Bias in Online Classes: Evidence from a Field Experiment

Rachel Baker[1], Thomas Dee[2], Brent Evans[3], June John[4]

[1]University of Pennsylvania Graduate School of Education; 3700 Walnut St, Philadelphia PA 19104; rbbaker@upenn.edu

[2]Stanford University Graduate School of Education; 520 Galvez Mall CERAS Building, 5th Floor, Stanford CA 94305; 650-723-6847 and NBER; tdee@stanford.edu

[3]Department of Leadership, Policy, and Organizations, Peabody College, Vanderbilt University; 230 Appleton Place, PMB 414, Nashville TN 37203; 615-322-6491; b.evans@vanderbilt.edu

[4] Meta, 1 Hacker Way, Menlo Park CA 94025

Disclosure Statement

This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B140009 to the Board of Trustees of the Leland Stanford Junior University. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

Abstract

Asynchronous, online interactions are increasingly common, particularly in education, but relatively little is known about the influence of social identity in these environments. We test for the presence of race/place-of-origin and gender biases among students and instructors in asynchronous online post-secondary classes by measuring responses to discussion comments posted in the discussion forums of 124 different massive open online courses (MOOCs). Each comment was randomly assigned a student name connoting a specific race/place of origin and gender. We find evidence that assumed identities influenced the likelihood of both instructor and peer responses. The comparative effects by identity indicate that instructor responses consistently privileged White males who were, on average, 94% more likely to receive a response than other students. We also find that White female students were particularly likely to receive a peer response. We discuss the implications of these findings for understanding social-identity dynamics in classrooms and the design of online learning environments.

Keywords: Racial Bias, Gender Bias, Online Education

I. Introduction

An increasingly large number of basic economic interactions (e.g., workplace communication, consumer purchases) occur in digitally mediated, asynchronous environments. This is uniquely true in education and especially at the postsecondary level where online learning environments were becoming increasingly common even prior to the dramatic changes brought on by the COVID-19 pandemic. For example, earlier evidence indicates that a third of postsecondary students took an online course and that an online format had negative academic consequences, especially for students with lower levels of prior achievement (McPherson and Bacow, 2015; Bettinger et al., 2017). The character of instructor and peer engagement in online classrooms is likely to contribute to their comparative effectiveness and may raise fundamental issues of fairness. Given that social identity theory suggests that membership in social groups such as a specific gender and race affects self-concept and the perceptions of others, examining how engagement is related to race and gender in educational settings is critically important (Tajfel & Turner, 1979). In particular, a long-standing descriptive literature (e.g., American Association of University Women, 1992; Tenenbaum & Ruck, 2007) indicates that teachers in conventional classrooms appear to exhibit biases against females and racial minorities (e.g., directing engagement and encouragement disproportionately to White and male students). Similarly, evidence (Bettinger et al., 2016) suggests students participating in asynchronous discussion forums in online classes at a degree-granting institution exhibit preferences for engaging with similar peers and that such engagement improves learner outcomes.

Ex ante, it is not clear whether online classrooms would mitigate or increase the prevalence of such biases. The comparative anonymity of these entirely digitally mediated interactions, which provide fewer visual clues of race/place of origin or gender, could attenuate biases by reducing the tendency towards racial and gender-based categorizations. Alternatively, the stylized character of online classrooms may increase the psychological salience of the limited identity cues that are available as well as reduce the social incentives for self-control. Using a field experiment, this study provides novel evidence of the possible presence of racial and gender biases among instructors and students in online courses. This experimental study is situated in the discussion forums of 124 postsecondary Massive Open Online Courses (MOOCs). In such large-scale, online learning environments, these forums provide the primary, and often the only, opportunity for instructors and students to interact. These interactive message boards also perform vital educational functions as students rely on the discussion forums to ask questions about the course content and structure and to receive answers and encouragement from fellow students and course instructors. We tested for the presence of biases in these settings by creating fictional student identities with race/place-of-origin- and gender-connotative names, placing randomly assigned comments in the discussion forums using these fictional student identities, and observing the engagement of other students and instructors with these comments.

We believe this study makes at least two broad and distinct contributions. First, it provides important new evidence on the character of digitally mediated interactions. The nature of such interactions has particular relevance in the context of online classrooms. These learning environments have become increasingly common but face serious challenges in terms of supporting student engagement and human-capital accumulation. While evidence exists on bias in several other settings, we are unaware of any existing studies specifically assessing instructor bias in online educational environments. Second, the empirical evidence from this field experiment also informs an important theoretical ambiguity. An active and growing body of evidence indicates that a race or gender-congruent instructor (i.e., "a teacher like me") results in improved student outcomes (e.g., Dee, 2004, 2005; Fairlie, Hoffman, & Oreopoulos, 2014; Gershenson, Holt, & Papageorge, 2016; Lindsay & Hart, 2017; Van den Bergh, Denessen, Hornstra, Voeten, & Holland, 2010). However, the existing reduced-form evidence cannot distinguish between effects due to active instructor biases and those due to how student performance responds to an instructor's identity (e.g., stereotype threat and role-model effects). Because this study relies on experimentally constructed student identities, it unambiguously isolates the effects that are instructor-centered (e.g., implicit and explicit biases).

