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 widespread 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 im-

pact 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. Consequently, in non-English countries, our sample is more representative of advanced courses, such as postgraduate programs or disciplines where English is the primary language of instruction.

The resulting dataset comprised 5.4 million syllabi from approximately 4,000 higher education institutions across fifteen countries from 1990 to 2019. These syllabi were classified into 69 top-level fields derived from the U.S. Department of Education's CIP code classification.[2] 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.

Each syllabus lists one or more instructors, with 76% of the syllabi listing a single instructor,

[1] The documentation available at: https://docs.opensyllabus.org
[2] https://nces.ed.gov/ipeds/cipcode/browse.aspx?y=55

16% listing two, 4% listing three, and another 4% listing more than three instructors. Instructor gender was determined automatically 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% unknown gender.

Outcome variables

We defined the following key outcome variables to analyze the impact of the instructors' gender composition and collaboration on interdisciplinarity and readings selections.

```
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
```

```
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
```

```
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
```

```
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
## Warning in attr(.knitEnv$meta, "knit_meta_id"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
```

```
## Warning in attr(x, "align"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
```

```
## Warning in attr(x, "format"): 'xfun::attr()' is deprecated.
## Use 'xfun::attr2()' instead.
## See help("Deprecated")
```

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”,

Table 1: Outcome variables

Name	Definition
Interdisciplinarity	Percentile rank of a course's interdisciplinarity score for the given year.
Number of References	Total readings listed in a syllabus.
Age of References	Percentile rank of the publication age of readings listed in a syllabus.
Ratio of Female Authors	Proportion of female authors cited in the readings.

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 readings selection, we calculated three critical dimensions: the breadth, novelty, and female-authors representation. First, we define *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. Second, we define *Age of References* 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.[3] 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.

[3] 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.

Testing Gender Preferences in Teaching Collaborations

To test the hypothesis of gender preferences shaping teaching, especially co-teaching practices, we employ a montercalo approach. Drawing from the 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, using a gender team composition network. In this network, for a given institution, field, and academic year, all instructors were switched using Monte Carlo simulation. See Supplementary Information (Section SI-1) for details. This approach matches our assumption that mobility between institutions and fields is limited within a single academic year, while the within-institution mobility is possible.

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 instructors 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.

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, shape course outcomes across academic years, $t = 1999, \cdots, 2020$. The outcome of interest include (1) interdisciplinarity, (2) the average age of references, (3) the proportion of cited female authors, and (4) the

total number of references. 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 the 2000 unique institutions,
- $\epsilon_{j,t}$ is the residual error term.

3 Results

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)

