9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Intersectional Inequalities in Online Communication Among Postsecondary Learners

Keywords: Online Education; Discussion Forums; Inequality; Intersectionality; Natural Language Processing

Extended Abstract

Online and hybrid learning has been and will continue to be an integral part of postsecondary education, especially after the COVID-19 pandemic. However, research has suggested that compared to in-person classrooms, online learning produces greater achievement gaps between learners from marginalized populations and their peers [1]. To improve equity in online and hybrid learning, a variety of instructional strategies and technological tools have been adopted, and asynchronous discussion forums are one of the most commonly used tools due to their affordance to support idea sharing and exchange at any time and ease of use for both instructors and learners [2]. In theory, asynchronous forums can facilitate both collective knowledge building and foster a sense of belonging, thereby contributing to a connected and engaged educational experience [3]. But in practice, learners' experience in forums might vary due to individual differences in their backgrounds and dispositions. It would be especially concerning if learners from marginalized populations have marginalized experiences in discussion forums when the tool is intended to improve interpersonal connection and close achievement gaps.

In order to systematically understand disparities in online communication experiences across different learner populations, this study draws on large-scale, longitudinal text data from discussion forums in postsecondary education and leverages natural language processing techniques to quantify the nuances of participation and interaction patterns. It provides one of the largest pieces of empirical evidence on sociodemographic gaps in learner online communication experiences, extends the understanding of educational inequality to an actionable behavioral level, and informs digital learning researchers of equity concerns in practice. The analysis here includes both pre-COVID and COVID-affected periods; findings here can shed light on the digital divide under different conditions. In addition, this study examines inequality across intersectional groups to account for the interaction of multiple marginalized identities as they shape individual learner experiences.

We rely on the content and metadata of all discussion forum posts from the Canvas learning management system deployed at a public minority-serving institution in the United States between September 2019 and June 2022. After removing non-English posts and courses with fewer than 10 students or less than 50% students who post, the data encompasses 2,096,133 posts created by 43,438 undergraduate students enrolled in 3,291 courses offered across 9 academic terms, including in-person, online, and mixed modalities caused by COVID-19. Admin records also document students' background information and transcripts. To understand systematic inequalities in online communication experiences, we focus on gender and racial identities and their intersection, as these identities are somewhat more visible than others and therefore may directly influence their experience. The gender and racial composition in each course are shown in Figure 1. We estimate whether, compared to their peers, minority groups 1) post less, 2) receive fewer peer responses, and 3) receive lower-quality peer responses. For response quality in 3), we use three established measures: `GloVe Aligned`, the average pairwise

cosine similarity of GloVe [4] word embeddings between tokens of the original post and their most similar tokens in the response; `% In R`, token overlap between the original post and the response; and the `analytic` psycholinguistic lexical feature from LIWC [5] of the response. We regress each outcome variable separately on indicators of racial, gender, and intersectional groups to estimate the magnitude of inequality. Across these models, course fixed effects are always included to control for contextual differences, and when the outcome is an attribute of responses, post characteristics are also controlled for.

Our results indicate several inequalities. First, non-male learners participated slightly more in discussion forums than their peers, manifested by higher volumes of posts, while underrepresented racial minorities (URM) participated less compared to non-URMs (Figure 2). Second, after controlling for posting behavior, non-males received more peer responses than their male counterparts while URM learners received fewer responses (Figure 3). Because of these opposite gaps experienced by non-males and URMs, the intersectional group of non-male URMs had no different participation and response patterns from male non-URM learners. Third, controlling for volumes of received responses, non-males got "shallower" peer responses because they had higher semantic overlap with the original posts but lower levels of analytical thinking (Figure 4). Similar patterns were seen among URMs but only when they are not male. Finally, we explore whether these findings vary across different courses and find that the foregoing inequalities are concentrated on fully online and large courses. To put all in a nutshell, there is some evidence for marginalized learner populations being further marginalized in forum-based online communication, especially in learning contexts where discussion forums play a more important role. Gender and racial identities function slightly differently in this process: non-male learners have more subtle "marginalization" experience because on the surface they participate more and get more peer attention, but the attention is of lower quality than what males get. Intersectional groups therefore have even more complicated experience.

While discussion forums are typically employed to improve connectedness and engagement in online and hybrid education, our results highlight potential unintended consequences and the need to adopt additional or alternative instructional strategies to create more inclusive and interactive learning environments. Our understanding of intersectional behavioral inequalities can also help educators pinpoint sources of achievement gaps and design policy-level changes or interventions that specifically target traditionally underserved learner populations and reduce inequities in the post-pandemic education experience.

References

[1] Di Xu and Ying Xu. "The Ambivalence About Distance Learning in Higher Education". In: *Higher Education: Handbook of Theory and Research*. 2020, pp. 1–52.

[2] Anna Sun and Xiufang Chen. "Online education and its effective practice: A research review." In: *Journal of Information Technology Education* 15 (2016).

