




Department of
Social Policy

Working Paper
November 2022



Effects of team diversity on performance, perceptions, and predictions:

Experimental evidence of gender composition
and language

Valentina Contreras^a, Chiara Orsini^{b,a}, Berkay Ozcan^a, and Johann Koehler^a

Department of Social Policy
London School of Economics and
Political Science Houghton Street
London WC2A 2AE

Email: soc.pol.webteam@lse.ac.uk

Telephone: +44 (0)20 7955 6001

lse.ac.uk/social-policy



Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

To cite this paper: Contreras, V., Orsini, C., Ozcan, B., Koehler, J. (2022) Effects of team diversity on performance, perceptions, and predictions: Experimental evidence of gender composition and language, Social Policy Working Paper 06-22, London: LSE Department of Social Policy.

Effects of team diversity on performance, perceptions, and predictions: Experimental evidence of gender composition and language

Valentina Contreras^a, Chiara Orsini^{b,a}, Berkay Ozcan^a, and Johann Koehler^a

^aThe London School of Economics and Political Science

^bThe University of Sheffield

November 21, 2022

Abstract

We test how an academic study group's gender composition and its share of non-native English speakers affects students' (1) academic performance, (2) self-perceptions, (3) leadership aspirations and behaviours, and (4) predictions of academic performance. We randomise post-graduate students into study groups at a leading UK university that boasts a significant proportion of non-native English speakers (close to 50%) and a large share of women (80%). Exogenous variation in group composition following randomisation permits the estimation of a linear-in-shares model to observe the causal effects of study group gender and linguistic diversity. Using the university's administrative data and original survey data, we find that an increased share of women in a team improves students' grades and their perception of whether they felt heard during study group deliberations. We also find that non-native speakers outperform their native-speaking counterparts. However, the achievement gap diminishes when more women and non-native peers are in the team. Finally, we find no detriment for men who occupy a minority position in a study group. Conversely, men in groups with more women are more likely to report taking leadership roles.

Keywords: Peer effects, Higher education, Gender, Linguistic diversity

Introduction

In recent decades, women’s and immigrants’ participation in higher education and labour markets has increased. In 2019, OECD countries received on average eight new migrants per thousand inhabitants, and on average, international students accounted for 13% of all enrolments in master’s programmes and 22% of PhD enrolments OECD (2020). According to UNESCO, women’s participation rate in tertiary education almost doubled from 2000 and 2014. Similarly, in the labour force, the ratio of female to male participation in OECD countries has increased from 70% in 2000, to 77% in 2021 (World Bank 2022).

Diversity’s individual- and group-level effects could puzzlingly point in mixed directions. For firms, increases in gender, cultural, and linguistic diversity might bring the benefits of a broader range of views and skills, or they might generate communication difficulties and other forms of friction (Lazear 1999). Likewise for university students, similar positive externalities might arise from exposure to a diverse student body, or it might disturb instruction pace and flow in ways that hinder student academic achievement (Diette & Uwaifo Oyelere 2012).

This paper aims to shed light on that puzzle by estimating the causal effect of a team’s gender- and linguistic diversity on students’ performance, perceptions, and predictions. We randomize the gender- and linguistic composition of small teams of postgraduate students at a top UK University, and we estimate diversity’s impact on individual performance with respect to academic outcomes; we estimate its impact on perceptions of dynamics of group deliberation like ‘voice’ and ‘leadership’; and we estimate students’ predictions of their own academic promise.

The study took place in two consecutive iterations of a compulsory one-term course with an annual cohort of around 180 Master’s students (average age 25). We administered the study in two cohorts in 2020-2021 and 2021-2022). 80% of the students in each cohort were women and 49% were non-native English speakers. In this course, university officials assigned students into seminar groups (called seminars) in a way that was exogenous to their demographic characteristics. After this first random assignment, within each seminar group, we then randomised students a second time into smaller teams. We combine three data sources. First, we use administrative data with rich information on detailed student background from admission records, course selection, marks, course teachers’s characteristics (gender and native English speaker status) and academic advisor. Second, for each cohort we collected survey data before the course commenced and again at its conclusion. Third, we gathered individual-level data on

course performance from an end-of-term examination that accounted for 100% of the course grade.

Our answers to the research questions capitalize on three features of the study setting. First, there was no self-selection into the seminars and teams that we used to estimate compositional effects. Consequently, there is exogenous variation in the gender composition and in the distribution of native versus non-native English speakers (as well as any other demographic characteristics of students across groups). Therefore, our estimates of the effects of gender composition and the share of native speakers can be interpreted as causal.

Second, the course enrolls a high percentage of international students. UK institutions are attractive to international students looking to complete or continue studies in an English-speaking country. According to OECD (2020), The UK is the second receiving country of international students, only after the US. This course is not an exception; around 80% of the students are foreign-born, and around half are non-native English speakers. Thus, the setting maximized interactions between students with different language backgrounds.

Finally, the postgraduate programme and the course boast around 80% female participation. This allows us to study the effects of gender composition in teams in a female-majority setting. Previous literature has focused on the effects of gender composition on interactions and outcomes in male-dominated settings (i.e., STEM fields).¹ This is often done to understand the causes of low representation of women in these fields which pay on average higher wages than female-dominated fields. It is hypothesized that a high fraction of men in fields where women are a minority discourage other women from choosing these fields, and hence lead to a low representation of women. However, studying the effect of gender composition in settings where women are the majority is equally important. For example, these settings can help to clarify if findings of studies in male-dominated areas are explained by men's attitudes toward women or attitudes of a majority toward a minority regardless of the dominating sex.

We present estimates for three sets of outcomes. First, we report performance estimates for overall course mark. We find that, on average, native English speakers perform around 4% better than non-native English speakers even after controlling for other factors. However, we do not find evidence that the share of non-native speakers in

¹For instance, Oosterbeek & Van Ewijk (2014), and Shan (2022) study gender per effects for economics undergraduate students, where women constitute around 30% of the student body. Stoddard et al. (2020) study gender dynamics among students from an accounting undergraduate programme with a similar percentage of women participation.

the group affects individuals' exams marks. Regarding gender, we detect a statistically significant effect for non-native speakers of an increase in the share of female peers on exam marks, equal to 2% higher for 1 more woman in the team. Second, we report perception estimates for measures of team dynamics. We find that non-native speakers feel less heard when in groups with more native speakers, as one more native speaker in the team causes a 2 points (out of 10) decrease in non-natives' self-assessment of the influence of their voice. We also observe that an increase in the proportion of women in the seminar group causes an increase in women's perception of being heard in their group. The effect is equal to 1 point (out of 10) for each additional woman in the seminar group. In terms of leadership, we find that men are more likely to consider themselves as leaders of their team when in groups with more women, which shows that men do not suffer negative effects from being in a gender minority. Finally, we present prediction estimates for changes in students' expectations of their own future performance. We find that students in groups with more native speakers lower their expectations. They become more pessimistic about their future academic achievement.

The findings are relevant for universities and employers who deal with internationalisation, diversity, and inclusion in the workplace in settings where women are the majority. Non-native speakers benefit from having more diverse peers, and native speakers are not harmed by increased shares of peers with different linguistic backgrounds. Men and women benefit from having more female peers even in female-majority settings. This evidence supports policies that promote internationalisation, and helps inform course designs for academics and teams' formation for employers.

This paper contributes in various ways to a growing literature on diversity in educational settings and in teams at the workplace. In particular, to the strands that investigate the effects of gender composition and presence of non-native speakers. In educational settings, previous research has focus mainly on primary and secondary education. At this level, the evidence on the effect of diversity in language and country of origin on pupils outcomes is mixed. For instance, in the USA, Diette & Uwaifo Oyelere (2012) find that the effect of non-native speakers on the native peers are heterogenous. They report small positive effect for the native students in the bottom and middle part of the achievement distribution, and small negative effects among those at the top. In contrast, Geay et al. (2013) finds no negative effect of presence of non-native English speakers in UK schools.² In terms of gender diversity, there is overall agreement in

²Several studies investigate the effect of the share of immigrant children in students' academic outcomes. The results are also mixed. For instance, Gould et al. (2009) and Jensen & Rasmussen (2011) find evidence of negative effects of the share of immigrants in classrooms with data from

the literature that there are positive effects of higher shares of girls in the classroom on achievement of boys and girls.³ However, much of these findings from primary and secondary education might not be relevant in settings with adults students.

