***This manuscript was published in the Journal of Experimental Social Psychology in November 2020. The article is available open access at** *https://www.sciencedirect.com/science/article/pii/S0022103120303668*

The Impact of Classroom Diversity Philosophies on the STEM Performance of Undergraduate Students of Color

Jessica J. Good

Davidson College

Kimberly A. Bourne

University of Washington

R. Grace Drake

Washington University in St. Louis

Author Note

Jessica J. Good, Department of Psychology, Davidson College; Kimberly A. Bourne, Department of Psychology, University of Washington; R. Grace Drake, Department of Psychology, Washington University in St. Louis.

This research was supported by a Research in Undergraduate Institutions grant from the National Science Foundation (grant ID #1650290). Special thanks to Drs. Donna Molinek, Durwin Striplin, and Mario Belloni for designing lessons, the members of the Davidson Intergroup Studies Lab, and to all the faculty around the country who helped to recruit participants. Correspondence regarding this article should be addressed to Jessica Good, Box 7136, Davidson College, Davidson, NC 28035. Contact: jegood@davidson.edu

Abstract

Using a large, nationally representative sample of first year undergraduate students we tested whether instructors' use of diversity philosophies could impact the learning of new math and science content among Students of Color and White students. Participants (N = 688) were randomly assigned to one of nine simulated online course environments using a 3 (diversity philosophy: Multicultural, Colorblind, Control) x 3 (lesson: Chemistry, Physics, Math) x 2 (participant race: Students of Color, White students) between-participants experimental design. After listening to an audio welcome message from the instructor and reading the course syllabus, both of which contained the embedded diversity philosophy manipulation, participants watched a novel 10-minute lesson, completed a comprehension quiz, as well as measures of belonging and perceived instructor bias. Students of Color showed greater comprehension of the math/science lesson in the multicultural condition compared to the colorblind condition. Students of Color also perceived the instructor to be more biased in the multicultural condition compared to the colorblind condition. White students tended to either be unaffected or oppositely affected by the diversity philosophy manipulation. Overall, results suggest that college instructors' use of multicultural (or colorblind) language sends a signal of inclusion (or exclusion) to Students of Color, affecting not only their social experience in the class but also their learning potential.

Keywords: diversity philosophy; STEM; multiculturalism; colorblindness; classroom; education

The Impact of Classroom Diversity Philosophies on the STEM Performance of Undergraduate Students of Color

A critical issue facing the United States at present is the underrepresentation of women and People of Color in Science, Technology, Engineering, and Math (STEM) fields (NSF, 2019), as well as the performance gap in favor of White men in these domains (Cohen, Garcia, Apfel, & Master, 2006; Haak, HilleRisLambers, Pitre, & Freeman, 2011). Although the reasons for this gap are multifaceted – including historical, economic, and cultural factors – social psychological research has identified both direct (e.g., active discrimination; Milkman, Akinola, & Chugh, 2015; Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012) and indirect signals of exclusion (e.g., lack of role models, field-specific ability beliefs; Dasgupta & Stout, 2014; Leslie, Cimpian, Meyer, & Freeland, 2015; Ramsey, Betz, & Sekaquaptewa, 2013) for people with underrepresented or marginalized identities. We propose that explicit diversity philosophies, used by both educational institutions and instructors, are an additional signal of exclusion or inclusion that can impact students' sense of belonging and ultimately their performance in STEM.

It is particularly important to study the use of diversity philosophies in educational contexts because of their widespread adoption; as of 2012, 65% of educational institutions had an official diversity statement (Wilson, Meyer, & McNeal, 2012). These statements often feature an underlying colorblind (CB) or multicultural (MC) philosophy; CB promotes racial and cultural assimilation by emphasizing similarities between groups, while MC promotes racial and cultural inclusion by valuing differences between groups (Plaut, 2010). These institution-level diversity statements have been shown to predict student performance. For example, women undergraduates of color who read a university brochure featuring a MC statement expected greater racial and gender diversity and performed significantly better on a standardized math test than those who read a brochure featuring a CB statement (Wilton, Good, Moss-Racusin, & Sanchez, 2015). Likewise, a content analysis of existing diversity policies within Belgian middle schools showed that MC policies were associated with smaller gaps in sense of belonging and academic achievement between majority and minority ethnic groups, while CB policies were associated with greater achievement gaps over time (Celeste, Baysu, Pahlet, Meeussen, & Kende, 2019).

