though there was no difference across genders at baseline. Third, test-takers might face more difficulties communicating with certain type of helpers. This can lead them to debate how to interpret a hint or to misunderstand the suggestions of a helper and lower their performance. While we cannot disentangle these mechanisms with our experimental design, we discuss how well they can rationalise our data.

Our results have important policy implications for workplace diversity initiatives. First, in

situations where this is possible, allowing women to select advisors or teammates of the opposite gender can be a low cost lever in the hands of organizations to improve both information flow and productivity. Alternatively, when allowing workers to choose co-workers is not feasible, managers can benefit from forming teams around the affinity of workers, rather than simply skill complementarities. This situation can arise in common team formation scenarios in public and private organizations. For example, a junior worker might need to be assigned a senior co-worker who can help and guide them on how to prepare a report. If the manager knows that both senior workers have the same ability but one has more affinity and past experience collaborating with the junior, then our results suggest that the second might be better. Our results suggest that there exists a trade-off between assigning high ability seniors with no affinity versus lower ability seniors with better affinity with junior worker. It is possible that the lower ability of the senior worker can be outweighed by better communication of information between higher affinity senior-junior pair.

Secondly, since our results suggest that preferred male co-workers have less negative impact than randomly-assigned male co-workers, then team-building activities can improve performance as they can effectively turn randomly-assigned co-workers into friendly co-workers.

A third policy implication is that the recent push to increase diversity in pink-collared jobs (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 can mitigate the negative effect of being paired with an unknown co-worker of a different gender. Our results show that the returns to different types of co-workers depend on the nature of the task. If the perceived abilities of different types of co-workers differs from their actual abilities (as we observe in the cooking category), then training that addresses potential misperceptions of abilities across genders can have an impact on performance.

Related literature. Our paper 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; Breza et al., 2018; Cohn et al., 2014; Blanes i Vidal and Nossol, 2011); or peer pressure or mutual monitoring (Knez and Simester, 2001; Jones and Kato, 1995; Kandel and Lazear, 1992). Ashraf and Bandiera (2018) give an excellent survey of the key theoretical and empirical papers in the social incentives at workplace literature. Some studies have shown that friendships amongst co-workers matters for performance in organizations (Bandiera et al., 2010), and attending a business training with a friend improves business activity for women with restricted mobility (Field et al., 2016). A few studies investigate the effects of diversity in peers on performance. Rasul and Rogger (2015) describe that there is a positive association between increased ethnic

diversity and performance in the public sector in a developing country. 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.

This paper contributes by investigating gender peer effects, and focusing in particular on information exchange and how that can be facilitated by allowing choice over co-workers to women. Calder-Wang et al. (2021) is the only existing study that opens up the discussion along these lines. They use 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 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 (Caeyers and Fafchamps, 2016; Angrist, 2014). Moreover, by randomly allocating hints across helpers, we are able to directly investigate whether the effect of helpers on performance depends on the availability of different sources of information and whether helpers affect the demand for information from hints. By varying the gender stereotype of the task, we are also able to test how gender and preference for helpers interacts with the gender stereotype of the task.

Since helpers act in an advising capacity and aid test-takers perform, 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 (Fairlie et al., 2014); female students benefit from being taught by female professors (Carrell et al., 2010; Hossain, 2021); a higher share of women board of directors results in higher proportion of female top executives the following year (Matsa and Miller, 2011); and male teachers prefer working with male principals (Husain et al., 2018).[5] Results in our paper show that male helpers can lower performance and information exchange for female test-takers. We contribute to the literature by showing that these problems are only there for those male helpers that are randomly chosen. Therefore, our paper shows that whether working with opposite gender co-worker matters for women depends on whether they can choose who acts as an advisor. Moreover, our paper also highlights the important of gender stereotypical nature of the task.

The paper also relates to the broader literature on how diversity affects economic development.[6]

[5]Based on same-type mentorship being more beneficial for minorities, theoretically, Athey et al. (2000) and Muller-Itten and Oery (2021) describe the long run effects of different affirmative action policies. Blau et al. (2010) and Ginther et al. (2020) show that women academics benefit from mentorship.

[6]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

Based in a developing country, our paper provides micro evidence on how gender interactions take place in organizations and how those can impact information exchange across genders. Our paper contributes to this literature by evaluating co-worker choice as a form of low cost policy to enhance performance for women in gender diverse teams.

2 Context and research design

2.1 Context

During 2020, we collaborated with a university in Punjab, Pakistan. The university operated a morning and evening shift. For logistical ease, the experiment was conducted with male and female undergrad students that studied in the morning shift. Students pursuing degrees in Business Administration, Commerce, and Public Administration in semesters one, three, five, and seven, as well as degrees in Industrial Management (IM) in semester seven, were chosen. These degrees were useful since all students studied Economics, and within the local context Economics is viewed as a gender neutral category. That provided us with a useful contrast against gender stereotyped categories like Cooking and Sports. The experiment took place in September in the university's auditorium. Since there was sufficient space to ensure that there was no spillovers across dyads of test-takers and helpers, the entire experiment was conducted in one go, on the same day.

2.2 Experimental design

Our lab-in-the-field experiment aims to assess whether gender and preference over co-worker matters for performance and whether the channel through which this effect is mediated is information exchange. We carried out a lab-in-the-field experiment especially because it allowed us a level of control over the helpers and the nature of the task that was essential to understand communication frictions across people of different genders.

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 performed on tests and were the main participants of interest, while helpers helped test-takers perform. 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 can maximize their performance. Since we were constrained in the number of female students, this rotation design was also helpful for power.

et al., 2012); reduces social solidarity and social capital (Putnam, 2007); negatively impacts choice of public policy and economic growth Easterly and Levine, 1997; reduces firm productivity (Parrotta et al., 2014). Chattopadhyay and Duflo, 2004 focus on gender diversity and show that reservation of a council seat for women affects the 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.

At baseline, we stratified on degree-year-gender-GPA and randomly sampled two groups from within each stratum. From within these two groups, we randomly picked some participants as test-taker and others as 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 A3 for distribution of test-takers and helpers sampled across degrees). Note that these 154 female students were 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.

The exact match between helpers and test-takers had to happen the day of the experiment as we were not sure which students from our list would show up. We began by randomly seating the test-takers. Before the start of the experiment we placed folded pieces of papers with numbers on them in a bowl and we asked the group of test-takers to select one. Once everyone had a number we asked them to stand in a line in an ascending order based on their number. We started seating test-takers beginning from the bottom extreme right corner, the next was seated on bottom extreme left and then the next on back-center (details can be seen in appendix figure A1). Then the pattern was repeated for all test-takers in a degree-year. Test takers remained seated for the whole experiment, while helpers moved.

2.2.1 Random Helpers

For the random helper treatment for each degree-year, we asked the students in the helper group to select papers from a bowl with numbers on them. Like the test-takers, we then asked them to stand in ascending order. Gender was never mentioned throughout the experiment as we did not want to make gender salient and potentially change behaviours owing to that (cf. Hoff and Pandey, 2006; Benjamin et al., 2010; Boschini et al., 2012; Bordalo et al., 2012; Bertrand and Duflo, 2017). The enumerators were trained to start from the bottom right corner allocating a random female helper and then a random male helper to the bottom left corner and leave the center test-taker alone (control) in the first round. And then to switch patterns across rounds. These patterns were agreed to by the research team before the experiment. After each test-round, helpers were asked to come to the front of the room. Test takers remained seated. Helpers were asked to choose papers from the bowl again and then the whole exercise was repeated. With the difference that for each test-taker a different helper type was allocated across rounds. This generated random variation in the helper types that test-takers worked with across tests.

2.2.2 Preferred Helpers

For the preferred helper treatment, we started with first taking the preferences for helpers from the test-takers. This was done before the start of the experiment and right after consent. We already had our group of helpers that we had sampled before the start of the experiment from within each

degree-year. We gave these lists to test-takers and told them that since we would like to consider their preferences when allocating helpers to help them perform on tests, they could rank these students in order of preference to work with them as helpers for the study. Preferences were not taken by gender and neither was gender mentioned on the list on which test-takers rank-ordered helpers. As discussed this was done as we did not want to make gender salient. We used a random serial dictatorship to allocate preferred helpers across test-takers. The bottom right test-taker got their first preference, followed by the bottom left and then back. The random serial dictatorship allocation mechanism has been shown to not only be Pareto efficient, but also strategy-proof (i.e., it cannot be manipulated by misrepresenting preferences) (Abdulkadiroğlu and Sönmez, 1998; Breza and Chandrasekhar, 2019). Second, it is easy to implement in practice and therefore policy relevant. Third, this design is easy to analyze given the randomization of the choice order. Table 3 describe whether test-takers worked with their most preferred helpers in the preferred treatment. 57% female and 50% male test-takers worked with their most preferred helpers, while 87% female and 78% male test-takers worked with their top 3 most preferred helpers. 91% test-takers (male and female) worked with their top 5 most preferred helpers.

For logistical ease as well as to ensure that test-takers are 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. In the third, fourth, and fifth tests, test-takers were randomly matched with male, female helpers while the rest did their tests alone (see figure 3 for an example of the allocation across rounds 2-6).

Table 2 describes the proportion of male and female helpers that were either random or preferred and were allocated to male and female test-takers. The probability of a female helper given that they are preferred was 75% for female test takers, while 25% of female test-takers worked with preferred male helpers. 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% male test-takers worked with preferred helpers that were females 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'.[7] In the case of gender, Ibarra, 1992 shows that the structure of social networks depends on gender.

2.2.3 Hint availability

To investigate whether there were any communication frictions between test-taker and helpers, we cross-randomized the availability of hints for questions. Hints were prepared by the research

[7]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. (cf. McPherson et al., 2001)

team.[8] These were simple cues to help answer the question and were randomly available for 80% of questions per test. We made hints available through hint sheets which could be scratched to uncover hints for any question. Providing hints through hint sheets was useful as it allowed us to observe the exact questions in which information was accessed by test-takers.

Table below summarizes our treatments. Across test rounds, each test-taker was randomly matched with a random or preferred helper of either gender. The counterfactual was test-takers 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 sample of test-takers with different treatments across test rounds.[9] Appendix table A4 and A5 describe the proportion of test-takers with helpers of different types across different test rounds. Our plan was that each test-taker should get a treatment once. However, there were a few instances when a few test-takers either got a treatment more than once or didn't get either the random or preferred helper treatment. Appendix table A8 to A17 presents balance tables testing that these test-takers were not systematically different across rounds. Last, in 75 test rounds, some 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 tests were dropped from the analysis.

Incentives. Incentives were kept fixed throughout the experiment. Both test-takers and helpers was given PKR 100 as a participation fee. Test-takers were paid piece-rate according to the number of correct answers and were incentivized to maximize their performance. For them, one test was randomly selected out of the six and PKR 15 was given for each right answer. 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 could receive a maximum of PKR 600. In addition, all participants could play a game and then increase their PKR 100 participation fee by two and a half times. Through

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

[9] Recall that 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 treatment and control in the usual sense. Balance tables testing whether people who left were systematically different from those who stayed are in appendix table A7 and A13. Appendix table A7 show that for female test-takers those who left were not systematically different from those who stayed. However for male test-takers appendix table A13 shows that those who were younger and had a higher GPA were more likely to leave the test.

a lucky draw 6 participants were given mobile phone power banks. The stakes were high enough that we received many requests later to repeat the experiments so they could earn more money.