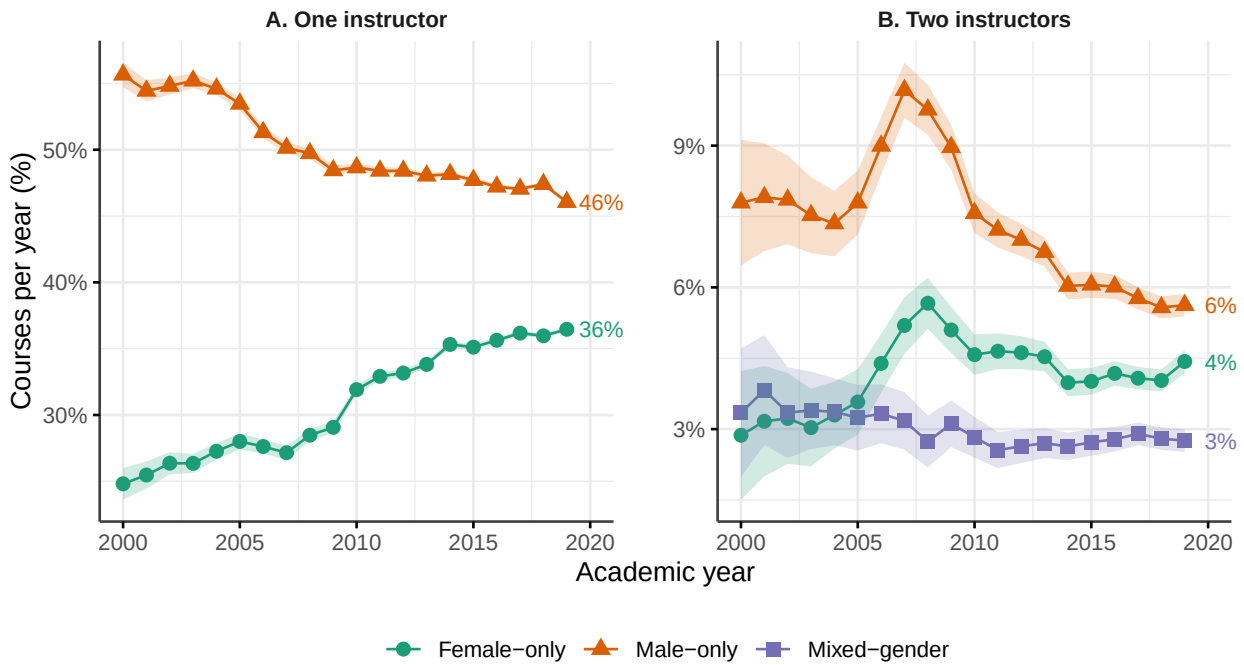

Figure 1: **Evolution of Teaching Teams by Gender Composition and Size.** Trends in course gender composition from 2000 to 2019 reveal a steady increase in women's participation (both in solo-taught and two-instructor courses), while the proportion of mixed-gender teams has remained relatively stable 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.

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.[4] Notably, mixed-gender teaching teams (3%) are consistently less common than same-gender teams (10%).

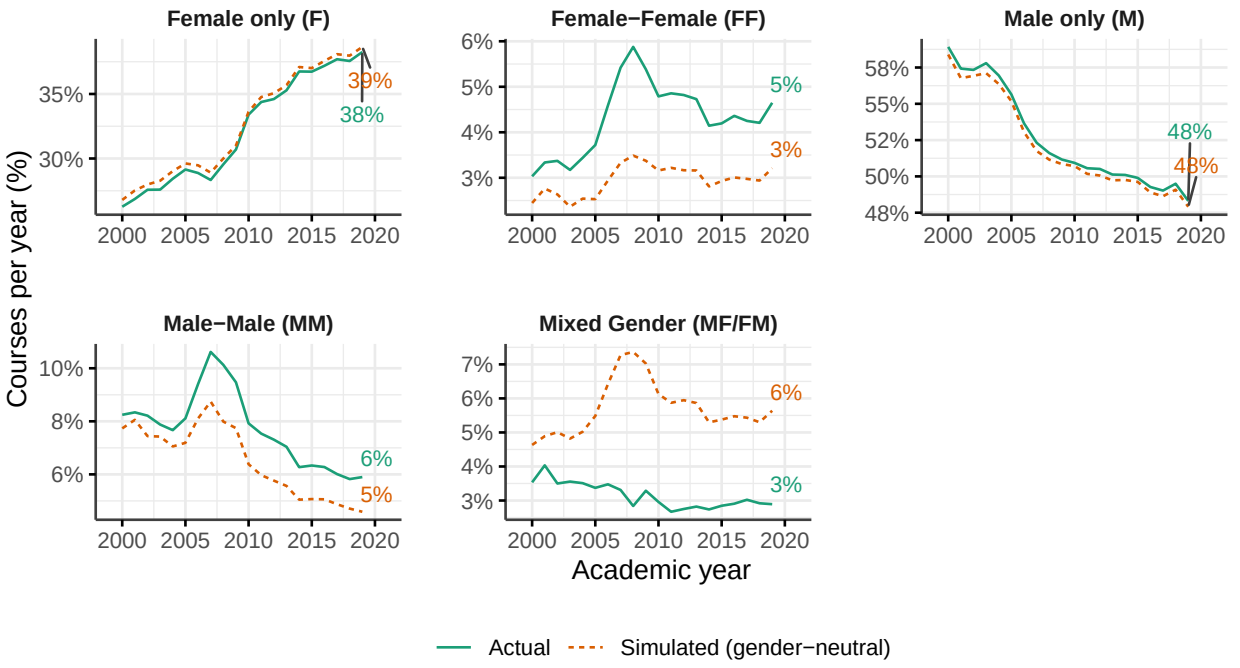

Figure 2: Comparison of gender composition between actual courses and courses simulated with Monte Carlo. Simulations ensure gender neutrality while keeping constant the institutional constraints (i.e., number of teams by size, institution, academic year, and 69 academic fields). Simulated mixed-gender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

To assess possible gender-based barriers to co-teaching teams, Figure 2 compares the actual percentages of different instructor gender compositions of 2000-2019 university courses (solid green lines) with simulated courses (dashed orange lines) generated from a Monte Carlo model. This model randomly shuffles instructors while preserving the overall

[4]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

distribution of course loads per year by institution and academic field, but ignoring gender-based preferences or constraints (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%), representing a 50% difference. These findings suggest that gender composition is shaped by social or institutional dynamics beyond chance, including possible gender homophily, mentoring networks, or departmental assignment practices that reinforce same-gender pairings, especially among women.