Institutions set the tone for how diversity is to be managed, but individual instructors vary in the way they discuss (or avoid) diversity within the classroom. Teachers with greater MC beliefs report higher motivation, lower agreement with negative stereotypes, and more willingness to adapt their teaching to differing student needs (Hachfeld, Hahn, Schroeder, Anders, & Kunter, 2015). Similarly, faculty endorsement of MC predicted greater adoption of inclusive teaching practices whereas faculty who endorsed CB to a greater extent were less persuaded to adopt inclusive teaching practices (Aragón, Dovidio, & Graham, 2016). Although this research shows that instructors' individual diversity philosophies impact their teaching practices, we do not know the impact of these individual philosophies on students' actual learning. Regardless of the top-level institutional diversity messaging, students may be affected by their individual instructor's stated views on diversity. Given that the majority of the U.S. STEM professoriate is White (Hess, Gault, & Yi, 2013), and that White teachers may be especially likely to endorse a CB philosophy (Johnson, 2002; Lewis, 2001; Sleeter, 2001), the message most students receive in their STEM courses is likely CB in content. This instructor-specific diversity messaging may contribute to inequities in class performance for Students of Color as compared to White students.

However, to our knowledge no prior published literature has systematically investigated the impact of instructors' use of CB and MC messages on student performance, particularly for newly learned material. Prior evidence suggests that Students of Color may view MC language as a signal of inclusion and show higher performance (Wilton et al., 2015), whereas White students may feel threatened (Dover, Major, & Kaiser, 2016) or excluded (Plaut, Buffardi, Garnett, & Sanchez Burks, 2011) when MC language is used. In the present study, we experimentally test whether instructors' use of diversity philosophies (CB, MC, or Control) impacts Students of Color and White students' learning of novel STEM content and performance in that STEM domain. We argue that introductory level undergraduate STEM courses are an important point of intervention given that Students of Color and White students intend to major in STEM disciplines at comparable rates, but Students of Color are disproportionately more likely to switch out of STEM majors (Riegle-Crumb, King, & Irizarry, 2019; Syed, 2010). We provide an ecologically valid investigation by having first year college students from around the U.S. enroll in a simulated online STEM course. We manipulate the diversity philosophy espoused by the instructor before having students watch a novel video lesson and complete a comprehension quiz, as well as measures of belonging and perceived instructor bias.

In our initial hypotheses, we predicted that instructors' diversity philosophies would impact student outcome variables differently depending on whether students were members of underrepresented racial minority groups (URM; Black, Latino/a/x, Native American) or overrepresented groups (ORG; White, Asian) based on U.S. national statistics (NSF, 2019; Ogunwole, Drewery, & Rios-Vargas, 2012). We organized our predictions according to racial and ethnic group representation (relative to overall U.S. population estimates) in higher education broadly (Ogunwole, et al., 2012) and in STEM specifically (Ferrare & Lee, 2014;

NSF, 2019), as well as common comparison groups within the literature on STEM retention (see Dierker, Alexander, Cooper, Selya, Rose, & Dasgupta, 2016). However, upon post hoc examination of our data and further reflection on the literature, we realized that Participants of Color (POC), including Asian and Asian American students, tended to show greater similarity in their responses than did our initial comparison groups. We therefore chose to alter our analytic plan and compare the performance of Participants of Color (POC) to White students rather than grouping participants based on national STEM representation statistics. In general, we tested whether MC, compared to CB, would serve as a signal of inclusion for POC, resulting in greater lesson comprehension and feelings of belonging, and lower perceptions of instructor bias.

Method

All measures, manipulations, and exclusions are reported in text or SOM; no data were analyzed prior to concluding data collection. All audio files, videos, materials, data, preregistration, and SOM are available at https://osf.io/qyna9/?view_only=f0b56e602a4c4d7791ea413e1ea40116.