Timing. The experiment progressed as follows: After consent, each participant filled out a survey that asked about their preferences for helpers. They were also given a short quiz on understanding how to answer questions on confidence in their performance on tests. Next, all participants (test-takers and helpers) were given a baseline test to measure 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. Participants then played a lucky draw game in which they could win mobile phone power banks. After the experiment their scores were calculated and payments were made through mobile money ('easypaisa'), which is quite common in Pakistan (see figure 2 for details of the sequence of activities).

Differences in research design to estimate peer effects. Our research design relies heavily on the latest developments in empirical estimation of peer effects and departs from previous peer effect studies in at least two important ways. First, we only study 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 subject 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 allocate 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 two groups of participants at baseline. Test-takers and helpers were chosen without replacement from different pools. This helps overcome what has been called 'exclusion bias' in the estimation of peer effects in the literature, a bias that cannot be overcome just through random assignment of peers (Guryan et al., 2009; Caeyers and Fafchamps, 2016; Angrist, 2014).[10].

3 Theoretical framework

We conceptualise helpers and hints as two independent sources of information who 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 $X \in \{A, B\}$ the correct

[10]Exclusion bias arises from the fact that subject i and i's peers all come from the same pool and "assignment of peers is done without replacement i.e. i cannot be his own peer....the exclusion of i from the pool of i's peers creates a small sample negative relationship between i's characteristics and that of his peers: if i is above average,the average of those remaining in the pool is lower than i; conversely, if i is below average, the average of those remaining in the pool is higher than i. Hence i's characteristics are negatively correlated with the expected value of the remaining peers in the pool." (Caeyers and Fafchamps, 2016, p.2)

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: $a = A$ or $b = B$ and a payoff of 0 otherwise.

The 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: $S_A = \mathbb{P}(X = A) \times 1 + \mathbb{P}(X = B) \times 0 = \mu$.

The helper shares a message $m \in \{m_a, m_b\}$ correlated with the correct answer. We assume that that message is informative and its precision is $\mathbb{P}(m_X|X) = p_m > \frac{1}{2}$. Similarly, 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 helper therefore indicates the correct answer with probability p_m, but can also suggest the wrong answer with probability $1 - p_m$. Similarly, the hint is helpful because following it makes it more likely to get the correct answer, but it does not guarantee the right answer (it could be misinterpreted for example). 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.

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 $x(m)$ 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. 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. We do not model explicitly the way the test-taker evaluates this opportunity cost, but we think that this is the main source of cost.

To summarise, the timing is as follows:

1. If a hint is available, the test-taker evaluates her expected score with and without hint and decides whether to take the hint.

2. If both a helper is available, and if a hint is available and uncovered, the test-taker receives the two independent signals.

3. The test-taker updates her beliefs given the signals and chooses the most likely answer.

3.1 Expected value of hint, 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 hint to obtain the test-taker's expected value of taking the hint. We then evaluate whether the test-taker would uncover the hint at different levels of the cost of uncovering κ. Finally, we compute the expected score in a treatment, conditional on the hint uncovering behavior of the test-taker.

Let S_A denote the expected score when the test-taker 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 when she has access to the helper only. Finally, S_{MH} denotes the score when she has access to a helper and a hint (and chooses to use the hint). We then have:

$$
\begin{aligned}
S_A &= \mathbb{P}(X = A) = \mu \\
S_H &= \mathbb{P}(X = A) \sum_{h \in \{h_A, h_B\}} \mathbb{P}(h|A)\mathbb{1}\{x(h) = a\} + \mathbb{P}(X = B) \sum_{h \in \{h_A, h_B\}} \mathbb{P}(h|B)\mathbb{1}\{x(h) = b\} \\
S_M &= \mathbb{P}(X = A) \sum_{m \in \{m_A, m_B\}} \mathbb{P}(m|A)\mathbb{1}\{x(m) = a\} + \mathbb{P}(X = B) \sum_{m \in \{m_A, m_B\}} \mathbb{P}(m|B)\mathbb{1}\{x(m) = b\} \\
S_{MH} &= \mathbb{P}(X = A) \sum_{m \in \{m_A, m_B\}} \sum_{h \in \{h_A, h_B\}} \mathbb{P}(h|A)\mathbb{P}(m|A)\mathbb{1}\{x(m, h) = a\} \\
&\quad + \mathbb{P}(X = B) \sum_{m \in \{m_A, m_B\}} \sum_{h \in \{h_A, h_B\}} \mathbb{P}(h|B)\mathbb{P}(m|B)\mathbb{1}\{x(m, h) = b\},
\end{aligned}
$$

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

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. 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.

Given these expected scores, we can obtain the value of uncovering the hint when alone as $S_H - S_A$ and 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 $S_H - S_A > \kappa$ and in the presence of the helper if $S_{MH} - S_M > \kappa$.

Finally, let S_H^* and S_{MH}^* the expected scores of the test-taker (with and without 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, $S_H^* = S_A$. 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 behaviour) for all possible combinations of parameters. For clarity, we restrict attention to the more interesting cases in which the helpers and the hint are valuable to the test-taker. In addition, we assume that the hint is more informative than the helper, who in turn is better-informed than the test-taker. That is, we assume that $p_H > p_M > \mu$.

3.2 Testable predictions

We can use the stylised 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 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 hint when facing different types of helpers, and on 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 hint should only depend on the relative precision of the hint and the helper.

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 emphasising 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 if the helper makes the test-taker less likely to uncover the hint. This arises 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 of 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 $S_A = 0.2$, with just a hint $S_H = 0.6$, with the helper

$S_M = 0.5$, and with both helper and hint it is $S_{MH} = 0.7$.[11] 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 $S_H - S_A = 0.4 > 0.3$. When the test-taker considers taking the hint in the presence of the helper, she no longer uncovers the hint since $S_{MH} - S_M = 0.2 < 0.3$. 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.*

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 by six rounds of twenty-five-minute tests. A baseline test was administered to both test-takers and helpers. Test-takers then 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 total questions 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 are no section order effects in performance. Second, 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. Cooking and Sports category questions were sourced from the internet while Economics questions were taken from Economics textbooks (see appendix subsection 11.1 for details).

4.2 Measuring hint uncovered

Hints were available on a question-by-question basis. 80% of questions had hints available. Test-takers were given a separate sheet that had scratch-off panels under which hints were written for

[11] 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.

each question. An example of the sheet is shown in Figure A4. The number of panels scratched off could then be counted to determine how many hints the test-takers utilized and for which question. 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, 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. Coffman, 2014 p.1633 describes the method: *"Participants were told there are 100 robots, each with integer probability on [1,100] of answering a question correctly. That is, there was a robot who would answer correctly 1% of the time, a robot who would answer correctly 2% of the time, ... and a robot who would answer correctly 100% of the time. They were told that for each question, one robot would be drawn at random who could answer the question for them. They have to decide which robots they would allow to answer the question for them; they submitted an "accuracy threshold"—a number, Y, between 1 and 100 such that if robot answered correctly less than Y% of the time, they would prefer to have their own answer submitted instead." A correct answer submitted, whether it was theirs or the robot's, earned full points. Based on this mechanism, after each question test-takers were supposed to ans: "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."*

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

The top panel of Table A6 presents test-taker belief in baseline test (without helpers or hint) that the answer is correct. This is presented both for male and female test-takers. Overall men are more confident than women that their answer is correct. If we break it down by subjects, the gender gap in confidence can be seen in Economics and Sports. Sports is the category with the largest gender gap in confidence and as we will see later that is where there is a gender gap in performance and demand for information. 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 women more than men believe that they performed better than the class in Economics and Cooking, however the gap in not very large. In the case of Sports, compatible with the other measure, the gender gap in self confidence is quite large.

Figure 5 shows the gender gap (female minus male test-takers) in hints uncovered and baseline test score by subjects (Economics, Cooking and Sports). 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 women than men). This remains the case by subjects as well. Irrespective of gender

gap in baseline test scores, the gender gap in hints is positive. Second, the biggest knowledge gap for women lies 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 men. Sports is the category in which women seek the most new information. Therefore, working with a male helper can benefit women the most in Sports.

4.4 Ability of preferred versus random helpers

We first look at survey data to understand the stated reasons for test-taker preference of helpers. Table 5 describes these for the sample of matched test-takers and helpers. 71% female test-takers gave a helper their top rank stating close friend as a reason, while 21% stated high GPA as a reason. Female test-takers also stated female (4%) and 'nice' (4%). Male test-takers are similar. 47% gave a helper a top rank stating close friend as a reason, while 13% stated high GPA as a reason. Male test-takers also stated female (13%), 'nice' (13%), class fellow (7%), caste (7%) and younger than me (7%).

Next we use a regression framework to test whether preferred and random helpers are systematically different on ability. This allows us to understand whether a difference in performance of test-takers across preferred and random helpers is because of a difference in ability. We measure ability of helpers using two sources of data. First is helper GPA, that is from the administrative data of the university collected at baseline. The second source is the baseline test. This allows us to obersve helpers' ability in all these three disciplines i.e Economics, Cooking and Sports. Table 6 presents results. It shows that preferred helpers with either female and male test-takers are not systematically selected on ability. This is the case, irrespective of how ability is measured. These results suggest that if we observe a change in performance of test-takers across preferred or random helpers, it is not due to ability.

5 Results

5.1 What is the effect of gender and choice over helpers on performance?

5.1.1 Estimation

The estimation for a test-taker (i), in question (q), subject (s) and test-round (t) is as follows:

$$\begin{aligned} Correct\ ans_{iqst} = \pi\ Random\ female\ helper_{iqst} + \phi\ Preferred\ female\ helper_{iqst} \\ + \omega\ Random\ male\ helper_{iqst} + \theta\ Preferred\ male\ helper_{iqst} + \alpha_i + u_{iqst} \end{aligned} \quad (1)$$

where $Correct\ ans_{iqst}$ is a dummy for whether the answer to a question was correct. $Random\ female\ helper_{iqst}$

is a dummy variable that takes value 1 if a test-taker is allocated a female helper chosen randomly. $Preferred\ female\ helper_{iqst}$ on the other hand is a dummy variable that takes value 1 if a test-taker is allocated a female helper that they ranked among 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 coefficient of interest are π, ϕ,ω and θ and comparison of interest is between π and ϕ and ω and θ.

5.1.2 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 decrease 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%, relative to working alone. The effect is statistically significant and 6% of the mean of 0.51. This overall negative effect is significantly different from the effect of preferred helper. An *F*-test at the bottom of the table testing equivalence of the two effects has a p-value of 0.02. This difference remains true in all categories except Cooking.

In Column (4), female test-takers have a 6% higher probability of a correct answer when they work with their preferred helper. The effect is also economically significant at 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 have a lower performance when paired with helpers of either type, relative to when they are alone. An *F*-test at the bottom of the table testing similarity of the effect has a p-value of 0.6. These results suggest that working with preferred helpers have benefits for women when the category is not gender-stereotypically female. In gender-stereotypically female cases, it is better for performance of women to be alone rather than work in a team with men.