II. Experimental Design

This field experiment occurred within 124 Massive Open Online Courses (MOOCs). Critically, other online courses frequently share both the basic design features of these MOOCs (e.g., asynchronous engagement, recorded lectures, discussion forums) and their postsecondary content. Furthermore, despite the cycle of early hype and then cynicism around MOOCs (McPherson and Bacow, 2015), these free classes remain a widely used form of online learning. In 2019, more than 900 universities offered 13,500 unique MOOCs, and 120 million students signed up for at least one course (Shah, 2019). In addition to the dramatic increase in online education at traditional postsecondary institutions, early evidence also indicates that the COVID-19 pandemic dramatically increased the uptake of MOOCs. For example, in May of 2020, Coursera announced it had recently registered 5 million new users and 10 million course enrollments, a 644 percent increase over the prior year (Dee, 2020).

A. Study Sample

We identified our experimental sample of MOOCs by compiling the universe of MOOCs offered by a major provider that started between August 1 and December 31, 2014.[1] We screened the available courses and included those that met the following criteria: a course length of five weeks or longer, postsecondary content, the presence of a general discussion forum, and a course not taught by an instructor that was included in our small preceding pilot. Additionally, we only included one course

[1] As part of our human-subjects protocol, we do not identify this provider nor do we provide the titles of the classes or the exact text of the comments we placed.

per instructor. When instructors taught more than one course, we decided which course to include based on date (i.e., taking earlier courses over later ones) and length (i.e., taking longer rather than shorter classes). When all else was equal, we selected the course that was listed first alphabetically. The 124 MOOCs in our sample covered a diverse range of subjects, including accounting, calculus, epidemiology, teaching, and computer programming. All had content appropriate for students seeking an associates or baccalaureate degree, though none were explicitly part of a degree program, and were open to anyone. Most (94) were offered by four-year not-for-profit institutions of higher education in the United States. These 94 colleges were, on average, very selective. Those courses that were offered by international institutions were taught in English.

Using fictive student identities, we placed eight discussion-forum comments in each of the 124 MOOCs. Within each course, eight student accounts were used to place one comment each. The eight student accounts each had a name that was connotative of a specific race/place of origin and gender (i.e., White, Black, Indian, Chinese, each by gender).[2] Each of the eight possible social identities was used once per class. Our random-assignment procedure, which we describe below, was designed to ensure that the student name, the comment they placed, and the order in which each comment was placed were random. We placed comments in the "General Discussion" or similar sub-forum and we timed comments to be spaced out roughly equally over the duration of the course, from the beginning of the course to two weeks before the end of the course. We observed all replies to each comment for the two weeks after placement.[3] By observing the responses to our comments by instructors and by

[2] We chose to separate Indian and Chinese names instead of using the more common racial category "Asian" because of the large number of MOOC participants from each of these respective countries. For this reason, we describe these categorizations as race/place of origin rather than just race.

[3] Our small exploratory pilot study that preceded the experiment indicated that this window would capture the vast majority of responses to all placed comments.

students in the course, we can identify any difference in the number of responses received by our accounts of fictive students that were assigned different race/place-of-origin and gender identities.

B. *Student Names and Comments*

Drawing from several hundred actual student comments placed in a variety of MOOCs during a preceding pilot study, we constructed a list of 32 generic discussion-forum comments that would be applicable across all types of courses. To ensure that we could use our comments across courses, which is important for our experimental design, our comments did not ask questions related to the content or deadlines of a specific course. Our comments, based on actual comments placed in other MOOCs, focused on general topics such as questions about studying and issues of course difficulty that could be sensibly placed in any course regardless of the subject matter. More specifically, some of the comments focused on issues directly related to course procedures and completing the course (e.g., specific questions about due dates or questions about how to complete assignments). We refer to this set of comments as "completion-focused." The remaining set of comments were declarative statements that might catalyze conversation (e.g., a comment that the course was easier than the student expected), questions about other students in the class (e.g., asking where people are from or why they are taking the class), or questions about academics outside the particular classes (e.g., asking what courses to take next). We refer to this set of comments as "advising/social." We list all 32 comments in Table A1 with slight adaptations to the text in order to preserve anonymity. In the experiment, the frequency of student and instructor responses to these comments was similar to that of the real student comments on which they were based, which suggests that these comments were representative and realistic.

Within each participating course, we randomly paired these comments to fictive student accounts with our race/place-of-origin and gender- evocative names. To create a bank of names, we drew from Anglo-American, African-American, Indian, and Chinese names that were recently used in other audit experiments that experimentally manipulated perceptions of race/place of origin and gender (Bertrand & Mullainathan, 2004; Milkman, Akinola, & Chugh, 2015; Oreopoulos, 2011). We identified a set of four first names and four last names for each gender in each race/place of origin (16 unique names for each of the 8 possible social identities, 128 unique names in total). Each posting used a first and last name, which is a common practice by actual students in MOOC forum postings, to maximize the chance of being identified with the appropriate race/place-of-origin and gender profile.

C. Randomization

In each MOOC, we had one of each of our eight possible social identities place one randomly assigned comment. This within-course design allows us to control flexibly for all the unobserved course-specific traits that may influence commenting within the course. However, to avoid other potential confounds, we also adopted procedures that would create random variation in the social identity of the poster, the comment placed (i.e., which of the 32 comments), and the order in which it was placed in the course (i.e., 1st through 8th). First, to choose the sequencing of social-identity profiles within each course, we established an initial random ordering of the sequence of the eight possible profiles and did so in a manner that ensured that no same-gender or same-race/place-of-origin identity appeared consecutively. For the first course in the study, we randomly assigned one of the 16 possible names appropriate for social identity of each poster. We then randomly assigned a comment to these profiles in this randomly ordered sequence (i.e., 1, 2, 3,…, 8).

