Team Size and Diversity*

Brais Álvarez Pereira,† Shan Aman-Rana‡ and Alexia Delfino§

March 29, 2024

Abstract

We analyse the relationship between performance, team diversity and size. We first propose a model with knowledge spillovers in production, which predicts that the effect of a diverse team member - relative to the impact that a non-diverse team member - increases with team size. We experimentally test the model by randomly assigning students to solve knowledge questions in teams of different sizes, with or without a person with a diverse knowledge set. In line with the model, we find that the benefit of having a diverse rather than a same-skill colleague is greater in larger relative to small teams. These results have implications for how organizations can design their teams to maximize knowledge flows and performance.

JEL codes: J1, J15, J16, M50, O15
Keywords: Gender, diversity, team performance, information, communication

* We are grateful to seminar participants at multiple universities, conferences, and workshops, including the European University Institute, ASSET 2018, 43rd Spanish Economic Association Conference, Loyola Workshop on Economic Behavior, NOVAFRICA Conference 2022. We thank comments by Marco Casari, Andrea Mattozzi, Mathijs Janssen, Giulio Schinaia, Georgios Manalis, Matthias Schmidtblaicher, Ameet Morjaria, Andrea Ichino, David K Levine, Stefano Gagliarducci, Pablo Brañas-Garza, Alex Coutts. Lucia Solda, Jair Fernandes Martins, Moon Kim, Da Lin, Hengzhou Li, Weining Wang, Yanwen Wang, Laure Anique and Elizabeth Schroppe provided excellent research assistance. BELAB and Universidade Lusófona da Guiné-Bissau provided generous logistical support during the implementation of the lab experiment. All mistakes are our own. AEA Trial Registry RCT ID: AEARCTR-0007988.

†Nova School of Business and Economics (Nova SBE) NOVAFRICA, Nova School of Business and Economics, Universidade Nova de Lisboa, Campus de Carcavelos Rua da Holanda, Nº 1 2775-405 Carcavelos - Portugal. Email:brais.pereira@novasbe.pt

‡Corresponding author. Department of Economics, University of Virginia, 248 McCormick Rd Charlottesville, VA 22904-4182.Email:sa8ey@virginia.edu

§Department of Economics, Bocconi University, Via Sarfatti, 25 Milano - PI 03628350153. Email:alexia.delfino@unibocconi.it

1 Introduction

The desire for organizational diversity is at an all-time high (Deloitte, 2021). Across sectors and geographies, the quest for diversity is shaping business strategy and challenging organizational design. On the one hand, organizations hope to reap the benefits of complementarities between diverse skill sets. On the other hand, diversity may lead to greater communication and coordination costs (Lazear, 1999; Prat, 2002; Hong and Page, 2004; Kahane et al., 2013; Garicano, 2000). The blooming or perishing of diverse teams ultimately depends on how diversity gets integrated within an organization (Shore et al., 2009) or social group (Casari and Tagliapietra, 2018).

This paper focuses on one particular organizational design problem: should diversity be introduced in small or large teams? As a motivating example, think about a hospital. Nurses mainly work with patients, but sometimes they are asked to solve management issues (e.g., optimal scheduling). To help with these tasks, the hospital hires an economist. Where would this new *diverse* hire be more useful, in a small or large team of nurses? Our focus is on the policy relevant case of an organization who hires one diverse colleague and has a choice to allocate them into teams of different sizes, compared to the counterfactual case in which the new hire has a non-diverse skillset.[1]

Empirically, studying the effect of having one diverse member in teams of different sizes can be challenging. First, we need to observe the performance of a large number of teams which differ exclusively in size, keeping constant tasks and incentives, among other things. Second, we require variation in team composition which is as good as random, avoiding endogenous selection into teams. We overcome this challenge by means of a lab experiment, where we are able to randomly allocate participants to teams of different sizes who are facing the same tasks and incentives.

Our main empirical object of interest is the marginal impact that the economist can have in a large compared to small team of nurses. However, relative to small mixed teams, the improvement in performance in large mixed teams may simply come from the larger number of nurses. To rule this out, we use experimental variation in a differences in differences framework: we compare the increase in performance in large versus small teams when an economist rather than a nurse is added.

Our main finding is that the benefit of a skill-diverse teammate rather than a skill-homogeneous one is increasing in team size. Moreover, in larger teams, the gains from the diverse expertise of a new member extend beyond their specialized knowledge area and encompass multiple domains. This evidence suggests that organizations aiming to enhance performance through skill diversity should consider the influence of team size in affecting their workforce dynamics.

We first propose a theoretical framework which draws on Lazear (1999) to explore the effect of skill diversity on performance in teams of different size.[2] There are two types of tasks and two types of workers, each able to solve only one type of task. When a task is randomly drawn, each worker has to solve it individually, but can improve her performance by talking to colleagues who are potentially better than her at solving it. The main prediction of the model is that the contribution of one diverse worker to their colleagues' performance should be increasing in team size. The intuition

[1] Another possibility is to compare teams of different sizes with the same exact share of diverse members. However, this would not take into account hiring constraints which may be binding in the real world. We discuss the implications of our design choice in the discussion section.

[2] We focus on the case in which diversity matters through differences in skills or knowledge. However, heterogeneous within-type and between-type costs of communication in the theoretical model make our findings also relevant for the study of discrimination (differences in inteam and outteam behavior), typically associated to ethnic, religious, gender or national diversity (Alesina and La Ferrara 2000, 2005; Bertrand and Duflo 2017).

is that adding workers with the same skill set to a team increases their colleagues' performance, but at a decreasing rate (since their skills become redundant as their number increases). Thus adding a person with diverse knowledge, instead of an extra worker with the same skill-set, can have a bigger impact in a larger team, where the returns to the same-type knowledge are declining.

We design our experiment to closely map the theoretical setting. Our participants are Nursing and Economics students —nurses and economists from now on. Each of them solves 25 multiple-choice questions, first in isolation and then in teams where communication is encouraged. To allow for gains from having a skill diverse teammate, each test contains questions in Nursing, Economics and General Knowledge. To get variation in team diversity and size, we assigned one economist or one nurse to teams with already one or three nurses. Nurses are the treated participants, while economists provide a source of exogenous variation in diversity within teams. We measure performance as the percentage of correct answers that changed between the team test and the one in isolation, thus controlling for individual baseline ability.

Our results are consistent with the model predictions. First, we find that nurses' average performance is the same in small and large teams made only of nurses. This reflects declining marginal returns from adding people with the same skill set in a team, as formalized in the model. Second, the marginal effects of having a skill diverse teammate depends on team size. In teams of size 2, relative to a homogeneous team, being paired with an economist increases average individual performance by 3.5 percentage points ($p < 0.10$). On the other hand, an economist rather than a nurse in teams of size 4 leads to an additional 6 percentage points higher performance for nurses ($p < 0.05$). This shows that the benefit of having a skill diverse teammate is higher in larger teams.

While our theoretical framework sheds light on the overarching impacts of diversity within teams of varying sizes, it does not forecast the precise domains or subjects where these effects might become apparent. However, we explore the variations in the effects of skill diversity within differently sized teams across various knowledge areas. At baseline, economists are better than nurses in Economics as well as General Knowledge questions, suggesting that an improvement in nurses' performance from working with an economist can manifest in either or both of these knowledge domains. We find that economists improve nurses' performance in general knowledge by 8.8 percentage points ($p < 0.10$) in large relative to small teams. Moreover, economists improve nurses' performance in Economics in both small and large teams and to the same extent. This underscores that having a diverse colleague in a larger team can facilitate leveraging the varied expertise of team members across different knowledge domains.

We conclude our analysis by investigating whether demographic diversity inhibits or enhances the effect of skill diversity in teams. While our framework focuses on skill diversity, skills often correlate with dimensions such as gender, ethnicity and socio-economic background.[3] This complicates the organizational trade-offs involved in managing and coordinating a diverse workforce. While including new members with a diverse skillset may improve organizational performance, the multi-faceted nature of diversity may also increase communication or coordination costs. For instance, female nurses in a hospital may have more difficulties understanding or trusting a male economist than a female economist (e.g., if he uses a different communication style or non-verbal behaviors). We explore this issue by considering the case in which diverse skills are also associated with different gender, ethnic, religious or urban backgrounds. Such intersections may potentially

[3]Despite substantial progress in the last decades, human capital accumulation across different areas of knowledge is still characterized by segregation along the lines of gender, ethnic and socio-economic background (Hoxby and Avery, 2012; Van de Werfhorst, 2017; Thompson, 2021; Hsieh et al., 2019).

exacerbate communication challenges within teams.

To study the impact of combining skill and gender diversity, our experimental design introduces variation in the gender composition of teams in a controlled way. We manipulate whether the economist in a team is a man or a woman, implying that skill-diverse teams in our experiment may be gender homogeneous or not. We find that the positive interaction between skill diversity and team size is only weakly affected by gender composition. Nurses reap the largest benefits from being in a larger team with an economist of the same gender, but the effect is similar to when the economist is of a different gender.

To explore whether the impact of skill diversity is influenced by team diversity in ethnicity, religion or urban origin, we exploit the fact that both nurses and economists were randomized into teams.[4] We confirm our main result on the positive interaction between skill diversity and team size when teammates share the same religion, ethnicity or urban origin. However, our data also show evidence that the gains from skill diversity in large teams may be mitigated when there is also more demographic diversity. This suggests that communication frictions between colleagues with different educational backgrounds may be worsened in large groups, but only when they are also combined with demographic differences.

For organizational design, the main takeaway from our paper is that both the size and composition of teams should be taken into account when introducing new skill sets within the firm. Previous work has highlighted the trade-off between enhanced productivity and communication costs in diverse teams. Our paper contributes by predicting where we expect the productivity effect to dominate and focuses on a choice variable (i.e. size) that firms are able to manipulate.

Our paper contributes to the vast Economics literature on the impact of diversity on productivity (Haltiwanger et al., 1999; Hoxby, 2000; Hansen et al., 2006; Gagliarducci and Paserman, 2012; Ghosh, 2022; Marx et al., 2021; Hjort, 2014; Hoogendoorn et al., 2012; Feld and Zölitz, 2017; Azmat and Boring, 2020; Aman-Rana et al., 2021). Close to our skill-diversity variation, Hoogendoorn et al. (2012) find that cognitive ability dispersion has an inverted U-shape relationship with team performance. We contribute to this literature by highlighting how the same kind of diversity can have different effects on performance depending on the size of the team considered. Our theoretical framework builds on Lazear (1999) who describes that the optimal size of a diverse team depends on the disjointness and relevance of information as well as the costs of communicating between diverse members of the team. This paper complements Lazear (1999) by considering the effects of diversity in more general teams of size N. Our theoretical framework investigates the organizational design problem of the allocation of k diverse workers to teams with m homogeneous members. Our primary objective is to analyze and contrast the performance of a diverse team with that of a same-sized homogeneous team, identifying how this difference compares in teams of different sizes.

Studies in management have also highlighted the importance of team diversity for team outcomes (see Horwitz and Horwitz (2007) for a review). Close to the focus of this paper Krammer (2021) and Hu et al. (2021) describe the mediating role of diversity in the relationship between team size and research and innovation. Using both theory and computations, Hong and Page (2004) study the trade-off between diversity and ability. The authors argue that as group size grows a randomly-formed team of intelligent problem solvers can, under certain conditions, outperform a group made of the highest-performing problem solvers. The intuition is that the former group

[4] Given that demographic diversity (except for gender) was not explicitly varied by our design, these results should be taken with caution. In addition our statistical power is limited for some comparisons.

will have higher functional diversity (i.e., "differences in how people represent problems and how they go about solving them" (Hong and Page, 2004, p. 16385)). Page (2007, 2008, 2019) further build on these pivotal insights. We contribute to this body of research by directly examining the introduction of diverse workers into small versus large teams, and present results that are in line with this body of work. Our study offers estimates derived from an experiment conducted in a closely controlled environment.

Recent studies have investigated empirically communication in teams (e.g. Bloom et al., 2014; Sandvik et al., 2020; Battiston et al., 2021; Menzel, 2021; Espinosa and Stanton, 2022) and, in particular, uncovered frictions that arise when different genders or ethnic groups work together (Sharma and Castagnetti, 2023; Coffman, 2014; Bordalo et al., 2016, 2019; Coffman et al., 2021; Shan, 2020; Marx et al., 2021; Hjort, 2014). While most of this work has focused on studying one aspect of diversity at a time, we try to understand how the combination of skill and demographic diversity may add complexity to team dynamics. Close to our variation in gender composition, in an experimental study with business students Apesteguia et al. (2012) find that teams formed by three women under-perform when compared with any other gender combination and teams of two men and one women perform best. Our study complements their findings by incorporating variation in group size.

2 Theoretical framework

In this section we present a model that delivers precise predictions on gains in expected performance of participants in skill diverse and homogeneous teams of different sizes. All proofs can be found in Appendix A.1. In Appendix A.2, we introduce an extension of the model encompassing both skill and demographic diversity. We illustrate this by considering a scenario where diversity in skills involves a member of a different gender, although this framework can readily accommodate various other forms of demographic diversity, such as religion or ethnicity, as long as communication between members of different demographic groups incurs additional costs. The main result is that additional costs of inter-group communication may mitigate (or enhance) the relationship between team size and skill diversity.

2.1 Set-up

We draw on Lazear (1999) to explore the role that team size plays in the relationship between diversity and performance.

Each worker is assigned a task Z, a random variable that requires Nursing knowledge with probability p (task Z_1) and Economics knowledge with probability $1 - p$ (task Z_2). Two types of workers with different skills can each solve one of the tasks, but not the other. Type-1 workers (nurses) have a stronger background in Health and type-2 workers (economists) in Economics, but they are otherwise identical (e.g., in motivation). There is also heterogeneity within each type of worker, as some are more skillful than others.

Both nurses and economists have the same probability of facing task Z_1 or Z_2. For simplicity, we consider the extreme case in which the knowledge of nurses and economists is fully disjoint: a nurse never knows the solution to an economics task Z_2 and an economist never knows the solution to a health task Z_1.

We build on Lazear (1999) by considering the effects of diversity in more general teams of size N. While he focuses on comparing the expected performance of two homogeneous size-1 teams with that of a diverse size-2 team, we derive our main insights from exploring expected performance when including k diverse workers in teams with m skill-homogeneous members. Our main focus is to compare how the diverse team performs in comparison with a same-size homogeneous team, identifying how this difference compares for teams of different sizes. It is important to note that our theoretical analysis focuses on individual performance for a type-1 worker rather than team performance. However, in many relevant contexts team performance can be viewed as the aggregation of individual performances, thus the key lessons of the paper can be applied to a wider range of contexts.

Tasks are solved individually by each member, who may gain from being part of a team because of communication with others. We assume that agents become as good in solving a given task as the best colleague they have talked to. Communication is costly, and talking to a colleague with a different skill set is more costly than talking to a similarly-skilled colleague. We interpret communication costs as the opportunity cost of time (Garicano, 2000). As people with different technical language take longer to understand each other, we assume that there are higher communications costs between people holding different knowledge sets.

