Gender in Teaching: Insights from Five Million Syllabi on Collaboration, Interdisciplinarity, and Reading Selections

Andrea Blasco* Elisa Orlandi†

Abstract

This study examines academics' preferences for teaching alone versus co-teaching, and how these choices impact teaching practices. Using a massive dataset of over five million syllabi, we explore how gender preferences shape instructors' team size and gender composition. Our findings reveal a strong tendency for gender homophily in co-teaching resulting in a significant underrepresentation of mixed-gender teams—half as many as expected under gender-neutral preferences. We also analyze the relationship between team gender composition and key course characteristics, such as interdisciplinarity, breadth and novelty of materials, and the gender of cited authors. Our results show that mixed-gender teams are significantly more interdisciplinary than same-gender teams. Additionally, classes taught by male-only teams cite fewer female authors than those taught by all-female teams, with mixed-gender teams not bridging this gap. Furthermore, solo female instructors and teams with at least one woman tend to reference more novel materials. These findings highlight new mechanisms contributing to gender bias in higher education and suggest that promoting mixed-gender co-teaching could enhance interdisciplinarity and lead to more balanced gender representation among academic authors.

Classification: TBA

Keywords: education | gender bias | teams | interdisciplinarity | novelty | syllabi

*European Commission, Joint Research Centre, Brussels, Belgium; (andrea.blasco@ec.europa.eu)

†University of Bologna, Bologna, Italy, and European Commission, Joint Research Centre, Brussels, Belgium.

1 Introduction

Universities, vocational schools, and other higher-education institutions are crucial in disseminating knowledge and developing human capital. Instructors shape this role through their choices in teaching subjects, assigned readings, and co-teaching arrangements. These decisions have a significant impact on student learning outcomes and students' career trajectories (1–3). However, they are not always optimal. Instructors can prefer working alone or with colleagues from similar backgrounds, limiting course diversity and interdisciplinarity, reinforcing gender stereotypes on certain subjects (4–6), or even limiting instructors' exposure to different teaching methodologies, hindering their professional growth (7). Despite its significance, research on the underlying factors driving instructors' choices in teaching remains limited primarily due to a lack of data.

While prior research has examined academics' decision-making in research collaborations, less is known about how these decisions shape teaching practices and, in turn, student outcomes. This study contributes to this literature by analysing a massive dataset of English course syllabi taught between 1990 and 2019 across several countries. This dataset allows for a systematic, quantative analysis of trends over multiple years (8). Each syllabus provides detailed information about the instructors, including their gender, the institution, and the course's content, including its field, assigned readings, and an overall description of its content. This data allows us to systematically examine how instructors of different genders form teams, shedding light on gender preferences and institutional constraints in teaching collaboration. Additionally, by computing various metrics related to teaching content, this data lets us assess how different configurations of the teaching teams — gender composition and size — associate with teaching practices, particularly interdisciplinarity and the assigned readings, controlling for institution, field, academic year, and other potential drivers of content.

Specifically, we aim to address two key questions. First, how often do academics choose to teach alone versus co-teach with a colleague, and to what extent are co-teaching teams composed of mixed-gender pairs? Second, how do the size and gender composition of teaching teams relates to key aspects of teaching: (1) the course's interdisciplinarity, (2) the novelty of assigned readings, and (3) the gender representation of the cited authors. Previous research has shown that these aspects shape students' learning outcomes and career trajectories (1–3), while also influencing instructors' professional development and promotion within their institution (7).

Our results reveal a significant role for gender in co-teaching practices, with mixed-gender teams occurring consistently less frequently than same-gender teams across institutions and fields. Furthermore, mixed-gender teams occur much less frequently than expected under a null model that forms teams in a gender-neutral manner while keeping fixed field-related and institutional constraints. These findings underscore a consistent and wide-spread underrepresentation of mixed-gender collaborations in teaching, that is not fully explained by unobserved differences across fields or institutions.

Our analysis further reveals a significant and strong association between the gender composition of a teaching team and the course's interdisciplinarity, with mixed-gender teams being more interdisciplinary than all-male teams or courses taught by individual instructors. We also find a significant association of team configurations with the fraction of cited female authors, with courses taught by a female instructor alone citing a higher fraction of female authors compared to courses taught by male instructors, with mixed-gender teams falling in between. Finally, we also find a trend in novelty, with female instructors assigning more recent readings than men, regardless of team size.

Previous research has shown that interdisciplinarity research is less likely to be funded (9), tends to attract fewer citations when it is highly interdisciplinary (10), and is correlated with the probability of publication in academic journals (11). Furthermore, students attending colleges with more interdisciplinary courses tend to earn higher earnings after graduation

(3, 12). We extend this work by looking at the association between gender composition and interdiscipilnarity in teaching, showing how mixed-gender teams occur less frequently but tend to be more interdisciplinary, suggesting that removing barriers to mixed-gender team formation may increase interdisciplinarity in teaching.

Our findings also contribute to the literature on gender dynamics in academia, particularly research team formation (13–15), which has shown significant gender homophily – a tendency to collaborate with colleagues of the same gender (16). It also examines gender differences in citation patterns, a driving factor of the persistent gender bias in academia (17), including tenure promotion (18), grant success (19), co-authorships (14), and peer recognition (20, 21). While prior research has largely focused on these areas, we shift attention to the citation gap in teaching. This issue may not only reinforce existing gender bias in academia but also shape students' learning outcomes and future career choices (22) – effects that are less well understood.

We also build on prior research on the underrepresentation of female-authored works in university curricula, which has been shown within specific fields (psychology and international relations) and at a small scale (23–25). Our results reveal a consistent and significant gap in cited works between female and male instructors, only partially addressed within mixed-gender co-teaching.

Finally, studies about team formation are especially relevant to our work. These studies have found a tendency of mixed-gender teams to perform better in various settings (26, 27). In research collaboration, for example, mixed-gender teams often obtain more citations, produce more novel research, publish in more prestigious journals and are more interdisciplinary (26, 28). Although we do not measure team "performance" as we lack data on students outcomes, our results show that mixed-gender co-teaching teams are consistently less likely to form, but, once created, they tend to deliver different outcomes, especially a higher interdisciplinarity, that previous literature has suggested having an impact on students outcomes as discussed above.

2 Materials and Methods

2.1 Data

We examined a corpus of over six million documents compiled by Open Syllabus (New York, US). This dataset was created through web extractions that identified syllabi from university websites, with a median confidence level of 99.8%. A tagging algorithm extracted key course details, such as the title, field, description, academic year, duration, and language, a list of anonymised instructors, and the assigned readings.[1] While the original dataset included syllabi in 49 languages, most documents (96%) were in English. For simplicity, we focused exclusively on these documents.[2]

The resulting dataset comprised 5.4 million syllabi from approximately 4,000 higher education institutions across fifteen countries from 1990 to 2019. OpenSyllabus classified these syllabi into 69 top-level fields derived from the U.S. Department of Education's CIP code classification.[3] About 2.9 million syllabi (53% of the total) listed readings matched with bibliographic sources, providing additional metadata about authorship information, journal, and publication year. The institution was matched to a list of more than 22,000 entities from the Research Organisation Registry, providing further metadata including the instutition's country and enrollment figures — the institutions in our sample account for over 35 million enrolled students today.[4]

[1] The documentation available at: https://docs.opensyllabus.org

[2] Our focus on English-language courses means that, while we have comprehensive data for English-speaking countries such as Canada, Ireland, the United States, and Great Britain, the sample in non-English-speaking countries tends to be more representative of internationally oriented universities. These are typically institutions offering programs in English or advanced-level courses, such as postgraduate programs or disciplines where English is the primary medium of instruction within traditional universities.