These 8 initial comments were randomly selected without replacement from the total list of 32 comments. When a second eligible course opened, we randomly selected 8 comments from the remaining pool and assigned them to social-identity profiles in a sequence that was rotated by one position (i.e., 2, 3,...,8, 1). As subsequent courses opened, we randomly selected matched comments until the pool of 32 was exhausted. After every four courses, our procedure returned to the full set of 32 comments. Similarly, we continued rotating the sequence in which social-identity profiles appeared and re-randomized when a full rotation was achieved (i.e., every 8 courses). We also relied on random selection of names without replacement and then re-randomized every 16 times so that names were balanced in the design of the study.

This process has several important features. First, it guarantees, for all participating courses, within-course variation in the student identities placing comments (i.e., the "treatments" of interest in our experiment). Second, by design, it also provides random variation for each student identity posting *within* courses in both the comment placed and the order in which it was placed. Finally, our approach ensures across all the courses a balanced representation of all the identities, names, and comments used in our study. We observe this balance in our final data set. For example, **e**ach particular social-identity profile (e.g., White male) was used exactly once per course, so each was used 124 times. The number of times a particular name in social-identity profile was used ranged from 6 to 8. The number of times each of the 32 comments was used across the entire study ranged from 29 to 32 with each social-identity profile placing each comment an average of 3.9 times.[4]

In a conventional experimental study, an important check on the study design is to examine whether the observed traits of the participating subjects are well balanced across treatment and control conditions. The issue of covariate balance has less relevance in our study because our observations (i.e., the placed comments in these online classes) have no covariates beyond our randomly assigned

[4] The slight imbalance in the frequency of names used and comments relative to what our design would imply is due to the fact that we dropped two courses in which we had begun placing comments. One course was dropped because our monitoring of student comments raised concerns that the existence of our study might be uncovered. A second course was dropped because, unlike other courses, it ceased accepting new registrants during the course progression, so we could not create new student accounts throughout the course in order to place comments. Including the data that we did collect from these courses does not influence our findings.

treatments of interest (i.e., the race/place-of-origin and gender-based social identity) and the categorical traits (i.e., course, comment type, comment sequence) for which we provide unrestrictive controls through the use of fixed effects. Nonetheless, to provide further evidence on the covariate balance in our design, we regressed a dummy variable representing each race/place-of-origin and gender identity on courses fixed effects and a set of fixed effects for comment type and the comment order (i.e., sequence). For 7 out of these 8 auxiliary regressions (Table A2), F tests indicate that we cannot reject the null hypothesis that these comment and sequence fixed effects have no "effect" on the assigned racial and gender identity of the poster.[5]

III. Empirical Framework

Our analytical strategy closely parallels our experimental design. That is, we regress our key outcomes (i.e., instructor and student responses) on seven social-identity indicators (i.e., using White male as the reference category) and condition on course, comment, and sequence fixed effects. Our preferred specification is:

$$Y_{ijkt} = \alpha + \sum_{i=2}^{8} \beta_i R_i + \theta_j + \delta_k + \mu_t + \epsilon_{ijkt}$$

where Y_{ijkt} is the outcome for posting i of comment k placed in the discussion forum of course j in the t^{th} position of the sequence of our comments in that course. R_i refers to the assigned social-identity (i.e., race/place of origin and gender) of the comment. The term, θ_j, is a course fixed effect. The term, δ_k, is a comment fixed effect, and μ_t is a sequence fixed effect for the order in which the comment appeared. We allow the error term, ε_{ijkt}, to reflect the nestedness of the comments within courses by

[5] The one exception is for female Chinese identities. A closer inspection revealed that this spurious correlation is due to our randomization causing the female Chinese identities to be linked to some comments as few as 0 times and other comments as often as 8 times. However, it should be noted that our analysis conditions on comment fixed effects; also, we observe similar findings when we drop all female Chinese observations.

clustering the resulting standard errors at the course level. We estimate this specification by ordinary least squares but find that logit estimates for binary outcomes and negative binomial estimates for count-data outcomes produce similar results.

These course, comment, and sequence fixed effects account unrestrictedly for the natural heterogeneity in outcomes by the course, sequence order of the comment, and text of the comment. That is, they control for all variables that are constant within a course (e.g., general frequency of discussion forum activity), the average number of responses each particular comment receives across all courses, and the average effects of placing a comment earlier or later in a course. While the randomization we describe above should control for any concerns about differences in response rates across courses, comments, or the timing of comments, these fixed effects further ameliorate the consequences of a potentially poor randomization. For example, if the Black female profiles were by chance assigned to place the first comment (which is often more likely to receive a response) in those classes with unusually active discussion forums, these fixed effects will control for effects related to being in an active course as well as effects related to placing the first comment.

Our outcome measures consist of three variables from two domains (i.e., instructor and student responses); our analysis focuses on whether an instructor replied to the comment, whether at least one student replied to the comment, and the total number of student replies to a comment. We also note that the choice of White male as the reference category for this analysis was a theoretically and empirically natural one. Our expectation was that most instructors would be both White and male, as proved the case (Table 1), and prior studies have identified educationally important effects linked to the demographic congruence of students and teachers. We also note that our analysis for each outcome involves estimating the effects of 7 different social identities. A concern about false discovery in the presence of multiple tests within an outcome domain would argue for privileging a single test of the joint significance of the social identities (i.e., H_0: $\beta_2=\beta_3=\ldots=\beta_8=0$). We report the results of such F tests. However, we also present the individual p-values adjusted for multiple testing using the updated Romano-Wolf procedure (Clarke, Romano and Wolf, 2020). This procedure produces stepdown adjusted p-values that control for the familywise error rate (FWER) and accommodate dependence among the p-values through bootstrap resampling. We find that inferences based on these corrections closely resemble those based on conventional, cluster-adjusted p-values (Table A3).