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 the barriers to form 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 is a pervasive structural feature rather than one driven by field-specific gender balances.

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. For example, mixed-gender teams are relatively rare in Great Britain (1%) compared to Canada (5%). However, in the simulations, the proportion of mixed-gender teams in both countries increases substantially—to

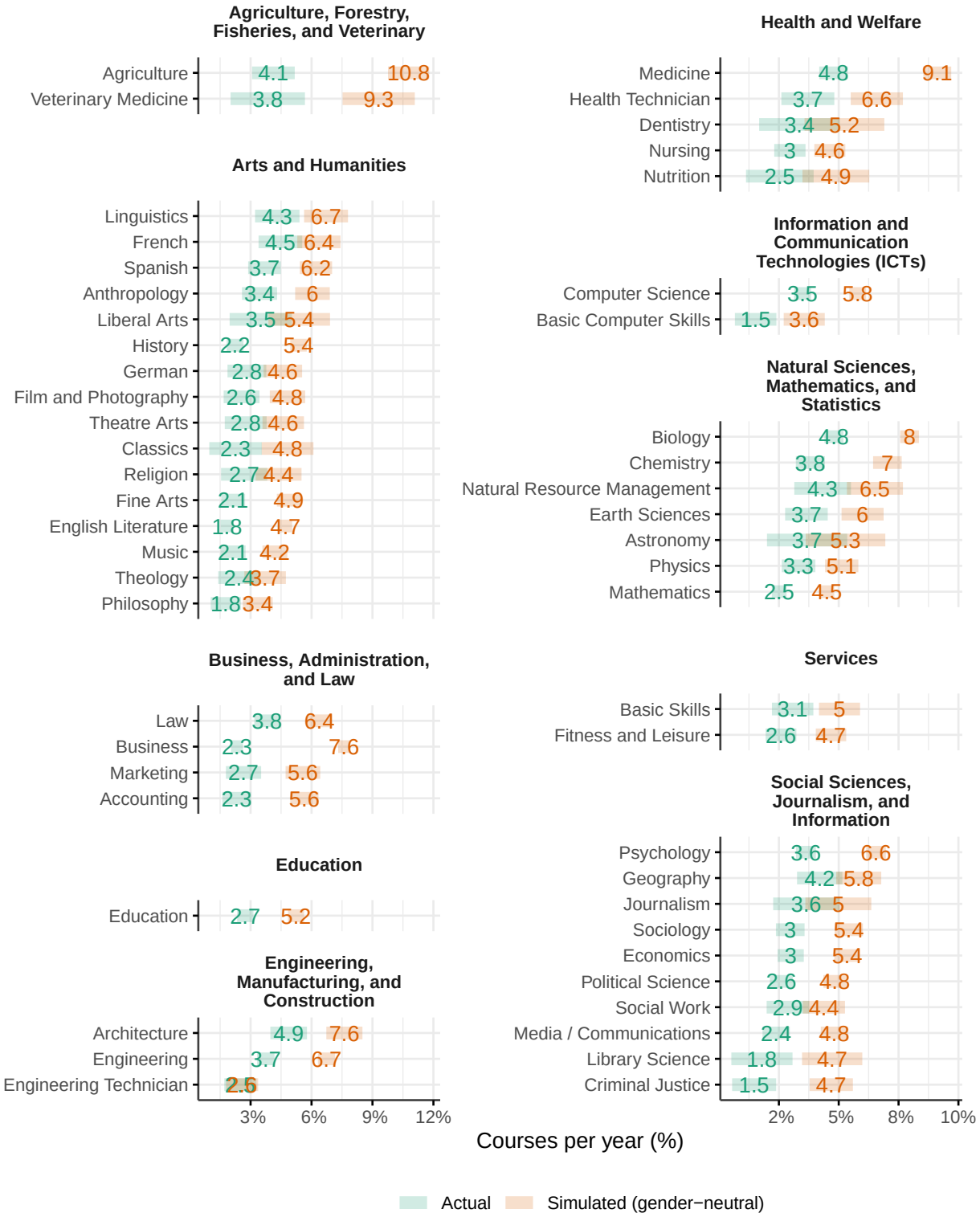

Figure 3: This figure disaggregates the comparison of proportions between actual (green) and simulated (orange) mixed-gender (MF/FM) courses by academic field. The shaded bar indicates 95% confidence level of the proportion.