[3] The original CIP classification is available online: https://nces.ed.gov/ipeds/cipcode/browse.aspx?y=55

[4] For further information on how OpenSyllabus classified and matched the data, the related documentation is available online at https://docs.opensyllabus.org.

Table 1: Outcome variables

Name	Definition
Interdisciplinarity	Percentile rank of the course's interdisciplinarity score for the year.
Age of References	Percentile rank of the average publication age of assigned readings.
Ratio of Female Authors	Proportion of women authors in the assigned readings.

Each syllabus lists one or more instructors, with 76% of the syllabi listing a single instructor, 16% listing two, 4% listing three, and another 4% listing more than three instructors. Instructor gender was determined automatically by OpenSyllabus based on names, resulting in 52% male, 37% female, and 11% unknown categories. After excluding syllabi with unknown gender, the distribution was 58% male and 42% female instructors, which aligns closely with the 45% of female academic staff reported in OECD countries (29). The same inference method was used to determine the gender of the authors listed in the readings, resulting in 32% female and 56% male authors, with only 12% of unknown gender.

Outcome variables

We defined the following key outcome variables to analyse how different team configurations, such as gender and team size, relate to interdisciplinarity and characteristics of assigned readings, such as their publication age and the inclusion of female-authored works.

Interdisciplinarity

To estimate interdisciplinarity, we measured field overlap using course descriptions written in the syllabi. Following Evans et al. (3, 12), we assigned each course an interdisciplinarity score based on its description. This approach converts descriptions into "bags of words", where word frequencies are normalised with the inverse ratio of the term frequency to document frequency metric. A correlation matrix is then generated across different academic fields to measure the distance between fields. The interdisciplinarity score for each syllabus is computed by taking one minus the weighted average of the pairwise correlations with other syllabi, with weights equal to the conceptual proximity of different fields. This method ensures that syllabi associated with distant fields –either academically or conceptually– are considered more interdisciplinary. To scale this approach for millions of documents, we optimized for efficiency by using random subsamples for academic fields across academic years. The final interdisciplinarity score was averaged across multiple subsamples for robustness. See Supplementary Information (Section SI-2) for details. To ensure robust comparisons in our analysis, we computed the percentile rank of the interdisciplinarity score for each syllabus i:

$$\text{Interdisciplinarity}_i = \text{PR}_{yr}(\text{Interdisciplinarity Score}_i),$$

where PR_{yr} represents the percentile rank function applied to all syllabi within a given year yr.

Readings Selection

To investigate characteristics of the assigned readings, we calculated the following critical dimensions: the age of readings (a measure of "novelty") and the proportion of female authors in the assigned readings. First, we compute the number of references, N_i, as the total articles, books, and chapters listed in a syllabus i. This variable serves as a broad measure of a course's "breadth," as more assigned readings may indicate a more extensive or comprehensive curriculum. Then, we define the *Age of References* variable as the difference between the syllabus year (Year_i) and the average publication year of each assigned reading k:

$$\text{Age of References}_i = \text{PR}_{yr}\left(\text{Year}_i - \sum_{k=1}^{N_i} \text{Publication Year}_k / N_i\right),$$

where PR_{yr} represents the percentile rank function applied to all syllabi within a given year yr. This variable gives a proxy of how recent, or "novel," the readings are.[5] Finally, we define the *Ratio of Female Authors* as the proportion of female authors among all authors in the assigned readings:

$$\text{Ratio of Female Authors}_i = \frac{\text{Female Authors}_i + 1}{\text{Female Authors}_i + \text{Male Authors}_i + 2}.$$

Here, we add two pseudo-observations (one for each gender) to stabilize the ratio, preventing extreme values in cases with very few authors. This metric allows us to investigate whether gender and collaboration relate to the representation of female-authored work in teaching.

Simulated Teaching Collaborations

To examine whether the gender composition of co-teaching teams deviates from what one would expect under gender-neutral matching, we employ the following Monte Carlo approach. Drawing from a methodology developed elsewhere (30), we counted the frequency of courses taught individually and the frequency of gender combinations (male-male, female-male, etc.) of those taught by two instructors, disaggregating these data per field, institution, and academic year. Then, we compared these combinations against those expected by chance in a gender-neutral composition network, where all instructors are switched randomly within a given institution, field, and academic year. This approach matches our assumption that forming teams between institutions and fields is limited, at

[5] While more sophisticated methods to measure novelty are available (30, 31), we opt for a simpler metric. Teaching innovation tends to be more incremental, and computationally intensive novelty indicators are impractical for large-scale datasets like ours.

least within a single academic year, while forming teams within the same field and institution is attainable.

The switching algorithm preserves the total gender counts and the distribution of teams. This ensures that a course with a given number of instructors in the original data will have the same number of instructors in the randomised network. Similarly, an institution with a given number of male and female instructors teaching in each field will have the same number of male and female instructors. The only difference between the randomised and the original data will be the gender composition of the teams. Therefore, in the randomized network, instructors form teams as if they were unaware of the gender. See the details in the Supplementary Information (Section SI-1).

Regression analysis

We analyse how different teaching team configurations, $j \in \{F, M, MM, MF/FM, FF\}$, where F = female alone, M = male alone, MM = two males, FM/MF = mixed gender, FF = two female instructors, associate with course outcomes across academic years, $t = 1999, \cdots, 2020$. The outcomes of interest include: (1) interdisciplinarity, (2) the average age of references, and (3) the proportion of cited female authors. To estimate these effects, we employ the following linear mixed-effects model:

$$Y_{j,t} = \alpha_t + \text{Team}_{j,t} + \text{STEM}_t + \eta_t + \text{Country}_t + \text{Enroll}_t + \delta_t + \epsilon_{j,t}.$$

Where:

- $Y_{j,t}$ is the outcome variable for team configuration j in year t,
- α_t is a fixed effect for the academic year,
- $\text{Team}_{j,t}$ is a fixed effect for team configuration,
- $STEM_t$ is a fixed effect for STEM courses,

- η_t is a random intercept for each of the 69 academic fields,
- Country_t and Enroll_t are categorical fixed effects for the institution's country and enrollment size, respectively,
- δ_t is a random intercept for each of approximately 4000 unique institutions,
- $\epsilon_{j,t}$ is the residual error term.