IV. Results

The experimental design resulted in a total of 992 postings (i.e., 8 individual comments placed across 124 courses), each of which was assigned one of eight identities (i.e., based on race/place of origin and gender). We received a total of 3,588 replies, made by 2,976 unique users. Table 1 provides descriptive statistics for responses to our comments. Instructors replied to 7.0% of our comments. At least one student responded to 69.8% of our comments with an average of 3.2 student replies to each of our comments. The variance in the number of student replies to each comment is large with comments garnering between zero and 213 replies. The next panel of Table 1 provides descriptive characteristics of the courses and comments in the study. STEM courses comprise 56.5% of the 124 courses in the sample. Fifty-eight percent of the courses in our sample were taught by either one White male instructor or a teaching team of exclusively White men. We consider 43.6% of the comments to be focused on course completion with the remainder categorized as general advising or social comments (see Table A1). The "poster identity" rows in Table 1 demonstrate that we had balance across each race/place-of-origin and gender combination; each social identity profile posted exactly one comment in each course.

Figure 1 provides a visual illustration of how the unconditional instructor (panel (a)) and peer (panel (b)) response rates varied by the randomly assigned identity of the fictive student posting a comment. The horizontal lines in these figures represent the overall sample mean. The results in panel

(a) indicate that White males were substantially more likely than all other student groups to receive a response from course instructors. That is, over 12 percent of comments assigned a White male identity received an instructor response while the overall sample mean was 7 percent. Comments assigned an Indian male identity were the least likely to receive an instructor response (i.e., 3.2 percent) followed by Chinese males and White females (i.e., 5.6 percent). Similarly, the results in panel (b) indicate that comments assigned an Asian (i.e., Indian or Chinese) male identity were those least likely to receive a response from a student peer (i.e., roughly 66 percent). However, in contrast to the pattern in instructor responses, comments assigned a White female identity were particularly likely to receive a response from another student (i.e., 80.6 percent).

Table 2 presents the key results from regressions that examine these patterns in specifications that condition on fixed effects unique to each course, to each comment, and to each order in the sequence of comments within a course (i.e., 1 through 8). The results in column (1) focus on instructor responses and unrestrictedly allow each of 7 poster identities to have a unique effect relative to the reference category (i.e., White males). A single F-test of the joint significance of these identities (i.e., H_0: $\beta_2=\beta_3=\ldots=\beta_8=0$) indicates that they have a weakly significant effect on the likelihood of an instructor response (i.e., p-value = 0.092). The point estimates indicate that, relative to comments assigned a White male identity, comments assigned to *each of* the other identities were less likely to receive an instructor response. These estimated differences range from 3.7 percentage points (i.e., with respect to Chinese females) to 9.0 percentage points (i.e., with respect to Indian males). Three of these estimated differences (i.e., Indian males, Chinese males, and White females) are statistically distinguishable from White males at the 95-percent level while another (i.e., Black males) has a weakly significant difference (i.e., p-value = 0.058). This pattern of statistical significance remains when using p-values adjusted for multiple testing (Table A3). We note that the evidence of instructors favoring White males over Asian males suggests implicit bias or taste-based discrimination over statistical discrimination as an explanation for these results. Specifically, there is evidence that instructors view male, White, and Asian students as more able and higher achieving than other groups of students (Ferguson, 2003; Hsin & Xie, 2014; Kao, 1995; Riegle-Crumb & Humphries, 2012; Tiedemann, 2002; Wong, 1980). If the evidence of instructor bias were related to statistical discrimination (i.e., instructors wanting to engage with students they may perceive to be more capable and motivated), we would not have expected the gaps between White and Asian males to be so distinctly large.

A second F test in column 1 of Table 2 fails to reject the null hypothesis that the seven coefficients share a common value (p-value = 0.477). The regression results in the second column of Table 2 impose this restriction and examine the impact of a White male identity relative to all other identities. These results indicate that comments randomly assigned a White male identity are 5.8 percentage points more likely to receive an instructor response (p-value = 0.022).[6] Given the instructor reply rate of 6.2 percent for non-White male posters, the White male effect represents a 94 percent increase in the likelihood of instructor response. In supplementary analyses, we examined the heterogeneity in this estimated effect across different course and comment features (Table A4). The results, though statistically imprecise, indicate that the evidence of bias in favor of White male students is larger in courses taught by White males and with respect to comments that are social or advising in nature and not in those that narrowly involve course completion.

The remaining results in Table 2 examine the engagement of students in these courses with the experimentally manipulated comments. In this outcome domain, we focus on both the extensive margin (i.e., whether a comment received a student reply) and the intensive margin (i.e., the number of student replies received). F tests in these regressions (i.e., columns 3 and 5) suggest that there is a

[6] The results of models examining the effect of poster race and gender on probability of a response from an instructor are similar (in sign, significance, and magnitude) when we limit the sample to MOOCs that were offered by a U.S. institution (N=94).

weakly significant effect of the identities assigned to the comments on the likelihood of receiving a peer response (i.e., column 3) but not on the number of replies (i.e., column 5). On closer examination, the heterogeneity in these effects with respect to specific identities is consistent with the results in Figure 1. Comments assigned a White female identity were particularly likely to receive a peer response (e.g., a 12.9 percentage point difference with respect to White males; p-value = 0.016). This statistically significant difference is robust after adjusting for multiple testing (i.e., 7 hypothesis tests conducted simultaneously across two outcome measures; Table A3).