Only a few studies investigate the role of a diverse environment in higher education settings in educational attainment. These studies have mostly focused on larger peer groups rather than in small teams. For instance, Braakmann & McDonald (2018) focus on the diversity among undergraduate students using the entire university cohort as the relevant peer group, and find heterogeneous results. Treating the entire classroom as the peer group, Chevalier et al. (2020) study the effect of ethno-linguistic composition of a classroom on academic outcomes for undergraduate students in the UK. They find that Non-native speaker's benefit from greater linguistic diversity. Oosterbeek & Van Ewijk (2014) find no significant effect of the gender of peers at the classroom level on academic outcomes. Our study is the first to analyse the effects of peers' diversity in both small groups (e.g., 4-5 students) and middle-size groups (e.g., 8 to 16) as well as among adult students at the postgraduate level.

This is important because, by studying peer effects in small groups of postgraduate students, we can learn about team dynamics, and gain insights into the role of diversity along gender and native English versus non native English speakers in the workplace, where causal evidence is scarce. Apart from Battaglini et al. (2022), that studies the effect of working with female colleagues on federal judges' decisions of hiring female law clerks, the few causal evidence on team diversity comes from laboratory settings with undergraduate students.⁴

Israel and Denmark respectively. Meanwhile, studying the impact of immigration in Austrian schools, Schneeweis (2015) finds negative effects on the academic outcomes of migrants, but no effect on native students.

³Lavy & Schlosser (2011) and Hoxby (2000) have studied the effect of peer's gender, and peer's gender and race (respectively) on pupil's academic outcomes. Both studies find that student's marks increase with the proportion of girls in the classroom. Some studies have explored longer term outcomes. For instance, Schneeweis & Zweimüller (2012), and Anelli & Peri (2019) have studied the effect of gender of school peers on student's choices of field of study. Similarly, Black et al. (2013) study the role of high school peers, particularly in terms of peer's gender and social class in determining student's future labour market, and other longer run outcomes. They find heterogeneous results, while teenage girls tend to benefit from higher proportion of girls, boys do not.

⁴For instance, Hoogendoorn & Van Praag (2012), and Hoogendoorn et al. (2013) study diversity in teams consisting of undergraduate students in business studies that start up a venture as part of their assignments. They study ethnic and gender diversity on teams performance on business sales and profits.

1 Background

1.1 Course setting

The experiment was nested in two consecutive iterations of an eleven-week post-graduate course module in the LSE’s Department of Social Policy across the 2020 and 2021 academic years. The module, *Understanding Policy Research (SP401)*, is mandatory for all students who enroll on the one-year interdisciplinary MSc programme in International Social and Public Policy. The Master’s programme attracts an internationally diverse student profile to study “how states and societies respond to global challenges of social, demographic and economic change, and of poverty, migration and globalisation.” The SP401 curriculum equips students with the tools to evaluate applied policy research critically, through a grounding in concepts that draw on qualitative and quantitative research, the policy process, and applied social policy problems.

SP401 coursework comprised three components, each of which was repeated weekly throughout the MSc programme’s first term. First, students viewed a pre-recorded lecture that instructors uploaded online at the start of each week.⁵ Second, students completed a team-based activity in their own time. Third, students undertook additional activities during a weekly in-person seminar meeting, whereupon their team convened with one to three other teams. Lecture topics ranged from introductions to quantitative and qualitative research on one hand to substantive overviews of key research issues in policy analysis and policy-making on the other. Seminar activities required students to apply concepts from the week’s lecture to specific policy problems.

1.2 Data Collection

The study draws on three sets of data. First, we gathered administrative data from the university’s Registrar on student characteristics. Those data encompassed self-report information on gender and language, as well as other demographic characteristics and admissions information.

Second, students completed two surveys. Students completed a baseline self-report questionnaire in the first week of the term. The baseline survey contained items relating to respondents’ prior methodological training, their expectations for their own future performance in the course, and their self-assessed approaches to navigating professional group dynamics. Second, students completed a questionnaire in the term’s final week.

⁵Pre-recorded lectures were implemented as a response to the COVID-19 pandemic.

The endline survey duplicated items from the intake survey relating to expectations about performance in the course in particular and in the MSc in general. The endline survey also contained items that captured respondents' reflections about the group-work to which they had contributed, including self-assessments of their own sense of whether they had influenced the group-work.

Third, students completed a take-home online examination administered roughly one month after instruction concluded. The exam included many short questions, for which the expected answer ranged from a few words to a paragraph. The exam posed questions from throughout the term's materials and mimicked the format of questions that students had completed in their study groups and seminar meetings. The exam mark accounted for 100% percent of final mark on the course. However, even though the course was mandatory, a fail did not automatically lead to non-completion of the MSc. Still, stakes are high, because if the exam mark fell into the bad fail category, the student may retake the exam the following year, which delays graduation in at least a year.

2 Data

2.1 Administrative data

Table 1 presents summary statistics for the administrative data for both cohorts of students. Panel 1 of Table 1 shows that students' average age is 24.6, 79.5% of students are women, and 51% of the students are non-native speakers. Panel (2) presents statistics about students' previous academic backgrounds. Most students have a Bachelor's degree as their highest qualification, and 18% of students hold a prior Master's degree before starting the MSc. Panel 2 also shows that 33% of students have either a completed or are pending qualification from a university in the United Kingdom. The bottom panel of Table 1 presents average mark for women, men, native speakers and non-native speakers. Men and women performed similarly (around 68/100); however, there were significant differences by native language. Native English speakers earned an average mark of 73 while non-native speakers earned an average mark of 65. Marks corresponded to the following classifications: marks in the 0-39 range equated to a "Bad Fail", 40-49 was a "Fail", 50-59 was a "Pass", 60-69 was a "Merit", and 70-100 a "Distinction". The bottom of panel (3) presents statistics of the proportion of student falling into each category by gender and native language.

Table 1: Summary statistics: Administrative data

		All	Men	Women	Non-native Speakers	Native Speakers*
<i>(1) Demographic characteristics</i>						
Proportion (%)		-	20.5	79.5	51	49
Total (N)		376	77	299	180	173
Age	Mean	24.6	25.4	24.4	25.3	24.1
	S.d	3.9	4.4	3.7	4.6	2.9
<i>(2) Prior studies</i>						
Highest qualification	Bachelor (%)	81.9	83.1	81.9	75.4	89.0
	Master (%)	17.6	16.9	18.1	24.6	11.0
	Yes (%)	33.4	37.6	32.1	22.2	42.2
Studied in United Kingdom	No (%)	66.6	62.3	67.9	77.8	57.8
	S.d	3.9	4.4	3.7	4.6	2.9
	<i>(3) Academic outcomes</i>					
Mark	Mean		68.9	70.4	66.7	72.5
	s.d		13.2	12.6	12.8	12.1
	Bad fail (%)	2.4	3	2	2.8	1.2
Classification	Fail (%)	4.3	4	4	3.3	2.9
	Pass (%)	14.6	8	16	21.0	9.3
	Merit (%)	24.7	26	24	29.8	19.8
	Distinction (%)	54.0	60	53	43.1	66.9

Notes: *The LSE administrative dataset does not contain information on student's native language. Thus, summary statistics presented in the last two columns correspond to the sample of students who answer the baseline survey.

2.2 Survey data

2 presents summary statistics of the data we collected through the two surveys.

Baseline

The baseline survey asked questions about student's native language, usual role in group work, familiarity with relevant subjects, and their expectations of final mark on the course. In the top panel of 2, we present statistics for self-perception of leadership. Through the survey students reported what role described them best when working in groups. We provided four options, and included a brief description of what each role entailed. We were particularly interested in variation among students in their self-perception of leadership skills as these are linked to better labour market outcomes (see for example, Kuhn & Weinberger (2005)). Therefore, we derive the binary variable "Leadership role" which is equal to 1 if the student reported to be best described by the

leadership role, and 0 if they chose any other alternative. About 30% of the students considered themselves leaders before starting the course. The baseline measure allowed us to estimate the course’s effect on changes to students’ perception of themselves.