[3] Joanne M McInnerney and Tim S Roberts. "Online learning: Social interaction and the creation of a sense of community". In: *Journal of Educational Technology & Society* 7.3 (2004), pp. 73–81.

[4] Jeffrey Pennington, Richard Socher, and Christopher D Manning. "Glove: Global vectors for word representation". In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 2014, pp. 1532–1543.

[5] Ryan L Boyd et al. "The development and psychometric properties of LIWC-22". In: *Austin, TX: University of Texas at Austin* (2022), pp. 1–47.

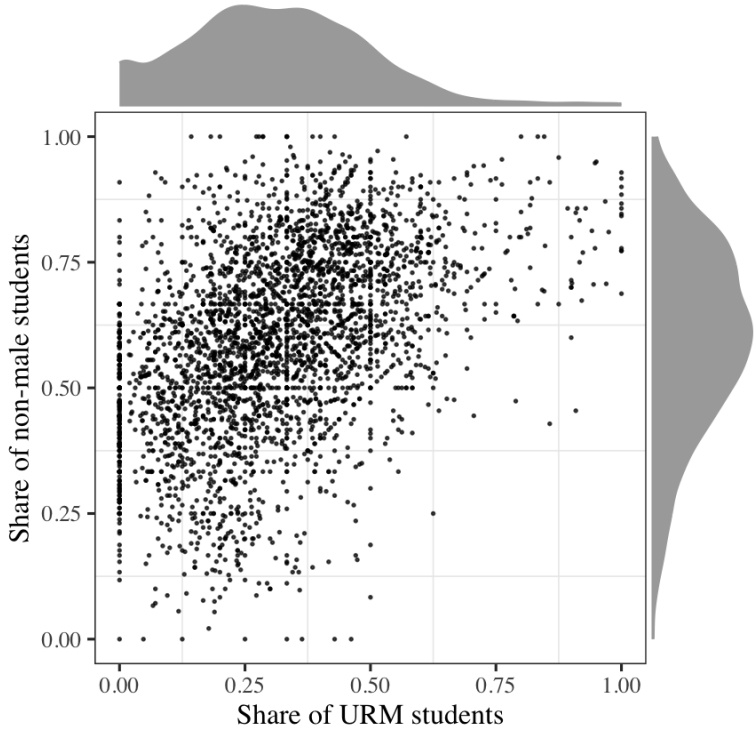

Figure 1: Demographic composition of individual courses. Each dot indicates a course, with the share of URM and non-male students on the x-axis and y-axis, respectively. URM: under-represented racial minority as is in the administrative records.

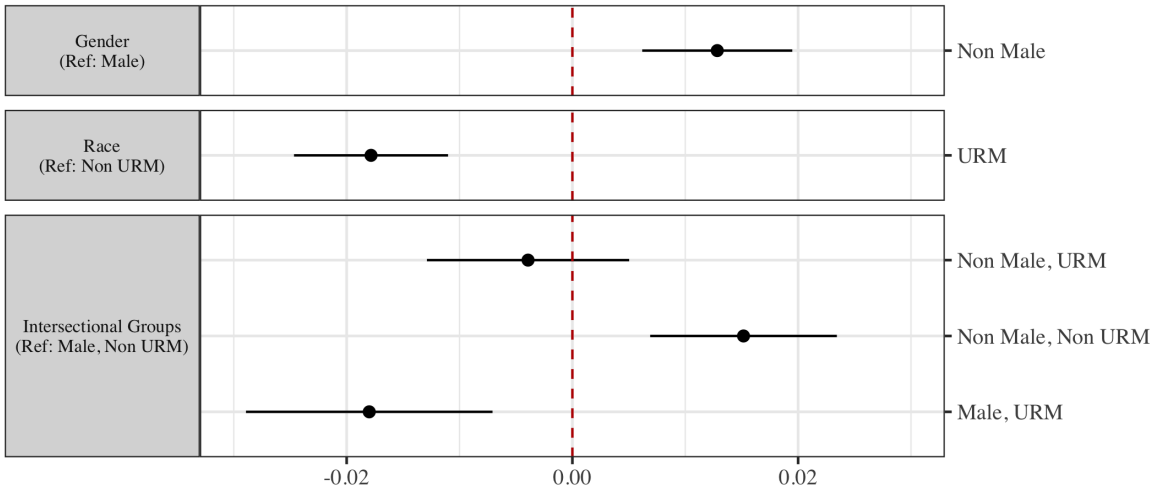

Figure 2: Estimated inequality in *the number of created forum posts*. Each panel shows results from a separate model at the learner-course level. Each model regresses z-scored post count on the corresponding demographic indicator(s), controlling for course fixed effects. Coefficient estimates with 95% confidence intervals are plotted.