Table 7, Columns (5)-(8) break down the random and preferred helpers by gender. Results suggest that the entire negative effect of random helper 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% (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 testing similarity of the effect across random and preferred male helpers has a p-value of 0.15. 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%, 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 reduces by 5% relative to being alone. Although the effect is not statistically significant, an *F*-test at the bottom of the table testing 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.

Results for gender-stereotypically female category like Cooking present a different picture. 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 testing 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 over 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 testing similarity of the effect across preferred and random female helpers fails to reject the null. This is true overall as well as by subjects.

We also explore whether choice matters more with male or female helpers. Table A19 reports results. Each cell in table A19 reports the coefficients from table 7 column (9). These report the change in probability of a correct answer when working with a helper of a certain type. The reference category remains the same i.e. test-taker alone without any helper. Male random helpers are worse than female random helpers by 6% points. The difference is significant at a 10% level. On the other hand, gender of helper is not important when the helper is a preferred one. The difference is 1% and not significant.

Choice matters more for male helpers than female helpers. Preferred female helpers are better than random female helper by 3%, however, the effect is not significant. On the other hand, preferred male helpers are better than random male helpers by 10% points. The difference is larger than for female helpers, but it is not significant. Overall, Table A19 shows that choice does matter more for male as opposed to female helpers. However, these differences are not statistically significant. Working with a random helper is worse than a preferred helper. This is the case irrespective of the gender of these random and preferred helpers. However, the fall in performance when working with a random male helper is much larger than a female random helper.

These results suggest two main ideas. First, choice over *male* co-workers is an important

determinant of performance of 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 fall in performance of female workers, relative to being alone.

5.1.3 Results for male test-takers

Table 8 presents results for male test-takers. Columns (1)-(4) show the effect of random or preferred helper of any gender on the performance of male test-takers. Overall results in Column (1) show that there is no differential effect of randomly chosen or preferred helpers for male test-takers. Both increase the probability of a correct answer by 4% (8% of control mean) relative to being alone. This similarity of effect also holds in the case of Economics (Column (2)), but not in the case of Cooking and Sports. In a stereotypically-female category like Cooking, working with a random helper increases the probability of a correct answer by 6% (12% of the control mean), relative to being alone. The effect is marginally significant. An *F*-test at the bottom of the table testing similarity of the effect across random and preferred helpers rejects the null (p-value=0.02). This suggests that for men, in a stereotypically-female category, it is better to work with random helpers than people they prefer. This effect reverses when we consider a gender-stereotypically male category like Sports. And an *F*-test at the bottom of the table testing similarity of the effect across random and preferred helpers rejects the null (p-value=0.03). This suggests that like female test-takers, gender stereotype of the task is important for men. However, unlike female test-takers, working with preferred helpers in gender stereotypically-male category increases performance. This is not the case in a gender stereotypically-female category.

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, this is not the case when we investigate the effect by subject. Results in Column (8) show that working with a preferred female helper increases the probability of a correct answer for male test-takers in the Sports category by 7% (13% of the control mean). This is despite the fact that men know more than women in Sports. The effect of a random female helper is negative and an *F*-test at the bottom of the table testing similarity of the effect for random and preferred female helpers rejects the null (p-value=0.01). This difference across preferred and randomly chosen helpers disappears for male helpers in Sports. Neither effects are significant and p-value of an *F*-test at the bottom of the table testing similarity of the effect for random and preferred male helpers is 0.49. Table 8, Columns (7) (as in Column (3)) shows that in the case of Cooking, random helpers of either gender are better than preferred. An *F*-test at the bottom of the table testing similarity of the effect for random and preferred female (male) helpers almost rejects the null (p-value=0.10 (0.12)).

As in the case of female test-takers we test whether choice matters more with male or female helpers. Table A20 presents results. Each cell in table A20 reports the coefficients from table 8 column (5). These report the change in probability of a correct answer when working with a helper of a certain type. The reference category remains the same i.e. test-taker alone without any helper. Table A20 suggests that unlike female test-takers, neither is choice over helpers important for male test-takers nor is the gender of any randomly or preferred helper important for performance. They perform equally well across any scenario.

Taken together results in table 7 and 8 show that the effect of mixed gender teams is different for male and female test-takers. What matters for both male and female test-takers is the gender stereotype of the category. In the case of female test-takers, in a gender-stereotypically female category like Cooking, working with any male helper reduces performance. On the other hand, in a stereotypically-male category like Sports, performance of male test-takers is better when working with their preferred female helpers. Given the gender gap in performance, for male test-takers (unlike female test-takers) a gender diverse team would not have been as helpful in terms of knowledge spillovers. In line with that we see that male test-takers never perform lower when they work in teams of any type. However, in some case, working in teams is better for their performance than working alone. This is very different from female test-takers.

6 Mechanism

The fact that preferred helpers improve the score of female test-takers in sports more than random helpers is puzzling in light of the model because the two types of helpers have the same average ability. The model suggests two reason why this can happen: (1) if the test-taker is less likely to take the hint in the presence of the random helper than the preferred helper, then that can reduce the benefit of adding the helper to the hint and as a result reduce the benefit of the helper as a whole; (2) if there are frictions in transmitting and processing information from the helper to the test-taker, then the test-taker might incorrectly put more weight on the wrong source of information. We discuss possible forms of frictions later on.

We investigate these mechanisms in two steps. First, we exploit the randomised allocation of hints to find out for which treatments and in which categories the presence of a helper reduced the value of a hint. Second, we use the data on hint uncovering to evaluate whether it differed across types of helpers and whether these differences can explain the differences in scores. We only present results disaggregated by hint availability and result on hint uncovering for female test-takers. The tables showing these results for males are available in appendix (Tables A1 and 8).

6.1 Value of helpers in the presence and absence of hints

In this sub-section we explore how the effect of the helper on the score changes between questions with hints and questions without hints. The model suggests that the presence of a helper can reduce the score of the test-taker if the presence of the helper makes the test-taker less likely to take the hint. If this is the mechanism driving the reduction in score in the presence of a male helper observed in the previous section, then we should observe that adding these male helpers only has an effect in questions with hints. If we measure an effect of the male helper in questions without hints, then another mechanism must be at play.

6.1.1 Estimation

To assess the difference in questions with and without hints, we estimate the following regression for a test-taker (i), in question (q), subject (s) and test-round (t):

$$Correct\ ans_{iqst} = \kappa_i + \gamma\ Hint\ avail_{iqst} + \sum_{\tau \in \{RF,PF,RM,PM\}} \pi_\tau\ Helper_{\tau it} + \theta_\tau\ Helper_{\tau it} \times Hint\ avail_{iqst} + e_{iqst} \quad (2)$$

where $Correct\ ans_{iqst}$ is a dummy that takes value 1 when a question was correctly answered. $Helper_{\tau it}$ is a dummy that takes value 1 if a test-taker works with a helper of type $\tau \in \{RF, PF, RM, PM\}$ where RF = random female, PF = preferred female, RM = random male and PM = preferred male. $Hint\ avail_{iqst}$ is a dummy for whether a hint was available for a question. All specifications include test-taker fixed effects (κ_i) so the variation exploited is within-person across helpers and hints. The reference category is test-takers without hints and helpers. e_{iqst} is a standard error clustered at a test-taker level, the level at which treatments were allocated (Abadie et al., 2017).

Returns from a particular helper type in the absence of hints are measured by π, while returns from a helper type in the presence of hints are given by $\pi + \theta$. If $\pi \geq 0$ but $\pi + \theta < 0$, then we can infer that it is indeed in the presence of the hint that the helper reduces the test-taker's score.

Table 9 shows the results for female test-takers. We first focus on Column (1) which shows the overall effect across subjects. In questions without hints, working with a preferred helper (either male or female) increases the probability of a correct answer. Working with a preferred female helper increases the probability of a correct answer by 9% (28% of control mean), while a preferred male helper increases the probability of a correct answer by 16% (50% of control mean), relative to working alone without hints. The positive effects in both cases are statistically significant.

This is not the case for random helpers. Random helpers of either gender have no effect on the performance of test-takers in questions without hints. These differential effects across preferred and random helpers arise despite the fact that preferred and random helpers are not systematically different in ability.

An *F*-test shows that the difference between the effect of a preferred male and that of a random effect is statistically significant with a p-value of 0.04. The difference between random and preferred helpers is larger for male helpers than female helpers, and the difference between the two types of female helpers is not statistically significant (p-value of 0.12). These p-values are presented at the bottom of the table.

These results suggest that the negative effect of male helpers does not arise in questions without hints. Instead, the effect seems to come from questions with hints. Indeed, looking at the effect of male helpers in the presence of hints, we see that random male helpers strictly reduce the score of the test-taker.

To see this, we need to sum the coefficients on the Helper dummy and the interaction to obtain $\pi_{RM} + \theta_{RM}$. This sum is equal to -0.08 for the random male helper treatment and is significantly different than zero (p-value 0.01). This indicated at the bottom of the table through the *F*-test testing *preferred helper* + *preferred helper* × *hint available* = 0.

Note that this negative effect is not due to the hint itself lowering the score, since the hint alone (in the absence of helpers) increases the probability of a correct answer by 24%, relative to the case of no hints and no helpers.

The effect of other types of helpers (preferred male and both types of female helpers) is not significantly different than zero in questions with hints.

We can investigate these results further by looking at the individual subjects. The positive effect of preferred helpers in the absence of hints is mostly driven by the Sports category shown in Column (4).

Instead, the negative effect of the random male helper in the presence of hint is driven by the cooking category, where the combined coefficients measuring the effect of a helper in the presence of hints ($\pi_{RM} + \theta_{RM}$) is equal to -0.15 and the *F*-test testing *random helper* + *random helper* × *hint available* = 0 has a p-value of 0.01.

Interestingly, this is also the case for the preferred male helper in the cooking category. The combined coefficient is equal to -0.18 and the *F*-test testing *preferred helper* + *preferred helper* × *hint available* = 0 has a p-value of 0.04.

These results point to the following conclusion. The negative effect of making a helper available only arises in questions with hints. This negative effect is only present overall for random male helpers and is present in the cooking category for both the preferred and random male helpers. These results are consistent with the predictions in the model insofar as the possibility of a helper reducing the test-taker's score is only possible when hints are available. However, for this mechanism to work, it also needs to be the case that the test-taker takes fewer hints in the presence of the helper than in his absence. We investigate this in the next subsection.

Separating the results with and without hints also gives insights into where the difference between random and preferred helpers come from. In particular, we find that in the absence of

hints, preferred helpers of both genders significantly improve the score of the test-takers, while random helpers have no effect. This is particularly the case in the sports category. Combining this observation with the fact that random male helpers actually reduce the score of the test-taker when hints are available in the cooking category drives the gap between preferred and random helpers further.

6.2 Hints uncovered

As noted in Section 3, the model generates some joint predictions on the rate of hint uncovering with and without helpers and on the expected score of the test-taker given hint and helper availability. We compare these predictions to the results to assess whether the difference in hint taking can explain the difference in scores across types of helpers.

Given the research design, we can directly observe whether fewer hints are uncovered when test-takers work with helpers of a particular type, relative to when they can access hints alone. This allows us to directly test for frictions in the flow of information. An advantage of this research design is that the counterfactual is test-takers accessing hints on their own without any helpers. This allows to us to see the effects relative to a counterfactual world where the demand for information is uninhibited and according to each person's own ability.