Additionally, for non-native English speakers, we derived a variable that indicated the “distance” between the student’s native language (as they reported in the baseline survey) and the English language. The variable allow us to capture heterogeneity between non native English speakers. We use Chiswick & Miller (2005)’s measure of linguistic distance which ranges from 1 to 3, in 0.25 increments, with three being the most similar to English. Among the sample, there was substantial diversity in students’ native languages; the language scores spanned the whole range from 1 to to 3, with a mean of 1.9.⁶

Then, we present statistics of two variables that contain information about familiarity with relevant subjects and mean and standard deviation of expected final mark on the course. Most students reported some experience with research methods; on average, they were more familiar with qualitative than quantitative methods.

End of year survey

The end of year survey asked students to reflect on team dynamics, students’ views of their future interactions in teams, and academic performance expectations. The bottom panel of Table 2 presents the mean and standard deviation by group for relevant end term survey data.

At the end of the course, we asked students three questions related to their perception of their “voice” in team interactions. Survey items probed student’s level of agreement with the statement “*My voice was heard during group discussions*”, and with two follow up items: “*Working in teams for SP401 made me more confident than before in voicing my view in future interactions*”, and “*Working in teams for SP401 made me more confident than before that my view will be heard in future interactions*”. Table 2 shows that most students in general agreed with the first statement, and both women and native speakers in particular were more likely than their counterparts to have agreed. For the two statements about confidence in future interactions, the average response was around 5 (neither agree nor disagree). Again, women and native speakers responded above 5 on average, which indicates that these students tended to agree more with the two follow up questions.

⁶See Appendix 6.1 for details on student’s country of origin, native language, and their correspondent measure of linguistic distance to English

Table 2 also presents statistics for two variables related to leadership. The first is a binary variable that we constructed using data from the end of year survey item that collected information on each student's perception of their role in their team. As for the baseline leadership variable, the variable took the value 1 if the student answered Manager/Leader and zero otherwise. We also used this leadership variable in combination with the baseline leadership variable to measure changes in students' assessment of their role in teams. Table 2's bottom panel displays that around 30% of students reported that they had a leadership role in their teams, which is consistent with what they reported in the baseline survey. However, we observed within group-differences across baseline and end of term team roles. While 20% of men reported taking leadership roles at the start of the course, 30% reported the same at the end of the term. The opposite was true for non-native English speakers, who were more likely to report that they identified with the leader role in teams before the start of the course. The second leadership variable was constructed using peers' input. For the 2020-2021 cohort, we asked students to match each of the roles with one or more team members. Then, we constructed a binary variable equal to 1 if a student was mentioned as taking the Manager/Leader role by at least one other team member and zero otherwise. 30% of students were nominated at least once as the leader of the group by their peers.

Finally, Table 2 also presents the average expected mark at the end of the term. At this time, average expected marks are lower than at the start of the course for all groups.

Table 2: Summary statistics: Surveys

Variable		All	Men	Women	Non-native Speakers	Native Speakers
<i>(1) Baseline Survey</i>						
Leadership role (0 if no, 1 if yes) *	Mean	0.3	0.2	0.4	0.3	0.4
	S.d.	0.5	0.4	0.5	0.4	0.5
	N	182	36	146	84	98
Language Score (1-3)	Mean	-	1.9	1.9	1.9	-
	S.d.	-	0.6	0.5	0.5	-
Familiarity with qualitative research methods (0-10)	Mean	5.9	5.6	5.9	5.8	6.0
	S.d.	3.4	2.1	2.5	2.4	2.5
Familiarity with quantitative research methods (0-10)	Mean	4.1	4.2	4.1	4.1	4.2
	S.d.	2.2	2.2	2.2	2.3	2.2
Expected Mark (0-100)	Mean	72.6	72.0	72.8	73.7	71.4
	S.d.	12.6	13.1	12.4	12.7	12.3
	N	355	71	284	182	173
Average response rate	%	93.9	92.2	94.3	-	-
<i>(2) End of term survey</i>						
My voice was heard during group discussions (Agreement 0-10)	Mean	8.7	8.3	8.8	8.6	8.9
	S.d.	1.8	2.1	1.7	1.7	1.6
Leadership role in team (0 if no, 1 if yes)	Mean	0.3	0.3	0.3	0.2	0.4
	S.d.	0.5	0.5	0.5	0.4	0.5
Expected Mark End (0-100)	Mean	71.0	70.0	71.3	71.5	70.6
	S.d.	8.8	8.5	8.9	9.0	8.7
	N	187	39	148	98	75
More confident in voicing my view in future interactions (Agreement 0-10)	Mean	5.1	4.8	5.2	5.3	4.9
	S.d.	3.1	2.9	3.1	3.1	3.1
More confident that my view will be heard in future interactions (Agreement 0-10)	Mean	5.2	5.1	5.3	5.5	5.1
	S.d.	2.9	2.5	3.1	2.9	3.0
	N	85	22	63	33	47
Peers nominate as leader in team (0 if no, 1 if yes)**	Mean	0.3	0.3	0.3	0.2	0.3
	S.d.	0.4	0.5	0.4	0.4	0.5
	N	187	39	148	98	75
Average response rate	%	59.8	58.4	60.2	63.9	60.7

Notes: *2021-2022 cohort only, ** 2020-2021 cohort only. We obtain the data on native language from the baseline survey; thus, average response rate for native vs non-native speakers is in relation to baseline response.

3 Empirical strategy

We aim to understand the effect of native language and gender composition on academic achievement, team dynamics, and expectations of future performance. Estimating peers' effects on student outcomes can be challenging, mainly because of a self-selection problem. University students often get to choose which classes to take, which can lead to endogenous peer group composition. The endogeneity arises when an unobserved characteristic is correlated with the outcomes of interest and with the probability of joining a group. This endogeneity can severely bias the estimation of peer effects, as individuals in the same group will have correlated outcomes even if peer effects do not exist.

There is no selection problem in our setting because group formation is exogenous to peers' characteristics and expected results at every level of interaction: course, seminar group, and team. First, there is no self-selection at the course level because the course is compulsory for all Social Policy master's students. Second, course administrators allocated students to seminar groups independently from students' characteristics including gender and native language. Students could request changes in their allocated seminar group only if they confronted a timetable clash. Where such a clash necessitated a re-allocation, the course administrator relocated the student to any non-clashing seminar and thus preserved the allocation's exogeneity. However, to identify causal effects from the team composition, we further randomised students within each seminar into smaller groups of approximately four students in each. As seminar size varied between cohorts, there were either two or four teams per seminar, but the total number of teams remained constant across cohorts. Still, we corroborate that the assignment to seminars is as good as random by using a regression based test.⁷ The tests look at the within-group correlation between the individual's own characteristics of interest (native language and gender) and the average characteristic of their peers in the reference group (seminar). Table 3 shows the average and standard deviation of the proportion of women and native speakers per team, and the number of seminar groups per cohort. Table 4 presents the test statistics of the regression-based test and (two-sided) p-values for assignment to seminars. Additionally we include the results for the test when applied to our randomisation to teams (withing seminars). P-values are large for all tests, indicating that there is not enough evidence to reject the null hypothesis of random assignment to seminars and teams.

⁷We use the test proposed by Jochmans (2022). The test is a modification of the within-group regression of Sacerdote (2001) that allows for variation in the size of the urn.

Table 3: Proportion of women and native speakers per team

Cohort	Proportion of women		Proportion of native speakers		Groups (N)
	Mean	S.d.	Mean	S.d.	
2020-2021	0.79	0.21	0.42	0.28	44
2021-2022	0.81	0.21	0.55	0.25	44
Total	0.80	0.21	0.48	0.27	88

Notes: Mean indicates the average proportion of women by classroom per year. s.d indicates the standard deviation.

Table 4: Random assignment test statistics and p-values

		Proportion of native speakers in seminar group	Proportion of native speakers in team
Native English	Test statistic	0.911	0.005
	p-value	(0.362)	(0.996)
	Urns (N)	2	44
		Proportion of women in seminar group	Proportion of women in team
Female	Test statistic	-0.735	1.346
	p-value	(0.462)	(0.178)
	Urns (N)	2	44

Note: The number of urns corresponds to the number of groups from which peers are drawn.