Participants

We conducted an a priori power analysis using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) to estimate the sample size necessary for main effects and interactions at 80% power using the between participants effect size $f = .15$ (based on Wilton et al., 2015). The analysis yielded an estimate of 536 participants; we aimed to recruit 600 participants in order to allow for potential exclusions. We set the following a priori exclusion criteria: skipped more than 10 percent of questions, indicated that they were not a first-year college student, total time spent on the quiz was +/- 3 SDs from the mean, did not answer all dependent variable questions, or wrote gibberish on free response questions. Post hoc, we also decided to exclude participants who had completed the survey multiple times due to a temporary security glitch, and who took more than 24 hours, or less than 10 minutes to complete the survey

We recruited first year students from 219 colleges and universities around the country by emailing professors with a request to send out our survey information to their first-year students. Of the 829 students who participated, 141 were excluded[1]. Within our final sample of 688 participants, 53.1 % identified as women, 45.2 % identified as men, and 1.7% identified as genderqueer, non-conforming, or other. The racial and ethnic composition of the sample was as follows (participants could select more than one group): 45.5% White, 21.8% Black or African American, 24.7% Hispanic or Latinx, 13.7% East Asian, 9.6% South Asian, 3.6% American Indian or Alaskan Native, 2.8% Middle Eastern, 1.6% Native Hawaiian or Pacific Islander, and 1.2% as other. Participants were all first-year undergraduate students between 18 and 27 years old (86.3% = under 20, 11.3% = 20, 2.3% between 21 and 27) and were currently attending historically black colleges or universities (1.6%), community colleges (7.7%), liberal arts colleges (11.8%), private universities (26.0%), and public universities (52.8%) in the U.S.[2] All participants were compensated with a $15 Amazon e-gift card.

Materials

Diversity philosophy manipulation. The manipulation of diversity philosophies (MC, CB, or Control) was administered to participants twice during the course of the study, first in the "welcome" audio recording on the course home page, and then within the behavior policy on the

[1] One participant wished to be removed from analyses. Thirty-one participants were excluded for taking the course multiple times. Five participants were removed because skipped more than 10% of questions, and 9 were removed for not being in their first year of college. Because survey duration varied so widely, we first excluded based on inattention or possible survey malfunction before calculating mean quiz time; 29 participants were excluded for spending over 24 hours on the study and 48 were excluded for spending less than 10 minutes on the study (since the lesson videos were 10 minutes, it was not possible to complete the study in less time). Eighteen participants were then excluded for having quiz times that were 3 standard deviations above the mean. After these exclusions, we did not need to exclude any participants for writing gibberish on free response questions.

[2] One person did not know the type of school they attended.

course syllabus. Because prior research has identified a positivity confound in common manipulations of MC and CB (authors, 2019), we focused our manipulation on only the central ideological difference between MC and CB (whether group differences vs. similarities are emphasized).[3] In the 80-second "welcome" message, the instructor gave a brief overview of the course. In the MC *(and CB)* conditions, the instructor additionally stated, "my philosophy is that throughout the course I would like all of you to keep in mind the ways that you are different from *(similar to)* each other. Our differences *(similarities)* allow us to collaborate well together and enrich our learning." In the Control condition the instructor simply said, "my philosophy is that throughout the course I would like all of you to be considerate and respectful of your peers".

The diversity philosophies were expanded upon within the behavior policy section in the syllabus. All syllabi were two pages long and included instructor information, course description, statement on academic integrity, behavior policy, statement on academic accommodations, a list of assignments, and the class schedule. Across the three STEM courses, only the specific topical concepts covered varied, and across the three diversity philosophy manipulations, only the behavior policy varied[4]. The behavior policy for the MC *(and CB)* conditions was:

> My classroom policy promotes learning to give all students an environment in which they can flourish. I encourage students to think about the ways in which we are different *(similar)* concerning race, gender, sexuality, religion and ethnicity, since valuing the

[3] Though diversity philosophies have been manipulated in various ways (see Whitley & Webster, 2019 for a review), our manipulations are consistent with the focus on valuing differences (MC) vs. emphasizing similarities (CB) that has been used in much of the literature (e.g., Apfelbaum et al., 2017; Cho et al., 2017; Purdie-Vaughns et al., 2008; Wilton et al., 2015).

[4] The syllabi were pilot tested on Amazon Mechanical Turk ($N = 105$), with the MC syllabus rated as valuing differences in the classroom more so than the CB syllabus ($t(38) = 4.17$, $p < .01$, 95% CI [.85, 2.45], $d = 1.31$), and the CB syllabus rated as valuing similarities in the classroom more so than the MC syllabus ($t(38) = -3.73$, $p < .01$, 95% CI [-2.16, -.64], $d = 1.17$). There were no differences in perceived difficulty, positivity of the syllabus, perceived effort, comfort, success, fairness of course, or realism across the three types of courses (Chemistry, Physics, Math), or between the three diversity philosophy conditions (MC, CB, Control). See SOM for details.

ways in which we are different *(similar)* can enhance the classroom experience. Focusing on the ways in which our backgrounds and identities are different *(similar)* can increase feelings of empathy and foster better student interactions. It is my hope that we can use this philosophy to work together towards a common learning goal.