Although these results provide little evidence that, on average, students in online classes differentially engage with comments posted by students of different race/place of origin and genders, that does not preclude the possibility of students preferring to respond to comments posted by people who share their race/place of origin and gender (i.e., homophily, the concept that similar people have greater social ties with each other than dissimilar people (McPherson, Smith-Lovin, & Cook, 2001)). We explored this possibility by observing the public online profiles and names of the real students who responded to our comments and identifying, when possible, their race/place of origin and gender.[7] Using these data, we identified whether our assigned comments received race/place-of-origin- and/or gender-congruent responses and estimated the effects of the randomly assigned identities on the prevalence of such homophilic responses. The results (Table A5) indicate that the only large and statistically significant result is among White female students responding to comments

[7] We determined real student race/place of origin and gender in three sequenced steps. First, we observed the public profiles of respondents to our comments. If a race and gender were provided in that public profile, we rely on the stated race and gender. Second, if the public profile did not state a race and gender but provided a picture, we use the picture to guess race/place of origin and gender. Third, in the absence of other information, we use student first and last names, which are commonly affiliated with discussion forum postings, to guess the student's race/place of origin and gender. Members of our research team coded the race/place of origin and gender of each name using their best judgment and publically available lists of names. Our research team was able to guess the gender and race/place of origin of 64% of repliers to our comments using these sources of information. Interrater reliability was high for those profiles that were double- or triple-coded. 36% of respondents did not provide enough information for the research team to guess a race/place of origin and gender.

assigned a White female identity. Specifically, we find that random assignment to a White female identity increases the likelihood of a response by a White female student by 10.3 percentage points.

V. Conclusions

In this study, we report novel field-experimental evidence that the concerns about biased personal interactions that are widely examined in the context of conventional classrooms also exist in open online courses. In other words, online learning environments, even when asynchronous, are still social environments in which social identities can have salience. We situated our field experiment in the discussion forums of MOOCs. Because online courses are often asynchronous, these forums provide a uniquely important venue for instructor-to-student and student-to-student engagement. Our field experiment produced suggestive evidence that the comparative anonymity granted by asynchronous, digitally mediated interactions in online discussion forums does not eliminate bias among instructors. In particular, we found a sizable bias in favor of White male identities which were nearly twice as likely to receive a discussion-forum response from the instructor compared to other student identities. This finding is broadly consistent with prior experimental evidence (Milkman, Akinola, & Chugh, 2015; Moss-Racusin et al., 2012) for the existence of faculty discrimination against women and racial minority applicants in interactions outside the classroom in which only applicant names were available (e.g., applying for lab positions and doctoral programs). We also found evidence that responses from other students were particularly likely for comments assigned a White female identity, an effect that appears to be due to responses from White female peers. These experimental results are consistent with other recent evidence of homophily in online learning environments (Bettinger et al., 2016).

We believe our findings also make an important contribution to the broader and quite active literature on the effects of race and gender-congruent instructors. These studies generally suffer from a limitation that attenuates their specific guidance for policy and practice. That is, these studies cannot cleanly identify the extent to which the effects of a "teacher like me" are due to student-centered effects (e.g., role model effects, stereotype threat) and/or instructor-centered effects (e.g., bias). Because our study relies on experimentally constructed student identities, it provides unambiguous evidence for the existence of effects that are instructor-centered. Furthermore, we also note that the patterns in our results (e.g., a large instructor-response gap between White and Asian males) is consistent with the hypothesis that these instructor behaviors reflect implicit or taste-based biases rather than statistical discrimination. While this evidence does not preclude the relevance of student-centered effects, it does suggest that teacher-facing interventions that reduce biased behaviors are likely to be both well targeted and effective in supporting student engagement.

Despite the advantages of our field-experimental approach, at least three caveats are notable. First, we intentionally chose names based on their clear affiliation with a race/place of origin and gender profile. Students with names less easily associated with a specific race/place of origin and gender may face less discrimination. Second, our study was situated in a specific form of online education (i.e., free, open, asynchronous courses not associated with a specific degree program) and it is not clear that these findings would be observed in face-to-face courses or in online courses with more opportunities for instructors and peers to gain insight into students' engagement and ability. Third, because our forum posters are fictive, we cannot assess the effects that the biases we observe may have on student performance or persistence in the course. Because the instructor and peer-engagement measures we study are in all likelihood important mediators of learning outcomes, we suspect that such effects exist. However, examining the effects of bias on student outcomes in online settings will require further and different study.

For example, one broad and possibly compelling direction would be to design, implement, and evaluate alternatively designed online learning environments that are effective in promoting equitable forms of engagement. Understanding the determinants of student engagement in online settings is particularly relevant because these environments, especially Massive Open Online Courses (Evans, Baker, & Dee, 2016; Perna et al., 2014), often suffer from low in-course persistence. Relative to conventional classrooms, online environments are uniquely amenable to such design innovations, in part because they can be implemented at scale with both fidelity and relatively little cost. One obvious and simple approach would be to structure these classrooms in a manner that kept student identities strictly anonymous (e.g., removing names and photos). However, we also note that such extreme anonymity may have unintended consequences. A more sophisticated approach would be to structure online environments that guide instructors to engage with students in more equitable ways (e.g., dashboards that provide real-time feedback on the characteristics of their course engagement or short, embedded professional-development modules). The design features of online learning environments can also be adapted to either reduce homophily among students or to promote it when it aligns with educational goals. Regardless, our field-experimental study suggests such design innovations merit careful consideration given the evidence of biases our study uncovered.