3.1 Estimation

We use a linear model to estimate the causal impact of the proportion of women and proportion of native speakers on all the outcomes of interest. Manski (1993) introduced the original model to estimate peer effects, which attributes outcomes to individual characteristics and the characteristics of a group to which a student belongs. We extend and develop that model to capture the effects on each of the outcomes of interest of variation in gender composition and share of native english speakers as follows:

$$Y_{igs} = \alpha + \beta_1 W_g + \beta_2 NS_g + \beta_3 W_{gs} + \beta_4 NS_{gs} + \gamma X_{ig} + C_i + \epsilon_{igs} \quad (1)$$

where Y_{igs} is the outcome of interest for student i in group g and seminar s , W_g is the gender composition of group g , NS_g is the percentage of native speakers in group g . W_{gs} and NS_{gs} are the percentage of women and of native English speakers in seminar s , but excluding students in student i 's own group g . X_{ig} is a vector of control variables

including age, familiarity with subjects relevant to the course, dummies for English as first language and gender, student’s highest level of education, previous UK studies, and teacher and academic advisers’ characteristics. Additionally, when we estimate equation 1 for the subsample of non-native speakers, we include Chiswick & Miller (2005)’s language distance scores in X_{ig} . C_i is a dummy for the student’s cohort.

Note that groups are a sub-unit of the seminar. Therefore, the composition of the seminar varies with the composition of the groups. Thus, excluding students in the same team for the measures W_{gs} and NS_{gs} helps to avoid multicollinearity problems. More importantly, as students can neither self-select into seminar groups nor into study groups, W_g , W_{gs} , PNS_g and PNS_{gs} are all exogenous to the outcomes Y_{igs} . Therefore, all the coefficients of interest, that is β_1 , β_2 , β_3 , and β_4 , provide causal information about compositional effects on the outcomes of interest.

We estimate equation 1 for three sets of outcomes: *performance* measured by academic achievement, *perceptions* of team dynamics, and *predictions* of academic expectations. Performance outcomes correspond to examination marks, for which we estimate equation 1 using ordinary least squares. We cluster standard errors by cohort and seminar level.

To estimate the effects on perceptions and predictions we use endline survey data. The key team dynamics outcomes comprised responses to items such as “Voice was heard”, which was measured on a scale of 0-10; and “Leadership role”, which was binary. The main variable on expectations is “Expected mark” which is measured from 0-100. For the continuous outcomes, we estimate the effects with a least squares regression, and a probit for binary outcomes. Additionally, because the endline survey has a 60% response rate, adjust for survey non-response before running the regressions on these variables. We adjust for nonresponse using the inverse probability weighting method.⁸ First, we classify all individuals (responders and non-responders) into cells based on observable characteristics that predict whether an individual responds or not; we use student’s cohort, seminar group, gender, native English status, and information on previous studies in the UK. Then, we calculate the response probability conditional on this set of characteristics and assign responders a weight corresponding to the inverse response probability of their cell.

For a detailed description of outcomes of interest see Appendix 6.3.

⁸For a detailed description of the Inverse probability weighting method see Hernán & Robins (2016)

4 Results

4.1 Performance on Academic outcomes

Table 5 presents the estimated coefficients of equation 1 for exam marks and for the “Distinction” outcome. Column (1) presents estimates for the regression on marks for the whole sample, and columns (2) to (5) for the regression on each group. We find no significant gender differences in exam marks, but native English speakers have significantly higher marks than non-native English speakers even when controlling for other factors. These results might be attributable to the difficulties that non-native English speaking students confront in learning in a second language (e.g., Bernhofer & Tonin (2022)).

In terms of effects of gender composition, we find that an increase in the percentage of women in the team and seminar caused an increase in exam marks for non-native English speakers. For instance, column (5) shows that for non-native speakers an increase in 1% of the percentage of women in the team caused an increase in 0.08 points in exam marks. Thus, in a group of 4, 1 more woman in the group (25% increase in percentage of women), caused an average increase of 2 points.

Table 5: Regression coefficients: Exam Marks

	(1) All	(2) Women	(3) Men	(4) Native Speakers	(5) Non-Native Speakers
Previous UK Studies	0.81 (1.55)	-0.26 (1.84)	7.52* (2.89)	0.43 (2.25)	0.73 (2.62)
Gender (Female=1)	-1.46 (1.45)			-1.81 (1.69)	-0.60 (2.20)
Native English	4.20** (1.45)	4.18* (1.67)	2.22 (2.52)		
Percentage Native Speakers (Team)	-0.00 (0.02)	0.00 (0.02)	-0.04 (0.04)	0.01 (0.03)	-0.01 (0.03)
Percentage Native Speakers (Seminar-O)	0.02 (0.03)	0.01 (0.04)	0.08 (0.05)	0.02 (0.04)	0.03 0.01 (0.03)
Percentage of Women (Team)	0.01 (0.03)	0.01 (0.03)	-0.03 (0.04)	-0.05 (0.04)	0.08* (0.03)
Percentage of Women (Seminar-O)	0.06 (0.03)	0.06 (0.03)	0.03 (0.10)	0.06 (0.05)	0.09 (0.05)
Language Score					4.39 (2.33)
Constant	84.21*** (5.52)	81.66*** (6.32)	91.78*** (16.82)	92.65*** (16.03)	59.28*** (11.87)
N	343	273	70	167	155

Notes: All models include controls for: age, education level, experience with quantitative methods, experience with qualitative methods, seminar leader gender, seminar leader native English, adviser gender, adviser native English. Exam Marks are in scale from 0-100. Seminar - O refers to other members in the seminar group (excluding individual's own team). Data source: Administrative records for cohorts 2020-2021, and 2021-2022. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, cluster standard errors in parentheses.

4.2 Perceptions of team dynamics

We now analyse the perception of team dynamics by testing the causal role of linguistic and gender diversity on voice and leadership.

4.2.1 Voice in teams

Table 6 presents the regression estimates for the item “My voice was heard during group discussions” (measured as level of agreement from 1-10). Column (1) shows that an

increase in the percentage of women in the team and in the seminar group caused an increase in students' level of agreement. Estimates by subsample suggest that women and non-native speakers benefited from having more women in the group, as an increase in the percentage of women either in one's own team or in the seminar group caused an increase in the extent to which students agreed with the statement. For instance, for non-native speakers an increase in 1% of the percentage of women in the classroom caused an increase in the level of agreement of 0.02 points. Thus, in a group of 4, 1 more woman in the group (25% increase in percentage of women), caused an average increase of 0.5 points in level of agreement with the statement. On the other hand, an increase in the percentage of native speakers had the opposite effect on non-native English speakers, who reported that they felt less heard when working in teams with more native English speakers. The effect was similar in size to the effect of the share of women, but in the opposite direction.

Table 6: Regression coefficients: “My voice was heard during group discussions”

	(1)	(2)	(3)	(4)	(5)
	All	Women	Men	Native Speakers	Non-Native Speakers
Previous UK Studies	0.08 (0.26)	-0.19 (0.24)	-0.02 (0.65)	-0.56 (0.28)	0.38 (0.39)
Gender (Female=1)	0.48 (0.30)			0.32 (0.35)	0.78 (0.54)
Native English	0.54* (0.21)	0.37 (0.22)	0.59 (1.22)		
Percentage Native Speakers (Team)	-0.01 (0.00)	-0.01 (0.00)	0.00 (0.01)	0.00 (0.01)	-0.02* (0.01)
Percentage Native Speakers (Seminar-O)	0.00 (0.01)	0.00 (0.01)	-0.01 (0.01)	-0.00 (0.01)	0.01 (0.01)
Percentage of Women (Team)	0.01* (0.00)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.02* (0.01)
Percentage of Women (Seminar -O)	0.01* (0.01)	0.02* (0.01)	-0.00 (0.02)	0.01 (0.01)	0.02 (0.01)
Language Score					0.71 (0.36)
Constant	4.02* (1.73)	6.00*** (1.16)	9.61* (4.31)	10.42*** (2.35)	1.42 (2.19)
N	218	174	44	104	101

Notes: All models include controls for: age, education level, experience with quantitative methods, experience with qualitative methods, seminar leader gender, seminar leader native English, adviser gender, adviser native English. The variable “My voice was heard during group discussions” measures the level of agreement with the statement in a scale of 0-10. Seminar - O refers to other members in the seminar group (excluding individual’s own team). Data source: End year survey, cohorts 2020-2021, and 2021-2022. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, cluster standard errors in parentheses.