3 Results

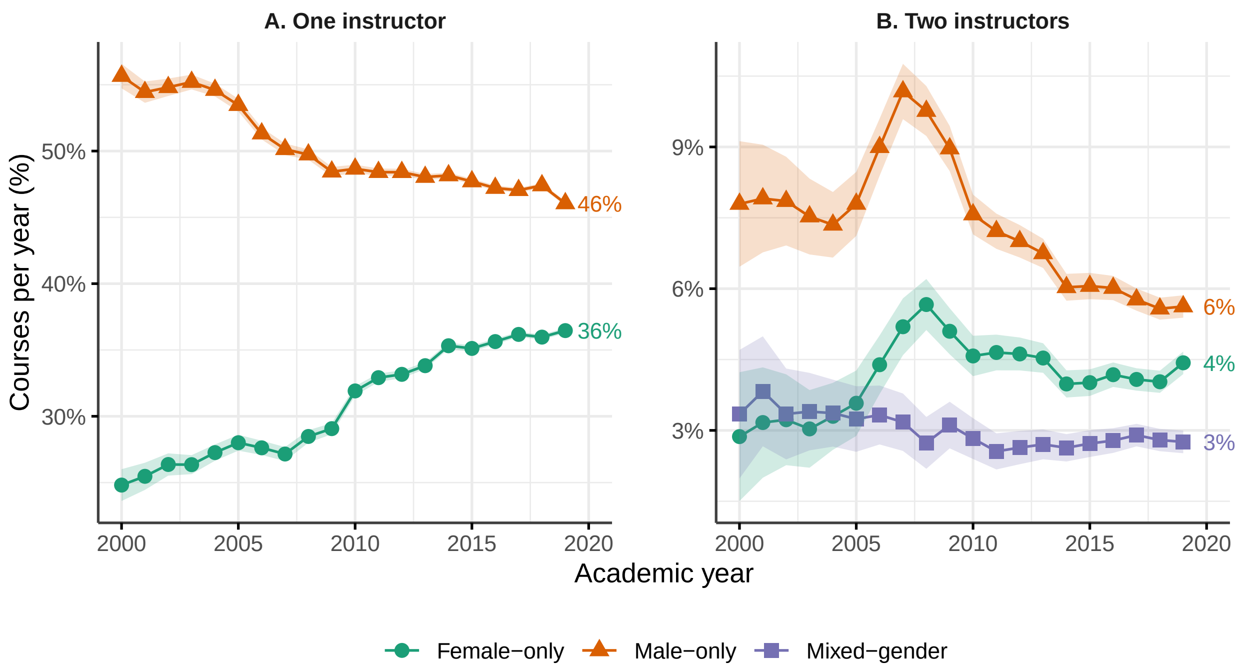

Figure 1: Evolution of Teaching Teams by Gender Composition and Size. Trends in course gender composition from 2000 to 2019 reveal a steady increase in women's participation (both in solo-taught and two-instructor courses), while the proportion of mixed-gender teams has remained relatively stable over time. (A) Percentage of solo-taught courses by instructor gender over time. (B) Percentage of two-instructor courses by gender composition over time. N = 5.1 million courses. Shaded area represent 95% confidence interval of the proportion per academic year.

Figure 1 presents trends in the instructors' gender composition and team size in university courses from 2000 to 2019 for courses with one or two instructors. Panel A shows courses with one instructor, where male-only courses (orange triangles) consistently dominated but declined from around 60% in 2000 to 46% in 2019. Over the same period, female-only courses (green circles) increased from roughly 25% to 36%. Panel B shows courses with two instructors, where male-only courses also decreased—from about 8% in the early 2000s to 6% in 2020, while female-only (green cricles) and mixed-gender (purple squares) instructor teams remained relatively stable, each accounting for approximately 4% and 3%, respectively, in recent years. Together, these results reveal a gradual but consistent shift toward greater gender diversity in course instruction over time, particularly a decline in male-only instruction and a rise in female-only instruction, especially in single-instructor courses.[6] Notably, mixed-gender teaching teams (3%) are consistently less common than same-gender teams (10%).

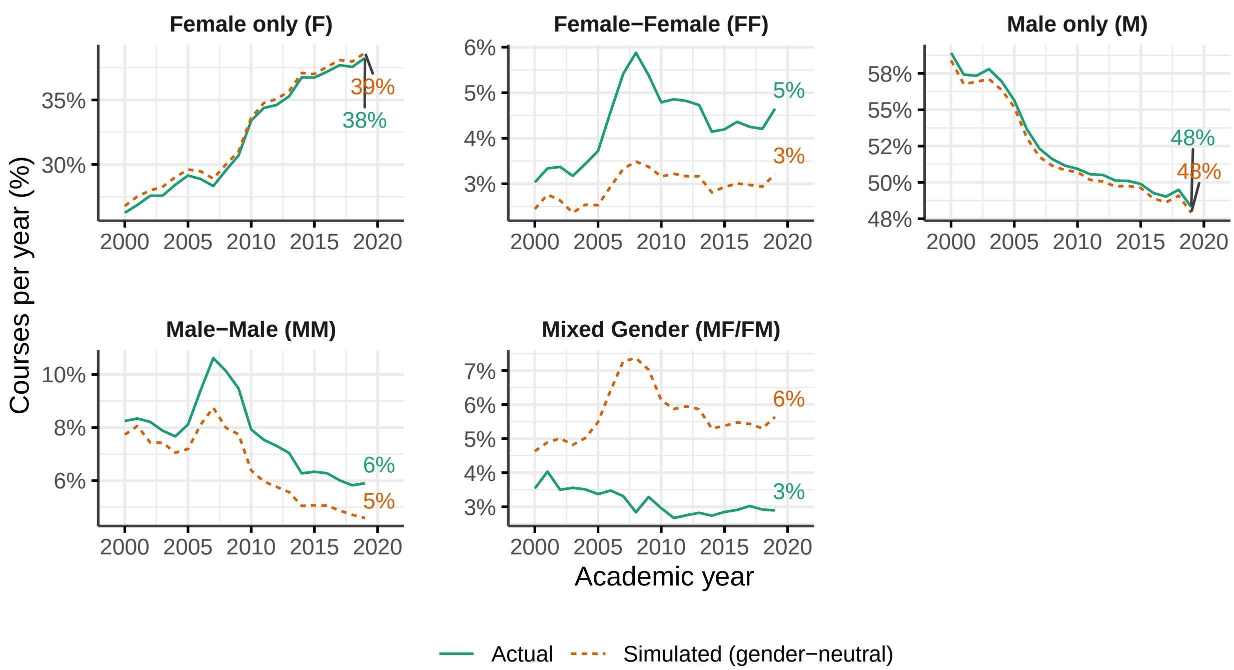

Figure 2: Comparison of gender composition between actual courses and courses simulated with Monte Carlo. Simulations ensure gender neutrality while keeping constant the institutional constraints (i.e., number of teams by size, institution, academic year, and 69 academic fields). Simulated mixed-gender teams consistently exceed observed proportions, highlighting persistent barriers to forming gender-diverse teaching teams.