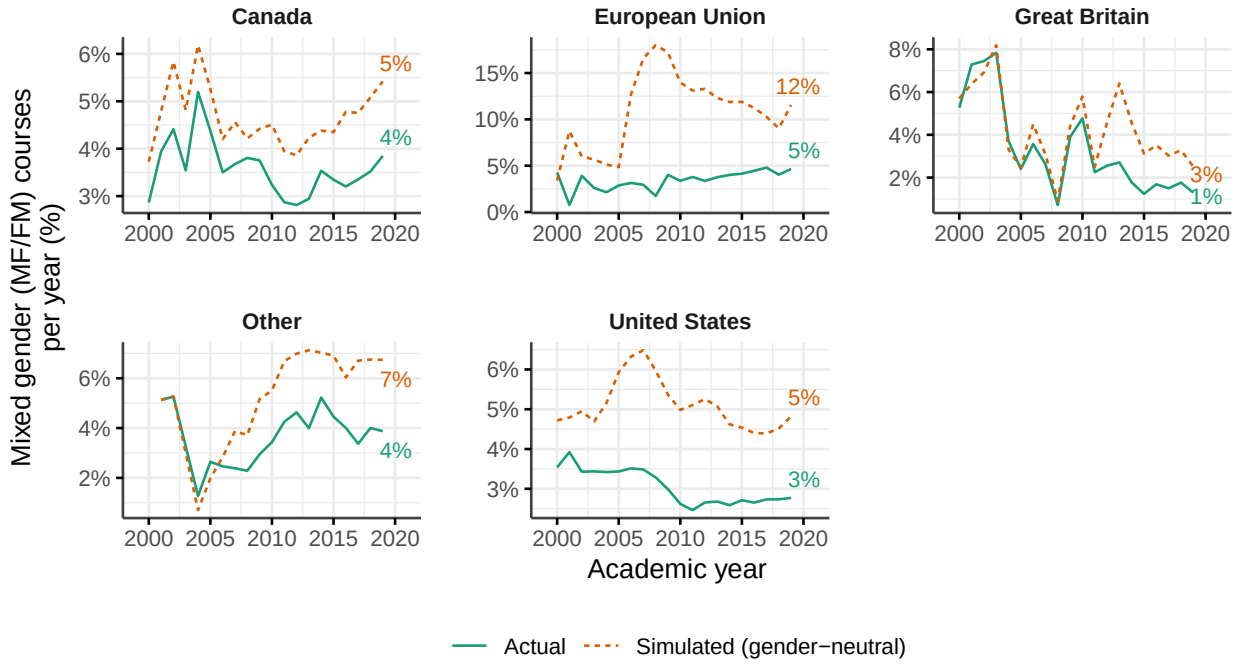

Figure 4: TBA

XX% and XX%, respectively. 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 assess the impact of teaching team size and gender composition on course materials, we analysed their relationship with four key metrics: the percentile rank of the interdisciplinarity score per year ("interdisciplinarity"), the average publication year of the readings ("age of readings"), the share of female authors cited in the readings ("share of female authors"), and the total number of readings included ("course breadth"). We used multiple linear regressions to account for differences across academic fields, institutions, years, and other relevant controls.