2.2 Performance in isolation

A worker's performance when working in isolation is determined by the task at hand, her type and ability. If the task is of class Z_1, a type-1 worker i will perform the job with quality $z_{1i} \in [0, \hat{Z}_1]$, drawn from the density of production possibilities $f(z_1)$. If the task is of class Z_2, a type-1 worker in isolation will produce 0. Similarly, a type-2 worker j will perform the job with quality $z_{2j} \in [0, \hat{Z}_2]$, drawn from $g(z_2)$ whenever the task is of class Z_2, and 0 otherwise. It follows that the expected production of workers i and j in isolation is respectively given by $E(z_i) = pE(z_{1i})$ and $E(z_j) = (1-p)E(z_{2j})$.

2.3 Performance in skill homogeneous teams

In a team setting, each individual might talk to her colleagues and improve her own performance. We assume that team size and communication costs are small enough for each worker to find it optimal to communicate with everyone else.

Consider type-1 worker i being part of a team with $m > 1$ workers with the same knowledge set. If i communicates with each of her $m-1$ colleagues at a cost c, her expected performance is:

$$E(z_i|m) = pE(z_{1i}|m) + (m-1)c = pZ_{1m} - (m-1)c, \tag{1}$$

where Z_{1m} is the expected highest order statistic from a sample of m draws from $f(z_1)$, which is a concave function of the sample size for any continuous distribution (de la Cal and Cárcamo, 2005). That is, each worker performs her task as well as the best worker she communicates with, minus the incurred communication costs.[5]

[5] Better workers may not want to share their superior knowledge once the task and abilities are revealed. There are incentives for workers to communicate their knowledge as long as part of individual payment depends on team performance, as in our experiment.

The concavity of the expected higher order statistic implies that, for homogeneous teams composed of same-skill workers:

Proposition 1. *The effect of team size on expected performance is concave.*

Intuitively, this result is due to the redundancy of same-type workers' knowledge, which increases with the number of workers of this type. In the model we assume per-person communication costs c to be linear or constant for simplicity, but this result holds as long as the per-person communication costs do not decrease with team size by more than the expected new information gained from having an extra team member of the same type.

The lower gains in expected performance resulting from having additional workers of the same type in larger teams has implications for the effect of diversity.

2.4 Performance in skill diverse teams

A team is now composed of m type-1 workers (nurses) and k type-2 workers (economists). A type-1 worker can potentially improve her expected performance (with respect to working alone) by communicating with both types. Communicating with a worker of a different type costs $c_o > 0$, and we normalize the cost of within-type communication to 0.

In a diverse team, the expected performance of a type-1 worker i is:

$$E(z_i|m,k) = pZ_{1m} + (1-p)Z_{2k} - kc_o. \tag{2}$$

When the team is homogeneous and only has $m+k$ type-1 workers, i's performance is instead:

$$E(z_i|m+k,0) = pZ_{1m+k}. \tag{3}$$

Denote as $\Delta_k(m)$ the difference between worker i's performance in the diverse and homogeneous team of a given size $m+k$:

$$\Delta_k(m) = E(z_i|m,k) - E(z_i|m+k,0) = p(Z_{1m} - Z_{1m+k}) + (1-p)Z_{2k} - kc_o. \tag{4}$$

For a given team size, $\Delta_k(m)$ is the additional contribution of k diverse workers to the expected performance of each of their m type-1 colleagues. The concavity of the expected highest order statistic, together with the knowledge of type-2 workers contributing the same in larger and smaller teams, produces the following result:

Proposition 2. *For a type-1 worker, the gain in expected performance from being placed in a diverse team with k type-2 workers - instead of a same-sized homogeneous team - is increasing in team size.*

The result follows from the concavity of the expected highest order statistic. Proposition 2 implies that the impact of skill diversity can vary across teams of different sizes. As the number of type-1 workers increases within a team, their individual knowledge becomes redundant. Consequently, a diverse worker is relatively more valuable in larger teams compared to the potential impact of a non-diverse team member. Figure 1 illustrates the expected performance of type-1 workers, both in isolation and in teams of different size and diversity.

The finding that the impact of a skill-diverse worker on the performance of members within a homogeneous team grows with team size is a relevant finding. For instance, it can accommodate situations in which the same type of diversity can have a detrimental effect on performance in smaller teams (e.g., if it crowds-out useful knowledge brought by common skills), while yielding a positive impact in larger ones.

3 Context and research design

We carried out a lab experiment with university students studying Nursing and Economics in the largest university in Guinea Bissau. Students were drawn from their second, third and fourth (final) year in their respective degrees.

3.1 Experimental design

We designed the experiment to simultaneously study three factors that might affect individual performance in teams: size, skill diversity, and gender diversity. With this aim, participants were randomly assigned to teams of size 2 and 4, composed of only nurses or of nurses with one economist of either gender. Our treatments are illustrated in Figure 2. From a total of 248 students (61 economists and 187 nurses) we created 52 teams composed of two members and 36 teams composed of four members, for a total of 88 teams. Out of the teams with two members, 37 teams had one economist. Out of the teams composed of four members, 24 teams had one economist. In approximately half of the teams with an economist, the economist had the opposite gender from the rest of the team (see Appendix D for details of the recruitment and logistics).

All the participants that started the experiment completed it to the end. While most participants took part in the study once, 23 participants (19 nurses and 4 economists) out of a total of 248 students took part in the study in more than one session. Since the tests were unique per session, we consider these participants as fresh draws from the pool of students and include them in the main results. However, we perform a variety of tests to address any effect that the experience of such participants can have on the results. First, our main specifications cluster standard errors at both the team and participant level to account for repeaters' correlated performance across sessions, and include an indicator for participants who repeated the experiment more than once. We also include a control for the average baseline performance of the team to hold fixed the effects of repetition on team-level performance. Second, we test whether repetition is correlated with any of the treatment arms. Table B.1 shows that that is not the case. Third, we show that the main results are robust to the use of alternative strategies to account for the presence of repeaters, such as keeping only the first occurrence for each participant, only clustering at the team and participant level, or only controlling for a dummy for participating in more than one experimental session (see B.2 and B.3 for these robustness tests).

3.2 Task and incentives

Each participant had to solve a knowledge test twice. The test consisted of 25 multiple-choice questions with five options and a unique correct answer. For a nurse (economist), ten questions were general-knowledge, ten questions were from Nursing (Economics) and five were questions from the other discipline.[6] Since each student faced 20% questions from the other discipline,

[6] See the Online Appendix for more details on the source and characteristics of the questions.

the task required both Nursing and Economics skills. We thus embedded the need for skill complementarity within the design, as in the model.

Students had to first answer the test in isolation. This was followed by a second round of testing in teams, using the same exact questions that each participant had completed individually. Participants were distributed into teams and got to know about the composition of their team only after completing the individual test. In teams, participants were allowed to communicate and discuss their answers with the other members. Every participant in each session studying the same subject answered the exact same test, had 30 minutes to complete it in each round and there was no penalty for incorrect answers in both rounds.

In the first round, students received 80 FCFA (0.12$) per right answer.[7] For the second round of testing, in order to incentivize participants to share their knowledge, one of the tests from the team was randomly chosen and all team-members were paid according to the number of correct answers in that test (80 FCFA for each correct answer).

3.3 Balance

Nurses assigned to different types of teams (i.e. small vs large, skill diverse vs homogeneous) were very similar on a variety of observable characteristics (Figure C.1). These checks support balanced assignment of treatment groups in pre-treatment covariates. Figure C.1 shows that there is only one statistically significant differences across treatment groups. Appendix Figure C.2 presents an equivalent graph for economists. There is only 1 significant coefficients out of more than 20 tested.[8]

3.4 Baseline differences in skills between nurses and economists

To understand the skill differences that exist across nurses and economists we present the baseline skill gap in Figure 3. This figure shows the average percentage difference in the score of nurses and economists in each of the three knowledge areas tested.

As expected, students perform better in the subject in which they specialize. Economists perform 25% higher than nurses in Economics questions, while nurses perform almost 20% better than economists in Nursing questions. This reassures that our empirical variation captures differences in skills, and it holds true for both male and female nurses (as shown in Figure C.3).

We also find differences in general knowledge, with nurses performing 6% worse than economists in this category. Thus we should expect that nurses placed in the same team as economists can experience benefits in either or both Economics and General Knowledge domains.

[7]Students were paid between 500 and 5000 FCFA (between 0.76 and 7.58$), depending on performance. These incentives were sizeable: the minimum daily wage is around 600 FCFA.

[8]The variable which is not balanced for economists is whether they have above-median grades. However, our main specification controls for the baseline team performance (including the economist's scores), effectively taking care of possible differences in economists' ability across treatment groups. In graphs available upon request, we also confirm that assignment of the additional gender-diversity treatment was balanced on most of the given observables.

4 Main results

4.1 Empirical strategy

We report estimates restricting attention to the sample of nurses, and briefly discuss results on economists in Section 4.5. While the main results presented in the paper are at the nursing student level, Appendix Table B.5 replicates the key results collapsing the data at the team level. All the main results continue to hold at the team level. Our preferred specification is the one at the nursing student level since this is more closely aligned with our theoretical framework and also allows us to cleanly identify the source of variation in skill diversity (economists) and study their effects on nurses.[9]

To test the effect of a skill diverse teammate in teams of different size we use the following estimation. For nursing student i we estimate:

$$\Delta Score_i = \alpha SkillDiversity_i + \gamma TeamSize4_i + \beta SkillDiversity_i \times TeamSize4_i + \theta X_i + u_i \quad (5)$$

where $\Delta Score_i$ is the difference in the percentage of correct answers between each individual i and their team test (the same test is used in both rounds).[10] $SkillDiversity_i$ is a dummy variable equal to 1 if a skill diverse teammate (i.e., an economist) is allocated to nursing student i, $TeamSize4_i$ is a dummy variable that takes value 1 if the nursing student is allocated to a team of size four. X_i include as control variables a dummy variable that takes the value 1 if a participant is a male, each team's average ability measured by their baseline performance in the individual test, and a dummy for taking part in multiple experimental sessions. The reference category comprises of participants in skill homogeneous teams of size 2. Standard errors u_i are clustered at the team and participant level.

Coefficient α is the effect on performance of a nurse working with one economist, relative to their performance when working with another nurse. γ is the effect of allocating a nurse to a homogeneous-skill team of size 4 relative to size 2. The main coefficient of interest is β, which captures the effect of being allocated to work with an economist relative to a nurse in a team of size 4 relative to size 2. Following Proposition 2, we expect $\beta > 0$.

4.2 Is skill diversity beneficial?

We first present evidence that our treatments affect the distribution of changes in score (%) between the first round (answering alone) and the second (answering in teams), in line with the presence of intra-team communication.

Figure 4 shows the distribution of the average test score change between the individual and the team test by skill diversity (working with an economist or not) and size (2 or 4). The distribution of the average change in test scores is concentrated on positive values irrespective of team size or diversity suggesting that people are communicating within their teams.[11] Second, Figure 4

[9] Such a bifurcation of the participants and the source of variation in team characteristics helps overcome mechanical correlations in outcomes that have been highlighted as a problem in the estimation of peer effects in the literature (Manski, 1993; Angrist, 2014).

[10] Following the model, we focus on individual performance rather than team performance. We think this is relevant as many organizations use individual incentives even when employees are organized in teams.

[11] The change in scores is the joint effect of doing the test twice and intra-team communication. We maintain the assumption

suggests that skill diversity can benefit performance, as the gains from team communication are greater when an economist is allocated to the team. For both team sizes, the distribution of performance is shifted to the right in skill-diverse teams (red solid lines). While team size may also positively impact performance, evidence from Figure 4 is not conclusive as the variance in performance for teams of four nurses is larger than in teams of two nurses.

More formally, Columns (1) and (2) of Table 1 show the average gains from size and from working with an economist. The specification uses Equation 5, but estimates only coefficients α and γ separately. Pooling both small and large teams, Column (1) shows that there are positive and statistically significant gains from working with an economist. The average gain is 7.6 percentage points in skill-diverse teams with respect to skill-homogeneous ones (16% of average individual test scores, $p < 0.01$). This confirms that peers learn from each other within our teams. Column (2) shows that size is also beneficial to team performance, independently of team composition. The possibility of asking advice to a larger set of people improves a participants' own performance by 4.5 percentage points (9% of average individual test scores, $p < 0.05$), an effect which is statistically indistinguishable from the one of being matched with a skill-diverse teammate ($p > 0.10$).

Taken together, these results may suggest that organizations have two different policy levers available to increase performance: enlarging teams, or making employees interact with people with a diverse skillset. However, if the benefits of skill diversity vary with team size, the policy problem is more complicated and may require pulling both levers at the same time. In the next section, we explore whether skill diversity has a differential effect in teams of different sizes.

4.3 When is skill diversity most beneficial?

This section presents the key result of the paper: a skill diverse teammate is most beneficial in large teams, relative to the impact that a same-skill person would have had. This result emerges already from the raw data plotted in Figure 4. While working with an economist improves the score of nurses in all teams, such improvement is much larger in a team of size 4 than size 2.

Column (3) of Table 1 formalizes this intuition by using the full specification of Equation 5, which interacts team size with skill diversity. The estimated coefficient α shows that there are positive, but smaller gains from working with an economist in small teams. The average gain is 3.5 percentage points (7% of average individual test scores, $p < 0.10$) in skill-diverse pairs with respect to skill-homogeneous pairs. The positive impact of diversity described in the previous paragraphs is thus attenuated in small teams. Even more strikingly, we see that working in a large team does not improve scores with respect to a small one when skills are homogeneous among teammates. This is consistent with the decreasing marginal returns of teammates with the same skills described by Proposition 1 of the model.

Crucially, we find that the coefficient on the interaction between skill diversity and team size 4 (β) is positive, significant and large. The coefficient is 6.1 percentage points, representing 13% of the average individual test scores ($p < 0.05$). Relative to the addition of another nurse, an economist is most beneficial in larger teams compared to small ones. This result is consistent with our main theoretical prediction (Proposition 2), which states that the addition of a person with diverse knowledge should have the largest marginal impact on each member's performance in

that the change due to repetition is the same across treatments. Empirically, this effect seems small: when assigned to working with only one economist, nurses do not improve their performance in the 5 Nursing questions that are not present in the test for economists.

larger teams. Compared to the results in Columns (1) and (2), Column (3) thus provides evidence that team size interacts with skill diversity and influences the impact that it can have.[12]

While our main theoretical and empirical results focus on the individual gain in performance in different types of teams, Appendix Table B.5 replicates the key results at the team level. We see that aggregate performance in large skill diverse teams is higher by 5.6 percentage points compared to small homogeneous teams ($p < 0.05$).