[6]This overall trend aligns with the increase in women holding academic positions worldwide. In 2000, women constituted approximately 35% of academic staff worldwide. By 2022, this figure had risen to about 44%, according to data from the World Bank; The data are available at: https://data.worldbank.org/indicator/SE.TER.TCHR.FE.ZS

To assess whether gender imbalances in co-teaching teams persist after accounting for differences across fields and institutions, Figure 2 compares the actual size and gender compositions of teaching teams in university courses from 2000 to 2019 (solid green lines) with simulated compositions (dashed orange lines) generated from a gender-neutral Monte Carlo model. This model randomly shuffles instructors while preserving the overall distribution of course loads per year by institution and academic field but ignoring gender-based preferences or constraints, as discussed before (see Section 2).

The results reveal that actual data systematically deviate from simulated expectations: female-female (FF) and male-male (MM) courses occur more freequently in actual data than in simulations by 66% and 20%, respectively, in 2019. By contrast, mixed-gender collaborations (MF/FM) are consistently underrepresented in the actual data (3%) compared to simulations (6%). These findings suggest that significant social or institutional dynamics—beyond chance—reinforce same-gender pairings, especially among women. Such dynamics may include gender-based homophily, mentoring networks, or departmental assignment practices.

Figure 3 disaggregates the comparison between actual and simulated mixed-gender (MF/FM) courses by academic field. Nearly all fields show clear underrepresentation of mixed-gender teams (relative to expectations) by two to four times. For instance, in Medicine (Health and Welfare), the actual share of mixed-gender collaborations is 4.8% versus an expected 9.1%, indicating a large imbalance. Similar gaps are evident in Law (3.8% actual vs 6.4% simulated) and Linguistic (4.3% vs 6.7%), underscoring the pervasiveness of barriers to mixed-gender teams across different academic fields. Remarkably, we see no differences in these patterns between male- vs female-dominated fields (e.g., Engineering or Accounting vs Nursing or Chemistry), suggesting that the underrepresentation of mixed-gender teams may be a structural feature rather than one driven by field-specific gender imbalances.

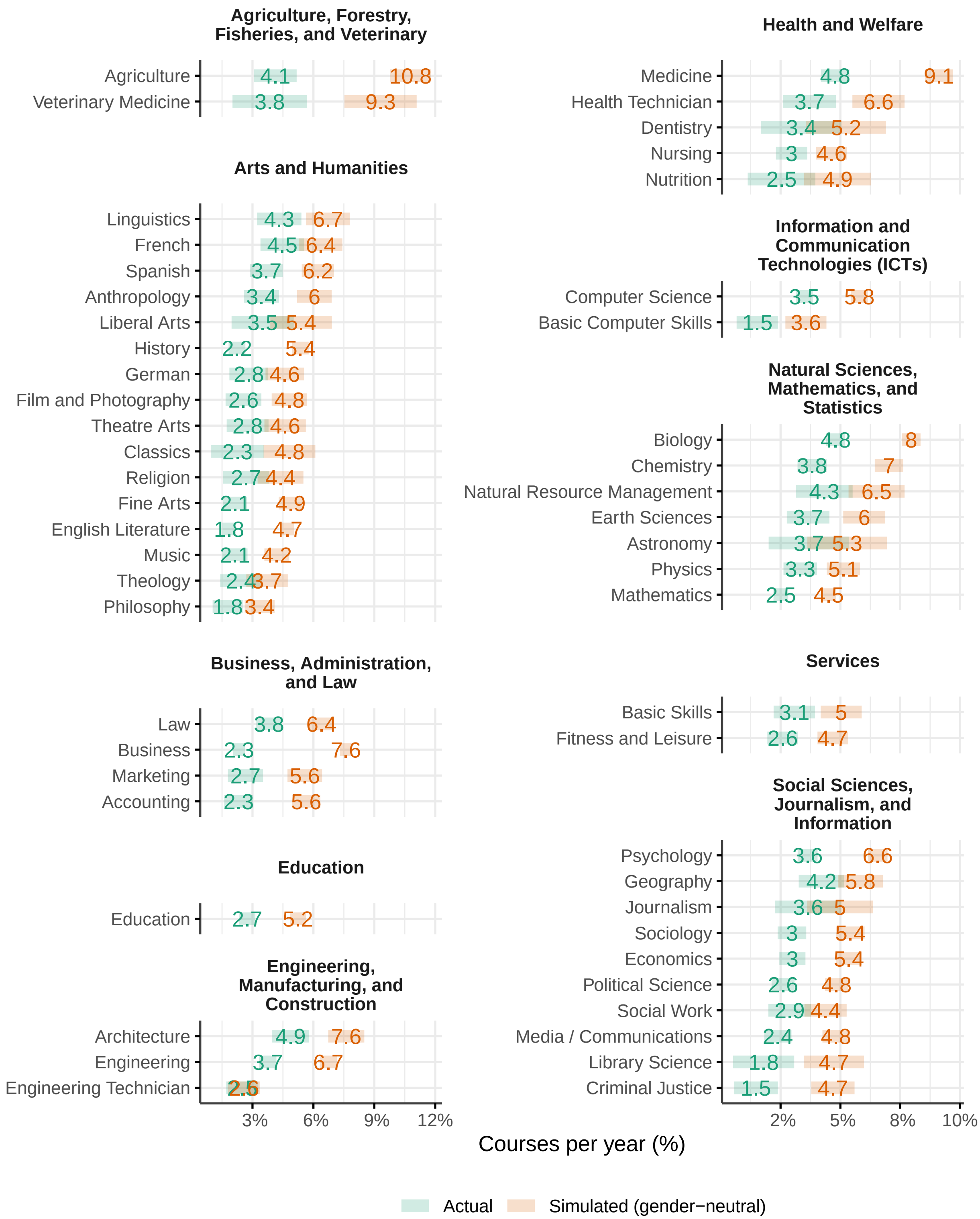

Figure 3: This figure disaggregates the comparison of proportions between actual (green) and simulated (orange) mixed-gender (MF/FM) courses by academic field. The shaded bar indicates 95% confidence level of the proportion.