The behavior policy for the Control condition simply emphasized respect for class members.

STEM lessons. Three original academic lessons (Chemistry, Physics, and Math) were created for the present research by college faculty in those respective disciplines. The lessons included material typically covered in introductory undergraduate classes and did not require prior knowledge of the discipline. The lessons were between 10 and 11 minutes in length and were presented as animated PowerPoint videos narrated by the instructor of the course[5].

Lesson quizzes. A 10-question multiple choice quiz was created for each lesson (Chemistry, Physics, and Math) by the same faculty who wrote each lesson[6]. We computed the percentage of questions answered correctly as our measure of student performance.

Belonging. Belonging was measured using 8 questions from Walton & Cohen (2007; e.g., "I would belong in this class" and "I would know how to do well in this class"). Participants were asked to think about what the online course was like for them and rate the statements on a scale of 1 (*strongly disagree*) to 7 (*strongly agree*). Mean scores were computed (Cronbach's α = .83).

Perceived instructor bias. Four self-report items (e.g. The professor evaluates all students impartially" and "The professor holds unconscious negative attitudes towards racial or

[5] We conducted pilot testing on Mechanical Turk (N = 138) to determine if there were any differences in enjoyment, engagement, difficulty, boringness, or challenge across the lessons. We found no differences on any of the variables; see SOM.

[6] In pilot testing with a Mechanical Turk sample (N = 138), time spent on the quizzes and quiz scores did not significantly differ across the three lessons.

ethnic minority students") adapted from Perry, Murphy, & Dovidio (2015) were rated on a scale of 1 (*strongly disagree*) to 7 (*strongly agree*). Mean scores were computed (Cronbach's α = .73).

Attention checks. Participants responded to an open-ended question ("What can you remember of the professor's behavior policy?"), followed by a multiple-choice question ("Which of the following does the professor believe will create the most beneficial learning environment?") with one answer choice corresponding to each of the 3 diversity philosophy conditions. The first question was administered in order to potentially exclude participants who were inattentive (wrote gibberish), and the second was included as an indicator of how well participants were able to explicitly remember the diversity philosophy presented.

Procedure

Students who responded to our recruitment email were directed to a consent form and eligibility survey hosted via Qualtrics[7]. Participants were told that the purpose of the study was to understand how to best present online course material to students for optimal learning and retention. After completing demographic questions, all participants who met inclusion criteria were randomly assigned to one of 9 possible online courses in a 3 (diversity ideology: CB, MC, Control) x 3 (lesson: chemistry, physics, math) between-participants design. Courses were created in Moodle (a common online learning platform) and were designed to mimic a typical online course (see Figure 1).

----

[7] Initially, all participants who indicated they were a first-year college student were eligible. As data collection continued, we adjusted inclusion criteria base on race and gender to ensure than our sample was sufficiently diverse.