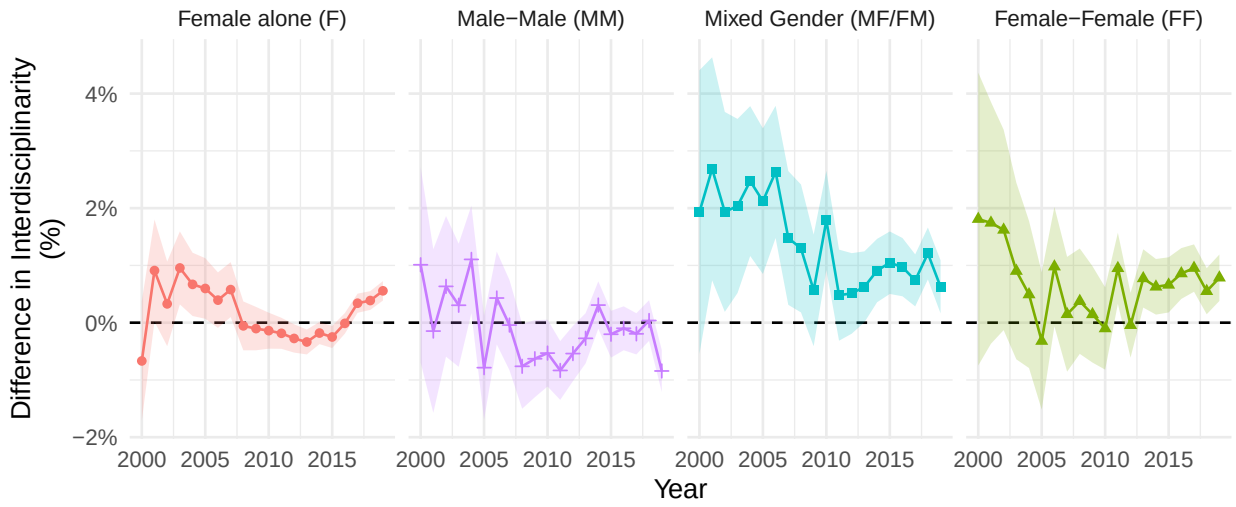

Figure 5: The figure illustrates the differences in interdisciplinarity between female-only (the baseline, set to zero), male-only, and mixed-gender courses. Although female-only courses exhibit slightly less interdisciplinarity than male-only courses (both individually and in teams), mixed-gender courses exhibt the highest levels of interdisciplinarity. These differences are obtained from separate linear regressions for each year and team size, controlling for country and academic field. Interdisciplinarity is measured in percentile ranks per year of the interdisciplinary score, with higher values indicating greater interdisciplinarity scores relative to other courses in the same cohort.

3.1.1 Interdisciplinarity

Figure 5 shows that, while female-only courses exhibit slightly less interdisciplinarity than male-only courses (individually and in teams), mixed-gender courses exhibit the highest levels of interdisciplinarity. Between 2000 and 2019, mixed-gender courses scored approximately 5% higher than male-only or female-only courses with similar characteristics. These findings suggest that, although minor gender-based differences in interdisciplinarity exist between female-only and male-only courses, gender diversity appears to have the most significant impact on fostering interdisciplinarity.

Disaggregating the analysis by academic field, with separate regressions by field accounting for year and country controls, shows that the impact of gender diversity is not homogeneous across fields (Figure **??**). Specifically, it appears to have a larger effect in specific fields, especially in Arts and Humanities (between 10% in 2005 and 5% in 2019) and Social Sciences, Journalism and Information (between 8% and 4% during the same

period).

3.1.2 Citing Women Authors

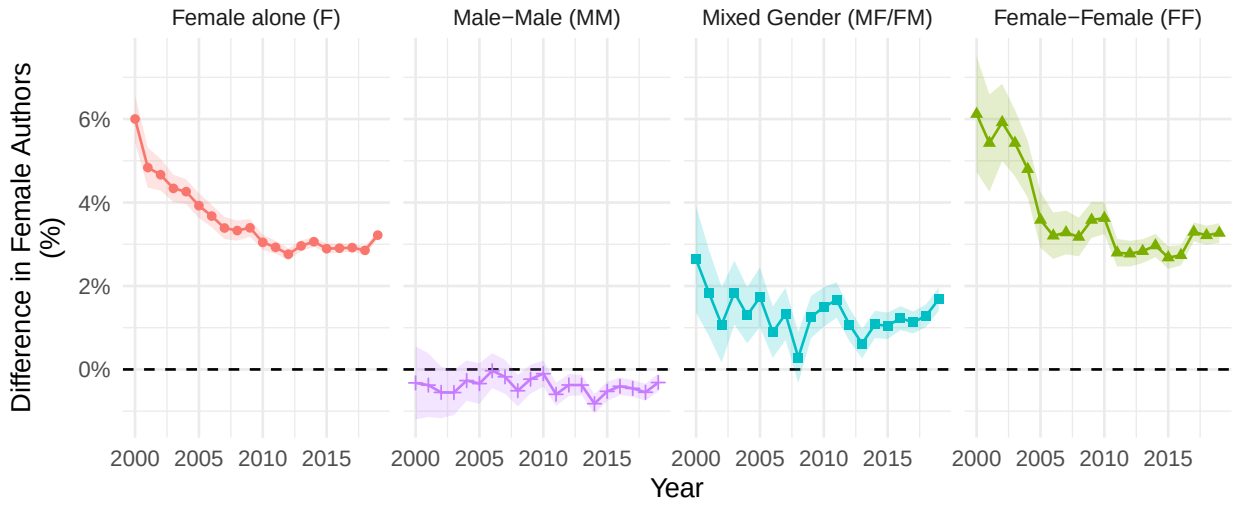

Figure 6: The figure illustrates the differences in the percentage of female authors cited in course readings between female-only (the baseline, set to zero), male-only, and mixed-gender courses. Male-only courses cite approximately 4% fewer female authors than female-only courses. While mixed-gender courses mitigate this effect to some extent, they do not reduced it by half, as one might expect if reading selections were equally influenced by both male and female instructors in a two-person team. This suggests that male instructors preferences continue to influence reading selections, even in mixed-gender courses. However, mixed-gender courses show a rising percentage of female authors in more recent years. These differences are represented as regression coefficients from separate linear regressions, controlling for total number of authors, academic year, team size, country, and academic field. Shaded areas represent 95% confidence intervals. The proportion of cited women is measured by adding two pseudo-observations to both male and female author counts, ensuring more conservative results.