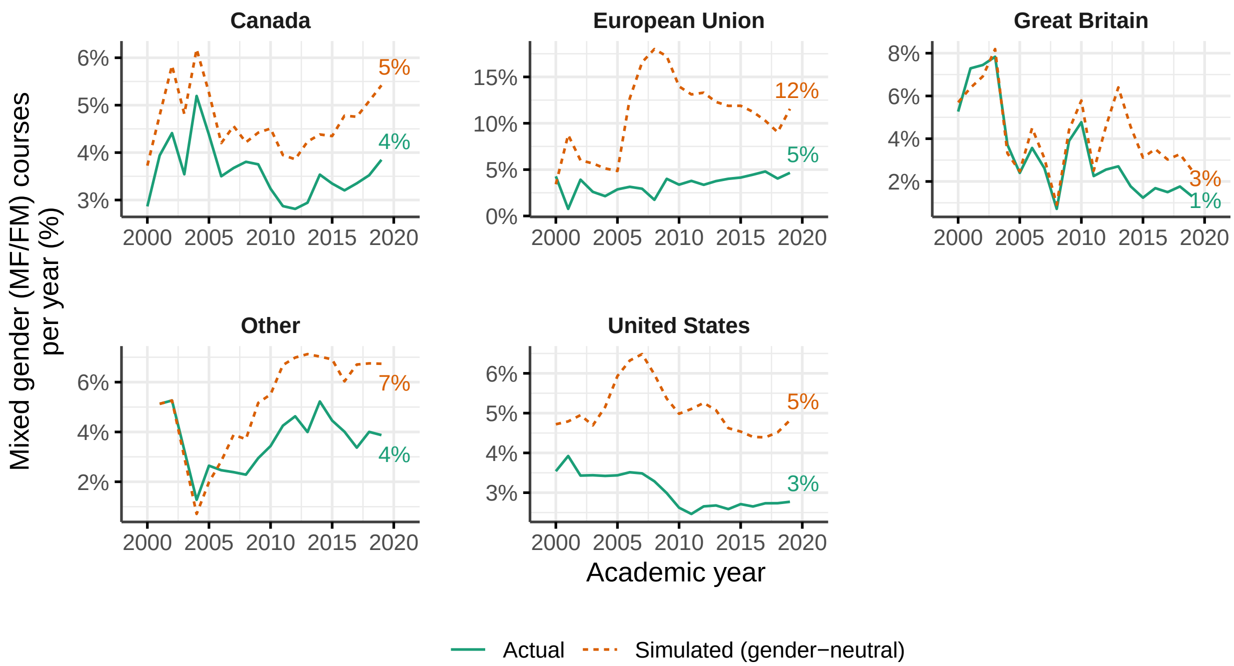

Figure 4: This figure disaggregates the comparison of proportions between actual (green) and simulated (orange) mixed-gender courses (MF/FM) by country. Simulated mixed-gender teams consistently exceed observed proportions in all countries. EU countries are: Austria, Denmark, France, Germany, Ireland, Italy, Netherlands, Poland, Portugal, Spain and Sweden.

Figure 4 illustrates that the proportion of mixed-gender teams in the simulations consistently exceeds the observed proportions across various geographic regions, despite notable variation in the magnitude of these differences. In 2019, for example, mixed-gender teams are relatively rare in Great Britain (1%) compared to Canada (4%). However, in the simulations, the proportion of mixed-gender teams in both countries increases substantially—1% vs 3% and 4% vs 5%, respectively. In 2019, the largest discrepancy between simulated and observed proportions is found in EU countries from 5% to 12%. Overall, these patterns underscore the robustness of the findings across diverse geographic contexts.

3.1 The Impact on Course Materials

To examine how teaching team size and gender composition are associated with important aspects of course materials, we analyse three key metrics: the percentile rank of the interdisciplinarity score per year ("interdisciplinarity"), the average publication year of the readings ("age of readings"), and the share of female authors cited in the assigned readings ("share of female authors"). As discussed before, we use linear mixed-effect regressions to study these relationships while accounting for unobserved differences across academic fields, institutions, time, and other relevant controls.

3.1.1 Interdisciplinarity

Figure 5 illustrates the differences in interdisciplinarity across team configurations. Our results indicate that there are no systematical differences in interdisciplinarity between female and male instructors when they are teaching alone: in recent years women tend to be more interdisciplinary, but this effect is not consistent over time. Conversely, mixed-gender teams tend to exceed in interdisciplinary all-male teams both individual and with two male instructors. This finding is consistent over time. At the same time, we find no consistent evidence of a difference between mixed gender and teams with two female

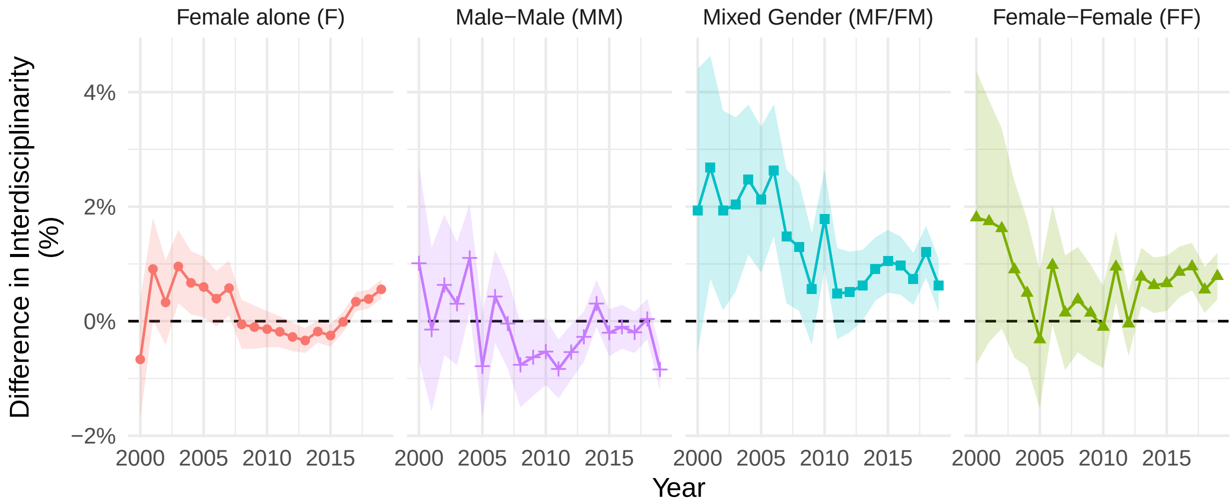

Figure 5: This figure illustrates the yearly difference in interdisciplinarity relative to courses taught by a single male instructor (baseline = 0) for each team configuration. The values are regression coefficients obtained separately for each academic year, with controls for country and field. Interdisciplinarity is expressed as the percentile rank of each course's interdisciplinarity score within its cohort; positive values therefore indicate a greater interdisciplinarity. The shaded area represents 95% confidence intervals.

instructors. These results point to a consistent association between gender diversity and interdisciplinarity.

3.1.2 Citing Women Authors

Figure 6 highlights significant and consistent gender association with the percentage of female authors over total authors cited in course readings per year, controlling for academic field and country.[7] Our results show that all-female courses consistently cite a higher fraction of female authors than all-male ones in the same year, with an effect that decreases over time going from 6% in 2000 to 3% in 2019 (with minor or insignificant differences between courses taught by one vs. two instructors). Mixed-gender courses are somewhat in between, as they cite a larger share of women than all-male courses, but less than all-female courses. Overall, these results indicate a consistent association between gender and cited patterns, suggesting that readings tend to cluster based on the gender of the

[7] This analysis excludes syllabi where no readings were matched with the available bibliographic sources and, within the matched readings, excluding the references where the authors' gender remained unidentified.