We also present estimates for two follow-up questions related to student perceptions of how interactions with their team would affect their future interactions. The first is “more confident in voicing their views” (measured as level of agreement from 1 to 10). In Table 7, column (4) shows that for non-native English speakers, an increase in the percentage of female peers caused an increase in their confidence in voicing their views.

The second follow up question pertained to students’ perception of how others would

receive their views, which was also measured on a scale from 1 to 10. Table 7, columns (5) to (8) present the regression coefficients of this item by subsample. Similar to the previous survey item, column (8) shows that an increase in the percentage of women in the team and in the rest of the seminar group caused an increase in reported confidence on how their views will be received in the future. Results from the voice items suggest that non-native speakers benefited from having peer groups with a higher percentage of women.

Table 7: Regression coefficients: “More confident in voicing my view” and “my voice will be heard”

	“More confident in voicing my view”				“My voice will be heard”			
	(1) Women	(2) Men	(3) Native Speakers	(4) Non-Native Speakers	(5) Women	(6) Men	(7) Native Speakers	(8) Non-Native Speakers
Previous	1.66*	2.38	1.87	1.43	0.69	1.10	0.62	0.27
UK Studies	(0.79)	(1.75)	(0.99)	(0.97)	(0.81)	(1.46)	(1.10)	(0.95)
Gender (Female=1)			-0.91 (0.98)	0.94 (0.97)			-1.17 (0.96)	1.92 (0.97)
Native English	-0.38 (0.97)	1.29 (1.63)			-0.75 (0.91)	1.40 (1.36)		
Percentage Native Speakers (Team)	0.00 (0.02)	0.01 (0.02)	0.01 (0.02)	-0.01 (0.02)	0.00 (0.02)	0.00 (0.02)	0.00 (0.02)	-0.01 (0.02)
Percentage Native Speakers (Seminar -O)	0.01 (0.02)	-0.03 (0.03)	-0.00 (0.02)	0.01 (0.03)	0.01 (0.02)	-0.04 (0.02)	0.01 (0.02)	0.00 (0.03)
Percentage of Women (Team)	-0.00 (0.02)	-0.02 (0.03)	-0.02 (0.02)	0.08** (0.02)	0.01 (0.02)	-0.02 (0.03)	-0.02 (0.02)	0.09*** (0.02)
Percentage of Women (Seminar -O)	0.02 (0.02)	0.04 (0.03)	-0.01 (0.02)	0.10** (0.03)	0.02 (0.02)	-0.01 (0.04)	-0.02 (0.02)	0.09*** (0.03)
Language Score				-0.78 (0.83)				0.16 (0.75)
Constant	-3.20 (4.28)	10.55* (4.37)	5.71 (6.20)	0.57 (4.27)	-1.71 (3.87)	12.66** (3.89)	6.47 (6.92)	0.53 (4.19)
N	59	21	47	30	57	20	45	30

Notes: All models include controls for: age, education level, experience with quantitative methods, experience with qualitative methods, seminar leader gender, seminar leader native English, adviser gender, adviser native English. The variable “More confident in voicing my view” measure the level of agreement (from 0 to 10) with the statement: “Working in teams for SP401 made more confident than before in voicing my view in future interactions”. Data source: End year survey 2021-2022. The dependent variable “My voice will be heard” indicates the level of agreement (from 0-10) with the statement “Working in teams for SP401 made more confident than before that my view will be heard in future interactions”. Seminar - O refers to other members in the seminar group (excluding individual’s own team). Data source: End of year survey 2021-2022. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, cluster standard errors in parentheses.

4.2.2 Leadership

In addition to voice, we present estimates of three measures of leadership. First, we present a binary measure of leadership based on self-reported data gathered in the end of year survey. The dependent variable is one if the student answered “Manager/Leader” to the question: “Which of these roles best describes your own role in your sub-group?” and zero if they chose any of the other alternatives (Sceptic/Thinker, Checker/Recorder, or Conciliator). Table 8 presents the estimated coefficients for this measure. Across the whole sample, native speakers were more likely to report that they adopted a leadership role, but there were no statistically significant compositional effects of language or gender. However, the estimated coefficients from the sub-sample regressions that appear in Table 8, column (3) show that men are more likely to consider themselves team leaders when they have more female peers.

Table 8: Probit regression coefficients: Identify as leader

	(1)	(2)	(3)	(4)	(5)
	All	Women	Men	Native Speakers	Non-Native Speakers
Previous UK Studies	0.20 (0.22)	0.13 (0.25)	-3.42 (2.73)	0.30 (0.30)	-0.09 (0.42)
Gender (Female=1)	0.12 (0.25)			0.48 (0.30)	-0.85 (0.47)
Native English	0.67** (0.22)	0.80*** (0.24)	10.03** (3.59)		
Percentage Native Speakers (Team)	0.00 (0.00)	0.00 (0.00)	0.06** (0.02)	0.00 (0.00)	-0.01 (0.01)
Percentage Native Speakers (Seminar -O)	0.01 (0.00)	0.01* (0.00)	-0.16** (0.05)	0.01 (0.01)	0.00 (0.01)
Percentage of Women (Team)	-0.00 (0.00)	-0.00 (0.00)	0.02 (0.01)	-0.00 (0.01)	0.01 (0.01)
Percentage of Women (Seminar -O)	0.00 (0.00)	-0.00 (0.01)	0.33** (0.10)	0.01 (0.01)	-0.01 (0.01)
Language Score					0.54 (0.33)
Constant	-2.79** (1.07)	-2.53* (1.15)	-52.62** (16.14)	-3.50 (1.96)	-2.95 (1.72)
N	218	174	44	104	101

Notes: All models include controls for: age, education level, experience with quantitative methods, experience with qualitative methods, seminar leader gender, seminar leader native English, adviser gender, adviser native English. The variable “Identify as leader” is a binary variable that indicates if a student indicates that the role that describes them best is Manager/Leader (value 1) or not (value 0). Seminar - O refers to other members in the seminar group (excluding individual’s own team). Data source: end of year survey cohorts 2020-2021, and 2021-2022. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, standard errors in parentheses.

In Table 9 we present estimates for two outcomes, a measure of change in the leadership role (columns (1) to (4)), and the outcome “Others identify as leader” (columns (5) to (8)). The first measure is constructed by comparing students’ responses in the baseline and end surveys. It is equal to 1 if the student reported not commonly taking the leadership position in their teams in the baseline survey but then reported to be the leader of their team in the end survey, 0 if they provided the same answer in baseline and end survey, and -1 if they reported taking leadership roles at baseline, but not a leadership role in endline. Thus, we interpret a positive coefficient as increasing

willingness to take on leadership roles. The results are similar to the first measure of leadership. For men (Table 9, column 2), an increase in the percentage of female peers causes an increase in the probability of changing their perception of their role in a team, towards being more likely to identify as a leader. More specifically, for men an increase in 1% of the percentage of women in the team, causes an increase on the dependent variable of 0.02 points. Thus, in a group of 4, an additional woman in the group would cause an average increase of 0.5 (or 25%) in the change of willingness to lead.

Table 9 also shows the probit regression coefficients for peers' perceptions of leadership. This outcome is a binary measure equal to 1 for student i if someone in their team reports that student i had the lead role in the team. For non-native English speakers, being in groups with more native speakers makes them less likely to be nominated by others as the team leader, while being in teams with more women makes it more likely for them to be mentioned as the team leader.