Figure 6 highlights significant and consistent gender effects on the percentage of female authors over total authors cited in course readings while controlling for academic field and country in the regression.[5] Courses taught by a single male instructor consistently cite about 3% fewer female authors than those taught by a female instructor in the same cohort. However, the gap decreased by 25% between 2001 and 2019, going from 4% to 3%. Mixed-gender courses do not offset this disparity, as they cite fewer women by a similar margin, showing little difference from male-only classes taught by two instructors.

[5]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.

However, since 2013, a gap has emerged between mixed-gender and male-only courses, indicating a positive trend in gender representation driven by mixed-gender teams.

3.2 Age of Readings ("novelty")

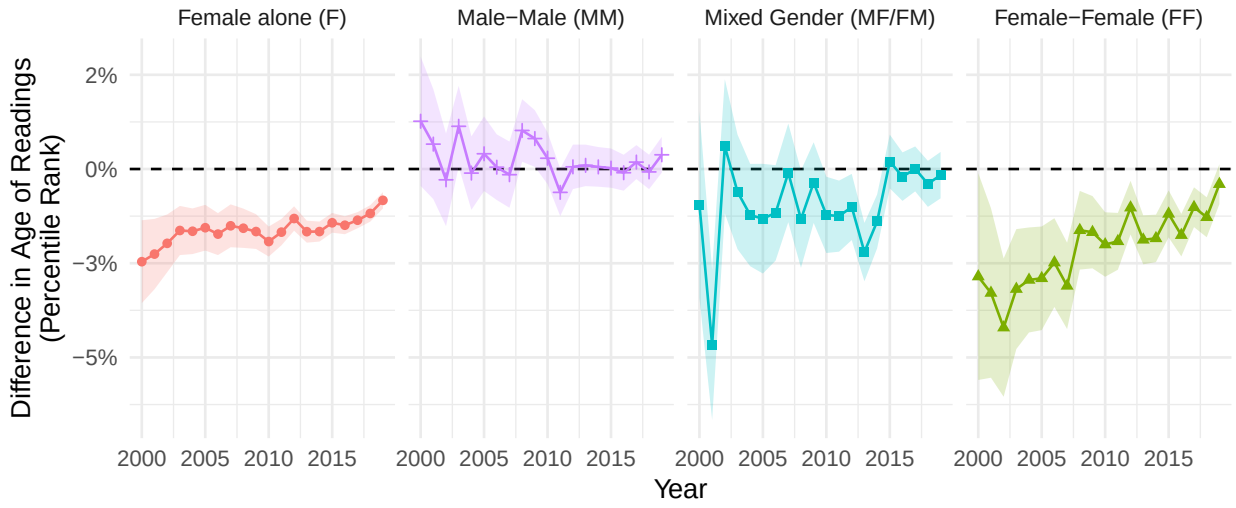

Figure 7: The figure illustrates the evoloving differences in publication age of the readings selected in the courses. While male-only teams tend to select older readings than female-only ones, mixed-gender teams do not show a consistent pattern.

Figure 7 illustrates significant gender differences in the publication age of the selected readings.[6] Courses taught by a single male instructor between 2000 and 2019 tended to choose readings about 2% older than those taught by a single female instructor, indicating significant gender-based differences. Similarly, during the same period, courses taught by two male instructors tend to list older readings than female-only ones. No consistent differences emerge between same-gender and mixed-gender courses. These findings broadly support that as more women hold academic positions, the materials taught to students tend to be more current. At the same time, gender diversity appears to have little impact on the age of readings.