References

American Association of University Women (AAUW). (1992). *How Schools Shortchange Girls.* Washington, D.C.: AAUW Educational Foundation.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review,* 94, 991-1013.

Bettinger, E., Liu, J., & Loeb, S. (2016). Connections matter: How interactive peers affect students in online college courses. *Journal of Policy Analysis and Management*, 35, 932-954.

Bettinger, E. P., Fox, L., Loeb, S., & Taylor, E. S. (2017). Virtual classrooms: How online college courses affect student success. *American Economic Review*, *107*(9), 2855-75.

Clarke, D., Romano, J. P., & Wolf, M. (2020). The Romano–Wolf multiple-hypothesis correction in Stata. *The Stata Journal*, *20*(4), 812-843.

Dee, T.S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, 86, 195–210.

Dee, T.S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 95, 158–165.

Dee, T.S. (2020) "VCs Are Pouring Money in the Wrong Education Startups," Wired, November 19, 2020. https://www.wired.com/story/vcs-are-pouring-money-into-the-wrong-education-startups/

Evans, B. J., Baker, R. B., & Dee, T. S. (2016). Persistence patterns in Massive Open Online Courses (MOOCs). *Journal of Higher Education*, 87, 206-242.

Fairlie, R. W., Hoffman, F., & Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review*, 104, 2567-2591.

Ferguson, R. F. (2003). Teachers' perceptions and expectations and the black-white test score gap. *Urban Education*, 38, 460-507.

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, *52*, 209-224.

Hsin, A. & Xie, Y. (2014). Explaining Asian Americans' academic advantage over whites. *Proceedings of the National Academy of Science*, 111: 8416-8421.

Kao, G. (1995). Asian Americans as model minorities? A look at their academic performance. *American Journal of Education*, 10: 121-159.

Lindsay, C. A., & Hart, C. M. (2017). Exposure to same-race teachers and student disciplinary outcomes for Black students in North Carolina. *Educational Evaluation and Policy Analysis*, *39*(3), 485-510.

McPherson, M. S., & Bacow, L. S. (2015). Online higher education: Beyond the hype cycle. *Journal of Economic Perspectives*, 29, 135-153.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415-444.

Milkman, K. L., Akinola, M., & Chugh, D. (2015). What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology*, 100, 1678-1712.

Moss-Racusin, C., Dovidio, J. F., Brescoll, V. L., Graham, M. J. & Handelsman, Jo. (2012). Science faculties subtle gender biases favor male students. *Proceedings of the National Academy of Science*, 109, 16474-16479.

Oreopoulos, P. (2011). Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes. *American Economic Journal: Economic Policy*, 3, 148-171.

Perna, L. W., Ruby, A., Boruch, R. F., Wang, N., Scull, J., Ahmad, S., and Evans, C. (2014). Moving through MOOCs: Understanding the progression of users in Massive Open Online Courses. Educational Researcher, 43, 421–432.

Riegle-Crumb, C. & Humphries, M. (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender & Society*, 26: 290-322.

Shah, D. (2019). By the Numbers: MOOCS in 2019 - Class Central. Retrieved from https://www.classcentral.com/report/mooc-stats-2019/

Tajfel, H. & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (eds.) *The social psychology of intergroup relations* (pp. 33-47). Monterey, CA: Brooks/Cole.

Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*(2), 253-273.

Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in Mathematics*, 50: 49-62.

Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, 47 (2), 497-527.

Wong, M. G. (1980). Model students? Teachers' perceptions and expectations of their Asian and white students. *Sociology of Education*, 53: 236-246.

Table 1 - Descriptive Statistics

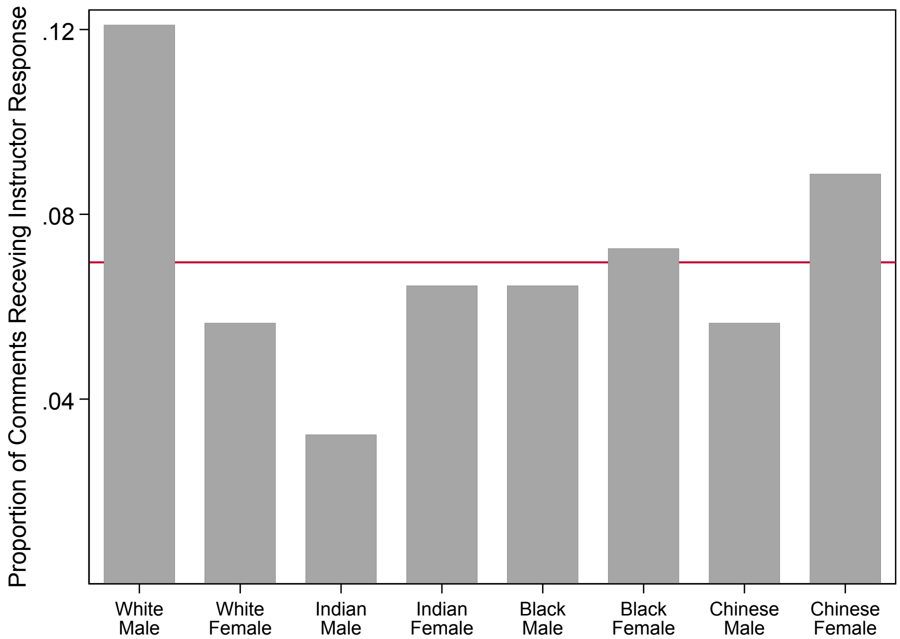
Variables	Mean	SD	Min	Max
Outcomes				
Instructor Replied (0/1)	0.070	0.255	0	1
Student Replied (0/1)	0.698	0.460	0	1
Number of Student Replies	3.205	9.817	0	213
Course/Comment Characteristics				
STEM Course	0.565	0.496	0	1
White-Male Instructor	0.581	0.494	0	1
Completion-Focused Comment	0.436	0.496	0	1
Poster Identity				
White Male	0.125	0.331	0	1
White Female	0.125	0.331	0	1
Black Male	0.125	0.331	0	1
Black Female	0.125	0.331	0	1
Indian Male	0.125	0.331	0	1
Indian Female	0.125	0.331	0	1
Chinese Male	0.125	0.331	0	1
Chinese Female	0.125	0.331	0	1