In particular, the model suggests that the negative effect of a random male helper on the score in questions with hints (as identified in the previous subsection) can be explained by the test-taker uncovering fewer hints in the presence of the helper (Result 3).

6.2.1 Estimation

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

$$\begin{aligned} Hint\ uncovered_{iqst} = \tau\ Random\ female\ helper_{iqst} + \alpha\ Preferred\ female\ helper_{iqst} \\ + \mu\ Random\ male\ helper_{iqst} + \chi\ Preferred\ male\ helper_{iqst} + \beta_i + \epsilon_{iqst} \end{aligned} \quad (3)$$

where $Hint\ uncovered_{iqst}$ is a dummy equal to 1 if a hint was scratched for a question, and the helper dummies are defined as in previous specifications. The regression includes test-taker fixed effects (β_i) and exploits within-person across helpers variations. The reference category is test-takers without helpers, accessing hints on their own. ϵ_{iqst} are standard errors clustered at a test-taker level.

Table 10 shows the probability of uncovering hints overall (Column (1)) as well as by subjects (Columns (2)-(4)) for each treatment. The data is restricted to questions in which hints are available.

Our main question is whether the reduction in performance from male helpers in questions with hints is accompanied by a reduction in hint taking as predicted by the theory (Result 3). Table

9 showed that this was the case for random male helpers overall and in the Cooking category.

Table 10 confirms that in Cooking (column (3)), test-taker did uncover significantly fewer hints in the presence of a random male helper. The probability of uncovering hints is 7% (9% of the mean) lower when test-takers are accessing hints through random males helpers, than when they access hints alone. This effect is marginally significant at the 10% level. This is not the case when the test-taker had other types of helpers, which is consistent with Result 3 in the theoretical predictions. However, we note that the effects for preferred male helpers are not very precise and an F-test at the bottom of the table testing similarity of the effect for random and preferred male helpers fails to reject the null (in line with the predictions of Result 2).

Finally, it is worth noting that we see no large or significant reduction in hint taking in the presence of the random male helper overall (Column (1)), even though we do observe a reduction in score due to the presence of the random male helper. This could be explained by the relatively larger rate of hint uncovering in Economics (Column (2)) that is not accompanied by a corresponding increase in score in Table 9. However, it is hard to clearly identify the causes as these coefficients are not statistically significant.

6.3 Some remaining puzzles

The previous section highlighted that the negative effect of random male helpers on the test-taker's score can be rationalised by a lower rate of hint uncovering, especially in the cooking category. The theoretical results suggest that this can arise when the test-taker faces a cost of uncovering hints and finds that the marginal gain of uncovering the hint in the presence of the helper is too low to justify the cost. There are, however, some remaining puzzles that are not consistent with the predictions of our theoretical framework. We discuss these in this section.

The first puzzling result is that we do not see a large or statistically-significant reduction in hint taking in the presence of the preferred male helper in Cooking (Table 10, Column (3)) even though we observe a significant reduction of the score in that category with this type of helper in Table 9 (Column (3)). 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.

The second is the reduction in hint taking with the random male helper in Sports (Table 10, Column (4)) which is not accompanied by a reduction in score Table 9 (Column (4)). The probability of uncovering hints is 6% (7% of the mean) lower when test-takers are accessing hints in the presence of a random male helper, than when they access hints alone. This effect is significant at the 10% level. This is not accompanied by a reduction in score in the presence of a random male helper in Table 9, in contradiction with Result 3. As discussed in section 3, a test-taker should only take fewer hints in the presence of the helper if she expects the marginal gain from the hint in the presence of the helper to be small. In particular, the combined gain from the hint and the helper must be

less than the sum of the two individual gains. If the helper's information is less precise than the hint, this can only arise when the score from having only a helper is strictly less than the score from having only a hint. We should therefore see that a reduction in hint-taking in the presence of a helper is accompanied by a strict reduction in score. This suggests that the test-taker might have either been driven to take fewer hints in sports for reasons unrelated to their expected score, or might have misperceived the expected value of hints relative to the helper in this category.

We suggest three classes of explanations for these observations: social image concerns, misperceptions of the quality of information or of the test-taker's own ability, and difficulty in communicating information between helpers and test-takers.

Social image concerns. Both the decision to uncover hints and the decision of whether to follow the helper's suggested answer can be affected by social image concerns.

For instance, uncovering a hint provides a signal to the helper that the test-taker is not very confident about an answer. A test-taker who is concerned about how the helper perceives her ability might therefore refrain from uncovering a hint when it would have been valuable. Alternatively, uncovering the hint could make the helper believe that the test-taker has doubts about the value of the helper's advice. This could explain the lower rate of hint taking with the random helper in Sports even though the difference in scores did not seem to justify taking fewer hints. This image concern can also be the source of the cost of taking hints in the first place. Given that Sports was a gender-stereotypically male category, it seems plausible that the female test-taker was reluctant to give the impression that they did not trust the advice of a male helper.

Social-image concerns can also affect the answer chosen by the test-taker and can rationalise the reduction in score in the Cooking category with the preferred helper, that is not explained by a reduction in hint taking. Choosing a different answer than the one suggested by the helper can lead the helper to form a negative opinion about 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. Picking a different answer can therefore signal low ability. 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 this gender-stereotypically female category where the female test-taker might have added pressure to appear knowledgeable (and in particular when faced with a helper that she knows personally).

An alternative but related mechanism is that the test-taker might rush into answering questions in this category, and deliberate less, in order to impress the helper.[12]

[12]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, to the marriage market.

Misperceptions of ability. Misperception of the helper's ability can lead to lower score without affecting the rate of hint taking. For instance, it is possible that the test-taker follows the helper's suggestion when she should have followed the hint when the two disagree. If she fails to anticipate this error, she would uncover the hint both alone and with the helper, anticipating that the hint would help, so indeed we should not see differences in rate of uncovering. However, if the helper is less precise than the hint, then ignoring the hint when the two disagree leads to a lower score.

Our data points to the existence of some misperception of ability. Females (65%) expected males to have lower knowledge of cooking even though there was no difference in performance across genders at baseline. However, misperception of ability does not appear to be at work behind the two puzzling results. First, the direction of misperception of ability does not support the negative results in cooking with the preferred helper. This misperception does not rationalize this puzzle since if women misperceived lower ability of men in cooking, they should have relied more heavily on hints than helper so their scores with helper should not be lower. Second, it cannot explain the second puzzle (lower hint taking and unchanged score with random male helpers in sports) as there does not seem to be significant misperceptions in sports. It is possible, however, that test-takers misperceived the ability of random helpers in sports but not that of preferred helpers.

Communication frictions. Finally, the presence of a helper can reduce the score of the test-taker if communication between the two distracts the test-taker. This could be particularly true in questions with hints if the presence of a hint lead 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, and would explain why this effect only arises with a preferred helper. However, these frictions still cannot justify lower hints and unchanged scores with random male helpers in sports.

Other explanations include discrimination in conveying information by the helper. For example, the helper could deliberately share incorrect information because they feel that they are competing with the test-taker or if there is a norm that the males should perform better than the females. However, it is unclear why such discrimination or supply side problems would materialize only in particular subjects.

Conclusion

Meaningful gender diversity at work remains an important goal of many organizations. This study is the first that highlights the question of choice over co-workers as 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 arises 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.

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

This study also opens up avenues for further research to explain the puzzles we observe. What can explain the poor performance of women with random male helpers in gender stereotypically female category like cooking? Why do they take less hints with random male helpers in sports? Most importantly, understanding the specific mechanism behind the results we observe can determine what policies are best suited to address the potential productivity losses we identify. If social-image concerns are behind the loss of information transmission, then creating opportunities for workers to interact and get to know each other better can improve performance. If misperceptions of the ability of workers of the opposite gender in certain tasks is driving it, then training to address these misperceptions can generate better outcomes. If performance is reduced due to issues in communicating, then communication training for workers can help alleviate these issues. Finally, if these issues are more deeply-rooted, then it is important to understand whether team incentives or alternative reward schemes can help reduce such frictions.

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.

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.

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.

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.

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.

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.

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.

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

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.

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.

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.

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. (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.

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.

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).

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

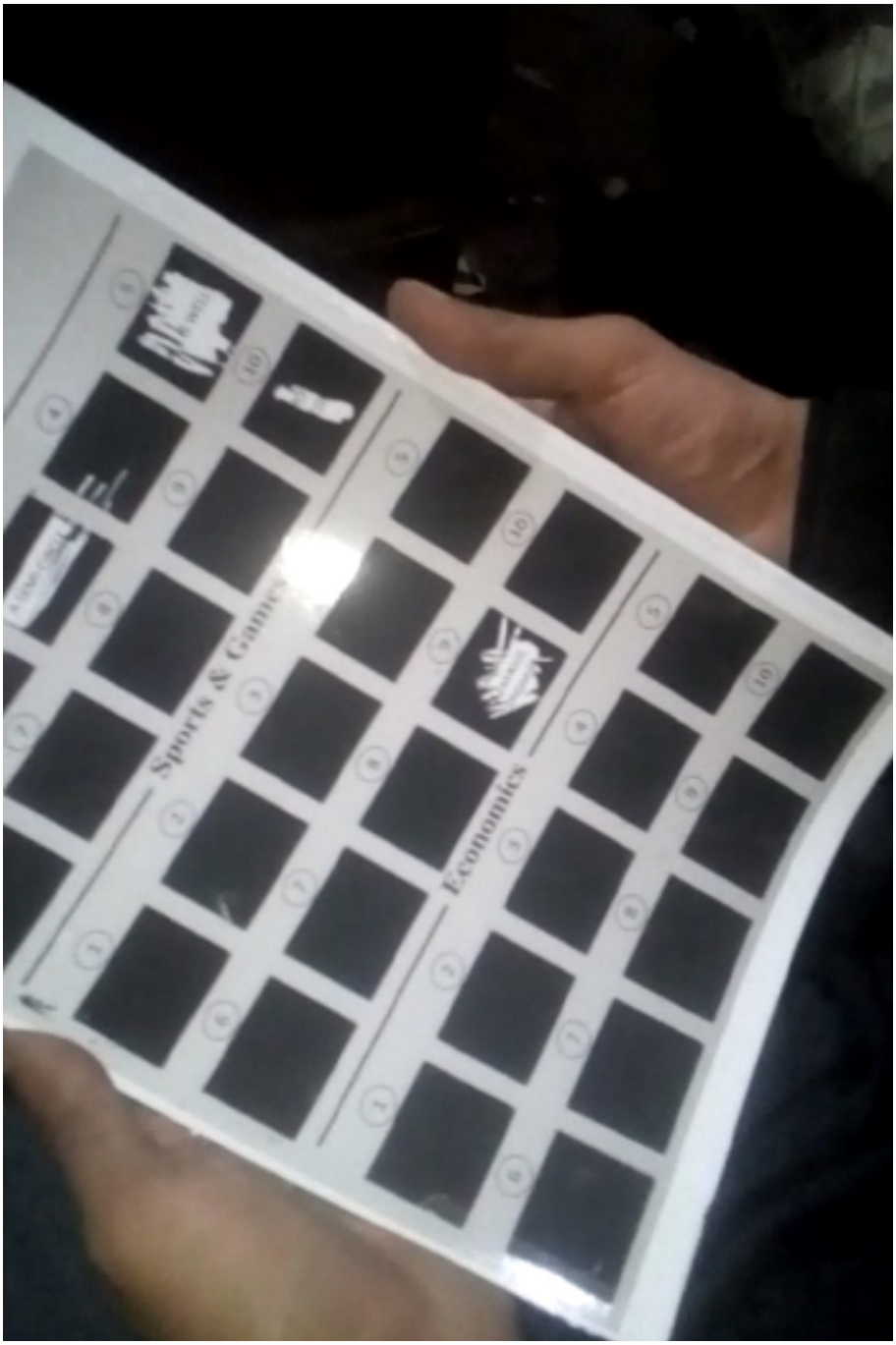


Figure 1: Hint sheet used by test-takers.