Results by knowledge domain. How do the benefits of skill diversity realize in teams? And does this differ across knowledge areas? As described in Figure 3 nurses perform worse than economists in both General Knowledge and Economics questions. Therefore, the effects of introducing an economist (relative to a nurse) can increase with team size in either or both of these two subjects.

Although our theoretical framework illuminates the overall effects of diversity in teams of different sizes, it does not clearly differentiate the specific domains (general knowledge vs economics) in which these effects may manifest. Nonetheless, Table 2 digs deeper into these questions by looking at participants' performance gains in different knowledge areas. The first row of Table 2 shows the effect of being in a small diverse team compared to a small homogeneous team (coefficient α). Being paired with an economist - compared to working with one other nurse - improves a participant's performance in economics by 28 percentage points ($p < 0.01$), and decreases performance in nursing by 6.6 percentage points ($p < 0.05$). This suggests that in a small team, on net placing nursing students with an economist is more useful than pairing them with another nurse.[13] Diversity is thus beneficial for performance. In contrast, the last row of Table 2 shows that being in a large team made of only nurses does not lead to improvements in any subject compared to a small team of nurses. While the coefficient on "Team Size 4" is positive and sizable for nursing (3.4 percentage points), it is not statistically significant. This result directly speaks to the skill-redundancy dynamics shown in Proposition 1 and Figure 1.

The middle row of Table 2 shows the estimates of coefficient β in different subjects. We find that economists differentially improve nurses' performance in general knowledge by 8.8 percentage points ($p < 0.10$) in large relative to small teams (Column 1). On the other hand, economists improve nurses' performance in economics in both small and large teams and to the same extent (Column 2). This highlights how the presence of a diverse colleague in a larger team can aid in harnessing the diverse expertise of team members across various knowledge domains.

To summarize, in this Section we showed that an economist's knowledge is most beneficial in large teams, relative to the impact that a nurse would have had in a similarly-sized team.

In the next Section, we ask whether the impact that economists have on nurses' performance depends on whether they differ in other characteristics beyond skills.

[12]Students' characteristics may also influence whether teams are able to benefit (or not) from including an expert in another discipline. For instance, previous work has highlighted the challenges that women face in interacting in team settings, especially in sharing of ideas (Coffman, 2014; Born et al., 2018; Heikensten and Isaksson, 2019). Similarly, students from lower socio-economic backgrounds or minorities may be more - or less - likely to learn in a team setting depending on their traits, peers, and whether their minority background is made salient (Deckers et al., 2015; Hoff and Pandey, 2006; Steele, 1998; Shan, 2022). Appendix Table B.4 explores whether there are differences in the types of students who benefit from being matched with skill-diverse teammates. We split our sample by nursing students' gender (Columns 1 and 2), religion (Columns 3 and 4), ethnicity (Columns 5 and 6) and city of origin (Columns 7 and 8). Overall, we find no clear heterogeneity in treatment effects based on these characteristics, however, our statistical power in this exercise is also limited.

[13]The model includes a prediction that skill homogeneous small teams may perform better than skill diverse small teams. However, empirically this depends on the extent to which certain skills matter for individual performance.

4.4 Does demographic diversity inhibit the effect of skill diversity?

This section investigates whether demographic diversity inhibits the effect of skill diversity. We first exploit our experimental variation by interacting the $Skill\ Diversity$ regressor in Equation 5 with a dummy for whether the diverse teammate is of the same or of a different gender to the rest of the team. This model is shown in Table 3.[14]

We first focus on the effects of working with an economist of the same gender as nurses (coefficients α and γ). In this case, Table 3 shows almost identical results to the ones discussed in the previous section on the average effects of working with a skill diverse teammate. We see that skill diversity has no impact in small teams, but significantly improves the average scores in larger teams. The coefficient on the interaction between "Team Size 4" and "Skill diverse" is 7.3 percentage points for teams in which the economist is of the same gender as the other teammates (representing 15% of the mean individual score), and it is statistically significant at the 90% level. Overall, the effects of introducing a skill diverse teammate in a gender homogeneous team goes in the direction suggested by the model.

Coefficients β and δ explore the joint effect of multi-dimensional diversity, in skills and gender, on performance. A nurse paired with one economist experiences a substantial benefit when the economist is of the opposite gender (5.6 pp increase in score). This benefit is even larger than being matched with an economist of the same gender (coefficients $\alpha = \beta$, $p = 0.05$). Figure 5 provides a plausible explanation for such higher performance in gender-diverse than gender-homogeneous pairs: gender diverse teams have a larger gap in skills between nurses and economists (in both general knowledge and economics) than gender homogeneous teams.

When economists are of a different gender relative to nurses, moving from pairs to teams of four people also positively affects performance. Coefficient δ is indeed positive and significant (5.4 percentage points, or 11% of the mean individual score). We are also unable to reject the hypothesis that the interaction between skill diversity and teams of size four is the same in gender-homogeneous or gender-diverse teams ($p = 0.605$).

However, Table B.3 shows that the significance and magnitude of coefficient δ is more dependent on the exact strategy used to account for participants in multiple sessions. All in all, this indicates that the statistical power to detect differences between teams which are gender diverse in Table 3 is limited. Nevertheless, we recognize that the evidence so far may also point to the absence of the "size premium" for skill diversity when accompanied by gender diversity.[15]

We then explore whether the relationship between skill diversity and team size is affected by teammates' differences in other demographics variables. For each team, we compute the share of members who share the same religion (Christian vs non-Christian), the same ethnicity (Balanta vs non-Balanta) or the same urban origin (being from the capital city or not). We categorize teams as being diverse in a certain demographic characteristic - e.g., religion - if at least one member is different to the others. Figure 6 shows the number of nursing students (out of a total of 187) in teams with different types of demographic diversity: 89 nurses are in religion diverse teams, while

[14]Note that in our experiment gender diversity is only introduced in skill-diverse teams, since our aim was to study whether demographic diversity impacts the effect of skill diversity on performance.

[15]This could be explained within our framework by introducing inter-gender communication costs that are increasing with team size, as proposed by our extension in Appendix A.2. This effect can be explained, for instance, if the economist becomes less effective in sharing his/her knowledge within a large team of students of a different gender, or if the team pays less attention to him/her (Conlon et al., 2021). Social image concerns, that hinder nurses' willingness to ask for advice to the economist, may also be higher in larger teams and with an economist of a different gender (Bursztyn et al., 2017; Bendersky and Pai, 2018).

118 and 111 nurses are in ethnic and origin diverse ones.

Table 4 uses the same specification as Table 3, but interacts the *Skill Diversity* regressor of Equation 5 with dummies for whether the team is demographically diverse or not (along the dimension shown in the column heading). The table shows two important results. First, we find that the effect of introducing a skill diverse teammate in a homogeneous team is higher in larger than smaller teams. This result confirms our main theoretical prediction and it is aligned with the results on gender diversity previously discussed. Second, when teams are heterogeneous in terms of demographics, we find that a skill-diverse teammate benefits nurses' performance only in small teams and does not have any additional benefit in larger teams. This effect is found when considering either religious, ethnic or background diversity. The coefficient (δ) is negative and p-value from a test of $\gamma = \delta$ is 0.108 in the case of religious diversity (Column 1), 0.099 when we consider ethnic diversity (Column 2), and 0.019 in the case of origin diversity (Column 3). Moreover, it is again in line with the noisy effect of the interaction between skill diversity and team size found in gender diverse teams.

Through the lens of our framework, this evidence suggests that the costs of communicating in teams may vary with team size when teammates differ in more dimensions than just skills. For instance, in large teams minorities may feel less comfortable in asking advice to economists as this may send a negative signal about their competence to teammates who belong to different - and perhaps more advantaged - demographic groups (Hoff and Pandey, 2006; Steele, 1998). In small teams, this effect would be mitigated as the interaction with an economist happens privately without anyone else observing. While it is beyond the scope of our paper to pin down the exact mechanism of these results (which are exploratory in nature), we think they point to the importance of studying the particular challenges that introducing multi-faceted diversity in an organization may bring.

4.5 Discussion

Diversity share as a potential mechanism. Our experiment considers a policy-relevant situation in which one employee has to be allocated in teams that differ in their sizes. This design implies that the "share of diversity" is not held constant in small versus large teams, and this can be a potential mechanism behind the results. To see this, compute the probability that any two team members picked randomly have different skill sets. This probability is 1 in teams with one economist and one nurse, but it is 1/2 in teams with one economist and three nurses. Thus in our experiment the share of diversity falls as the size of skill-diverse teams increases. If communication costs increase with the share of diversity, rather than the absolute number of skill-diverse teammates, then our results may be explained by these differences in treatment intensity. Theoretically, this could be obtained in our framework by having the costs of communication increase as a function of the team diversity share. While our framework focused on the relative returns from adding a diverse skill member to teams of different size, such changes in relative communication costs might be another way of explaining our results.

Effects on the minority. While our focus is on the performance of nurses, our majoritarian group, the net benefit of diversity depends also on how minorities feel and perform. Abrupt policies that increase workplace diversity may have unintended consequences on the minority if there is limited integration and inclusion (Folke and Rickne, 2022). Unfortunately, the nature of team interaction in our experiment is too stylized and short to capture aspects of minority inclusion which are important to consider in reality. Nevertheless, the performance of our minority (economists) across team types gives us some insights into this issue. Figure C.4 shows that economists' performance improvement is greatest in a large team which is gender homogeneous. However, being assigned to a large team where economists are a minority both in gender and skills does not improve their performance with respect to working in pairs.[16] This result suggests that the combination of skill and gender diversity in large teams may create communication costs which are higher than in other combinations, at least for some participants.

Conclusion

The jobs of the future increasingly require workers to solve complex cognitive tasks within global teams (Lazear, 1999, 2000; Antràs et al., 2006, 2008; Edmondson, 2012). These trends make knowledge spillovers within diverse teams increasingly important determinants of firms' comparative advantage and competitiveness. How should organizations design their teams to maximise knowledge flows between members?

In this paper, we study two aspects of this design choice: skill diversity and team size. In particular, we are interested in whether including one person with a diverse skills set - relative to a person with common skills - is most beneficial for performance in small or large teams. We find support for a positive relationship between gains from skill diversity and team size.

We also find suggestive evidence that the positive relationship between skill diversity and team size is mitigated when teammates do not share the same demographic characteristics. This result is consistent with demographic differences affecting the costs of communicating in teams, but in a way which is varying with team size. This result speaks to emerging evidence on possible frictions to learning between individuals (Conlon et al., 2021) and highlights that environmental features (e.g, team size) may play a key role in limiting knowledge diffusion.

A promising direction for future research is studying how firms and institutions should design their workplaces to make the most of diversity, while mitigating its costs and avoiding lack of inclusion for minorities. Our results highlight that diversity is frequently multi-faceted and that acknowledging this multi-dimensionality is crucial to get a better understanding of communication flows between people working together. While our stylized experimental setting is not appropriate to measure all the possible trade-offs involved in modifying team composition, we believe that exploring the relationship between team size and multi-dimensional diversity in realistic field settings may be worthwhile for future research.

[16]The number of observations becomes too small, but this gap is driven by female economists in male-dominated teams, who significantly decrease their improvement.

References

Alesina, Alberto and Eliana La Ferrara, "Participation in heterogeneous communities," *The Quarterly Journal of Economics*, 2000, *115* (3), 847–904.

_ **and** _ , "Ethnic diversity and economic performance," *Journal of Economic Literature*, 2005, *43* (3), 762–800.

Aman-Rana, Shan, Alexia Delfino, Clement Minaudier, Shamyla Chaudry, and Brais Álvarez Pereira, "Gender and Choice: Experimental Evidence," *Available at SSRN 3958386*, 2021.

Angrist, Joshua D, "The perils of peer effects," *Labour Economics*, 2014, *30*, 98–108.

Antràs, Pol, Luis Garicano, and Esteban Rossi-Hansberg, "Offshoring in a knowledge economy," *The Quarterly Journal of Economics*, 2006, *121* (1), 31–77.

_ **,** _ **, and** _ , "Organizing Offshoring: Middle Managers and Communication Costs," *Harvard University Press*, 2008, pp. 311–39.

Apesteguia, Jose, Ghazala Azmat, and Nagore Iriberri, "The impact of gender composition on team performance and decision making: Evidence from the field," *Management Science*, 2012, *58* (1), 78–93.

Azmat, Ghazala and Anne Boring, "Gender diversity in firms," *Oxford Review of Economic Policy*, 2020, *36* (4), 760–782.

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

Bendersky, Corinne and Jieun Pai, "Status dynamics," *Annual Review of Organizational Psychology and Organizational Behavior*, 2018, *5*, 183–199.

Bertrand, Marianne and Esther Duflo, "Field Experiments on Discrimination," in "Handbook of Economic Field Experiments," Vol. 1, North-Holland, 2017, pp. 309–393.

Bloom, Nicholas, Luis Garicano, Raffaella Sadun, and John Van Reenen, "The Distinct Effects of Information Technology and Communication Technology on Firm Organization," *Management Science*, 2014, *60* (12), 2859–2885.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, "Stereotypes," *The Quarterly Journal of Economics*, 2016, *131* (4), 1753–1794.

_ **,** _ **,** _ **, and** _ , "Beliefs about Gender," *American Economic Review*, 2019, *109* (3), 739–73.

Born, Andreas, Eva Ranehill, and Anna Sandberg, "A man's world?–The impact of a male dominated environment on female leadership," *Working Papers in Economics, University of Gothenburg*, 2018.

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais, "'Acting Wife': Marriage Market Incentives and Labor Market Investments," *American Economic Review*, 2017, *107* (11), 3288–3319.

Casari, Marco and Claudio Tagliapietra, "Group size in social-ecological systems," *Proceedings of the National Academy of Sciences*, 2018, *115* (11), 2728–2733.

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

Coffman, Katherine, Clio Bryant Flikkema, and Olga Shurchkov, "Gender Stereotypes in Deliberation and team decisions," *Games and Economic Behavior*, 2021.

Conlon, John J, Malavika Mani, Gautam Rao, Matthew W Ridley, and Frank Schilbach, "Learning in the Household," Working Paper 28844, National Bureau of Economic Research May 2021.

de la Cal, J. and J. Cárcamo, "Inequalities for expected extreme order statistics," *Statistics and Probability Letters*, 2005, *73* (3), 219–231.

de Werfhorst, Herman G Van, "Gender segregation across fields of study in post-secondary education: Trends and social differentials," *European Sociological Review*, 2017, *33* (3), 449–464.

Deckers, Thomas, Armin Falk, Fabian Kosse, and Hannah Schildberg-Hörisch, "How Does Socio-Economic Status Shape a Child's Personality?," 2015.

Deloitte, "First Steps into the Labour Market 2021, International survey of students and recent graduates," *Report*, 2021.

Edmondson, Amy C, *Teaming: How organizations learn, innovate, and compete in the knowledge economy*, John Wiley & Sons, 2012.