Table 9: Regression coefficients: Change in Leadership self-perception and others identify as leader

	Change in Leadership self-perception				Others identify as leader (Probit)			
	(1) Women	(2) Men	(3) Native Speakers	(4) Non-Native Speakers	(5) Women	(6) Men	(7) Native Speakers	(8) Non-Native Speakers
Previous	0.03	-0.25	0.10	0.01	0.32	1.28	0.06	0.35
UK Studies	(0.12)	(0.20)	(0.14)	(0.13)	(0.31)	(1.01)	(0.39)	(0.44)
Gender (Female=1)			-0.02 (0.14)	0.06 (0.12)			0.48 (0.48)	-0.68 (0.49)
Native English	-0.03 (0.11)	0.44 (0.28)			0.32 (0.25)	-0.43 (0.63)		
Percentage Native Speakers (Team)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.02 (0.01)	0.00 (0.01)	-0.01* (0.01)
Percentage Native Speakers (Seminar -O)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-0.02 (0.02)	0.01 (0.01)	0.01 (0.01)
Percentage of Women (Team)	-0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)	0.01 (0.01)	-0.02 (0.02)	0.01 (0.01)	0.02 (0.01)
Percentage of Women (Seminar -O)	0.00 (0.00)	0.02*** (0.01)	0.00 (0.00)	0.00 (0.00)	-0.00 (0.01)	0.01 (0.02)	0.01 (0.01)	-0.00 (0.01)
Language Score				0.18* (0.08)				0.04 (0.09)
Constant	0.00 (0.53)	-1.50* (0.64)	-1.80* (0.82)	-0.19 (0.36)	0.312 (0.24)	1.949 (0.57)	-1.228 (-0.51)	-1.183 (-0.60)
N	101	28	70	53	130	34	71	80

Notes: All models include controls for: age, education level, experience with quantitative methods, experience with qualitative methods, seminar leader gender, seminar leader native English, adviser gender, adviser native English. The variable “Change in Leadership self-perception” is the difference between the two binary “Identify as leader” variables (at end of term minus at baseline). Thus, the variable is equal to zero if there is no change, 1 if student answers leader at end but not at baseline, and -1 if student answers leader at baseline but not at the end survey. Data source: Baseline survey and end of year survey cohort 2021-2022. The “Others identify as leader” variable is equal to 1 if a student was mentioned as taking the Manager/Leader by at least one member of their team, and zero otherwise. Seminar - O refers to other members in the seminar group (excluding individual’s own team). Data source: end of year survey 2020-2021. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, standard errors in parentheses.

4.2.3 Predictions of academic achievement

Finally, we present the regression coefficients for three outcomes related to students’ predictions of their academic performance on the course. The first of these outcomes is their expected mark on the course, according to what students reported at the end of

the term before they sat the examination. Table 10 presents the estimated coefficients for this model. The results suggest that the share of native speakers had strong effects on expected mark, especially for non-native English speakers. An increase in the share of native English speakers in the team or seminar reduced the expected mark that non-native speakers report at the end of the term. We also find that an increase in the percentage of women in the team caused an increase in expected mark. This result is in concordance with the results for actual marks (presented in section 4.1), as an increase in the share of women in the group caused an increase in exam mark. Thus, students predicted the direction of the favourable effect.

Table 10: Regression coefficients: expected mark at end

	(1) All	(2) Women	(3) Men	(4) Native Speakers	(5) Non-Native Speakers
Previous UK Studies	-2.96* (1.41)	-3.14 (1.60)	-1.91 (2.95)	-5.53*** (1.58)	1.91 (2.42)
Gender (Female=1)	0.97 (1.52)			-0.23 (1.89)	4.21 (2.54)
Native English	0.09 (1.38)	-0.93 (1.53)	3.30 (4.20)		
Percentage Native Speakers (Team)	-0.02 (0.02)	-0.02 (0.03)	-0.01 (0.05)	-0.01 (0.04)	-0.01 (0.03)
Percentage Native Speakers (Seminar-O)	-0.00 (0.02)	-0.02 (0.03)	0.06 (0.05)	0.06 (0.03)	-0.07* (0.03)
Percentage of Women (Team)	0.06* (0.03)	0.06* (0.03)	0.02 (0.05)	0.05 (0.03)	0.08* (0.04)
Percentage of Women (Seminar -O)	0.04 (0.03)	0.03 (0.03)	0.05 (0.07)	0.08* (0.03)	-0.02 (0.04)
Language Score				-2.69	(1.56)
Constant	62.18*** (5.94)	60.02*** (6.69)	72.96*** (15.56)	56.37*** (10.72)	72.85*** (7.83)
N	217	173	44	103	101

Notes: All models include controls for: age, Education level, Experience with quantitative methods, Experience with qualitative methods, Seminar leader gender, seminar leader native English, adviser gender, adviser native English. The dependent variable measures the student's expectation of their exam mark at the end of the course (on a scale from 0-100). Seminar - O refers to other members in the seminar group (excluding individual's own team). Data source: end year survey. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, cluster standard errors in parentheses.

We also analyse expected marks in comparison with student’s original expectations of academic performance, which we collected in the baseline survey. Table 11 presents the estimated coefficients of the regression with the difference between expected mark at baseline versus end of term (see columns (1) to (4)). We find that students in groups with more native speakers experienced a larger decline in their predicted performance. Non-native English speakers have a higher initial expected mark than native speakers (average of 73.7 vs average of 71.4), and the decrease is on average leading to a better prediction of actual mark

Lastly, Table 11 columns (5) to (8) present the estimates for the difference between the expected mark at the end of the course with the actual mark. We do not find any statistically significant effect of the share of women nor of the share of native speakers on the gap between expectations and real mark.

5 Conclusions

We implemented a field experiment to measure the role of gender composition and the share of native English speakers in small groups of postgraduate students. We study a cohort of students from a MSc programme in a top UK University. The programme is female dominated, and very international, which makes it an ideal setting to study how internationalisation and the increase in women’s participation in higher education affects group interactions as well as individual outcomes. To do so, we randomised students into small teams to generate exogenous variation in the demographic characteristics of students across groups. Then, we analysed the effects of group composition on three sets of outcomes: (1) *academic performance*, (2) *perceptions* of team dynamics, (3) and *predictions* of academic performance.

Regarding individuals’ academic performance, we find that non-native English speakers do worse than their native peers but that this gap closes when they have more female peers. For non-native English speakers, an increase in the proportion of women in the seminar group causes an increase in exam marks and on the probability of graduating with distinction.

We also find that group composition affects outcomes related to team dynamics. Firstly, women and non-native speakers were more likely to report their voices were heard when in groups with more women. Non-native speakers agreed less with the same statement when in groups with more English native speakers. Secondly, when looking at leadership measures, we find that men were more likely to identify as their

team leaders when they were in seminar groups with more women. Also, when we compare this result with students' answers in the baseline survey, we see that men were more likely to change their answers (from not feeling like a leader to being the leader of their group) when they collaborated with more women. We also find that non-native speakers were less likely to be signalled as the team leader by their peers when they collaborated with more Native English speakers.

Taken together, these results indicate that group composition, at least in the gender and linguistic dimension, plays a relevant role in both academic outcomes and students' self-perception of their group work interactions and related skills. More specifically, we show that non-native speakers benefit from having peers from diverse linguistic backgrounds. At the same time, we do not see any adverse effect of a higher share of non-native speakers on native speakers' outcomes.

Furthermore, in contrast to what previous studies have found for women in minority status, men do not seem to suffer from being in the minority in any outcome. Related literature has shown that women are less likely to take leadership roles when working in male-dominated environments. For instance, in a laboratory setting, Stoddard et al. (2020) finds that women are less likely to be considered influential or chosen as a spokesperson in predominantly male teams. However, it is unclear if the effect is caused by "tokenism" (the effect of being in the minority) or if it's caused by gender attitudes. Our analysis in a female-dominated setting suggests that the latter is true. Contrary to when women are in the minority, men's attitudes towards leadership are positively affected, i.e., they are more willing to take leadership roles when in groups with more women.

Our results are relevant for course design and team formations in educational settings, where students often need to work on group projects during their studies, and to the industry and the public sector. In terms of linguistic diversity, as non-native speakers' performance and perception of being heard improve when non-native speakers are in teams with a low presence of native speakers, higher education settings and workplaces should consider creating teams where there is a large enough proportion of non-native speakers, especially in settings where teams need to solve highly complex problems when everybody's opinion must be considered to lead to the best outcome. Additionally, this allow non-native speakers to practice and develop interpersonal skills that are highly valued in the labour market. Regarding gender composition of teams, our results suggest that women benefit from being in teams where there is a large fraction of women in them. Since the teams that we created had a low presence of males in them due to the female-dominated nature of the course we study, our findings suggest

that for universities and workplaces it may make sense to create female-only teams, especially when it is important that everyone’s opinion is heard, which is the case when complex problems are at hand.