Notes: The unit of observation is a comment placed in the discussion forums of online courses (i.e., 8 comments in each of 124 courses, N=992). The poster identity, the comment placed, and their sequencing were randomly assigned. See text for details. White-male instructor courses include single instructor courses taught by a White male and multiple instructor courses taught exclusively by White males. Non-completion-focused comments are comments labeled advising/social. See Appendix Table A1 for comment categorization.

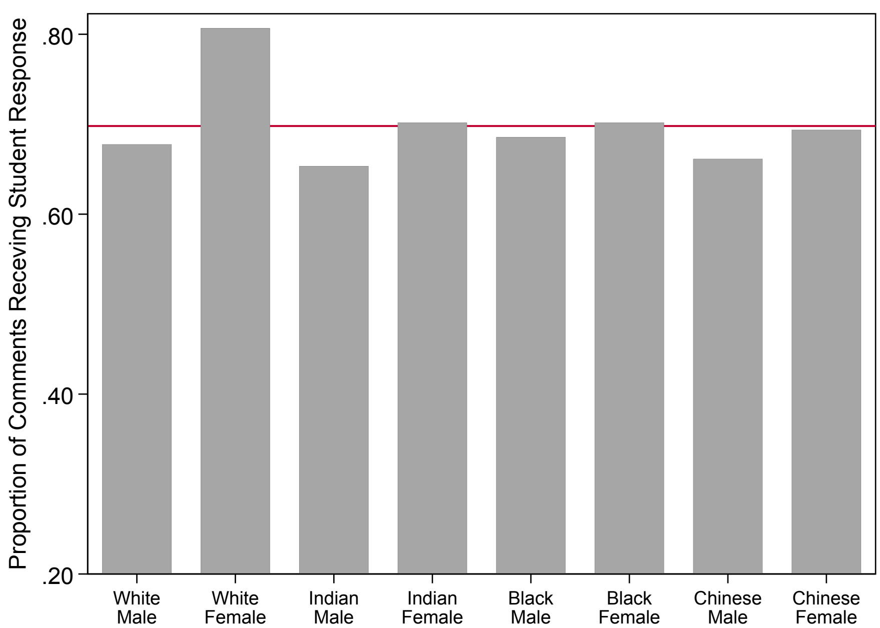
Table 2 - The Estimated Effects of Student Identities on Instructor and Peer Responses

Independent Variable	Instructor Replied		Student Replied		Number of Student Replies	
	Dependent Variables					
White Male	-	0.058*	-	-0.021	-	-0.590
		(0.025)		(0.041)		(0.495)
White Female	-0.069*	-	0.129*	-	1.391	-
	(0.034)		(0.053)		(0.937)	
Black Male	-0.055+	-	0.011	-	-0.281	-
	(0.029)		(0.054)		(0.614)	
Black Female	-0.046	-	0.035	-	0.158	-
	(0.032)		(0.055)		(0.638)	
Indian Male	-0.090**	-	-0.023	-	0.551	-
	(0.028)		(0.057)		(0.911)	
Indian Female	-0.059	-	0.013	-	1.601	-
	(0.036)		(0.061)		(1.580)	
Chinese Male	-0.055*	-	-0.036	-	-0.235	-
	(0.027)		(0.059)		(0.671)	
Chinese Female	-0.037	-	0.017	-	0.909	-
	(0.035)		(0.053)		(0.785)	
p-value: H_0: $\beta_2=\beta_3=\ldots=\beta_8=0$	0.092	-	0.089	-	0.369	-
p-value: H_0: $\beta_2=\beta_3=\ldots=\beta_8$	0.477	-	0.064	-	0.308	-
R^2	0.049	0.044	0.105	0.093	0.212	0.207

Notes: + $p < 0.10$, * $p < 0.05$ ** $p <0.01$. All analyses condition on course, comment, and sequence fixed effects. The p-values refer, respectively, to F tests for the joint significance of and the equivalence of the effects associated with the 7 non-white male poster identities. Standard errors, presented in parentheses, are clustered at the course level. The sample size is 992 (i.e., 8 comments posted in each of 124 courses).