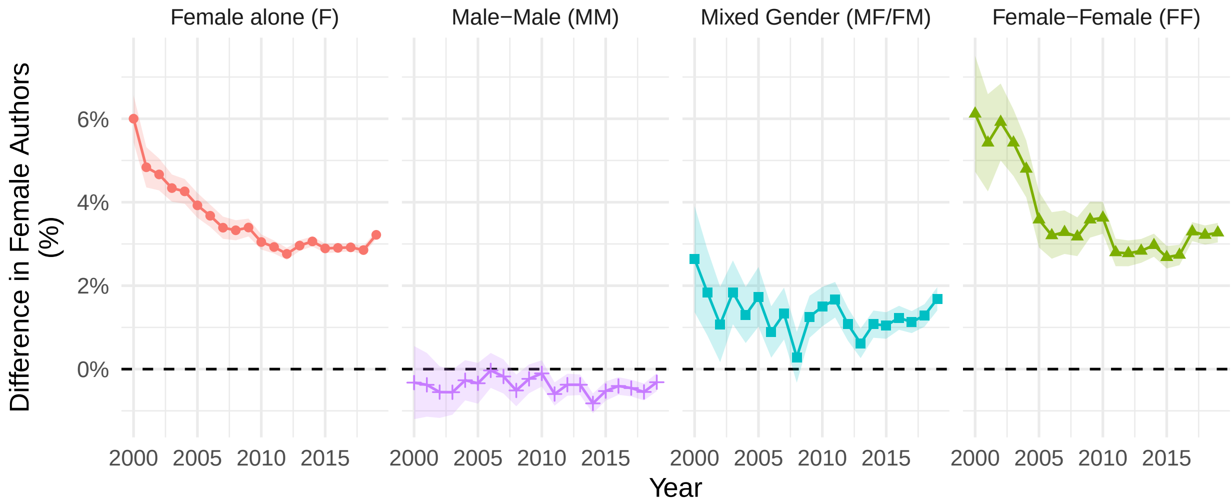

Figure 6: This figure plots the yearly difference in the share of female-authored readings relative to courses taught by a single male instructor (baseline = 0), for each team configuration. The values are regression coefficients obtained separately for each academic year, with controls for country and field. The outcome is the percentage-point gap in the proportion of female authors cited in the course readings; positive values therefore indicate a larger share of female authors over total authors associated with a given team configuration in that year. The shaded area represents 95% confidence intervals.

instructors and mixed-gender teams could promote greater diversity in assigned readings.

3.2 Age of Readings ("novelty")

Figure 7 illustrates significant gender differences in the publication age of the selected readings.[8] Between 2000 and 2019, courses taught by a single female instructor tend to assign newer readings than otherwise similar courses taught by a single man in that year, with a difference that goes from -3% in 2000 to -1% in 2019. Similarly, courses led by two women also assign newer material than those taught by two men, with a decreasing difference. Mixed-gender teams tend to assign newer readings than the baseline in few years, but not consistently in all the years. There are no significant differences between courses taught by two male instructors and the single-male baseline. Overall, these findings suggest that women tend to assign more novel readings, although this gender difference tend to decrease over time.

[8]This analysis includes only courses with matched bibliographic data.

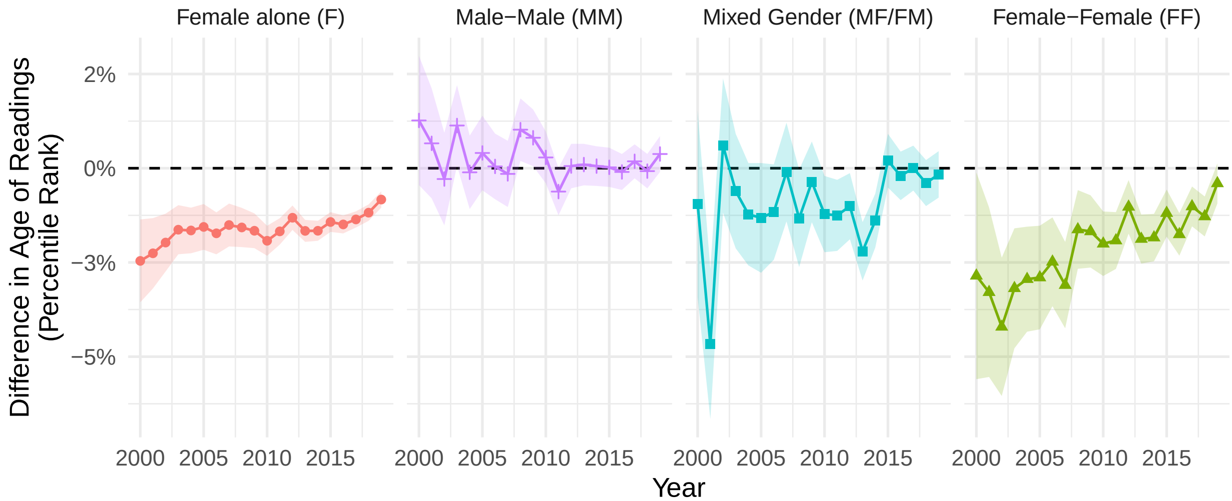

Figure 7: This figure illustrates the yearly difference in the percentile rank of the Age of References relative to courses taught by a single male instructor (baseline = 0), for each team configuration. The Age of Reference variable is the syllabus' year minus the average publication year of the assigned readings. The coefficients are estimated separately for each academic year with controls for country and broad field; positive values indicate that the team configuration assigns newer readings compared with the male-alone baseline in the same year. The shaded area represents 95% confidence intervals.

3.3 Robustness

To ensure the reliability and robustness of our results, we conducted several additional analyses. In addition to reporting results from separate regressions on subsets of data by year, we also estimated mixed-effects regressions using the full dataset. These results are broadly consistent with those presented in the paper.

Furthermore, while the reported model specification is our preferred one, we explored several alternatives. First, we log-transformed our dependent variables – the age of readings, interdisciplinarity score, and count of female authors — to explore potential non-linear relationships, particularly relevant for disciplines with very old readings (e.g., history)

Second, we employed a quasi-poisson mixed-effect model instead of the reported linear mixed-effect regression to better account for non-linearities in the count of female authors. We also trimmed the data to exclude outliers, such as courses with an unusually high number of readings. These changes yield similar results.

We further tested alternative model specifications such as considering the number of instructors as a numeric variable interacted with the proportion of female instructors to explore a linear association with the dependent variables. While we observe similar results, the linear association could be misleading when extrapolated beyond two instructors. So, we preferred to keep team configurations as fixed effects, as reported here.

Finally, we further explored specifications distinguishing between male-led and female-led courses, using the gender of the first listed instructor as a proxiy for seniority. Overall, results appear robust to these additional analyses.