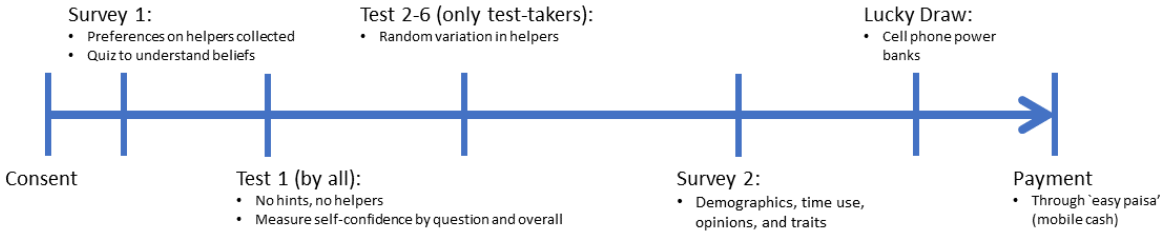


Figure 2: Sequence of activities

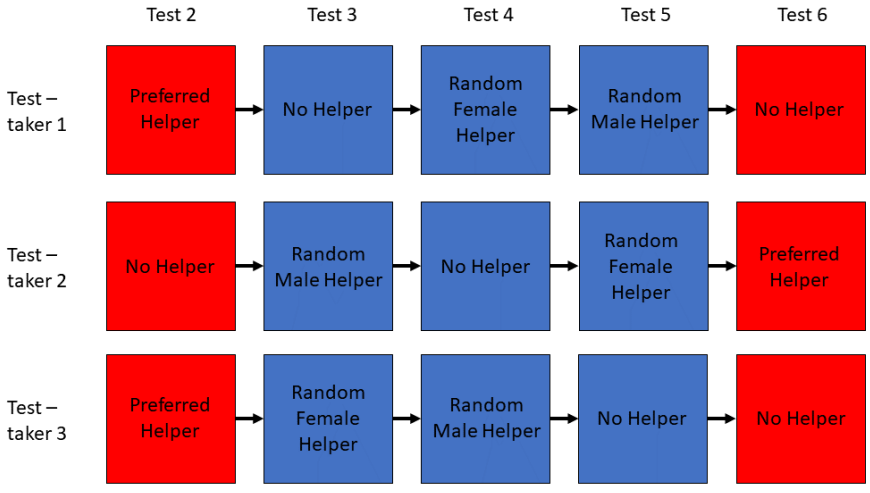


Figure 3: Example of random allocation of treatments per test-taker

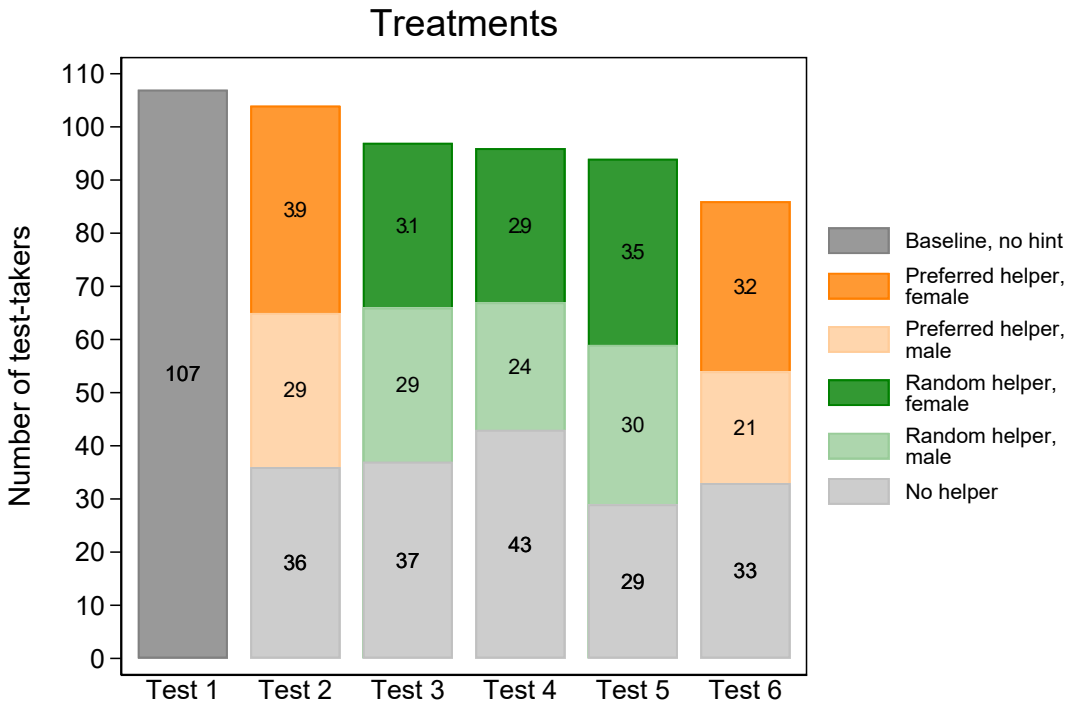


Figure 4: Sample of test-takers with different treatments

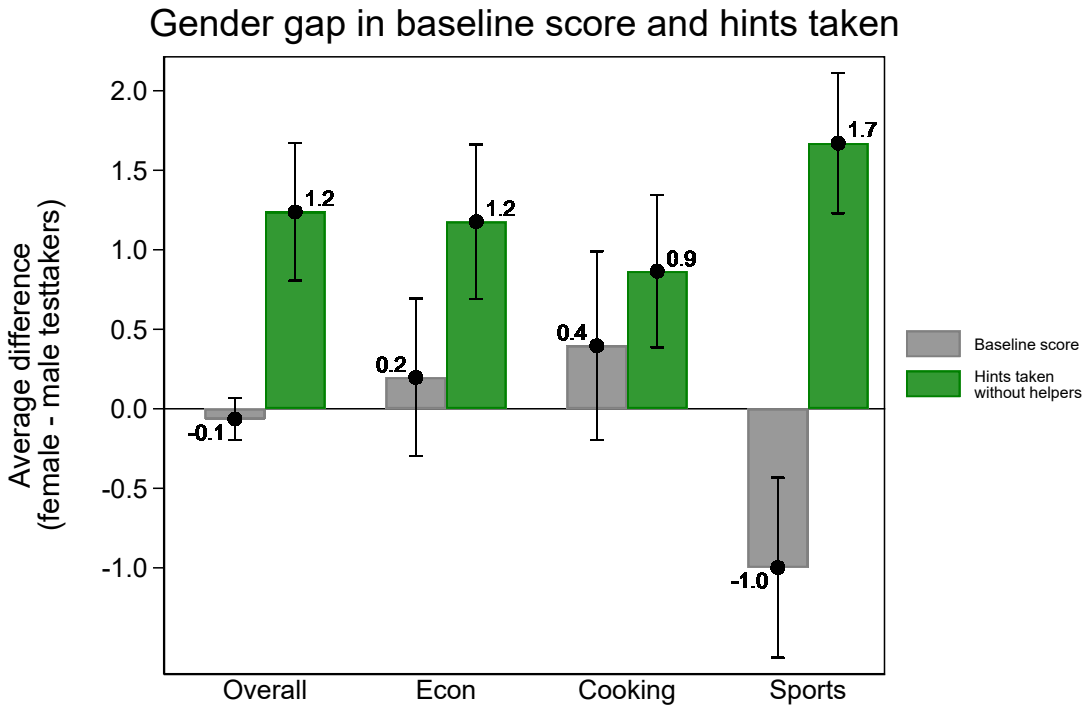


Figure 5: Gender gap in performance and hints taken by test-takers.

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	0.15	0.12	-0.09	0.53	-0.32					
	(0.12)	(0.54)	(0.32)	(0.38)	(0.24)					
Preferred Male Helper						-0.08	0.09	0.29	-0.51	0.31
						(0.18)	(0.73)	(0.48)	(0.45)	(0.48)
Person FE	No	No	No	No	No	No	No	No	No	No
Observations	126	124	124	124	124	126	124	124	124	124
Clusters	54	54	54	54	54	54	54	54	54	54
Control mean of outcome	3.22	10.42	2.52	4.39	3.52	3.22	10.42	2.52	4.39	3.52
Panel B: Helpers with male test-takers										
Preferred Female Helper	-0.07	-0.03	0.05	0.49	-0.56					
	(0.14)	(0.73)	(0.32)	(0.43)	(0.38)					
Preferred Male Helper						-0.06	0.11	0.03	-0.26	0.35
						(0.11)	(0.53)	(0.37)	(0.44)	(0.56)
Person FE	No	No	No	No	No	No	No	No	No	No
Observations	98	96	96	96	96	98	96	96	96	96
Clusters	40	40	40	40	40	40	40	40	40	40
Control mean of outcome	3.32	10.93	2.88	4.25	3.80	3.32	10.93	2.88	4.25	3.80

* 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. 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: Value of helpers in the presence and absence of hints**

Dependent Variables:	*Reference category: Test-takers without helpers, no hints*			
	Correct answer (dummy)			
	Overall	By Subjects		
		Econ	Cooking	Sports
	(1)	(2)	(3)	(4)
Sample: Female Test-takers				
Hint available	0.24***	0.13***	0.24***	0.36***
	(0.03)	(0.04)	(0.05)	(0.04)
Random female helper	0.03	0.04	0.04	0.02
	(0.03)	(0.06)	(0.06)	(0.06)
Preferred female helper	0.09***	0.02	0.10	0.16**
	(0.03)	(0.05)	(0.07)	(0.07)
Random male helper	0.01	0.11	-0.02	-0.07
	(0.04)	(0.09)	(0.07)	(0.07)
Preferred male helper	0.16**	0.04	0.14	0.29***
	(0.06)	(0.09)	(0.11)	(0.06)
Random female helper × hint available	-0.05	-0.05	-0.08	-0.02
	(0.04)	(0.07)	(0.07)	(0.06)
Preferred female helper × hint available	-0.09*	0.04	-0.15*	-0.16*
	(0.05)	(0.06)	(0.09)	(0.08)
Random male helper × hint available	-0.09*	-0.16*	-0.13	0.01
	(0.05)	(0.09)	(0.08)	(0.07)
Preferred male helper × hint available	-0.15*	0.04	-0.32***	-0.19
	(0.08)	(0.09)	(0.10)	(0.12)
P val: Random female = preferred female helper	0.12	0.67	0.41	0.07
P val: Random male = preferred male helper	0.04	0.44	0.19	0.00
P val: Random female + random female helper × hint = 0	0.42	0.91	0.15	0.95
P val: Preferred female + preferred female helper × hint = 0	0.89	0.17	0.20	0.83
P val: Random male + random male helper × hint = 0	0.01	0.24	0.01	0.31
P val: Preferred male + preferred male helper × hint = 0	0.99	0.29	0.04	0.24
Person FE	Yes	Yes	Yes	Yes
Observations	6270	2090	2090	2090
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. Hint available is a dummy variable that turns on 1 if a hint was available to help answer the question. 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 10: **Mechanism: Does choice of co-worker matter for uncovering hints?**