Espinosa, Miguel and Christopher T Stanton, "Training, communications patterns, and spillovers inside organizations," Technical Report, National Bureau of Economic Research 2022.

Feld, Jan and Ulf Zölitz, "Understanding peer effects: On the nature, estimation, and channels of peer effects," *Journal of Labor Economics*, 2017, *35* (2), 387–428.

Folke, Olle and Johanna Rickne, "Sexual Harassment and Gender Inequality in the Labor Market," *The Quarterly Journal of Economics*, 2022, *137* (4), 2163–2212.

Gagliarducci, Stefano and M Daniele Paserman, "Gender Interactions within Hierarchies: Evidence from the Political Arena," *The Review of Economic Studies*, 2012, *79* (3), 1021–1052.

Garicano, Luis, "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 2000, *108* (5), 874–904.

Ghosh, Arkadev, "Religious Divisions and Production Technology: Experimental Evidence from India," *Available at SSRN: 4188354*, 2022.

Haltiwanger, John C, Julia I Lane, and James Spletzer, "Productivity differences across employers: The roles of employer size, age, and human capital," *American Economic Review*, 1999, *89* (2), 94–98.

Hansen, Zeynep, Hideo Owan, and Jie Pan, "The Impact of Group Diversity on Performance and Knowledge Spillover–An Experiment in a College Classroom," Working Paper 12251, National Bureau of Economic Research May 2006.

Heikensten, Emma and Siri Isaksson, "Simon Says: Examining Gender Differences in Advice Seeking and Influence in the Lab," *Available at SSRN 3273186*, 2019.

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

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

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

Hoogendoorn, Sander, Simon C Parker, and Mirjam van Praag, "Ability dispersion and team performance: A field experiment," *Tinbergen Institute Discussion Paper 12-130/VII, Available at SSRN:2183095*, 2012.

Horwitz, Sujin K and Irwin B Horwitz, "The effects of team diversity on team outcomes: A meta-analytic review of team demography," *Journal of Management*, 2007, *33* (6), 987–1015.

Hoxby, Caroline, "Peer effects in the classroom: Learning from gender and race variation," Working Paper 7867, National Bureau of Economic Research August 2000.

Hoxby, Caroline M and Christopher Avery, "The missing" one-offs": The hidden supply of high-achieving, low income students," Technical Report, National Bureau of Economic Research 2012.

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow, "The allocation of talent and us economic growth," *Econometrica*, 2019, *87* (5), 1439–1474.

Hu, Die, Maoyan She, Lingfeng Ye, and Zhiwei Wang, "The more the merrier? Inventor team size, diversity, and innovation quality," *Science and Public Policy*, 2021, *48* (4), 508–520.

Kahane, Leo, Neil Longley, and Robert Simmons, "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*, 2013, *95* (1), 302–314.

Krammer, Sorin, "Research performance in Business and Management: The impact of team size and diversity," *Academy of Management Proceedings*, 2021, *2021* (1), 10981.

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

_ , "Diversity and immigration," in "Issues in the Economics of Immigration," University of Chicago Press, 2000, pp. 117–142.

Manski, Charles F, "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, 1993, *60* (3), 531–542.

Marx, Benjamin, Vincent Pons, and Tavneet Suri, "Diversity and team performance in a Kenyan organization," *Journal of Public Economics*, 2021, *197*, 104332.

Menzel, Andreas, "Knowledge exchange and productivity spill-overs in Bangladeshi garment factories," *Journal of Economic Behavior & Organization*, 2021, *185*, 721–746.

Page, Scott E, "Making the difference: Applying a logic of diversity," *Academy of Management Perspectives*, 2007, *21* (4), 6–20.

— , *The difference: How the power of diversity creates better groups, firms, schools, and societies-new edition*, Princeton University Press, 2008.

— , *The diversity bonus: How great teams pay off in the knowledge economy*, Princeton University Press, 2019.

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

Sandvik, Jason J, Richard E Saouma, Nathan T Seegert, and Christopher T Stanton, "Workplace knowledge flows," *The Quarterly Journal of Economics*, 2020, *135* (3), 1635–1680.

Shan, Xiaoyue, "Does minority status drive women out of male-dominated fields?," *Unpublished manuscript*, 2020.

— , "The minority trap: Minority status drives women out of male-dominated fields," *Unpublished manuscript*, 2022.

Sharma, Karmini and Alessandro Castagnetti, "Demand for information by gender: An experimental study," *Journal of Economic Behavior & Organization*, 2023, *207*, 172–202.

Shore, Lynn M., Beth G. Chung-Herrera, Michelle A. Dean, Karen Holcombe Ehrhart, Don I. Jung, Amy E. Randel, and Gangaram Singh, "Diversity in organizations: Where are we now and where are we going?," *Human Resource Management Review*, 2009, *19* (2), 117–133.

Steele, Claude M, "Stereotyping and its threat are real.," 1998.

Thompson, Owen, "Human Capital and Black-White Earnings Gaps, 1966-2017," Technical Report, National Bureau of Economic Research 2021.

Tables

Table 1: Skill diversity and team size

Dependent Variable:	**Change in score (%)**		
	(1)	(2)	(3)
Skill diverse team (α)	7.614***		3.505*
	(1.581)		(1.860)
Team size 4 (γ)		4.470**	0.619
		(1.700)	(2.167)
Skill diverse team × Team size 4 (β)			6.119**
			(2.937)
P val: Diversity + Diversity × Team 4 = 0			0.000
P val: Team 4 + Diversity × Team 4 = 0			0.001
P val: $\alpha = \gamma$			0.177
Mean of Individual Score (%)	47.7	47.7	47.7
Controls	Yes	Yes	Yes
Observations	187	187	187
Cluster	88	88	88

Notes. Change in score is the percentage change in score from the individual to the team test. Test remains the same for each person across the two rounds. Skill diverse team is a dummy variable that equals to one when there is an economics student in the team. Team size 4 is a dummy variable that equals one when there are four students in the team. All specifications include a control for the team's average baseline performance in the individual test, a dummy variable that takes the value one when the participant is a male student, and a dummy that takes the value of one if the participant attended multiple sessions. Standard errors are clustered at the student, team level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table 2: Skill diversity and team size by knowledge area

Dependent Variable:	**Change in score (%)**		
Subject:	General Knowledge	Economics	Nursing
	(1)	(2)	(3)
Skill diverse team (α)	1.384	27.981***	-6.613**
	(3.318)	(4.444)	(3.173)
Skill diverse team × Team size 4 (β)	8.807*	0.118	6.432
	(4.846)	(8.278)	(5.898)
Team size 4 (γ)	-0.899	-1.899	3.396
	(3.713)	(5.758)	(4.608)
P val: Diversity + Diversity × Team 4 = 0	0.006	0.000	0.970
P val: Team 4 + Diversity × Team 4 = 0	0.016	0.760	0.010
P val: $\alpha = \gamma$	0.553	0.000	0.026
Mean of Individual Score (%)	62.3	21.1	46.3
Controls	Yes	Yes	Yes
Observations	187	187	187
Cluster	88	88	88

Notes. Change in score is the percentage change in score from the individual to the team test. Test remains the same for each person across the two rounds. Skill diverse team is a dummy variable that equals to one when there is an economics student in the team. Team size 4 is a dummy variable that equals one when there are four students in the team. All specifications include a control for the team's average baseline performance in the individual test, a dummy variable that takes the value one when the participant is a male student, and a dummy that takes the value of one if the participant attended multiple sessions. Standard errors are clustered at the student, team level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table 3: Does gender diversity inhibit the effect of skill diversity?

Dependent Variable:	**Change in score (%)**
	(1)
Skill diverse × gender homogeneous (α)	0.767
	(1.995)
Skill diverse × gender diverse (β)	5.634**
	(2.276)
Skill diverse × gender homogeneous × team size 4 (γ)	7.314*
	(3.698)
Skill diverse × gender diverse × team size 4 (δ)	5.376*
	(3.207)
Team size 4	0.667
	(2.167)
P val: $\alpha = \beta$	0.047
P val: $\gamma = \delta$	0.605
Mean of Individual Score (%)	47.7
Controls	Yes
Observations	187
Cluster	88

Notes. Change in score is the percentage change in score from the individual to the team test. Test remains the same for each person across the two rounds. Skill diverse team is a dummy variable that equals to one when there is an economics student in the team. Team size 4 is a dummy variable that equals one when there are four students in the team. Gender diverse is a dummy that takes the value one if the economist is of a different gender from the nurses. Gender homogeneous is a dummy that takes the value one if the economist is of the same gender as the nurses. All specifications include a control for the team's average baseline performance in the individual test, a dummy variable that takes the value one when the participant is a male student, and a dummy that takes the value of one if the participant attended multiple sessions. Standard errors are clustered at the student, team level. Significance levels are denoted as: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

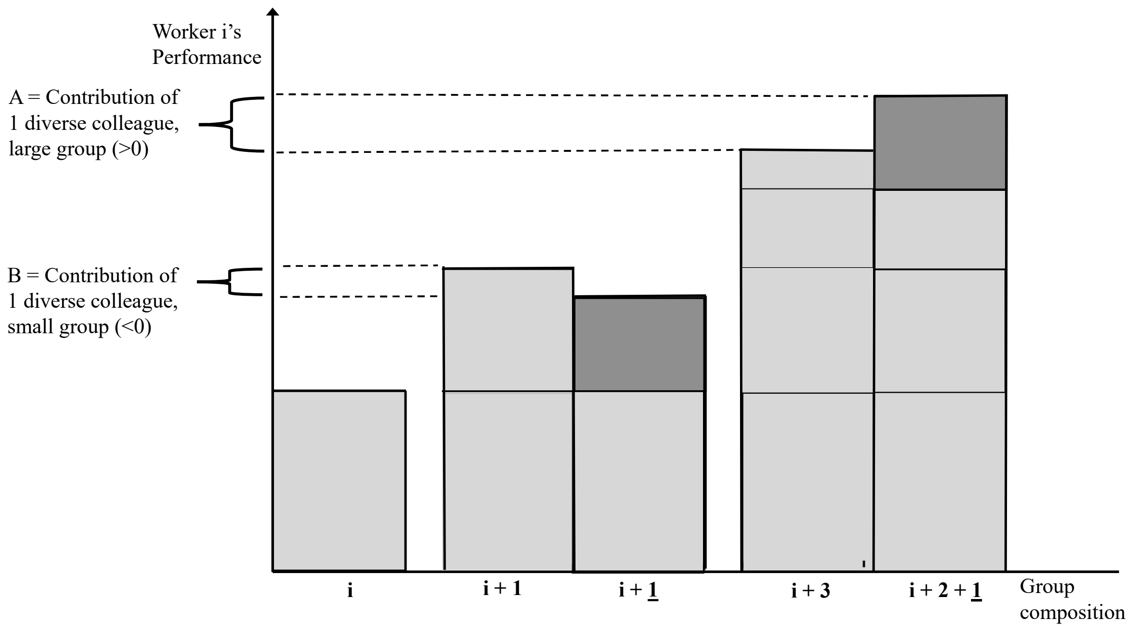
Table 4: Does demographic diversity inhibit the effect of skill diversity?

Dependent Variable:	**Change in score (%)**		
Demographic variable definition:	Religion	Ethnicity	Capital city
	(1)	(2)	(3)
Skill diverse × No demographic diversity (α)	0.831	0.840	2.272
	(1.959)	(2.098)	(2.633)
Skill diverse × Demographic diversity (β)	6.655**	4.760*	1.845
	(2.557)	(2.450)	(2.688)
Skill diverse × No demographic diversity × team size 4 (γ)	7.150*	7.769**	9.874***
	(3.926)	(3.865)	(3.536)
Skill diverse × Demographic diversity × team size 4 (δ)	-4.619	-3.940	-5.465
	(4.209)	(4.024)	(3.786)
Team size 4	0.268	0.466	0.259
	(2.398)	(2.388)	(2.402)
P val: $\alpha = \beta$	0.126	0.316	0.931
P val: $\gamma = \delta$	0.108	0.099	0.019
Mean of Individual Score (%)	47.7	47.6	47.7
Controls	Yes	Yes	Yes
Observations	187	186	187
Cluster	88	88	88

Notes. Change in score is the percentage change in score from the individual to the team test. Test remains the same for each person across the two rounds. Skill diverse team is a dummy variable that equals to one when there is an economics student in the team. Team size 4 is a dummy variable that equals one when there are four students in the team. Demographic diversity is a dummy variable that equals one when the team has any demographic diversity of the type as defined at the top of each column, remains zero otherwise. No demographic diversity is a dummy that equals one when the share of students in any demographic group is either zero or one. All specifications include a control for the team's average baseline performance in the individual test, and a dummy that takes the value of one if the participant attended multiple sessions. Standard errors are clustered at the student, team level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

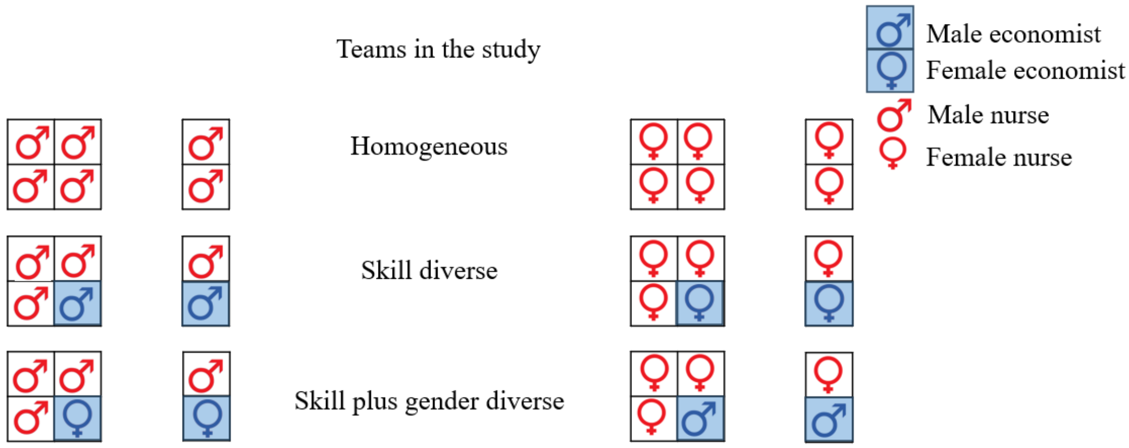
Figures

Figure 1: Illustration of the theoretical model for small teams.