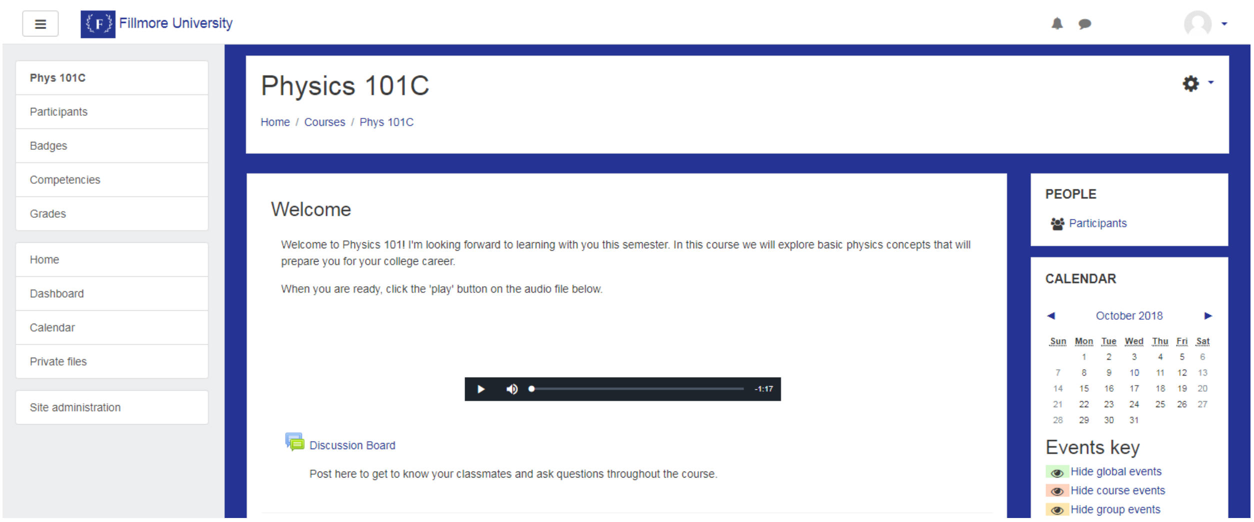

Figure 1. Screenshot of online course homepage.

Moodle courses were designed so that students felt like they were enrolling in a full course (a series of 10 chapters/lessons that "unlocked" as they completed activities) though we only built accessible content for "Lesson 1." The course homepage included welcome text that instructed participants to listen to an audio recording ostensibly created by the instructor of the course (Dr. Brandon Johnson). In reality, all audio content for the study was created by a professional voiceover artist (middle-aged male voice). The instructor gave a brief overview of the course, emphasized his class behavior policy (manipulation of CB, MC, or Control) and told the students to read the syllabus. Participants then navigated to the syllabus, followed by a 5-question syllabus quiz. Participants were required to correctly answer all questions before moving forward in the course (3 attempts allowed). After successfully completing the syllabus quiz, the "Lesson 1" module was then unlocked (STEM Lesson described above). Students were able to fast forward, rewind, and pause the lesson in order to take notes. After, participants completed the corresponding Lesson Quiz, followed by the measures of Belonging, Perceived

Instructor Bias, and attention checks[8]. They then filled out demographic questions and read a debriefing form. Once participants fully completed the experiment, they were asked to provide a valid university email address; e-gift cards were sent within 7 days of participation.

Results

Preliminary Testing

We conducted a sensitivity power analysis for the primary diversity philosophy condition by participant race interaction using G*Power 3.1 (Faul et al., 2007), with an alpha level of .05, 80% power, and sub-sample size of 452 (MC and CB conditions only). We had the power to detect an effect size of $f = .13$ ($\eta_p^2 = .017$).

For the full sample, quiz score and belonging were positively correlated, and both were negatively correlated with perceived instructor bias (see Table 1).

Table 1

Bivariate Correlations Among Dependent Variables

Variable	1	2	3
1. Quiz Score	--		
2. Belonging	.36**	--	
3. Instructor Bias	-.16**	-.27**	--

Note. $^{**}p < .01$

Data from the multiple-choice attention check question suggest that a fair number of participants were unable to explicitly remember the diversity philosophy to which they were exposed; 15.5%

[8] Participants also completed 4 items assessing their belief in a growth mindset and 2 items assessing their anticipated quiz performance. The race x diversity philosophy condition interactions were not significant for either variable; thus, for brevity these analyses are reported in the SOM.

in the MC condition, 4.7% in the CB condition, and 38.6% in the Control condition failed this attention check. Because of the large number of participants and the fact that they were not evenly distributed across conditions, we opted not to exclude participants who answered incorrectly. We did conduct analyses using whether or not they correctly answered the attention check as a covariate (see SOM), finding patterns of results that were strikingly similar to the uncovaried results. Ultimately, we decided to include all participants and present uncovaried analyses in the manuscript given that it was not clear whether participants who answered incorrectly actually failed to be affected by the manipulation. Because of the wording of the question ("Which of the following does the professor believe will create the most beneficial learning environment?"), it is possible that participants did not realize they were being asked to recall specific information rather than simply give their impression of the instructor's beliefs. This may be why students in the Control condition were especially likely to answer incorrectly.

Primary Analyses

As stated earlier, we pre-registered our hypotheses and analyses to compare under- and over-represented racial groups. However, upon examination of the data and further reflection, we realized that this group dichotomization did not accurately depict the data. We observed that the pattern of Asian participants' performance across the MC and CB conditions more closely mirrored that of other Participants of Color rather than White participants (see Figure 2).