	Reference category: Test-takers without helpers			
Dependent Variables:	**Hint uncovered (dummy)**			
	Overall	By Subjects		
		Econ	Cooking	Sports
	(1)	(2)	(3)	(4)
Sample: Female Test-takers				
Random Female Helper	-0.02	-0.02	-0.02	-0.03
	(0.02)	(0.03)	(0.03)	(0.03)
Preferred Female Helper	-0.00	-0.01	-0.00	0.01
	(0.03)	(0.06)	(0.04)	(0.04)
Random Male Helper	-0.02	0.06	-0.07*	-0.06*
	(0.03)	(0.06)	(0.04)	(0.03)
Preferred Male Helper	-0.02	-0.04	-0.01	-0.01
	(0.09)	(0.17)	(0.09)	(0.07)
P val: Random female = preferred female helper	0.54	0.82	0.77	0.30
P val: Random male = preferred male helper	0.95	0.59	0.55	0.47
Person FE	Yes	Yes	Yes	Yes
Observations	5016	1672	1672	1672
Clusters	54	54	54	54
Control mean of outcome	0.76	0.70	0.75	0.84

* 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.

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 behaviour

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: $S_A = \mu$, 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: $S_H - S_A = p_H - \mu > \kappa$
     - 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 $S_H - S_A = p_H - \mu > 0$ 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 $S_H - S_A = p_H - \mu > 0$ 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 $S_H - S_A = \mu - \mu = 0$ 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: $S_A = \mu$, 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: $S_H - S_A = p_H - \mu > \kappa$
     - 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 $S_H - S_A = p_H - \mu > 0$ 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 $S_H - S_A = p_H - \mu > 0$, $S_{MH} - S_H = p_M - p_H < 0$.
     (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 $S_H - S_A = \mu - \mu = 0$ 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 $S_H - S_A = 0$ 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: $S_A = \mu$, 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: $S_H - S_A = p_H - \mu > \kappa$
  - 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 $S_H - S_A = p_H - \mu > 0$, $S_{HM} - S_H = 0$.
  (b) If $\kappa > p_H - \mu$: $S_H = S_{MH} = \mu$ so $S_H - S_A = 0$ 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: $S_A = \mu$, 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: $S_H - S_A = p_H - \mu > \kappa$
  - 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 $S_H - S_A = p_H - \mu > 0$ 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 $S_H - S_A = \mu - \mu = 0$ 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 $S_H - S_A = 0$ 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: $S_A = \mu$, 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: $S_H - S_A = \mu - \mu > \kappa \Rightarrow$ 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 $S_H - S_A = \mu - \mu = 0$, $S_{MH} - S_H = \mu[1-(1-p_H)(1-p_M)] + (1-\mu)p_H p_M - \mu = (1-\mu)p_H p_M - \mu(1-p_H)(1-p_M) > 0$ given condition for high confidence.

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

6. **Very high confidence**

- **Conditions:** neither hint alone nor helper alone can change answer, and if helper and hint say B then still choose 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: $S_A = \mu$, 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:** $S_H - S_A = \mu - \mu = 0$ 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 $S_M - S_A$. 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 $S_M - S_A$ 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.

□

10 Tables

11 Appendix: For Online Publication

Table A1: **Mechanism: Value of helpers in the presence and absence of hints**

Dependent Variables:	*Reference category: Test-takers without helpers, no hints*			
	Correct answer (dummy)			
	Overall	By Subjects		
		Econ	Cooking	Sports
	(1)	(2)	(3)	(4)
Sample: Male Test-takers				
Hint available	0.20***	0.11***	0.18***	0.30***
	(0.03)	(0.03)	(0.04)	(0.05)
Random female helper	-0.01	0.07	0.02	-0.12*
	(0.04)	(0.07)	(0.08)	(0.07)
Preferred female helper	-0.00	0.12	-0.10	-0.02
	(0.06)	(0.09)	(0.10)	(0.11)
Random male helper	0.00	0.14	-0.08	-0.05
	(0.05)	(0.08)	(0.07)	(0.08)
Preferred male helper	0.09*	0.02	0.02	0.21**
	(0.05)	(0.07)	(0.07)	(0.10)
Random female helper × hint available	0.05	-0.00	0.05	0.10
	(0.04)	(0.08)	(0.09)	(0.07)
Preferred female helper × hint available	0.03	-0.09	0.08	0.11
	(0.07)	(0.10)	(0.11)	(0.13)
Random male helper × hint available	0.06	-0.09	0.17**	0.10
	(0.05)	(0.08)	(0.08)	(0.09)
Preferred male helper × hint available	-0.05	0.10	-0.07	-0.18
	(0.06)	(0.07)	(0.08)	(0.11)
P val: Random female = preferred female helper	0.90	0.68	0.31	0.36
P val: Random male = preferred male helper	0.08	0.22	0.25	0.00
P val: Random female + random female helper × hint = 0	0.14	0.07	0.13	0.63
P val: Preferred female + preferred female helper × hint = 0	0.26	0.54	0.82	0.05
P val: Random male + random male helper × hint = 0	0.07	0.26	0.13	0.33
P val: Preferred male + preferred male helper × hint = 0	0.26	0.00	0.42	0.59
Person FE	Yes	Yes	Yes	Yes
Observations	5790	1930	1930	1930
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. Hint available is a dummy variable that turns on 1 if a hint was available to help answer the question. 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 A2: **Mechanism: Does choice of co-worker matter for uncovering hints?**

	Reference category: Test-takers without helpers			
Dependent Variables:	**Hint uncovered (dummy)**			
	Overall	By Subjects		
		Econ	Cooking	Sports
	(1)	(2)	(3)	(4)
Sample: Male Test-takers				
Random Female Helper	0.07*	0.10**	0.01	0.11**
	(0.04)	(0.05)	(0.05)	(0.04)
Preferred Female Helper	0.10*	0.02	0.17**	0.12**
	(0.05)	(0.08)	(0.08)	(0.05)
Random Male Helper	0.03	0.10*	0.03	-0.04
	(0.04)	(0.06)	(0.05)	(0.05)
Preferred Male Helper	0.10**	0.09	0.05	0.16***
	(0.05)	(0.07)	(0.06)	(0.05)
P val: Random female = preferred female helper	0.59	0.28	0.10	0.75
P val: Random male = preferred male helper	0.10	0.90	0.63	0.00
Person FE	Yes	Yes	Yes	Yes
Observations	4632	1544	1544	1544
Clusters	48	48	48	48
Control mean of outcome	0.61	0.56	0.62	0.65

* 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.

Table A3: 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 A4: 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 A5: 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 A6: 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 A7: 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 A8: 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 A9: 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 A10: 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 A11: 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 A12: 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 A13: 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 A14: 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	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
(male = 1)	(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 =	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1)	(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:	0.22	0.22	0.27	0.14	0.00	-0.05	-0.05	0.08	0.08	0.13
private business	(0.44)	(0.44)	(0.47)	(0.38)	(0.21)	(0.20)	(0.20)	(0.21)	(0.21)	(0.21)
Father's occupation:	0.00	0.22	0.18	0.14	-0.22	-0.18	0.04	-0.14	0.08	0.04
kissan	(0.00)	(0.44)	(0.40)	(0.38)	(0.15)	(0.14)	(0.19)	(0.12)	(0.21)	(0.19)
Father's occupation:	0.11	0.00	0.18	0.00	0.11	-0.07	-0.18	0.11	0.00	0.18
government job	(0.33)	(0.00)	(0.40)	(0.00)	(0.11)	(0.17)	(0.14)	(0.13)	(0.00)	(0.15)
Father's occupation:	0.00	0.11	0.00	0.00	-0.11	0.00	0.11	0.00	0.11	0.00
private job	(0.00)	(0.33)	(0.00)	(0.00)	(0.11)	(0.00)	(0.10)	(0.00)	(0.13)	(0.00)
Mother's occupation:	0.67	0.67	0.73	0.14	0.00	-0.06	-0.06	0.52**	0.52**	0.58**
housewife	(0.50)	(0.50)	(0.47)	(0.38)	(0.24)	(0.22)	(0.22)	(0.23)	(0.23)	(0.21)
Mother's occupation:	0.11	0.00	0.09	0.00	0.11	0.02	-0.09	0.11	0.00	0.09
government job	(0.33)	(0.00)	(0.30)	(0.00)	(0.11)	(0.14)	(0.10)	(0.13)	(0.00)	(0.12)
Family's monthly income:	0.00	0.11	0.09	0.00	-0.11	-0.09	0.02	0.00	0.11	0.09
less than Rs. 10,000	(0.00)	(0.33)	(0.30)	(0.00)	(0.11)	(0.10)	(0.14)	(0.00)	(0.13)	(0.12)
Family's monthly income:	0.78	0.56	0.45	0.14	0.22	0.32	0.10	0.63***	0.41	0.31
Rs. 10,000 - Rs. 100,000	(0.44)	(0.53)	(0.52)	(0.38)	(0.23)	(0.22)	(0.24)	(0.21)	(0.24)	(0.23)
Family's monthly income:	0.00	0.00	0.27	0.14	0.00	-0.27*	-0.27*	-0.14	-0.14	0.13
greater than Rs. 100,000	(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	0.56	0.56	0.36	0.71	0.00	0.19	0.19	-0.16	-0.16	-0.35
than 10	(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 A15: 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 A16: 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 A17: 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 A18: 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.

Table A19: **What is the effect of gender and choice over helpers on performance?**

Sample: Female test-takers			
	Change in probability of a correct ans. with:		
	Random helper	Preferred helper	Difference
Male helper	-0.06**	0.03	-0.10
	(0.03)	(0.06)	(0.07)
Female helper	-0.01	0.02	-0.03
	(0.02)	(0.02)	(0.02)
Difference	-0.06*	0.01	-0.07
	(0.03)	(0.06)	(0.07)

* p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered at test-taker level in parenthesis.
Notes: Female helper is a dummy variable that turns on 1 if a test-taker is allocated a female helper. Male helper is a dummy variable that turns on 1 if a test-taker is allocated a male helper. 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. All specifications include test-taker fixed effects and so the variation exploited is within-person across helpers.

Table A20: **What is the effect of gender and choice over helpers on performance?**

Sample: Male test-takers			
	Change in probability of a correct ans. with:		
	Random helper	Preferred helper	Difference
Male helper	0.05	0.05	0.00
	(0.03)	(0.03)	(0.03)
Female helper	0.03	0.03	0.00
	(0.03)	(0.02)	(0.03)
Difference	0.02	0.02	-0.00
	(0.04)	(0.03)	(0.04)

* p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered at test-taker level in parenthesis.
Notes: Female helper is a dummy variable that turns on 1 if a test-taker is allocated a female helper. Male helper is a dummy variable that turns on 1 if a test-taker is allocated a male helper. 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. All specifications include test-taker fixed effects and so the variation exploited is within-person across helpers.