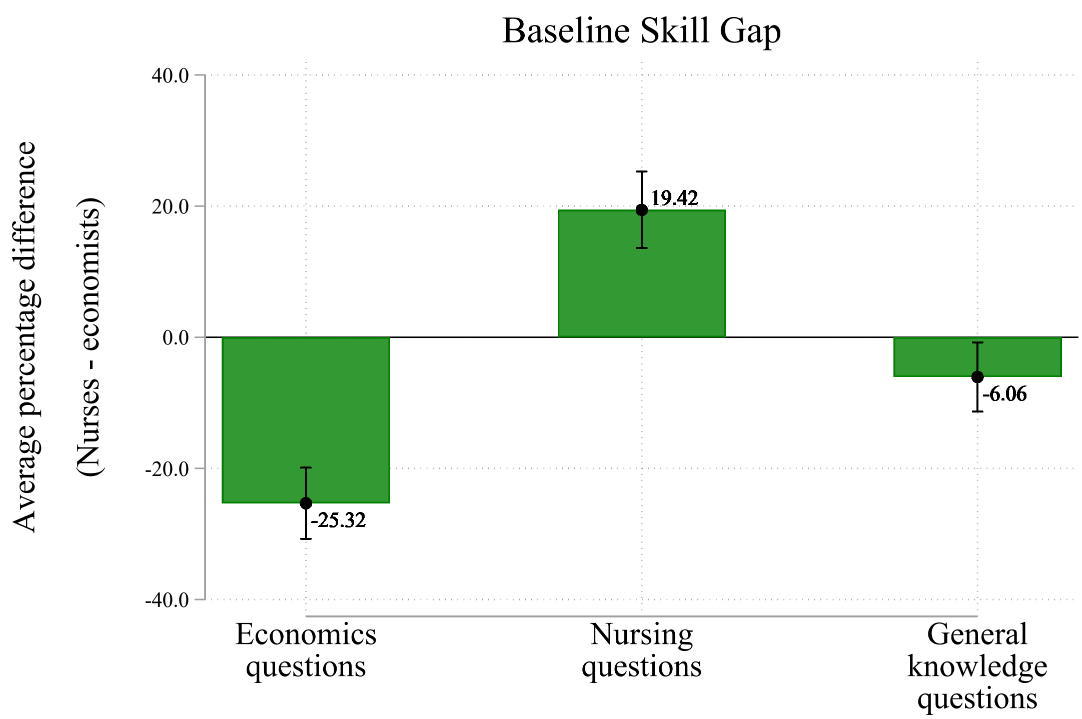
Notes. Each bar in this figure indicates the expected performance (net of communication costs) for an isolated type-1 worker *i* and for the same worker in different team compositions. The number after the plus (+) sign below each column indicate the number of i's teammates. The underlined number indicates a type-2 teammates.

Figure 2: Experimental design



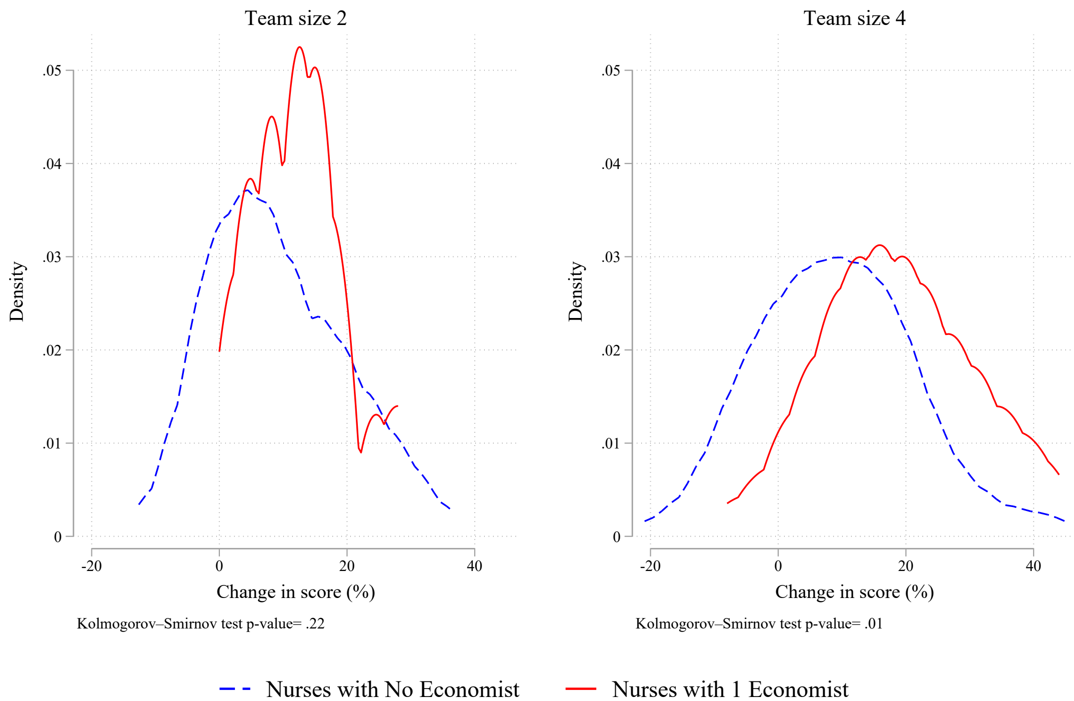
Notes. This figure shows the teams in the experiment. Gender diversity is only introduced in skill-diverse teams.

Figure 3: Skill gap in baseline test between Nurses and Economists



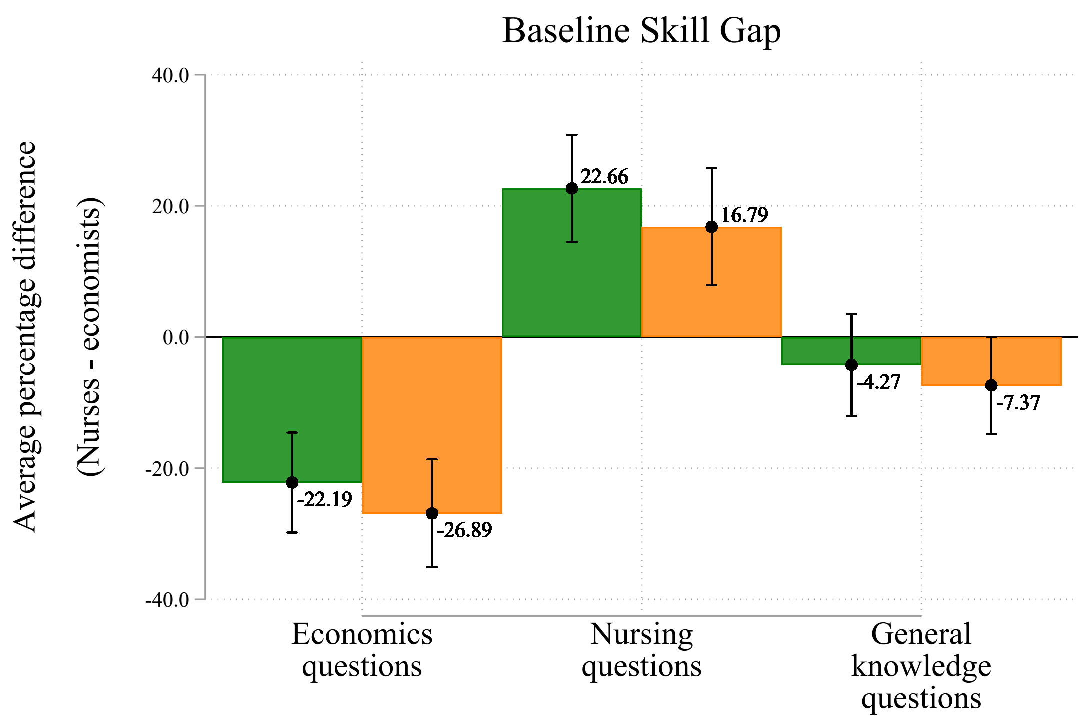
Notes. The figure is based on data from the individual test. No controls were added.

Figure 4: The effects of skill diversity on performance in small and large teams



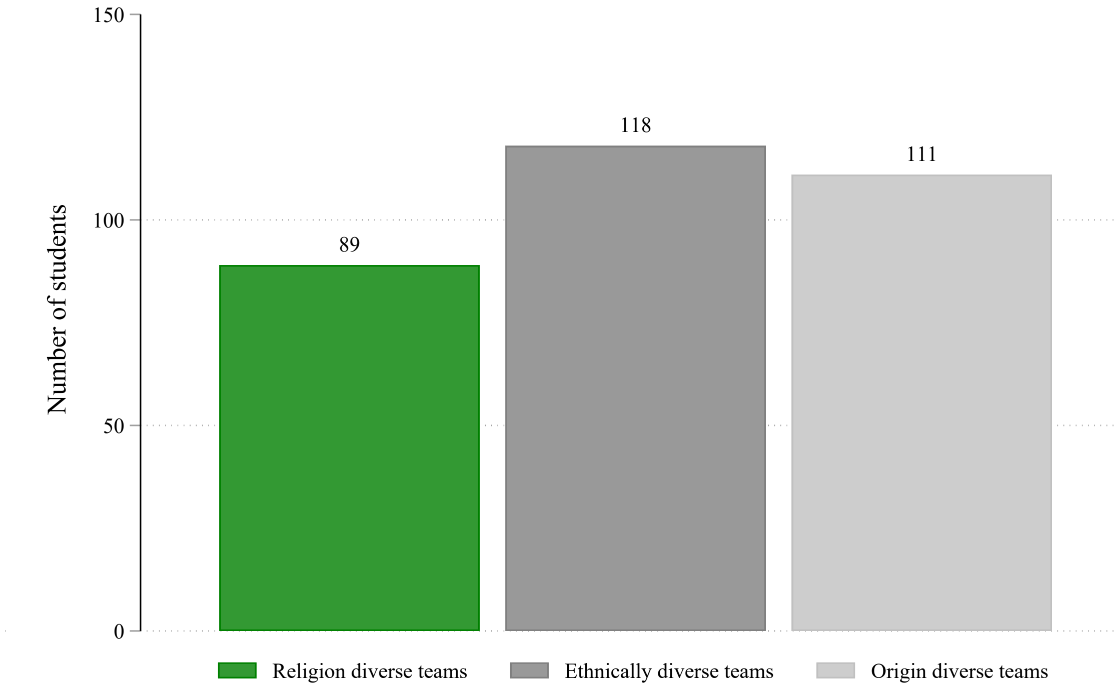
Notes. The effects of skill diversity on the percentage change in performance of Nursing students in small teams of 2 (left-hand side) & large teams of 4 (right-hand side). A Kolmogorov–Smirnov test of the equality of the two distributions has a p-value of 0.22 for teams of size 2, and a p-value of 0.01 for teams of size 4.

Figure 5: Skill gap in baseline test between Nurses and Economists by gender diversity of the team



Notes. The figure is based on data from the individual test. No controls were added. The green bar is the skill gap between the nurses and economist in a gender homogeneous team, while the orange bar is the skill gap between the nurses and economist in a gender diverse team.

Figure 6: Nursing students in teams with different types of demographic diversity



Notes. The figure shows the proportion of nursing students that were in demographically diverse teams of different types. Religious diverse teams are those that include Christians as well as students from other religious backgrounds. Ethnically diverse teams are those that include Balanta as well as students from other ethnicity such as Kasanga, Fulani and Pepel etc. Origin diverse teams are those that include students that are from the capital as well as other areas.

Appendices

A Appendix: Theory extension and proofs

A.1 Proofs

Proposition 1 *The effect of team size on expected performance is concave.*

Proof. Worker i's expected performance in a team of m workers is:

$$E(z_i|m) = pZ_{1,m} - (m-1)c$$

On the other hand, worker i's expected performance in a team of $m+1$ workers is:

$$E(z_i|m+1) = pZ_{1,m+1} - mc$$

Therefore, the gain in performance from having an additional worker in the team is:

$$\begin{aligned}\Delta(m) &= E(z_i|m+1) - E(z_i|m) \\ &= pZ_{1,m+1} - mc - (pZ_{1,m} - (m-1)c) \\ &= p(Z_{1,m+1} - Z_{1,m}) - c\end{aligned}$$

Since $Z_{1,m}$ is increasing and concave, it holds that $Z_{1,m+1} - Z_{1,m}$ decreases as m increases. Thus, $\Delta(m)$ decreases as m increases, which proves that $E(z_i|m)$ is concave. □

Proposition 2 *For a type-1 worker, the gain in expected performance from being placed in a diverse team with k type-2 workers - instead of a same-sized homogeneous team - is increasing in team size.*

Proof. The difference between worker i's performance in the diverse and homogeneous team of size $m+k$ is: $\Delta_k(m) = p(Z_{1,m} - Z_{1,m+k}) + (1-p)Z_{2,k} - kc_o$ On the other hand, the difference between worker i's performance in the diverse and homogeneous team of size $l+k$ is:

$$\Delta_k(l) = p(Z_{1,l} - Z_{1,l+k}) + (1-p)Z_{2,k} - kc_o$$

We have:

$$\begin{aligned}\Delta_k(m) - \Delta_k(l) &= p(Z_{1,m} - Z_{1,m+k}) + (1-p)Z_{2,k} - kc_o - (p(Z_{1,l} - Z_{1,l+k}) + (1-p)Z_{2,k} - kc_o) \\ &= p\,[(Z_{1,m} - Z_{1,m+k}) - (Z_{1,l} - Z_{1,l+k})]\end{aligned}$$

Assuming that we have $m > l$, since $Z_{1,m}$ is concave, it holds that $|Z_{1,m} - Z_{1,m+k}| < |Z_{1,l} - Z_{1,l+k}|$. Since $Z_{1,m} - Z_{1,m+k} < 0$, it holds that $Z_{1,m} - Z_{1,m+k} > Z_{1,l} - Z_{1,l+k}$. Thus, $\Delta_k(m) - \Delta_k(l) > 0$, thereby proving that the relative contribution of k workers is increasing in m. □

A.2 Extension: introducing demographic diversity

This extension of the model considers the case in which the skill diverse team member also belongs to a different demographic group compared to the rest of the team. We apply this extension to the case of gender diversity, but it can easily be applied to other types of demographic diversity like the ones explored in the paper (e.g., religion, ethnicity).

We assume that communication between members of a different gender has an additional cost $c_g(m)$, which we allow to vary with team size m. We make inter-gender communication costs a function of team size to capture the fact that in larger teams one gender may be discouraged from effectively participating in communication exchanges.

We assume that men and women have the same ability distribution within types, thus there is no additional gain or loss from interacting with a person with a diverse skill set who is of the opposite gender. If there are n workers of the opposite gender in the team, then worker i's expected performance is: $E(z_i|m,k,n) = pZ_{1,m} + (1-p)Z_{2,k} - kc_o - nc_g(m)$.