4 Discussion

Our analysis of approximately five million syllabi from over 4,000 universities revealed several key findings. First, we show that, while the share of classes taught by female instructors has been increasing over the last twenty years, the fraction of mixed-gender co-teaching classes has remained consistently low (3%) and significantly below that of same-gender teams (10%). This trend can be explained by a limited gender diversity in certain fields or administrative practices that restrict opportunities to form mixed-gender collaborations. But it can also arise from gender preferences in team formation, whereby instructors tend to form teams with same-gender partners ("homophily"), a tendency already observed in research collaborations and other academic domains as discussed in the Introduction.

We further show that the underrepresentation of mixed gender teams persists even after accounting for unobserved differences across fields, institutions, and years. Specifically, we employ Monte Carlo simulations that randomly form teams in a gender-neutral manner, allowing us to control for important institutional constraints, such as maintaining constant the course loads across institutions and across 69 academic fields within a given year. By comparing simulated with actual teams, we find that instructors tend to partner with same-gender colleagues more than twice as expected between 2005 and 2019. This result is robust across academic fields, although it is more pronounced in certain fields (Chemistry, Business, Medicine). Our findings also reveal that the underrepresention of mixed-gender teams is consistent across institutions in different countries, with the strongest effect in EU countries, where only 5% of courses are taught by mixed-gender teams, compared to an expected 12%.

Our findings further highlight how different team configurations are associated with key aspects of teaching: (1) mixed-gender teams tend to be more interdisciplinary than all-male teams, (2) all-female teams tend to assign more novel readings, and (3) mixed-gender teams tend to cite a higher share of female authors than all-male teams, but less than all-female teams. These associations are not driven by unobserved differences across fields, institutions, and years that we account for in the regressions. Furthermore, these findings highlight the importance of understanding the mechanisms driving team formation in teaching.

Our analysis bears several limitations. First, the simulations could be improved by more granular information on the course's subfields that go beyond the 69 top-level fields identified in the dataset. If male and female instructors tend to concentrate in different subfields (e.g., Econometrics and Macroeconomics, within Economics), the observed underrepresentation of mixed-gender teams may reflect a limited availability of instructors of the other gender, especially for advanced courses where teaching teams are more likely to form around narrower areas of expertise. From a policy perspective, this limitation is important as we currently cannot disentangle whether the underrepresentation derives from gender-based preferences towards subfields, limiting availability, or a tendency to seek same-gender partners, homophily. Further research is needed to fully disentangle these underlying mechanisms.

Similarly, another limitation is the lack of data on potential determinants of team formation that operate within institutions, such as instructor experience or academic rank. Hierarchical differences or administrative structures may contribute to the observed gap in cross-gender collaborations — for example, if newly appointed female instructors are less likely to be selected by higher-ranking male instructors. However, our results illustrate that the under-representation of mixed-gender teams persists even when co-teaching teams are disaggregated by male-led and female-led courses (as measured by the oreder of appearance in the syllabus) demonstrating that the trend holds regardless of hierarchical differences. Therefore, while we cannot fully exclude other institutional constraints, the observed gap in mixed-gender teams likely reflects strong gender-based homophily.

Our findings underscore a significant and consistent underrepresentation of gender-mixed teaching teams, carrying serious implications. Limited mixed-gender collaborations in teaching may limit the exchange of information among instructors of different gender and restrict students' access to diverse courses. These dynamics, in turn, could hinder interdisciplinary engagement or reinforce structural disadvantages for women within academic networks. Our work thus identifies a potential additional driver of gender bias, complementing existing work on gender disparities in various settings within academia, such as research collaboration, access to funding, and student gender stereotypes.

Overall, this study provides critical insights into the market of academics. As for research teams, teaching teams can impact instructors in multiple ways. First, junior faculty may have access to experience and advice from senior faculty, which could lead to a higher chance of promotion. Secondly, diverse teaching teams may inspire research collaborations, including interdisciplinary work. Finally, teaching assistants can actively search for mentors and may benefit from these collaborations when they apply for PhD programs or in the labour market. Therefore, understanding how teaching teams are formed can help explain career trajectories.

SI Supporting Information

SI-1 Simulating Gender-Neutral Courses

We employed a simlar methodology to that developed by (30) for the analysis of academic citations. We counted the frequency of gender combinations (male-male, female-male, etc.) of each syllabus per field and academic year. We compared these combinations against those expected by chance, using a gender team composition network.

In this network, for a given field, institution, and academic year, we switched all the instructors using a Monte Carlo algorithm. The switching algorithm preserves the total gender counts and the distribution of team size. This ensures that a course with n instructors in the original data will have the same number of instructors in the randomised network. Similarly, an institution with m male instructors and f female instructors teaching in a specific field will have the same number of male and female instructors. The only difference between the randomised and the original data will be the gender composition of the teams.

SI-2 Interdiscipline Similarity

We measured interdisciplinarity using text similarity between syllabi as in (12) and (3). We transformed text from course descriptions into "bags of words," with term frequences (TF) normalised using the inverse document frequency (IDF). For each year, we calculated the TF-IDF scores for all syllabi and for 69 academic fields, using concatenated descriptions for the fields. We then computed the weighted average of the cosine similarity, $\cos(i, f)$, between each course i and field f, where the weight is based on the similarity between field f and the course's closest field $f_{\max}$. Specifically, $w(f_{\max}, f) = \cos(f_{\max}, f)$, where $f_{\max}$ is the field that has the highest cosine similarity with course i (i.e., $f_{\max} = \arg\max_f \cos(i, f)$) in that academic year.

Thus, the interdisciplinarity score for course i is:

$$\text{interdisciplinarity score}_i = 1 - \frac{\sum_{f \in (1,69)} \cos(f_{\text{max}}, f) \cdot \cos(i, f)}{\sum_{f \in (1,69)} \cos(f_{\text{max}}, f)}.$$

To reduce computational costs, we "bootstrap" the field-by-field cosine similarity matrix by using a 10% random sample of syllabi for each academic year. We repeated the sub-sampling procedure ten times and averaged the results.