References

- Anelli, M. & Peri, G. (2019), ‘The effects of high school peers’ gender on college major, college performance and income’, *The Economic Journal* **129**(618), 553–602.
- Battaglini, M., Harris, J. & Patacchini, E. (2022), ‘Interactions with powerful female colleagues promote diversity in hiring’, *Journal of labor economics (forthcoming)* .
- Bernhofer, J. & Tonin, M. (2022), ‘The effect of the language of instruction on academic performance’, *Labour Economics* p. 102218.
URL: <https://www.sciencedirect.com/science/article/pii/S0927537122001099>
- Black, S. E., Devereux, P. J. & Salvanes, K. G. (2013), ‘Under pressure? the effect of peers on outcomes of young adults’, *Journal of Labor Economics* **31**(1), 119–153.
- Braakmann, N. & McDonald, S. (2018), ‘Student exposure to socio-economic diversity and students’ university outcomes—evidence from english administrative data’, *MPRA Paper No. 90351* .
- Chevalier, A., Isphording, I. E. & Lisauskaite, E. (2020), ‘Peer diversity, college performance and educational choices’, *Labour Economics* **64**, 101833.
- Chiswick, B. R. & Miller, P. W. (2005), ‘Linguistic distance: A quantitative measure of the distance between english and other languages’, *Journal of Multilingual and Multicultural Development* **26**(1), 1–11.
- Diette, T. M. & Uwaifo Oyelere, R. (2012), ‘Do significant immigrant inflows create negative education impacts? lessons from the north carolina public school system’, *IZA Discussion Paper* .
- Geay, C., McNally, S. & Telhaj, S. (2013), ‘Non-native speakers of english in the classroom: what are the effects on pupil performance?’, *The Economic Journal* **123**(570), F281–F307.
- Gould, E. D., Lavy, V. & Daniele Paserman, M. (2009), ‘Does immigration affect the long-term educational outcomes of natives? quasi-experimental evidence’, *The Economic Journal* **119**(540), 1243–1269.

-
- Hernán, M. A. & Robins, J. M. (2016), *Causal inference*, Boca Raton.
- Hoogendoorn, S., Oosterbeek, H. & Van Praag, M. (2013), ‘The impact of gender diversity on the performance of business teams: Evidence from a field experiment’, *Management Science* **59**(7), 1514–1528.
- Hoogendoorn, S. & Van Praag, M. (2012), ‘Ethnic diversity and team performance: A field experiment’, *Tinbergen Institute Discussion Paper* .
- Hoxby, C. (2000), Peer effects in the classroom: Learning from gender and race variation, Technical report, National Bureau of Economic Research.
- Jensen, P. & Rasmussen, A. W. (2011), ‘The effect of immigrant concentration in schools on native and immigrant children’s reading and math skills’, *Economics of Education Review* **30**(6), 1503–1515.
- Jochmans, K. (2022), ‘Testing random assignment to peer groups’.
- Kuhn, P. & Weinberger, C. (2005), ‘Leadership skills and wages’, *Journal of Labor Economics* **23**(3), 395–436.
URL: <http://www.jstor.org/stable/10.1086/430282>
- Lavy, V. & Schlosser, A. (2011), ‘Mechanisms and impacts of gender peer effects at school’, *American Economic Journal: Applied Economics* **3**(2), 1–33.
- Lazear, E. P. (1999), ‘Culture and language’, *Journal of political Economy* **107**(S6), S95–S126.
- Manski, C. F. (1993), ‘Identification of endogenous social effects: The reflection problem’, *The review of economic studies* **60**(3), 531–542.
- OECD (2020), *International Migration Outlook 2020*, OECD, Paris, France.
- Oosterbeek, H. & Van Ewijk, R. (2014), ‘Gender peer effects in university: Evidence from a randomized experiment’, *Economics of Education Review* **38**, 51–63.
- Sacerdote, B. (2001), ‘Peer effects with random assignment: Results for Dartmouth roommates’, *The Quarterly journal of economics* **116**(2), 681–704.
- Schneeweis, N. (2015), ‘Immigrant concentration in schools: Consequences for native and migrant students’, *Labour Economics* **35**, 63–76.

-
- Schneeweis, N. & Zweimüller, M. (2012), ‘Girls, girls, girls: Gender composition and female school choice’, *Economics of Education review* **31**(4), 482–500.
- Shan, X. (2022), ‘The minority trap: Minority status drives women out of male-dominated fields’, *Unpublished manuscript* .
- Stoddard, O., Karpowitz, C. & Preece, J. (2020), ‘Strength in numbers: A field experiment in gender, influence, and group dynamics’, *IZA Discussion Paper* .
- World Bank (2022), ‘Gender statistics’, <https://data.worldbank.org/indicator/SL.TLF.CACT.FM.ZS?locations=OE>.

6 Appendix

6.1 Student's country of birth and native language

Table 12: Number (N) of students by country and cohort

Country of birth	2020-2021 (N)	2021-2022 (N)	Total (N)	Country of birth	2020-2021 (N)	2021-2022 (N)	Total (N)
Albania	1	0	1	Mauritania	1	0	1
Argentina	2	2	4	Mexico	3	2	5
Armenia	0	1	1	Nepal	1	0	1
Australia	2	1	3	Netherlands	2	0	2
Bahrain	0	1	1	Nigeria	2	0	2
Bangladesh	1	2	3	Norway	0	3	3
Belgium	0	1	1	Pakistan	5	5	10
Brazil	2	1	3	Panama	1	0	1
Bulgaria	2	0	2	Paraguay	1	0	1
Burma (Myanmar)	0	1	1	Peru	0	1	1
Canada	5	7	12	Philippines	2	1	3
Chile	0	3	3	Poland	2	2	4
China	26	21	47	Qatar	0	1	1
Colombia	1	5	6	Romania	1	1	2
Dominican Rep.	1	0	1	Russia	0	1	1
England	38	38	76	Saudi Arabia	0	1	1
Eritrea	0	1	1	Scotland	2	0	2
FYR Macedonia	1	0	1	Singapore	0	3	3
Finland	1	0	1	South Africa	0	1	1
France	9	5	14	South Korea	4	5	9
Germany	1	5	6	Spain	2	1	3
Ghana	0	1	1	Sri Lanka	0	1	1
Greece	2	0	2	Sudan	1	0	1
Hong Kong	3	6	9	Sweden	0	1	1
Hungary	1	0	1	Taiwan	0	1	1
India	12	13	25	Thailand	0	2	2
Indonesia	2	0	2	Turkey	3	1	4
Ireland	0	2	2	USA	18	20	38
Italy	7	10	17	Ukraine	2	0	2
Japan	2	2	4	Uruguay	1	1	2
Jordan	1	0	1	Utd Arab Emts.	1	1	2
Kazakhstan	1	0	1	Venezuela	1	1	2
Kenya	1	1	2	Vietnam	1	0	1
Lebanon	1	0	1	Wales	1	1	2
Lithuania	1	1	2	Zimbabwe	0	1	1
Luxembourg	0	1	1				
Malaysia	1	0	1	Total	186	191	377

Source: University administrative data records for the course

Table 13: Non-native English speakers Language scores (Linguistic Distance)

Native Language	N	Score	Native Language	N	Score
Arabic	4	1.5	Mandarin	6	1.5
Bengali	1	1.75	Mandarinchinese	1	1.5
Bulgarian	1	2	Nepali	1	1.75
Burmese	1	1.75	Norwegian	1	3
Cantonese	5	1.25	Polish	5	2
Chinese	38	1.5	Portuguese	3	2.5
Dutch	2	2.75	Punjabi	1	1.75
French	14	2.5	Rumanian	2	3
German	5	2.25	Russian	1	2.25
Greek	1	1.75	Spanish	23	2.25
Gujarati	1	1.75	Swedish	3	3
Hindi	3	1.75	Tagalog	1	2
Hungarian	1	2	Tamil	3	1.75
Indonesian	1	2	Telugu	1	1.75
Italian	14	2.5	Thai	1	2
Japanese	4	1	Turkish	3	2
Korean	7	1	Vietnamese	1	1.5
Malayalam	2	1.75	Total	182	1.92

Notes: N represent the total number of students who reported each language as their Native language. Score is the Chiswick & Miller (2005) measure of linguistic distance from each native language to English. The measure ranges from 1 to 3, with three being the most similar to English.