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

3.3 Number of readings (“breadth”)

```
## Warning in add_figure(file.path(fig_dir, "lg_uniq_docs.pdf")): File not found:
## /Users/mrb/Documents
## local/Open-Syllabus/gender-teaching-prefs/output/figures//lg_uniq_docs.pdf
```

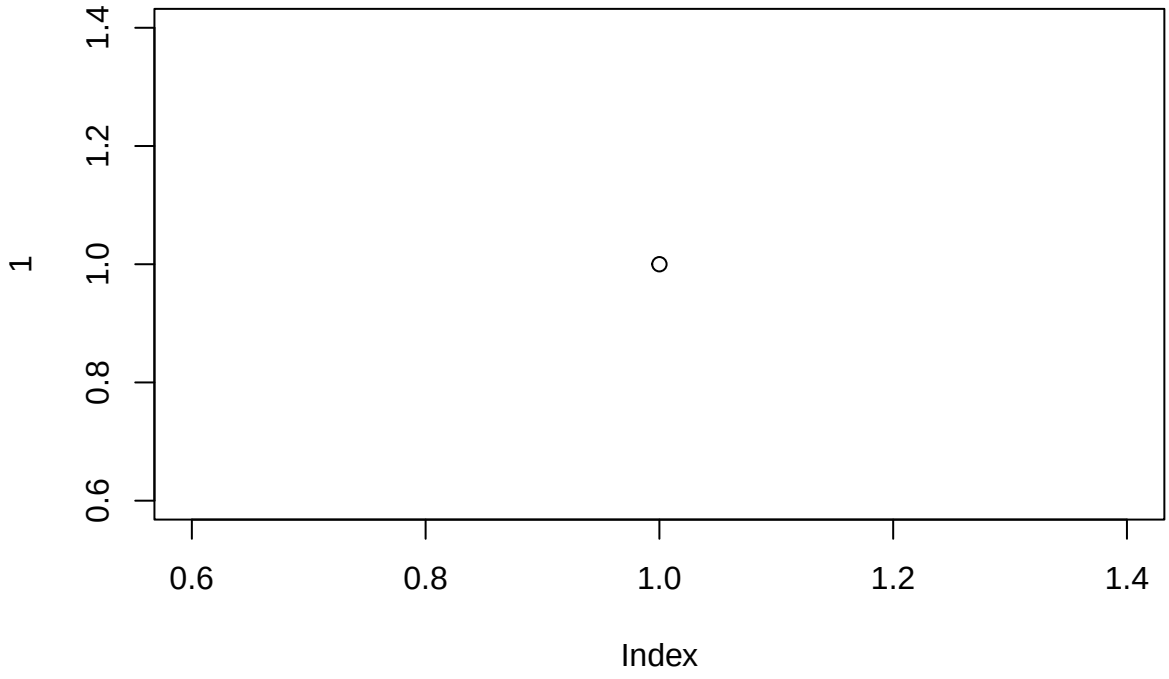

Figure 8: The figure illustrates the evoloving differences in number of readings across different team types over time. It shows that while gender-based differences vary in magnitude, mixed-gender teaching teams consistently outperform same-gender teams in the number of readings, suggesting a key role of gender diversity in delivering more comprehensive courses.

Figure 8 illustrates the evolving differences in the number of readings across different team types, serving as a direct measure of a course’s breadth. Courses taught by a single male instructor generally include more readings than those led by a single female instructor. However, the difference is greater between 2005 and 2015 and decreased to disappear in recent years. Mixed-gender courses exhibit significantly more readings than same-gender courses, though this difference varies over time, ranging from 20-30%

between 2005 and 2015 to around 10% in recent years. In contrast, courses co-taught by two male instructors tend to list approximately 5% more readings than all-female ones throughout the period. While these findings highlight fluctuating gender-based differences in course breadth, they also indicate strongly that mixed-gender teaching teams are crucial in delivering more comprehensive courses or those with broader curricula.

3.4 Robustness

To ensure the reliability and robustness of our results, we conducted several additional analyses. We included random effects for institutions in our regressions to account for potential correlations at the institution level, thereby controlling for unobserved heterogeneity across institutions. We trimmed the dataset to exclude outliers. Specifically, we removed courses with an unusually high number of instructors or implausible reading data (e.g., courses with an unrealistically low age of readings or an excessive number of readings). This step showed that the regression results were not influenced by outliers. To address potential non-linear relationships between the predictors and outcomes, we tested several transformations of the dependent variables. This approach allowed us to verify whether our results hold under different assumptions about the functional form of the relationships. We further disaggregated the analysis by academic field and by country to examine the robustness of the results across various contexts. This allowed us to assess whether our findings were consistent across institutional and geographical settings. 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. Over the last twenty years, while the share of classes taught by female instructors has been increasing, individually and in teams, 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 institutional constraints, like a limited gender diversity in certain fields, but it can also arise from gender preferences in team formation, with a tendency for instructors to form teams with same-gender partners ("homophily") as observed in research collaborations and other academic domains.