SI-3 Addtional Figures

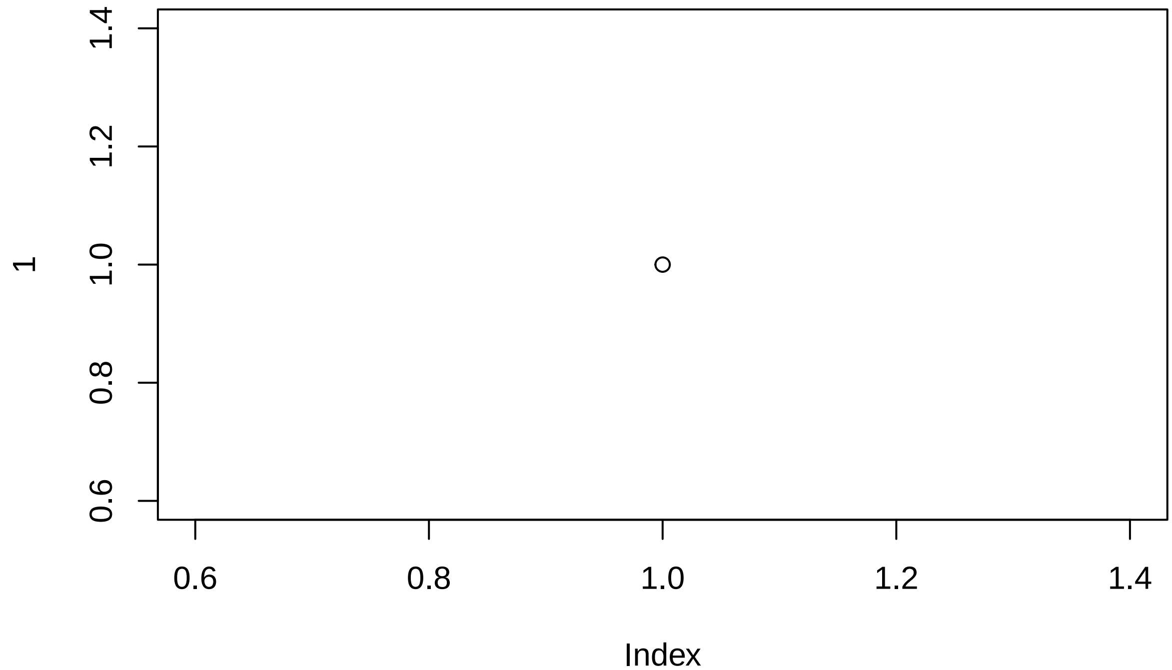

Figure SI-1: TBA

References

1. I. Bakkenes, J. D. Vermunt, T. Wubbels, Teacher learning in the context of educational innovation: Learning activities and learning outcomes of experienced teachers. *Learning and instruction* **20**, 533–548 (2010).

2. M. H. Vembye, F. Weiss, B. Hamilton Bhat, The effects of co-teaching and related collaborative models of instruction on student achievement: A systematic review and meta-analysis. *Review of Educational Research* **94**, 376–422 (2024).

3. S. Han, J. LaViolette, C. Borkenhagen, W. McAllister, P. S. Bearman, Interdisciplinary college curriculum and its labor market implications. *Proceedings of the National Academy of Sciences* **120**, e2221915120 (2023).

4. N. Ellemers, Gender stereotypes. *Annual review of psychology* **69**, 275–298 (2018).

5. P. Lockwood, “Someone like me can be successful”: Do college students need same-gender role models? *Psychology of women quarterly* **30**, 36–46 (2006).

6. S. Canaan, P. Mouganie, The impact of advisor gender on female students’ STEM enrollment and persistence. *Journal of Human Resources* **58**, 593–632 (2023).

7. L. R. Buckingham, A. López-Hernández, B. Strotmann, Learning by comparison: The benefits of co-teaching for university professors’ professional development in *Frontiers in Education*, (Frontiers Media SA, 2021), p. 776991.

8. A. Nowogrodzki, Mining the secrets of college syllabuses. *Nature* **539**, 125–126 (2016).

9. L. Bromham, R. Dinnage, X. Hua, Interdisciplinary research has consistently lower funding success. *Nature* **534**, 684–687 (2016).

10. V. Larivière, Y. Gingras, On the relationship between interdisciplinarity and scientific impact. *Journal of the American Society for Information Science and Technology* **61**, 126–131 (2010).

11. S. Xiang, D. M. Romero, M. Teplitskiy, Evaluating interdisciplinary research: Disparate outcomes for topic and knowledge base. *Proceedings of the National Academy of Sciences* **122**, e2409752122 (2025).

12. E. D. Evans, Measuring interdisciplinarity using text. *Socius* **2**, 2378023116654147 (2016).

13. S. Milojević, Principles of scientific research team formation and evolution. *Proceedings of the National Academy of Sciences* **111**, 3984–3989 (2014).

14. A. Boschini, A. Sjögren, Is team formation gender neutral? Evidence from coauthorship patterns. *Journal of Labor Economics* **25**, 325–365 (2007).

15. H. B. Love, A. Stephens, B. K. Fosdick, E. Tofany, E. R. Fisher, The impact of gender diversity on scientific research teams: A need to broaden and accelerate future research. *Humanities and Social Sciences Communications* **9**, 1–12 (2022).

16. M. Kwiek, W. Roszka, Gender-based homophily in research: A large-scale study of man-woman collaboration. *Journal of Informetrics* **15**, 101171 (2021).

17. N. Boivin, S. Täuber, U. Beisiegel, U. Keller, J. G. Hering, Sexism in academia is bad for science and a waste of public funding. *Nature Reviews Materials* **9**, 1–3 (2024).

18. K. Weisshaar, Publish and perish? An assessment of gender gaps in promotion to tenure in academia. *Social Forces* **96**, 529–560 (2017).

19. G. Bedi, N. T. Van Dam, M. Munafo, Gender inequality in awarded research grants. *The Lancet* **380**, 474 (2012).

20. H. Sarsons, Recognition for group work: Gender differences in academia. *American Economic Review* **107**, 141–145 (2017).

21. D. Card, S. DellaVigna, P. Funk, N. Iriberri, Gender differences in peer recognition by economists. *Econometrica* **90**, 1937–1971 (2022).

22. D. Card, C. Domnisoru, S. G. Sanders, L. Taylor, V. Udalova, "The impact of female teachers on female students' lifetime well-being" (National Bureau of Economic Research, 2022).

23. L. J. Skitka, Z. J. Melton, A. B. Mueller, K. Y. Wei, The gender gap: Who is (and is not) included on graduate-level syllabi in social/personality psychology. *Personality and Social Psychology Bulletin* **47**, 863–872 (2021).

24. J. Colgan, Gender bias in international relations graduate education? New evidence from syllabi. *PS: Political Science & Politics* **50**, 456–460 (2017).

25. J. K. Harris, M. A. Croston, E. T. Hutti, A. A. Eyler, Diversify the syllabi: Underrepresentation of female authors in college course readings. *PLoS One* **15**, e0239012 (2020).

26. Y. Yang, T. Y. Tian, T. K. Woodruff, B. F. Jones, B. Uzzi, Gender-diverse teams produce more novel and higher-impact scientific ideas. *Proceedings of the National Academy of Sciences* **119**, e2200841119 (2022).

27. J. B. Bear, A. W. Woolley, The role of gender in team collaboration and performance. *Interdisciplinary science reviews* **36**, 146–153 (2011).

28. A. Maddi, Y. Gingras, Gender diversity in research teams and citation impact in economics and management. *Journal of Economic Surveys* **35**, 1381–1404 (2021).

29. OECD, "What is the profile of academic staff and what is the student-academic staff ratio?" in *Education at a Glance 2022: OECD Indicators*, (OECD Publishing, 2022).

30. B. Uzzi, S. Mukherjee, M. Stringer, B. Jones, Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).

31. J. Wang, R. Veugelers, P. Stephan, Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy* **46**, 1416–1436 (2017).