Table A21: **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	Low Confidence Test-takers				High Confidence Test-takers			
Random female helper	-0.00	-0.03	0.00	-0.03	-0.03	0.05	-0.05	0.05
	(0.02)	(0.04)	(0.03)	(0.04)	(0.03)	(0.05)	(0.04)	(0.04)
Preferred female helper	0.02	0.03	-0.01	0.05	0.01	0.07	-0.03	0.02
	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.07)	(0.04)	(0.06)
Random male helper	-0.08***	-0.07	-0.14*	-0.06	-0.03	0.09	-0.11	-0.05
	(0.02)	(0.05)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
Preferred male helper	0.17*	0.11	-0.19**	0.26***	-0.08	0.03	-0.07	0.01
	(0.08)	(0.07)	(0.07)	(0.08)	(0.06)	(0.13)	(0.13)	(0.04)
P val: Random female = preferred female helper	0.41	0.27	0.87	0.09	0.28	0.87	0.71	0.59
P val: Random male = preferred male helper	0.00	0.01	0.58	0.00	0.42	0.64	0.77	0.38
Person FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4230	1300	1020	1310	2040	790	1070	780
Clusters	34	32	25	33	20	22	29	21
Control mean of outcome	0.52	0.46	0.57	0.51	0.50	0.40	0.61	0.49

* 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 A22: **Mechanism: Value of helpers in the presence and absence of hints**

Dependent Variables:	*Reference category: Test-takers without helpers, no hints*							
	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	Low Confidence Test-takers				High Confidence Test-takers			
Hint available	0.26***	0.14**	0.21**	0.31***	0.21***	0.12*	0.27***	0.44***
	(0.04)	(0.06)	(0.08)	(0.06)	(0.04)	(0.06)	(0.06)	(0.06)
Random female helper	0.01	-0.03	-0.03	-0.09	0.07	0.16*	0.11	0.20**
	(0.04)	(0.07)	(0.10)	(0.07)	(0.05)	(0.08)	(0.08)	(0.09)
Preferred female helper	0.10**	-0.06	0.10	0.22**	0.09	0.15*	0.10	0.04
	(0.04)	(0.06)	(0.10)	(0.08)	(0.06)	(0.08)	(0.10)	(0.09)
Random male helper	0.02	0.07	-0.10	-0.13	-0.02	0.22	0.04	0.02
	(0.05)	(0.10)	(0.11)	(0.09)	(0.10)	(0.18)	(0.10)	(0.11)
Preferred male helper	0.31***	0.10	0.04	0.32***	0.01	-0.04	0.22	0.29**
	(0.07)	(0.09)	(0.14)	(0.07)	(0.07)	(0.17)	(0.16)	(0.12)
Random female helper × hint available	-0.02	0.01	0.04	0.08	-0.12	-0.13	-0.20*	-0.19*
	(0.05)	(0.09)	(0.11)	(0.07)	(0.08)	(0.11)	(0.10)	(0.10)
Preferred female helper × hint available	-0.09	0.11	-0.14	-0.22**	-0.10	-0.10	-0.17	-0.02
	(0.06)	(0.08)	(0.14)	(0.10)	(0.08)	(0.06)	(0.12)	(0.12)
Random male helper × hint available	-0.13**	-0.17	-0.05	0.08	-0.01	-0.16	-0.18*	-0.09
	(0.06)	(0.11)	(0.14)	(0.09)	(0.09)	(0.17)	(0.11)	(0.11)
Preferred male helper × hint available	-0.17*	0.01	-0.28*	-0.08	-0.12	0.09	-0.36**	-0.35*
	(0.10)	(0.09)	(0.15)	(0.16)	(0.12)	(0.20)	(0.13)	(0.17)
P val: Hint + random female helper × hint = 0	0.00	0.08	0.00	0.00	0.22	0.85	0.48	0.01
P val: Hint + preferred female helper × hint = 0	0.00	0.00	0.52	0.27	0.14	0.81	0.31	0.00
P val: Hint + random male helper × hint = 0	0.02	0.73	0.18	0.00	0.03	0.76	0.40	0.00
P val: Hint + preferred male helper × hint = 0	0.36	0.03	0.60	0.11	0.40	0.26	0.42	0.54
Person FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4230	1300	1020	1310	2040	790	1070	780
Clusters	34	32	25	33	20	22	29	21
Control mean of outcome	0.31	0.35	0.41	0.25	0.33	0.30	0.40	0.14

* 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. Hint available is a dummy variable that turns on 1 if a hint was available to help answer the question. 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 A23: **Mechanism: Does choice of co-worker matter for uncovering hints?**

Dependent Variables:	*Reference category: Test-takers without helpers*							
	Hint uncovered (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	Low Confidence Test-takers				High Confidence Test-takers			
Random female helper	0.04	-0.02	0.08**	0.04	-0.13**	-0.03	-0.08*	-0.10*
	(0.02)	(0.04)	(0.03)	(0.03)	(0.05)	(0.06)	(0.05)	(0.06)
Preferred female helper	0.02	-0.06	0.07*	0.06	-0.03	0.11	-0.08	-0.05
	(0.03)	(0.07)	(0.03)	(0.05)	(0.08)	(0.09)	(0.06)	(0.05)
Random male helper	-0.04	-0.02	-0.03	-0.05	-0.09	0.07	-0.15**	-0.13***
	(0.04)	(0.07)	(0.03)	(0.04)	(0.06)	(0.10)	(0.06)	(0.04)
Preferred male helper	-0.01	-0.06	-0.03	-0.11**	-0.07	-0.07	0.00	0.12
	(0.08)	(0.19)	(0.06)	(0.05)	(0.12)	(0.23)	(0.12)	(0.12)
P val: Random female = preferred female helper	0.70	0.57	0.80	0.58	0.26	0.12	0.99	0.38
P val: Random male = preferred male helper	0.67	0.82	0.97	0.38	0.83	0.61	0.23	0.05
Person FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4230	1300	1020	1310	2040	790	1070	780
Clusters	34	32	25	33	20	22	29	21
Control mean of outcome	0.71	0.70	0.64	0.74	0.69	0.58	0.77	0.82

* 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.

Table A24: **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	Low Confidence Test-takers				High Confidence Test-takers			
Random female helper	0.02	0.07**	0.00	0.08	0.05	0.07	0.14**	-0.10*
	(0.02)	(0.03)	(0.06)	(0.07)	(0.05)	(0.06)	(0.06)	(0.06)
Preferred female helper	0.03	0.11***	-0.08	0.20***	0.03	-0.01	0.02	0.03
	(0.03)	(0.02)	(0.10)	(0.05)	(0.03)	(0.05)	(0.06)	(0.04)
Random male helper	0.04	0.06	0.01	0.21***	0.06	0.07	0.11	-0.05
	(0.04)	(0.06)	(0.07)	(0.06)	(0.05)	(0.05)	(0.07)	(0.06)
Preferred male helper	0.04	0.14***	-0.11***	0.16**	0.05	0.02	0.22	0.03
	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.05)	(0.18)	(0.08)
P val: Random female = preferred female helper	0.72	0.24	0.36	0.10	0.63	0.20	0.13	0.02
P val: Random male = preferred male helper	0.94	0.27	0.10	0.44	0.70	0.42	0.53	0.32
Person FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2730	1090	1120	590	3060	840	810	1340
Clusters	21	26	27	15	27	22	21	33
Control mean of outcome	0.51	0.36	0.53	0.53	0.46	0.43	0.47	0.55

* 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 A25: **Mechanism: Value of helpers in the presence and absence of hints**

Dependent Variables:	*Reference category: Test-takers without helpers, no hints*							
	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	Low Confidence Test-takers				High Confidence Test-takers			
Hint available	0.23***	0.10**	0.20***	0.42***	0.17***	0.12**	0.16***	0.25***
	(0.05)	(0.05)	(0.06)	(0.08)	(0.04)	(0.05)	(0.06)	(0.06)
Random female helper	-0.01	0.05	-0.06	0.10	0.01	0.11	0.13	-0.22**
	(0.05)	(0.07)	(0.11)	(0.12)	(0.05)	(0.13)	(0.11)	(0.08)
Preferred female helper	0.12*	0.23	-0.04	0.44	-0.07	0.02	-0.16	-0.14
	(0.06)	(0.14)	(0.14)	(0.33)	(0.09)	(0.09)	(0.14)	(0.11)
Random male helper	-0.00	0.13	-0.11	0.24**	0.02	0.15	-0.05	-0.17
	(0.07)	(0.12)	(0.10)	(0.09)	(0.07)	(0.12)	(0.10)	(0.10)
Preferred male helper	0.05	0.07	-0.05	0.36***	0.16**	-0.10	0.29	0.18
	(0.07)	(0.08)	(0.08)	(0.10)	(0.07)	(0.12)	(0.17)	(0.15)
Random female helper × hint available	0.03	0.02	0.08	-0.04	0.05	-0.04	0.02	0.15
	(0.04)	(0.09)	(0.11)	(0.11)	(0.06)	(0.15)	(0.14)	(0.09)
Preferred female helper × hint available	-0.11	-0.15	-0.06	-0.30	0.12	-0.05	0.23	0.22
	(0.07)	(0.17)	(0.10)	(0.36)	(0.11)	(0.10)	(0.19)	(0.14)
Random male helper × hint available	0.05	-0.09	0.15*	-0.03	0.06	-0.09	0.19	0.15
	(0.07)	(0.12)	(0.09)	(0.11)	(0.07)	(0.11)	(0.15)	(0.12)
Preferred male helper × hint available	-0.02	0.08	-0.08	-0.25*	-0.15**	0.15	-0.09	-0.18
	(0.08)	(0.09)	(0.10)	(0.14)	(0.07)	(0.11)	(0.12)	(0.17)
P val: Hint + random female helper × hint = 0	0.00	0.11	0.01	0.00	0.00	0.58	0.20	0.00
P val: Hint + preferred female helper × hint = 0	0.06	0.75	0.16	0.74	0.01	0.41	0.04	0.00
P val: Hint + random male helper × hint = 0	0.00	0.94	0.00	0.00	0.00	0.81	0.01	0.00
P val: Hint + preferred male helper × hint = 0	0.00	0.08	0.17	0.08	0.57	0.01	0.44	0.64
Person FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2730	1090	1120	590	3060	840	810	1340
Clusters	21	26	27	15	27	22	21	33
Control mean of outcome	0.32	0.28	0.38	0.19	0.32	0.34	0.34	0.35

* 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. Hint available is a dummy variable that turns on 1 if a hint was available to help answer the question. 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 A26: **Mechanism: Does choice of co-worker matter for uncovering hints?**

Dependent Variables:	*Reference category: Test-takers without helpers*							
	Hint uncovered (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	Low Confidence Test-takers				High Confidence Test-takers			
Random female helper	0.08	0.18**	-0.01	0.14*	0.07	0.04	0.04	0.06
	(0.05)	(0.08)	(0.06)	(0.08)	(0.07)	(0.04)	(0.08)	(0.05)
Preferred female helper	0.09**	0.17*	0.22*	0.27	0.14	-0.08	0.16	0.11**
	(0.04)	(0.10)	(0.11)	(0.19)	(0.08)	(0.10)	(0.12)	(0.05)
Random male helper	0.06	0.16	0.06	0.00	-0.01	0.02	-0.04	-0.05
	(0.05)	(0.09)	(0.06)	(0.10)	(0.07)	(0.06)	(0.09)	(0.05)
Preferred male helper	0.16**	0.15*	0.11	0.21**	0.02	-0.03	-0.10	0.16**
	(0.06)	(0.09)	(0.07)	(0.07)	(0.05)	(0.10)	(0.09)	(0.07)
P val: Random female = preferred female helper	0.91	0.86	0.06	0.52	0.47	0.27	0.37	0.35
P val: Random male = preferred male helper	0.02	0.99	0.40	0.10	0.68	0.62	0.54	0.00
Person FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2730	1090	1120	590	3060	840	810	1340
Clusters	21	26	27	15	27	22	21	33
Control mean of outcome	0.56	0.42	0.51	0.63	0.55	0.59	0.63	0.58

* 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.