If all type-2 workers are of the opposite gender with respect to type-1 workers, that is $k = n$, the expected performance of worker i in a skill and gender diverse team is:

$$E(z_i|m,k,k) = pZ_{1,m} + (1-p)Z_{2,k} - k(c_o + c_g(m))$$

The following proposition explains how worker's i expected performance changes in gender heterogeneous vs homogeneous teams of different sizes:

Proposition 3. *If $c_g(m)$ is independent of team size, the gain in expected performance for type-1 workers from introducing a same-gender team member should be the same in gender diverse and gender homogeneous teams. If $c_g(m)$ increases (decreases) with team size, then type-1 workers in gender diverse teams benefit less (more) from an increase in team size than those in homogeneous ones.*[17]

In the first case gender diversity does not influence the relationship between team size and skill diversity because it goes hand-in-hand with the costs of communicating with skill-diverse members.[18] This is not the case if the inter-gender costs of communication vary with team size. The intuition is the following: if cost $c_g(m)$ decreases as team size increases, it becomes less costly to communicate with others. As a result, not only do type-1 workers enjoy a gain in performance due to the larger team size, but they also enjoy a reduction in communication cost and can spend more time discussing questions that matter rather than solving communication issues. Since gender homogeneous teams do not enjoy any cost reduction from this increase in team size, the gain in performance is larger in gender diverse teams than gender homogeneous ones. On the other hand, if cost $c_g(m)$ increases as team size increases, then expanding team size also entails additional costs. As a result, because it is more costly to communicate with others, performance in gender-diverse teams will increase less in team size than in gender-homogeneous ones.

Proof. $c_g(m)$ **is independent from team size**

This implies that c_g is a constant. The change in expected performance from having an addi-

[17]One could also assume that the cost c_o of inter-type communication is a function of team size. A positive correlation with size would weaken the result in proposition 2, while a negative correlation would strengthen it. We leave this aside and focus on the empirically-relevant case of gender-related costs.

[18]The change in i's expected performance in a mixed-gender team is: $\Delta_k(m) = E(z_i|m+1,k,k) - E(z_i|m,k,k) = p[Z_{1,m+1} - Z_{1,m}]$. When the team is gender homogeneous ($n = 0$): $\Delta_0(m) = E(z_i|m+1,k,0) - E(z_i|m,k,0) = p[Z_{1,m+1} - Z_{1,m}]$.

tional type-1 worker of same gender as worker i in a gender diverse team is:

$$\begin{aligned}\Delta_k(m) &= E(z_i|m+1,k,k) - E(z_i|m,k,k)\\ &= pZ_{1,m+1} + (1-p)Z_{2,k} - k(c_o + c_g) - (pZ_{1,m} + (1-p)Z_{2,k} - k(c_o + c_g))\\ &= p(Z_{1,m+1} - Z_{1,m})\end{aligned}$$

On the other hand, we can obtain the change in expected performance from the same addition in a gender homogeneous team;

$$\begin{aligned}\Delta_0(m) &= E(z_i|m+1,k,0) - E(z_i|m,k,0)\\ &= p(Z_{1,m+1} - Z_{1,m})\end{aligned}$$

The change in expected performance from having an additional type-2 worker of same gender as worker i in a gender diverse team is:

$$\begin{aligned}\Delta_k(k) &= E(z_i|m,k+1,k) - E(z_i|m,k,k)\\ &= pZ_{1,m} + (1-p)Z_{2,k+1} - (k+1)c_o - kc_g - (pZ_{1,m} + (1-p)Z_{2,k} - k(c_o + c_g))\\ &= (1-p)(Z_{2,k+1} - Z_{2,k}) - (k+1-k)c_o\\ &= (1-p)(Z_{2,k+1} - Z_{2,k}) - c_o\end{aligned}$$

Again, we can obtain the change in expected performance from the same addition in a gender homogeneous team; $\Delta_0(k) = (1-p)(Z_{2,k+1} - Z_{2,k}) - c_o$ Therefore, it holds $\Delta_k(m) = \Delta_0(m)$ and $\Delta_k(k) = \Delta_0(k)$, that is that the change in expected performance from having an additional same-gender worker should be the same in both gender diverse and homogeneous teams.

$c_g(m)$ **increases with team size (**$c'_g > 0$**)**

The change in expected performance from having an additional type-1 worker of same gender as worker i in a gender diverse team is:

$$\begin{aligned}\Delta_k(m) &= E(z_i|m+1,k,k) - E(z_i|m,k,k)\\ &= pZ_{1,m+1} + (1-p)Z_{2,k} - k(c_o + c_g(m+k+1)) - (pZ_{1,m} + (1-p)Z_{2,k} - k(c_o + c_g(m+k)))\\ &= p(Z_{1,m+1} - Z_{1,m}) - k(c_g(m+k+1) - c_g(m+k))\end{aligned}$$

From there, we can obtain the change in expected performance following the same addition in a gender homogeneous team;

$$\Delta_0(m) = p(Z_{1,m+1} - Z_{1,m})$$

As a result, the difference in change in gender diverse and homogeneous teams is:

$$\begin{aligned}\Delta_m(k) &= \Delta_k(m) - \Delta_0(m)\\ &= p(Z_{1,m+1} - Z_{1,m}) - k(c_g(m+k+1) - c_g(m+k)) - p(Z_{1,m+1} - Z_{1,m})\\ &= -k(c_g(m+k+1) - c_g(m+k))\end{aligned}$$

The change in expected performance from having an additional type-2 worker of same gender as worker i in a gender diverse team is:

$$\begin{aligned}
\Delta_k(k) &= E(z_i|m,k+1,k) - E(z_i|m,k,k) \\
&= pZ_{1,m} + (1-p)Z_{2,k+1} - (k+1)c_o - kc_g(m+k+1) - (pZ_{1,m} + (1-p)Z_{2,k} - k(c_o + c_g(m+k))) \\
&= (1-p)(Z_{2,k+1} - Z_{2,k}) - c_o - k(c_g(m+k+1) - c_g(m+k))
\end{aligned}$$

From there, we can obtain the change in expected performance following the same addition in a gender homogeneous team;

$$\Delta_0(k) = (1-p)(Z_{2,k+1} - Z_{2,k}) - c_o$$

As a result, the difference in change in gender diverse and homogeneous teams is:

$$\begin{aligned}
\Delta_k(k) &= \Delta_k(k) - \Delta_0(k) \\
&= (1-p)(Z_{2,k+1} - Z_{2,k}) - c_o - k(c_g(m+k+1) - c_g(m+k)) \\
&- ((1-p)(Z_{2,k+1} - Z_{2,k}) - c_o) \\
&= -k(c_g(m+k+1) - c_g(m+k))
\end{aligned}$$

Thus, we see that if $c_g(.)$ increases with team size, we have $\Delta_m(k) < 0$ and $\Delta_k(k) < 0$, that is the increase in expected performance from an extra same-gender colleague is greater in gender homogeneous teams than gender diverse ones. The opposite holds if $c_g(.)$ decreases with team size. □

B Appendix tables

Table B.1: Whether repeated participation in the experiment is correlated with treatments

Dependent Variable:	Whether students participated in the experiment more than once
	(1)
Skill diverse team (α)	-0.0915
	(0.0977)
Skill diverse team × Team size 4 (β)	0.0315
	(0.137)
Team size 4 (γ)	0.0630
	(0.110)
Mean of Individual Score (%)	47.7
Controls	Yes
Observations	187
Cluster	88

Notes. The outcome is a dummy variable that takes the value of one if a student participated in the experiment more than once. 23 students (19 nurses and 4 economists) participated in the study in more than one session. Skill diverse team is a dummy variable that equals to one when there is an economics student in the team. Team size 4 is a dummy variable that equals one when there are four students in the team. All specifications include a control for the team's average baseline performance in the individual test and a dummy variable that takes the value 1 if the participant is a male. Standard errors clustered at study team levels. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table B.2: Skill diversity and team size: robustness to different specifications that take into account repeated participation in the experiment

Dependent Variable:	**Change in score (%)**			
	Keeping first session for all		Cluster at student, team level	Control for repeat students
	(1)	(2)	(3)	(4)
Skill diverse team (α)	3.111	1.987	3.943*	3.505*
	(1.935)	(2.097)	(2.041)	(1.888)
Skill diverse team × Team size 4 (β)	5.289	6.037*	5.968*	6.119**
	(3.207)	(3.380)	(3.070)	(2.971)
Team size 4 (γ)	0.854	0.836	0.318	0.619
	(2.368)	(2.576)	(2.326)	(2.171)
P val: Diversity + Diversity × Team 4 = 0	0.001	0.002	0.000	0.000
P val: Team 4 + Diversity × Team 4 = 0	0.005	0.002	0.002	0.001
P val: $\alpha = \gamma$	0.326	0.621	0.104	0.161
Mean of Individual Score (%)	47.3	47.3	47.7	47.7
Basic Controls	Yes	Yes	Yes	Yes
PDS Lasso Controls	No	Yes	No	No
Observations	168	168	187	187
Cluster	84	84	88	88

Notes. Change in score is the percentage change in score from the individual to the team test. Test remains the same for each person across the two rounds. Skill diverse team is a dummy variable that equals to one when there is an economics student in the team. Team size 4 is a dummy variable that equals one when there are four students in the team. All specifications include a control for the team's average baseline performance in the individual test, and a dummy variable that takes the value one when the participant is a male student. Column (2) contains additional controls chosen using PDS Lasso. Column (4) includes as an additional control a dummy that takes the value of one if the participant attended multiple sessions. In Column (3) standard errors clustered at student, team level. Standard errors are clustered at the study team level in the rest of the columns. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table B.3: Gender diversity and skill diversity: robustness to different specifications that take into account repeated participation in the experiment

Dependent Variable:	**Change in score (%)**			
	Keeping first session for all		Cluster at student, team level	Control for repeat students
	(1)	(2)	(3)	(4)
Skill diverse × gender homogeneous (α)	0.612	-0.491	1.569	0.767
	(1.982)	(2.047)	(2.085)	(2.006)
Skill diverse × gender diverse (β)	5.307**	3.967	5.832**	5.634**
	(2.433)	(2.596)	(2.495)	(2.278)
Skill diverse × gender homogeneous × team size 4 (γ)	6.762*	6.814*	7.313*	7.314*
	(3.935)	(3.911)	(3.700)	(3.736)
Skill diverse × gender diverse × team size 4 (δ)	4.078	5.542	5.037	5.376*
	(3.626)	(3.909)	(3.459)	(3.212)
Team size 4	0.879	1.182	0.327	0.667
	(2.394)	(2.653)	(2.342)	(2.172)
P val: $\alpha = \beta$	0.055	0.062	0.077	0.044
P val: $\gamma = \delta$	0.509	0.751	0.542	0.606
Mean of Individual Score (%)	47.3	47.3	47.7	47.7
Basic Controls	Yes	Yes	Yes	Yes
PDS Lasso Controls	No	Yes	No	No
Observations	168	168	187	187
Cluster	84	84	88	88

Notes. Change in score is the percentage change in score from the individual to the team test. Test remains the same for each person across the two rounds. Skill diverse team is a dummy variable that equals to one when there is an economics student in the team. Team size 4 is a dummy variable that equals one when there are four students in the team. Gender diverse is a dummy that takes the value one if the economist is of a different gender from the nurses. Gender homogeneous is a dummy that takes the value one if the economist is of the same gender as the nurses. All specifications include a control for the team's average baseline performance in the individual test and a dummy for being a male student. Column (2) contains additional controls chosen using PDS Lasso. Column (4) includes as an additional control a dummy that takes the value of one if the participant attended multiple sessions. In Column (3) standard errors clustered at student, team level. Standard errors are clustered at the study team level in the rest of the columns. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table B.4: Heterogeneity of treatment effects by students' demographic characteristics

Dependent Variable:	**Change in score (%)**							
Demographic Variable Definition:	Gender		Religion		Ethnicity		Origin	
	Male	Female	Christian	Non Christian	Major ethnicity	Minor ethnicity	Capital born	Not capital born
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Skill diverse team (α)	3.243	1.796	2.606	3.513	6.918*	0.570	4.030	3.005
	(2.211)	(2.957)	(2.340)	(4.688)	(3.637)	(2.127)	(4.103)	(2.201)
Skill diverse team × Team size 4 (β)	6.962*	6.450	6.307*	5.688	4.861	6.917	2.623	7.227**
	(3.538)	(4.646)	(3.475)	(4.977)	(4.650)	(4.207)	(6.468)	(3.218)
Team size 4 (γ)	-0.703	0.968	-0.642	3.204	-1.802	1.473	2.441	-0.047
	(2.723)	(3.399)	(2.743)	(3.558)	(3.812)	(3.109)	(5.135)	(2.502)
P val: Diversity + Diversity × Team 4 = 0	0.001	0.017	0.001	0.004	0.002	0.028	0.177	0.000
P val: Team 4 + Diversity × Team 4 = 0	0.025	0.015	0.011	0.095	0.370	0.002	0.244	0.001
P val: $\alpha = \gamma$	0.123	0.802	0.202	0.929	0.008	0.773	0.721	0.237
P val: $\alpha_{majority} = \alpha_{minority}$		0.853		0.977		0.00670		0.619
P val: $\beta_{majority} = \beta_{minority}$		0.731		0.666		0.671		0.316
P val: $\gamma_{majority} = \gamma_{minority}$		0.709		0.257		0.511		0.583
Mean of Individual Score (%)	51.6	43.8	47.4	48.2	48.4	47.2	47.1	47.8
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	93	94	149	34	65	120	54	132
Cluster	45	44	82	29	49	69	41	75

Notes. Change in score is the percentage change in score from the individual to the team test. Test remains the same for each person across the two rounds. Skill diverse team is a dummy variable that equals to one when there is an economics student in the team. Team size 4 is a dummy variable that equals one when there are four students in the team. Ethnic majority is Balanta, while minorities includes many different ethnicities such as Kasanga, Fulani and Pepel etc. All specifications include a control for the team's average baseline performance in the individual test, and a dummy that takes the value of one if the participant attended multiple sessions. Standard errors are clustered at the student, team level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table B.5: Team level estimates: When is skill diversity most beneficial?