We tested this hypothesis using Monte Carlo simulations and a well-established methodology that forms teams in a gender-neutral manner, while holding fixed important instititual constraints, such as maintaining constant the course loads across institutions in 69 different fields within a given year. After comparing simulated against actual teams, we find that instructors tend to partner with same-gender colleagues more than twice (50%) as expected between 2005 and 2019. This result is robust across academic fields, although it is more sizeable in certain fields (Chemistry, Business, Medicine), with no significant differences observed in STEM fields. The underrepresented share of mixed-gender teams is consistent across institutions in different countries, with the strongest effect in the selected EU countries, where only 5% of courses are mixed-gender, compared to an expected 12%.

While our analysis accounts for temporal effects as well as differences between fields and institutions, one 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.

[TBA – discussion of within field].

Our findings underscore the significant barriers to gender-mixed teams, which has significant implications. This insight is crucial as it extends existing research on familiar drivers of gender bias in various settings within academia, especially research collaboration, access to funding, and student gender stereotypes. Here, we highlight the impact of teaching collaborations, which may limit the exchange of information among instructors and restrict students' access to diverse courses. This, in turn, could hinder interdisciplinary engagement or disproportionately disadvantage women in academic networks.

Our analysis of the association between gender composition and interdisciplinarity highlights a strong impact of gender diversity on interdisciplinary teaching. Between 2000 and 2019, mixed-gender teams are consistently more interdisciplinary than same-gender teams (male-only or female-only), highlighitng the importance of gender diversity for interdisciplinarity. By contrast, women teaching alone are nearly as interdisciplinary as men. This highlights how gender preferences in team formation can impact teaching. By lowering the number of mixed-gender teams, gender preferences are also likely to reduce the amount of interdisciplinary course offerings. This finding adds a new potential barrier to the dissemination of interdisciplinary work in academia, which was overlooked previously.

Our analysis of gender composition and interdisciplinarity reveals a strong link between gender diversity and interdisciplinary teaching. Between 2000 and 2019, mixed-gender teams consistently exhibited greater interdisciplinarity than same-gender teams (male-only or female-only), underscoring the value of gender diversity in academia. Interestingly, women teaching alone are nearly as interdisciplinary as men, suggesting that gender preferences in team formation—not inherent differences in teaching styles—shape interdisciplinary outcomes. By reducing the prevalence of mixed-gender teams, these preferences likely limit interdisciplinary course offerings, introducing a previously overlooked barrier to

the spread of interdisciplinary work in academia.

Another key finding relates to how gender composition shapes the representation of women authors in college education. We find that female-majority teams consistently cite more female authors than male-majority teams, holding this pattern across all team sizes. For instance, courses with a single male instructor cite about 10% fewer female authors than those taught by a female instructor. In mixed-gender teams, male-lead teams tend to cite considerably fewer women authors than female-lead ones. This evidence points to an adverse effect of mixed-gender collaboration on citing women's materials for teaching. While this adverse effect is likely to impact female instructors, as their work is less visible among students, it may also affect students as they are exposed to less diverse work, which may impact their learning (e.g., through a lack of role models).

Finally, female instructors use more novel readings and materials in their syllabi when teaching alone than male instructors. Furthermore, all-female teams use more novel readings than women teaching alone. Additionally, on average, diverse teams of two instructors use readings older than those used by all-female teams but more recent than those used by all-male teams. This evidence suggests that women have a propensity for newer readings, which can benefit students who might be exposed to more content at the frontier. It is possible that, assuming women are younger on average, this gap in novelty could be driven by cohort effects. However, it is striking that the gap remains in diverse teams, suggesting a significant role in gender composition that goes beyond age.

This work is consistent with previous research showing an underrepresentation of gender-diverse teams in academia. However, contrary to research teams, we find that gender diversity harms women's work citations and novelty, suggesting that women teaching alone or in teams of all women have more freedom in pursuing more novel materials and give credit to women authors. This points to potential difficulties in how instructors of different genders view teaching, which in turn may drive the matching of teams and ultimately affect

students.

However, instructors may not always be able to choose their team members. For example, instructors at higher levels of the hierarchy may have more weight in deciding who matches with whom. We try to consider this effect by examining the order of appearance in the syllabus as a proxy of instructors' seniority. The analysis confirms strong gender preferences – if the first instructor is a man, then 66% of the following instructors are men. Similarly, if the first instructor is a woman, 62% of the following instructors are women.

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, $w(f_{\text{max}}, f) = \cos(f_{\text{max}}, f)$, where f_{max} is the field that has the highest cosine similarity with course i (i.e., $f_{\text{max}} = \arg\max_f \cos(i, f)$) 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

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).