11.1 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.[13] Cooking questions were also sourced from online sources.[14] All questions were tested in a pilot so ensure that they were easy to understand and there was only 1 correct answer per question.

11.2 Appendix figures

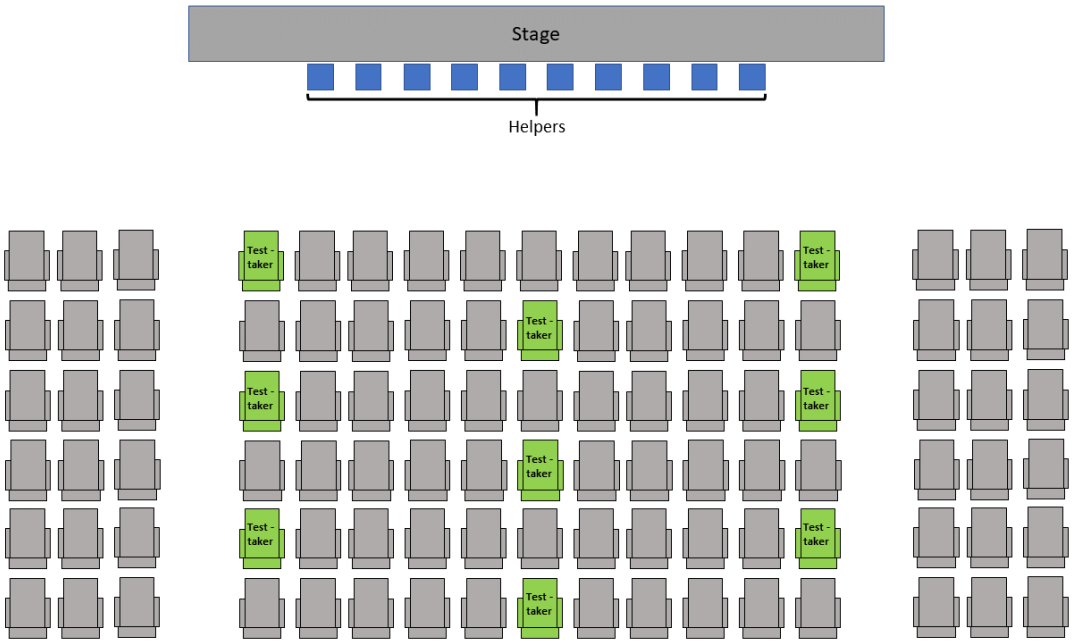


Figure A1: Seating arrangement

[13](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.

[14](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/.

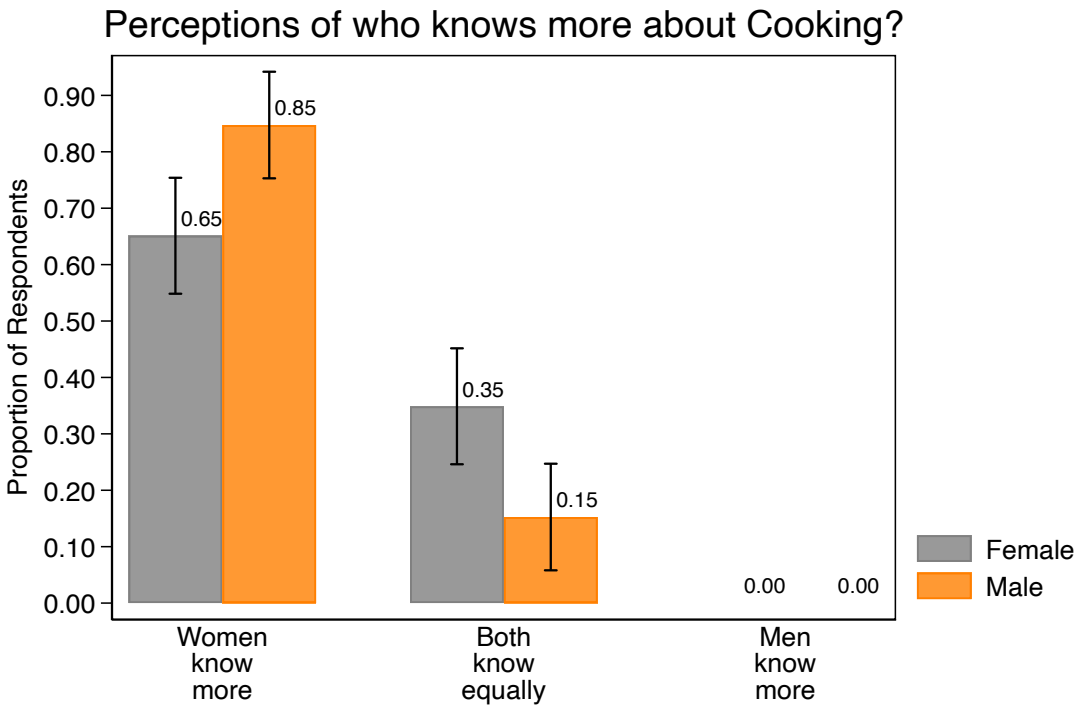


Figure A2: Perceptions of gender stereotypical nature of the different subjects

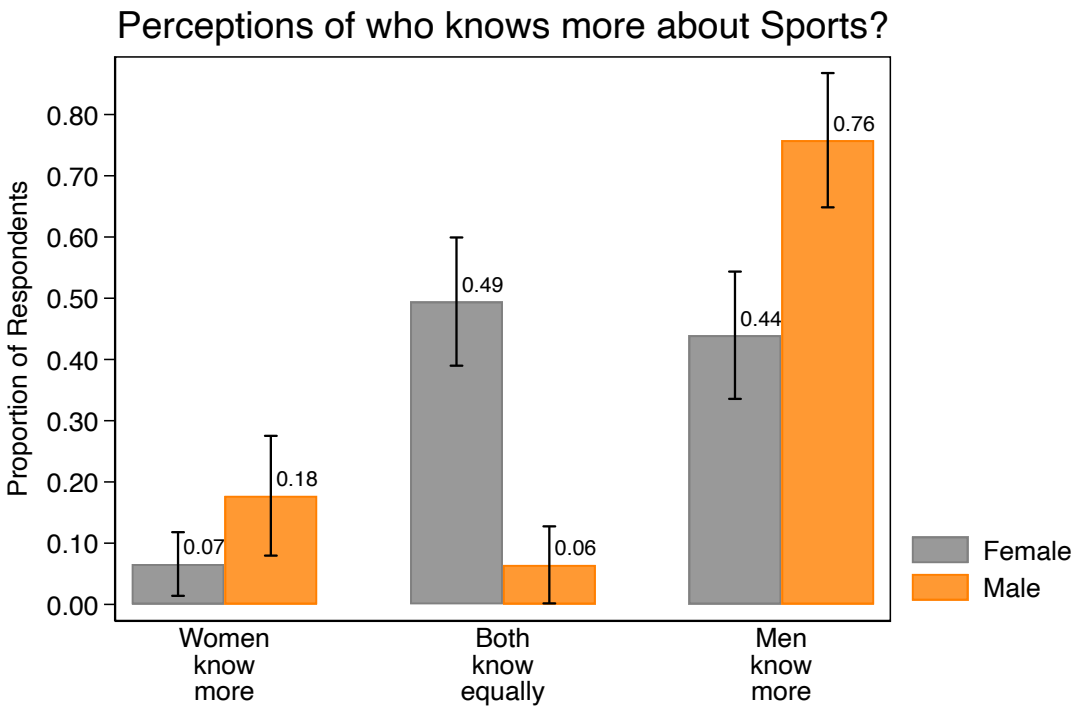


Figure A3: Perceptions of gender stereotypical nature of the different subjects

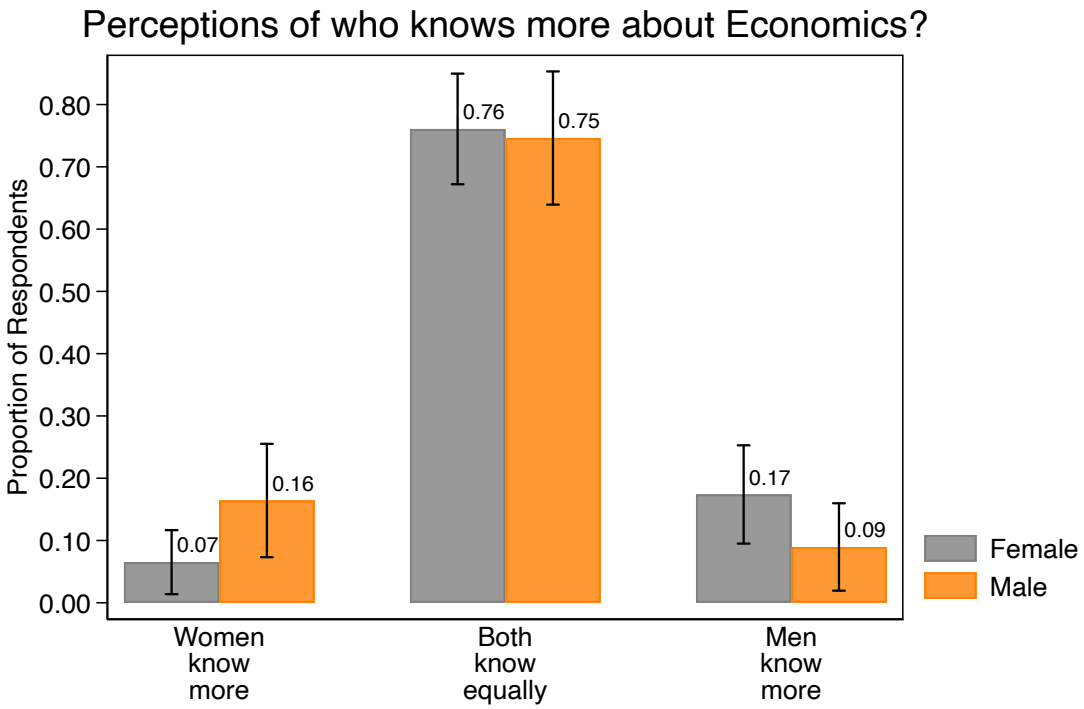


Figure A4: Perceptions of gender stereotypical nature of the different subjects