Dependent Variable:	**Change in score (%)**
	(1)
Skill diverse team (α)	2.235
	(1.701)
Skill diverse team × Team size 4 (β)	5.592**
	(2.690)
Team size 4 (γ)	0.656
	(2.237)
P val: Diversity + Diversity × Team 4 = 0	0.000
P val: Team 4 + Diversity × Team 4 = 0	0.000
P val: $\alpha = \gamma$	0.432
Mean of score (individual test)	12.2
Controls	Yes
Observations	88

Notes. Change in score is the percentage change in score from the individual to the team test. Test remains the same for each person across the two rounds. Skill diverse team is a dummy variable that equals to one when there is an economics student in the team. Team size 4 is a dummy variable that equals one when there are four students in the team. All specifications include a control for the team's average baseline performance in the individual test, gender of the team, and a team level average of whether participants took part in the study twice. Standard errors are clustered at the team level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

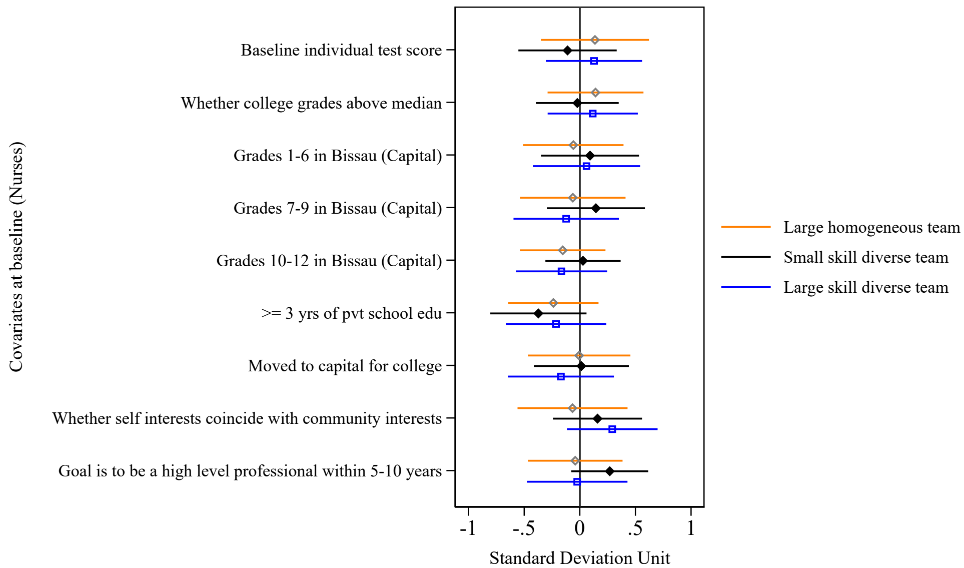
Table B.6: Skill diversity and team size (robustness including controls for gender and skill composition of the room)

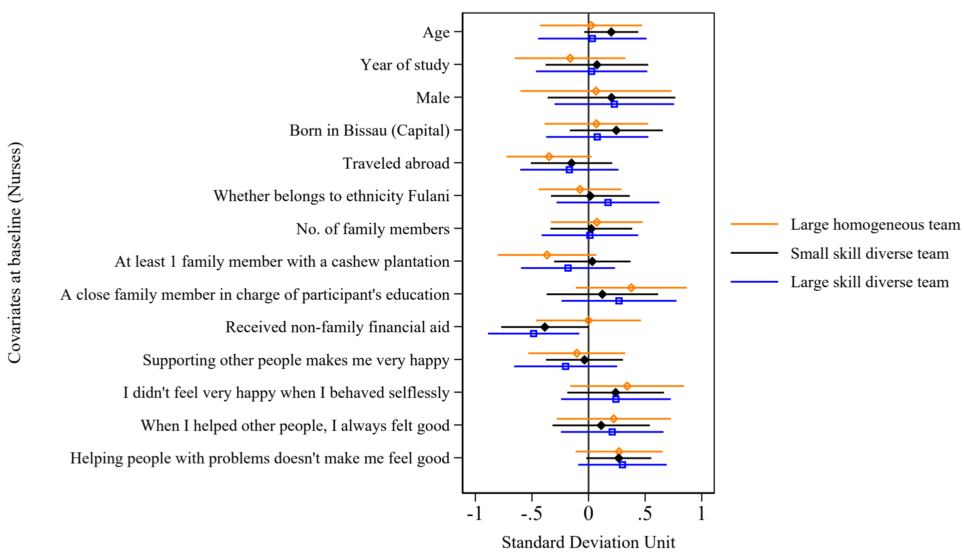
Dependent Variable:	**Change in score (%)**		
	(1)	(2)	(3)
Skill diverse team (α)	7.814***		3.894**
	(1.462)		(1.806)
Skill diverse team × Team size 4 (β)			5.883**
			(2.757)
Team size 4 (γ)		4.656***	0.875
		(1.649)	(1.948)
P val: Diversity + Diversity × Team 4 = 0			0.000
P val: Team 4 + Diversity × Team 4 = 0			0.001
P val: $\alpha = \gamma$			0.131
Mean of Individual Score (%)	47.7	47.7	47.7
Controls	Yes	Yes	Yes
Observations	187	187	187
Cluster	88	88	88

Notes. Change in score is the percentage change in score from the individual to the team test. Test remains the same for each person across the two rounds. Skill diverse team is a dummy variable that equals to one when there is an economics student in the team. Team size 4 is a dummy variable that equals one when there are four students in the team. All specifications include a control for the team's average baseline performance in the individual test, a dummy variable that takes the value one when the participant is a male student, a dummy that takes the value of one if the participant attended multiple sessions, and the average share of men in the room and the average skill of the students in the room. Standard errors are clustered at the student, team level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

C Appendix Figures

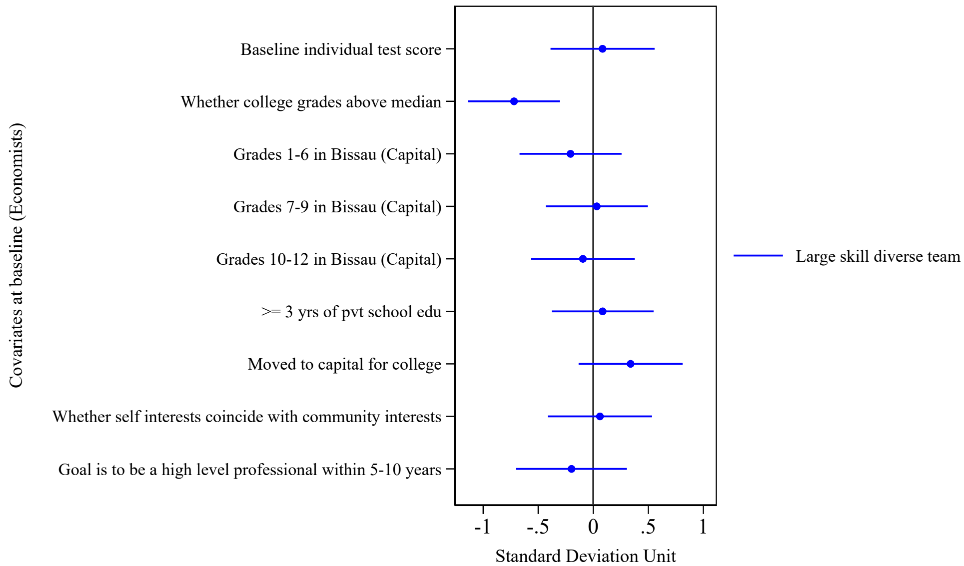
Figure C.1: Balance on characteristics of nurses in different types of teams

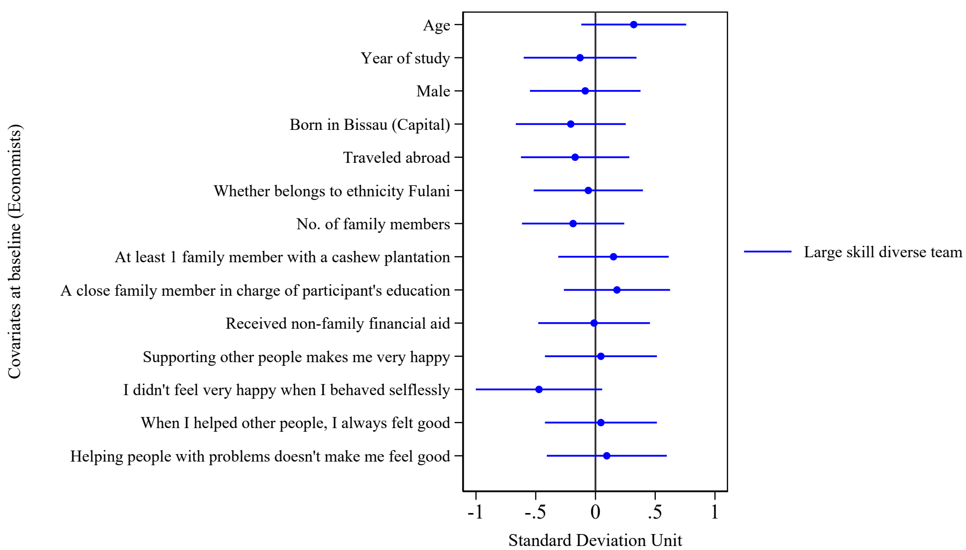




Notes. The figure is based on two sets of data: an endline survey of the participants and their performance in the baseline (individual) test. The sample keeps the first observation of those participants that took part in the study twice. The coefficients from a regression of the covariates on dummy variables for different types of teams are reported, alongwith the 90% confidence intervals. Whether the student received any financial aid is imbalanced.

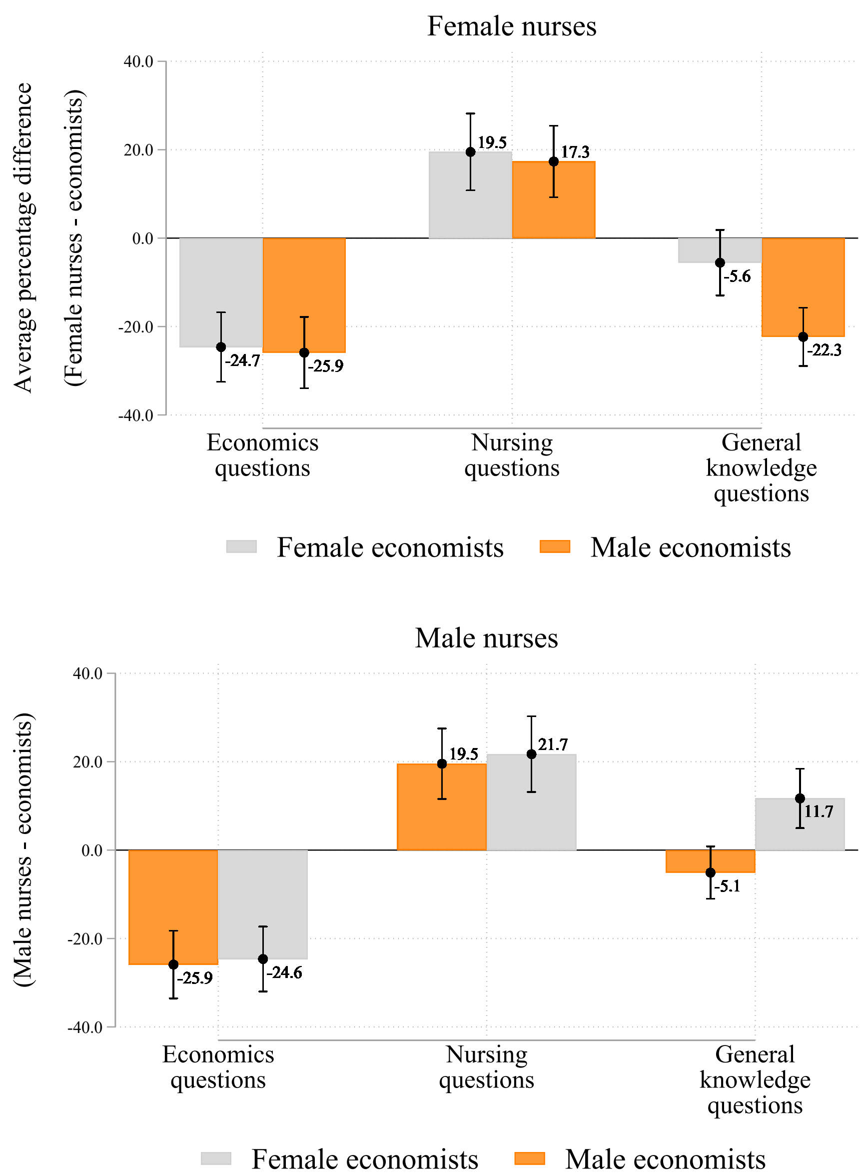
Figure C.2: Balance on characteristics of economists in different types of teams





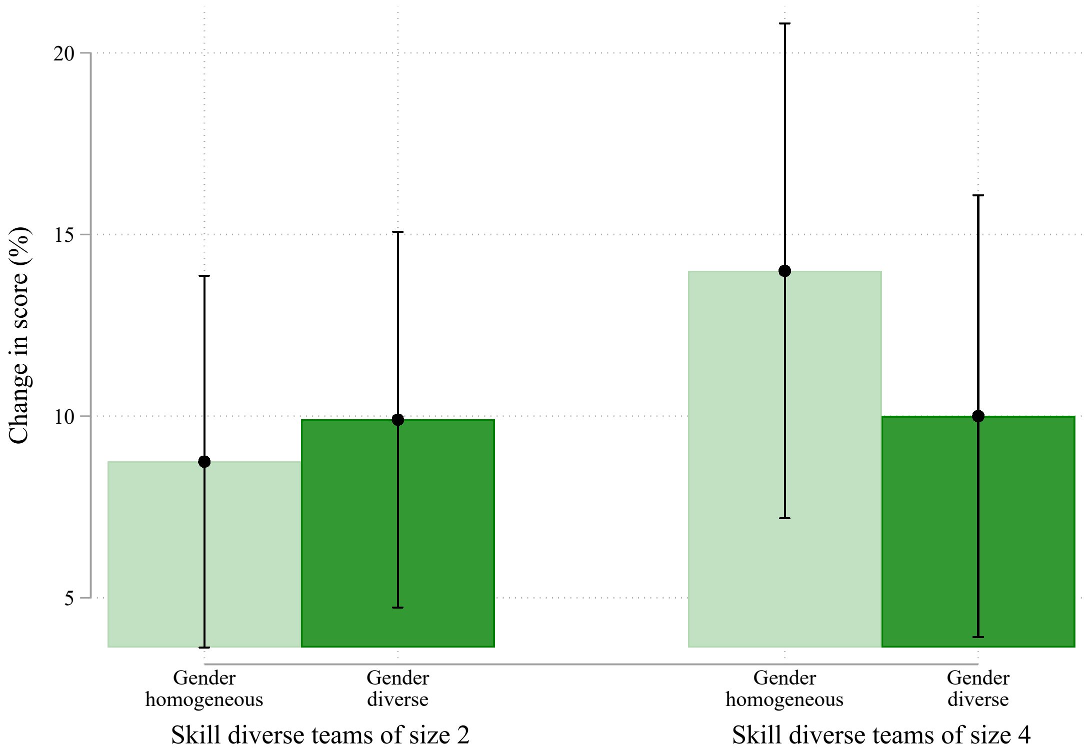
Notes. The figure is based on two sets of data: an endline survey of the participants and their performance in the baseline (individual) test. The sample keeps the first observation of those participants that took part in the study twice. The coefficients from a regression of the covariates on dummy variables for different types of teams are reported, alongwith the 90% confidence intervals. Economists are only part of skill diverse teams, thus balance can only be checked for their assignment to small or large teams. A dummy variable for whether college grades were above median is imbalanced.