(a) Instructor response



(b) Peer Response

Figure 1 - Probability of Instructor and Peer Responses by Student Identity

Table A1 – Forum Comments

Completion-Focused Comments
I joined this class late and am wondering if I missed anything that is important.
Are there links to other resources that could be helpful for the lessons?
I am putting off watching the lectures. Does anyone have any tips to help me not procrastinate?
Should I watch the videos all at once or one by one? What do others do?
I'm finding the lectures difficult to follow. Anyone else?
I haven't watched all of the lectures. I don't think I will be able to catch up -- what is the best lecture for me to watch?
How should I complete the assignments? Does anyone have tips on how to do them well?
What's the minimum percentage I have to get to pass?
How do I submit assignments? Can someone please explain this to me?
Do we just have to watch the lecture videos? Is there anything else we have to do?
Are the lectures the only homework assignments? Is there anything else?
How do I find out how well I am doing in this class?
What kinds of things do I need to know to do well in this class?
Advising/Social Comments
Does anyone use this course material for their job?
Is this class harder or easier than other classes in this field?
Anyone have any ideas on courses that would be good to take after this one?
What is the goal of this class? Is it mostly theoretical or is it also practical?
Do people like this class? I am not sure I can finish it, but I might take it later. Is it worth it?
I am learning lots from this class, even though it is a lot of work. Does anyone else feel this way?
I am falling behind in this course. How is the workload?
Where are people in this class from?
Are you taking this class for fun? Are you a student or are you working?
I am struggling in this class. Does anyone else find it to be hard?
I am feeling more confident about this class, though I struggled at first. Does anyone else feel this way?
This class is challenging, and I am really enjoying the challenge!
This class isn't as hard as I expected! I am enjoying it.
I don't find all of the lectures to be that helpful.
I am just starting week two of the class. Where are other people?
This class is great. It is perfect timing for me!
I don't have any prior experience. Will I do okay? What are the backgrounds of other people in the class?
Do you think I should put this class on my resume?

What do I need to do to unenroll from the class?

Table A2 - Testing the Balance of Student Identities
by Comment and Comment Order

Student Identity	*p*-value
White Male	0.1131
White Female	0.8776
Indian Male	0.6817
Indian Female	0.9985
Black Male	0.6890
Black Female	0.9042
Chinese Male	0.5891
Chinese Female	0.0065

Notes: Each row is based on a separate regression in which the race-gender profile is regressed on indicators for comment and comment order. The p-value is based on an F-test of the joint significance of comment and comment-order fixed effects. Each regression also conditions on course fixed effects.

Table A3 – p-values from Table 2 Adjusted for Multiple Hypothesis Testing

Treatment Arm	Cluster-corrected p-value	Stepdown Adjusted p-value
Dependent variable: Instructor Replied		
White Female	0.044	0.047
Indian male	0.002	0.001
Indian Female	0.104	0.105
Black Male	0.058	0.052
Black Female	0.150	0.162
Chinese Male	0.048	0.039
Chinese Female	0.295	0.313
Dependent variable: Student Replied		
White Female	0.016	0.014
Indian Male	0.682	0.746
Indian Female	0.838	0.830
Black Male	0.843	0.834
Black Female	0.523	0.684
Chinese Male	0.543	0.699
Chinese Female	0.755	0.740
Dependent variable: Student Replies		
White Female	0.140	0.093
Indian Male	0.546	0.746
Indian Female	0.313	0.630
Black Male	0.647	0.812
Black Female	0.804	0.721
Chinese Male	0.726	0.699
Chinese Female	0.249	0.334

Note: The cluster-corrected p-values are from the main results in Table 2. The stepdown adjusted p-values are based on the updated Romano-Wolf procedure (Clarke, Romano, and Wolf 2020), which controls for the familywise error rate (FWER) and allows for dependence among the p-values through bootstrap resampling. The adjusted p-values for the student responses reflect the testing of multiple hypotheses across two outcomes from the same outcome domain.

Table A4 - The Estimated Effects of a White Male Student Identity on Instructor Responses by Instructor, Course, and Comment Traits

Sample Construction	Instructor Replied	Sample Size
Full Sample	0.058*	992
	(0.025)	
White Male Instructor	0.075*	576
	(0.037)	
Non-White Male Instructor	0.049	416
	(0.035)	
STEM	0.048	560
	(0.037)	
Non-STEM	0.043	432
	(0.031)	
Completion-Focused Comment	0.025	433
	(0.029)	
Advising/Social Comment	0.060+	559
	(0.033)	

Notes: + p < 0.10, * p < 0.05 ** p <0.01. Each cell reports the estimated effect of a White-male poster identity relative to all other poster identities conditional on course, comment, and sequence fixed effects. Standard errors, presented in parentheses, are clustered at the course level. White-male instructor courses include single instructor courses taught by a white male and multiple instructor courses taught exclusively by White males. See Appendix Table A1 for comment categorizations.

Table A5 - The Estimated Effects of Student Identities on Race and Gender-Congruent Peer Responses

	Dependent Variables	
Independent Variable	Student Replied	Number of Student Replies
White	0.059+	0.286
	(0.035)	(0.363)
Black	-0.007	0.000
	(0.014)	(0.022)
Indian	0.042+	0.091+
	(0.022)	(0.055)
Chinese	-0.009	-0.005
	(0.015)	(0.021)
Female	0.045+	0.375+
	(0.026)	(0.201)
White Male	-0.031	-0.095
	(0.041)	(0.192)
White Female	0.103**	0.504
	(0.038)	(0.321)
Black Male	-0.011	-0.022
	(0.015)	(0.018)
Black Female	0.027	0.043
	(0.018)	(0.027)
Indian Male	0.007	0.043
	(0.026)	(0.077)
Indian Female	-0.005	0.009
	(0.013)	(0.027)
Chinese Male	-0.003	-0.001
	(0.015)	(0.020)
Chinese Female	0.005	0.006
	(0.014)	(0.017)

Notes: + $p < 0.10$, * $p < 0.05$ ** $p < 0.01$. Each cell reports the estimated effect of the poster identity from a unique regression in which the dependent variable is a reply (or the number of replies) from peers with the poster's race and/or gender identity. We identified the race and gender of student peers for 64 percent of repliers (see text for details). All analyses condition on course, comment, and sequence fixed effects. Standard errors, presented in parentheses, are clustered at the course level. The sample size is 992 (i.e., 8 comments posted in each of 124 courses).