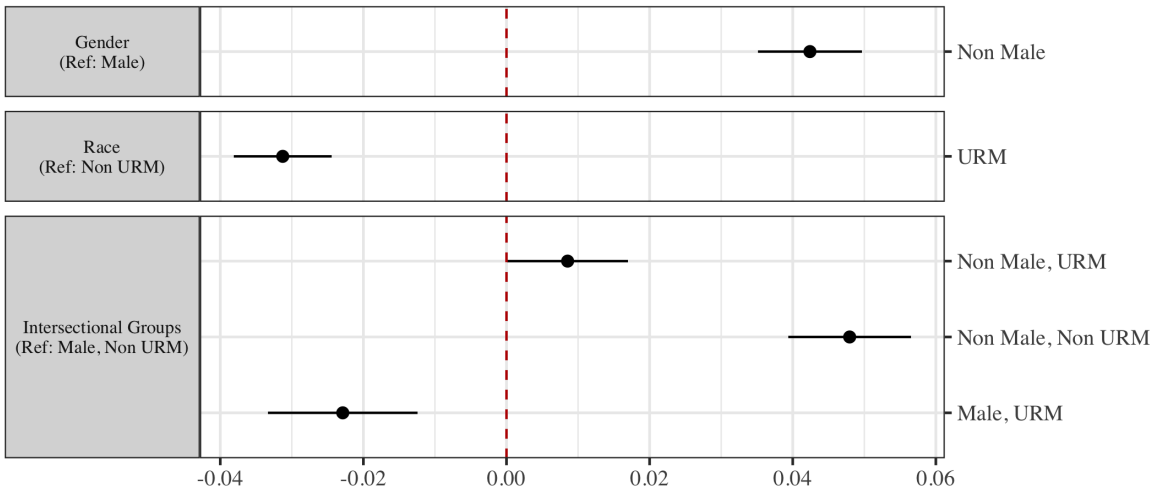

Figure 3: Estimated inequality in *the number of received peer responses*. Each panel shows results from a separate model at the learner-course level. A learner-course observation is included only if the learner made at least one post in the course. Each model regresses z-scored response count on the corresponding demographic indicator(s), controlling for course fixed effects and characteristics of the posts created. Coefficient estimates with 95% confidence intervals are plotted.

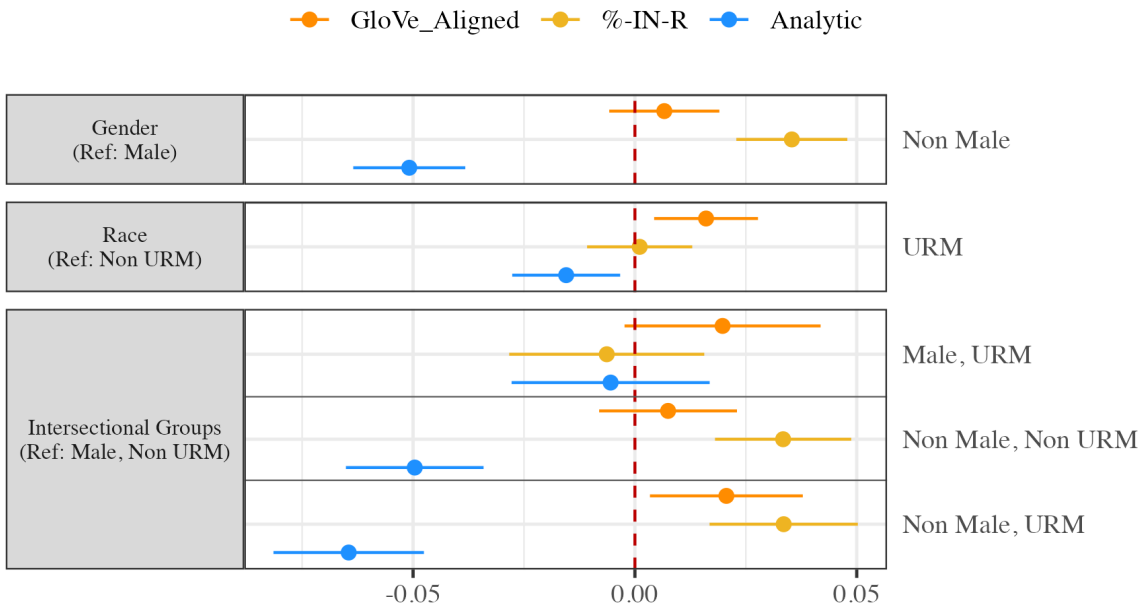

Figure 4: Estimated inequality in *the quality of received peer responses*. Each panel shows results from three separate models at the learner-course level, each based on one quality measure (colored). A learner-course observation is included only if the learner received at least one peer response in the course. Each model regresses a z-scored quality measure on the corresponding demographic indicator(s), controlling for course fixed effects and characteristics of the posts created. Coefficient estimates with 95% confidence intervals are plotted.