Figure C.3: Skill gap in baseline test between Nurses and Economists by gender



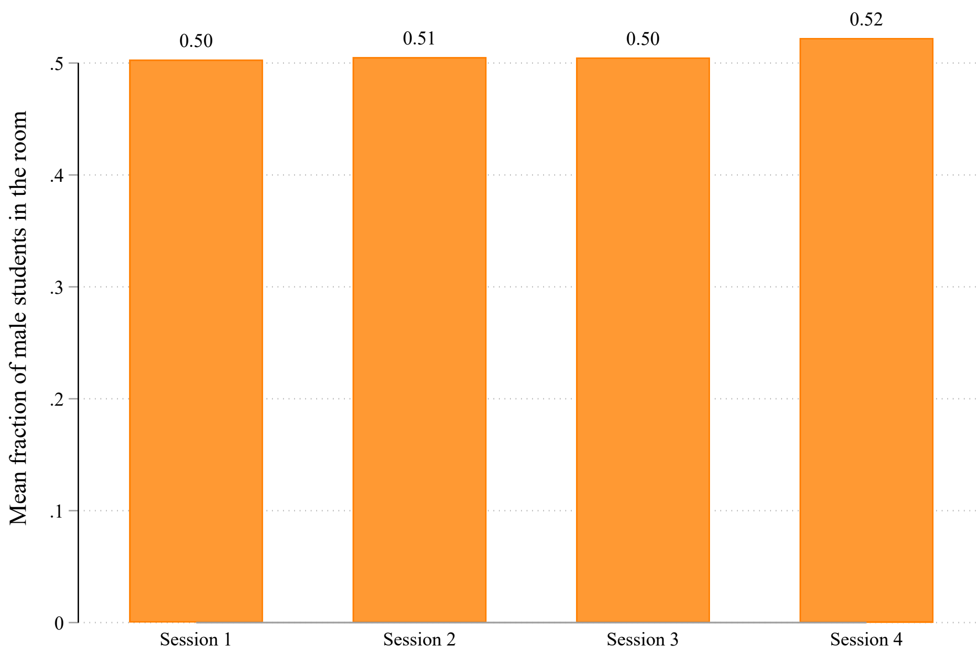
Notes. The figure is based on data from the individual test. No controls were added.

Figure C.4: Performance of minority students (economists) in different types of teams



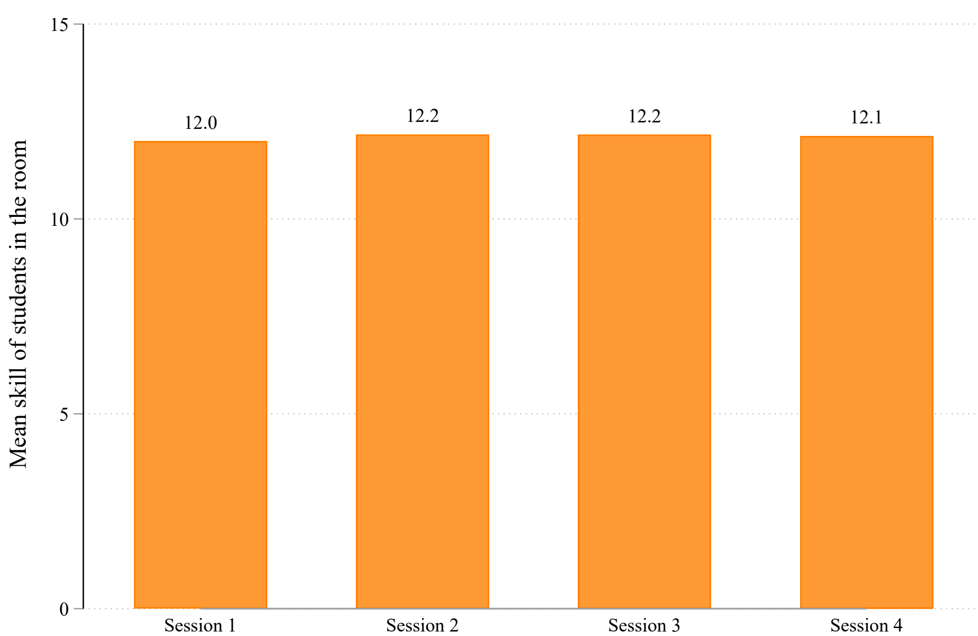
Notes. The figure is based on data from the performance of economists. Change in score is the percentage change in score from the individual to the team test. Test remains the same for each person across the two rounds.

Figure C.5: Gender composition of the room across different experimental sessions



Notes. The figure shows the average share of male students across rooms in different experimental sessions.

Figure C.6: Skill composition of the room across different experimental sessions



Notes. The figure shows the average skills of students across rooms in different experimental sessions. Skills are defined as their performance in the baseline (individual) test.

D Appendix: Recruitment and randomization

Recruitment All the students were recruited from the Bachelor's degrees in Economics and Nursing at the Universidade Lusófona da Guiné-Bissau (ULG). We excluded first year students because of the expected lower familiarity with their respective discipline. We chose students from these degrees for two reasons. First, Nursing was the degree with most students at the university, and the only one with enough female and male students to guarantee enough participants for the study. Economics students were chosen as the minority because the skill set of economists was disjoint from the skill set of nurses, suggesting a meaningful complementarity of skills in teams.[19] Of the 1104 registered students 503 signed up as available to participate in the experiment (387 Nursing and 116 Economics students). From this set, we randomly chose an identical number of female and male students from each degree and invited them to come on the day of the experiment. A total of 242 students participated in the study (62 economists and 180 nurses). The experiment took place in four sessions, within a single day, at the university campus. Each session was planned such that it would have at least one team of each type.

Payment After the main task, students completed a survey that asked them about their demographic and other information. Participants were incentivized to complete the survey by paying them 500 FCFA if the survey was completed. In addition to the above, each student received 500 FCFA (0.72$) for participating in the study, independently of performance.

The tests and survey were paper-based. Room supervisors were instructed to place students far enough from each other in the individual and team-based test, so that it was not possible for them to cheat or hear what others were discussing. No incidents were reported.[20] The tests were corrected on the spot by a team of graders in a separate room. Payments were made right after the experiment and participants received their payments privately in a room with no other participants. Grading of each test, including which questions participants got right and the right response for each question, were not revealed to anyone.

Randomization After consent and registration but before each session, participants were invited to wait in a room where all the participants for each session were randomly assigned to their teams at the same time. The mechanism for random assignment to teams was as follows. First, a unique ID with room and team assignment was created for each participant. This was printed on cards and placed in a bag. Since we wanted to stratify by gender of nurses and also randomize the gender of economists, we had to find a way to keep track of gender of participants but without making gender salient. To minimize gender salience, gender was never mentioned throughout the experiment. We asked participants (in no particular order and not by gender) to pick up their card from a bag one by one. Women and men chose their cards from different (but identically-looking) bags. Students were then guided towards their respective rooms as mentioned on their cards. A room supervisor then read out loud the instructions of the experiment. After finishing and

[19]In the recruitment announcement, it was stated that each participant would solve some simple tests for 2 hours and thirty minutes session. Students in both degrees were taking in-person exams in the weeks before the experiment. We took advantage of this opportunity to present the study to them and to register their interest and availability to participate in it.

[20]The rules were strict. A participant would be disqualified without any payment were she to break the rules, including talking to other participants (outside her team) or using her cellphone. In the team test the whole team would be disqualified if a single one of its members cheated.

handing-in the first test to their room supervisor, students were organized into teams as indicated by their ID card that they picked before the start of the experiment.

To clarify, the treatment variation was introduced within each session. Each session included at least one small team comprising either nurses paired with other nurses or nurses paired with economists. Similarly, each session also included large teams comprising of all four nurses or three nurses and an economist. Figure C.5 and C.6 shows the gender and skill composition of the room across different experimental sessions. We see that these characteristics were balanced across the experimental sessions. The rooms where the experiment was conducted were large, allowing us to seat students far from each other. Moreover, the presence of both treated and control teams in the same room suggests that any audience effect might be similar for the treated and control teams.

E Appendix: For Online Publication

E.1 Instructions to the participants - translated from Portuguese

Instructions for the supervisor – do not read to participants:
Wait for the participants of the session to arrive in the room: make sure that the correct room is indicated on the students' placards (green for economics and pink for nursing).
Make sure that all participants are seated as far apart as possible.
Complete the record sheet.
Read the instructions to the participants.

Instructions for the participants – read aloud:
Good morning! In this room we are going to perform 3 types of tests:

- The first will be individual and will have 25 questions.
- The second will be done in groups and will have 25 questions.
- Each question has only one correct answer.
- If a question is answered with more than one option, or with the answer chosen in a different way from that indicated in the document, the answer will be considered wrong.
- The third test consists of a survey to be completed with personal information.

The results of these tests will be used in an independent study and will not be handed over to the University, so it has no consequence on your academic results. All the information you provide will be completely confidential.
You will have a fixed payment per person of 500 FCFA, regardless of the test result.
Each of you will have to answer different questions: some are from the subject of your course of study, others from the subject of other colleagues participating in the activity, and some are general knowledge questions.
The first test will be taken individually, and you will be paid 80 francs per correct answer, up to a maximum of 2000 FCFA. The available time is 30 minutes. I will let you know when there are 10 minutes left until the time is over. When I announce that the time is up, everyone should stop writing, even if you haven't finished answering all the questions on the test. Then I will collect the sheets. You may not use phones or talk to colleagues; if anyone is caught breaking the rules, they will be expelled, and will forfeit their right to financial compensation. If anyone has any questions (but not about the content of the questions), you must call me and ask me in a low voice, and I will decide whether the question is pertinent and relevant to everyone or not. If that is the case, I will communicate the information to the whole room.

Instructions for the supervisor – do not read to participants:
Make sure all participants have their phones switched off.
Deliver the individual tests, different for nursing and economics students.
Before starting the test, all participants must write on pages 1 and 3 of their tests their university registration number and study ID number (which they find on their test papers). When everyone has finished recording their registration number and ID, the supervisor should start counting the 30 minutes for solving the individual test (register the starting time on the session sheet).

Caution! DO NOT SAY THAT IT IS BETTER TO ANSWER A QUESTION RANDOMLY IF THEY DO NOT KNOW THE ANSWER.
Warn students when there are 10 minutes left before the end of the test.
Collect individual tests. Record the time of completion on the sheet. Hand in tests to Jair (or his substitute) who will bring in the group tests.
Organize the room into groups; the members of each group must sit at the corresponding table. DO NOT GIVE INDICATIONS ABOUT THE DESIRABILITY OF MORE OR LESS COLLABORATION BETWEEN MEMBERS OF THE SAME GROUP IN THE GROUP TEST.
Read the instructions to the participants.

Instructions for the participants – read aloud:
Now the same test will be repeated in groups. Each person must complete his or her own personal sheet with the same questions, but in this test, you can communicate ONLY with your group mate(s). At the end of the group test, only one test of one of the group members will be randomly chosen and marked. All members of the group will be paid according to the number of correct answers in this test (80 francs per correct question, up to a maximum of 2000 FCFA). The time is 30 minutes. I will always tell you when there are 10 minutes left to go. When I announce that time is up, you must stop writing, even if you have not finished answering all the questions, and I will collect the sheets. You may not use phones or talk to colleagues from other groups or speak with a loud voice. Classmates from other groups in the same room should not be able to hear what the other groups are discussing. If one of you is found breaking the rules, all members of the group to which you belong will be expelled, without financial compensation. If anyone has a doubt (but this cannot be about the content of the questions), you must call me and ask me in a low voice, and I will decide whether the question is pertinent and relevant to everyone or not. If that is the case, I will communicate the information to the whole room.

Instructions for the supervisor – do not read to participants:
Deliver the tests in groups, with these being different for nursing and economics students.
Before starting the test, all participants must write on pages 1 and 3 of their tests their university registration number and study ID number (which they find on their test papers). When everyone has finished their registration, the 30 minutes to complete the test begin (register the starting time on the session sheet).
DO NOT SAY THAT IT IS BETTER TO ANSWER A QUESTION RANDOMLY IF THEY DO NOT KNOW THE ANSWER.
Warn students when there are 10 minutes left before the end of the test. Collect the group tests. Record the time of completion on the sheet. Hand in tests to Jair (or substitute) who will bring in the surveys.
Read the instructions.

Instructions for the participants – read aloud:
You will now complete a survey with personal information. You will have a payment of 500 francs if you complete the whole survey, in addition to the previous payments. You have 25 minutes to complete the survey. I will let you know when 5 minutes are left. When you have finished completing the survey, you may hand it in and go to the auditorium, where you will be called for your payments.

Instructions for the supervisor – do not read to participants:

Deliver the surveys, the same for all participants.

Before starting the test, all participants must write on pages 1 and 3 of their tests their university registration number and study ID number (which they find on their test papers).

When everyone has finished registering, start the 25 minutes to complete the survey (register the starting time on the session sheet). Help participants if they have any questions.

Warn the participants when there are 5 minutes to go.

When the participants have finished completing the surveys, they can leave. Record the finishing time of the last participant on the sheet. When everyone has finished, take the surveys and the personal sheet to Lucia. Then the supervisors should go to the auditorium for the end of work evaluation meeting.

E.2 Examples of questions from the experiment - translated from Portuguese

We included 120 different questions in the experiment: 40 general-knowledge questions were designed by professors at the partner university and the Ministry of Finance of Guinea-Bissau. Nursing questions and Economics questions were designed by professors from the respective departments. All questions were carefully checked for accuracy and then randomized into the different sessions and tests. As every question was only included in one session, no student saw or heard about the questions she or he faced in the experiment before taking the first test.

Examples for each group of questions include:

1. General-knowledge:

   - How many countries are there in Africa?
     (a) 72
     (b) 4
     (c) 46
     (d) 54
     (e) none of the previous answers
   - What is a water molecule composed of?
     (a) one hydrogen atom and one oxygen atom
     (b) two hydrogen atoms and one oxygen atom
     (c) one hydrogen atom and two oxygen atoms
     (d) two hydrogen atoms and two oxygen atoms
     (e) none of the above
   - Which of the following animals is impossible to find in Guinea-Bissau?
     (a) tortoise
     (b) chimpanzee
     (c) shark
     (d) giraffe
     (e) manatee

2. Economics:

   - The amount of currency in circulation in the economy is determined by...
     (a) the central bank
     (b) commercial banks
     (c) the General Direction for Taxes
     (d) financial markets
     (e) fractional reserve systems
   - The interest rate without discounting inflation is the...
     (a) real interest rate
     (b) nominal interest rate

(c) re-discount rate
(d) the tax pressure rate
(e) the discount rate

- The stock of equipment and structures used to produce goods and services is called...
  (a) technical knowledge
  (b) physical capital
  (c) innovation
  (d) current assets
  (e) human capital

3. Nursing:

- The following are factors that can produce pressure ulcers...
  (a) poor circulation and diabetes
  (b) urinary incontinence
  (c) generalized edema and paralysis
  (d) the three alternatives are correct
  (e) the three alternatives are incorrect
- The following is a basic measure for the control of contagious diseases...
  (a) vaccinate contacts
  (b) develop lectures
  (c) treat sources of infection
  (d) carry out immunological tests
  (e) carry out scientific research
- We consider "universal receiver" someone with the following blood type...
  (a) A-
  (b) B+
  (c) AB+
  (d) O-
  (e) the four alternatives are correct