6.2 Survey Attrition

Although most students answered the baseline survey, there is some attrition at the point of the end survey. Table 14 presents descriptive statistics for the group of end survey respondents (Response=1), and non-respondents (Response=0). Table 15 presents the odd ratios of the logistic regression used for the inverse probability weighting.

Table 14: Covariate's mean value for respondents and non-respondents

Covariates	Means	
	Response=0	Response=1
Age	24.7	24.5
Female=1	0.79	0.80
Native English=1	.50	0.47
UK Studies=1	0.35	0.32
Highest level of education (Master) = 1	0.20	0.19
Proportion Native Speakers (Team)	45.8%	51.4%
Proportion Native Speakers (Seminar -O)	42.1%	47.8%
Proportion of Women (Team)	78.2%	80.5%
Proportion of Women (Seminar -O)	78.2%	81.6%
Expected Mark	71.5	73.1
Experience with Quantitative methods	4.0	4.2
Experience with Qualitative methods	5.7	5.9
Adviser Gender (female=1)	0.54	0.46
Adviser Native Language (English=1)	0.64	0.55
Total (N)	153	226

Notes: Seminar - O refers to other members in the seminar group (excluding individual's own team). Data Source: Baseline survey and administrative records

Table 15: Odd ratios for end survey response

	Response
Cohort	0.82*** (0.24)
Year Born	0.01 (0.03)
Female =1	-0.05 (0.28)
Native English =1	-0.18 (0.23)
UK studies =1	-0.24 (0.25)
Seminar Group	0.03 (0.02)
Constant	-1668.34*** (466.69)
<i>N</i>	355

Notes: Odd ratios. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, standard errors in parentheses.

6.3 Description of dependent variables

Table 16: Description of dependent variables

Variable	Description	Range	Data Source	Questionnaire Item
Academic performance				
Exam Marks	Mark on the course's final exam. This is the only assessment in the course.	0-100	Administrative records	
Perceptions of team dynamics				
"My voice was heard during group discussions"	The variable indicates the level of agreement with the statement "My voice was heard during group discussions"	0-10	End of year survey	Level of agreement from 0 to 10 with the following statements: (1) "My voice was heard during group discussions" (2) "Working in teams for SP401 made me more confident than before in voicing my view in future interactions" (2021-2022 only) (3) "Working in teams for SP401 made me more confident than before that my view will be heard in future interactions." (2021-2022 only)
"More confident in voicing my view"	The variable indicates the level of agreement with the statement "Working in teams for SP401 made me more confident than before in voicing my view in future interactions"			(2) "Working in teams for SP401 made me more confident than before in voicing my view in future interactions" (2021-2022 only)
"My voice will be heard"	The variable indicates the level of agreement with the statement "Working in teams for SP401 made me more confident than before that my view will be heard in future interactions"			(3) "Working in teams for SP401 made more confident than before that my view will be heard in future interactions." (2021-2022 only)
Identify as leader	Binary variable that indicates if a student indicates that the role that describes them best is Manager/Leader (value 1) or not (value 0) We construct this variable by combining the end survey leadership variable plus a baseline survey leadership variable. The baseline leadership variable takes value 1 if the student answers Manager/Leader, and zero otherwise. Then we compute the change in leadership as the difference between the two binary leadership variables (at end minus at baseline). Thus, the variable is equal to zero if there is no change, 1 if the student answers leader at end but not at baseline, and -1 if the student answers leader at baseline but not at the end survey.	0,1	End of year survey	"Which one of these roles do you think describes you best when you do group work?" (Baseline) "Which of these roles best describes your own role in your sub-group?" (End of year) (a) Manager/Leader: provides leadership and direction for the group, (b) Sceptic/Thinker: ensure the group avoids premature agreement, push the group to explore all possibilities, (c) Checker/Recorder: check for consensus among group members, record the group's solutions, (d) Conciliator: resolve conflicts, ensure that members feel 'safe' to give opinions"
Change in Leadership self-perception	The variable is constructed using peers' input. We asked students to match each of the roles with one or more members of their team.	-1,0,1	Baseline and End of year survey	Match the following role with one or more members of your team. You can include yourself. Manager/Leader: provides leadership and direction for the group (2020-2021 only)
Others identify as leader	Then, we constructed a binary variable that is equal to 1 if a student was mentioned as taking the Manager/Leader by at least one member of their team, and zero otherwise.	0,1	End of year survey	
Predictions of academic performance				
Expected mark at end	This variable measures the student's expectation of their exam mark at the end of the course. We construct this variable as the difference between the student's expected mark at the end and the start of the course (Expected Mark at end - Expected Mark at start)	0-100	End of year survey	From 0 to 100, what do you expect your final mark in this course to be?
Change in mark's expectations (Start to end)	We construct this variable as the difference between the student's exam mark and their expected mark at the end of the course (Mark - Expected Mark at end)	(-100) - 100	Baseline and End of year survey	
Difference between actual mark and expected mark		34 (-100) - 100	Administrative records and end of year survey	

Table 11: Regression coefficients: Change in mark’s expectations and Difference between actual mark and expected mark

	Change in mark’s expectations (Start to end)				Difference between actual mark and expected mark			
	(1) Women	(2) Men	(3) Native Speakers	(4) Non-Native Speakers	(5) Women	(6) Men	(7) Native Speakers	(8) Non-Native Speakers
Previous	2.39	-2.46	1.21	5.64**	3.93	9.27	8.67**	-3.74
UK Studies	(1.23)	(5.38)	(1.60)	(1.86)	(2.56)	(5.34)	(2.60)	(4.12)
Gender (Female=1)			-1.54 (3.12)	1.50 (2.86)			0.39 (2.62)	-0.33 (4.85)
Native English	1.18 (1.29)	12.69 (9.47)			5.54* (2.47)	5.04 (5.25)		
Percentage Native Speakers (Team)	-0.07** (0.02)	-0.16 (0.08)	-0.06 (0.05)	-0.11** (0.03)	0.03 (0.04)	-0.03 (0.08)	0.01 (0.05)	0.02 (0.05)
Percentage Native Speakers (Seminar -O)	-0.06* (0.03)	0.10 (0.11)	0.04 (0.04)	-0.09* (0.04)	0.02 (0.05)	0.15 (0.10)	-0.03 (0.04)	0.11 (0.06)
Percentage of Women (Team)	0.05 (0.03)	0.07 (0.08)	0.06 (0.03)	0.06 (0.05)	-0.03 (0.05)	-0.01 (0.09)	0.00 (0.05)	-0.04 (0.06)
Percentage of Women (Seminar -O)	0.01 (0.03)	-0.13 (0.09)	0.03 (0.03)	-0.06 (0.05)	0.01 (0.04)	0.00 (0.12)	-0.01 (0.05)	0.05 (0.07)
Language Score				-2.93 (1.88)				9.58** (2.88)
Constant	5.36 (6.52)	-27.44 (35.93)	-26.62 (20.91)	7.34 (11.29)	16.76 (9.50)	-8.13 (20.99)	-0.21 (14.24)	-12.58 (13.97)
N	173	44	103	101	172	44	103	100

Notes: All models include controls for: age, education level, experience with quantitative methods, experience with qualitative methods, seminar leader gender, seminar leader native English, adviser gender, adviser native English. The dependent variable “Change in mark’s expectations” corresponds to the difference between the student’s expected mark at the end and start of the course (Expected Mark at end – Expected Mark at baseline). Data source: baseline survey and end year survey. The dependent variable “Difference between actual mark and expected mark” is the difference between the student’s exam mark and their expected mark at the end of the course (Mark - Expected Mark at end). Seminar - O refers to other members in the seminar group (excluding individual’s own team). Data source: administrative records and end year survey. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, cluster standard errors in parentheses.