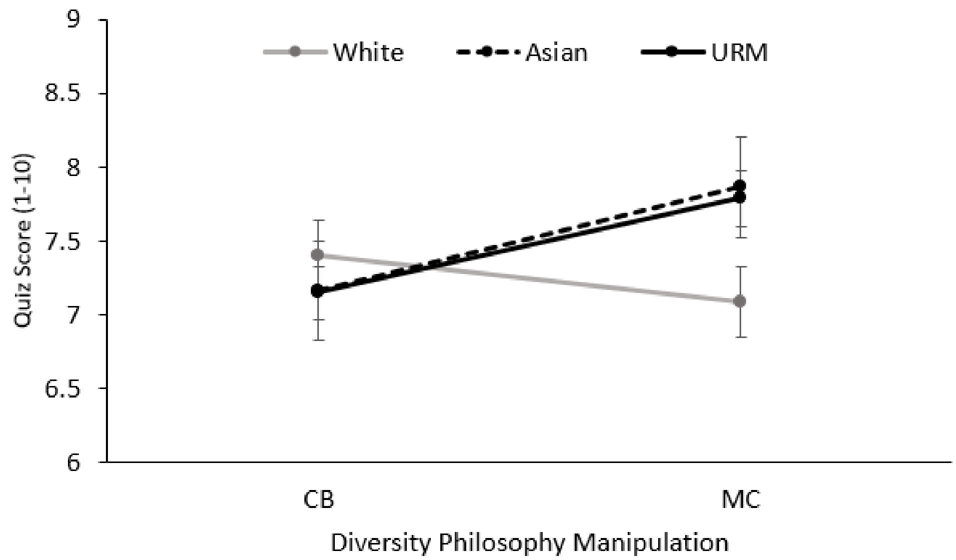

Figure 2. Mean quiz score in the CB and MC conditions, for White participants, Asian participants, and under-represented racial minority (URM) participants. Error bars represent standard error.

The following analyses represent a deviation from our pre-registered analysis plan in that we compare White participants to all Participants of Color (including Asian participants) rather than grouping White and Asian participants together. For participants who indicated more than one racial or ethnic identity, we asked them if they identified more strongly with one identity and grouped them based on that identity. If they did not identity more strongly with one identity and they listed both White and a racial or ethnic minority identity, they were grouped with Participants of Color. This yielded 231 White participants and 457 Participants of Color (POC).[9] All other aspects of our pre-registered analysis plan remain intact; in other words, we follow the

[9] Racial groups were approximately evenly distributed across diversity philosophy conditions: MC condition POC = 146, White = 73; CB condition POC = 159, White = 74; Control condition POC = 152, White = 84.

pre-registered analyses as planned, other than the change to the racial categorizations as described here.[10]

We formulated hypotheses concerning only the MC and CB conditions as we expected the clearest differences between these two conditions, although we included a Control condition in order to compare it to the other conditions in an exploratory way. Therefore, we first test for differences between the two experimental diversity philosophy conditions (CB and MC) as a function of participant race for all dependent variables. Next, we explore whether the Control condition differs from either experimental condition.

Quiz Score. We conducted a 2 (race: POC, White) x 2 (condition: MC, CB) x 3 (lesson: chemistry, physics, math) ANOVA with quiz score as the dependent variable. We found a significant main effect of lesson ($F(2, 440) = 5.17$, $p = .006$, $\eta_p^2 = .023$), which was qualified by a significant lesson by participant race interaction ($F(2, 440) = 5.25$, $p = .006$, $\eta_p^2 = .023$). Looking the effects of lesson separately by participant race, we found that the STEM lesson significantly impacted the scores of Participants of Color ($F(2, 299) = 16.94$, $p < .001$, $\eta_p^2 = .10$), but did not impact the scores of White participants ($F(2, 141) = 0.28$, $p = .755$, $\eta_p^2 = .004$). Participants of Color scored higher in the Math lesson ($M = 8.10$, $se = .19$; LSD $M_D = 1.60$, $se = .29$, $p < .001$, 95% CI [1.04, 2.17]) and the Chemistry lesson ($M = 7.83$, $se = .21$; LSD $M_D = 1.29$, $se = .30$, $p < .001$, 95% CI [0.70, 1.88]) as compared to the Physics lesson ($M = 6.51$, $se = .21$). There were no significant differences for Participants of Color between the Math and Chemistry lessons (LSD $M_D = 0.31$, $se = .29$, $p = .27$, 95% CI [-0.25, 0.89]).

The primary condition by race interaction was not significant ($F(1, 440) = 2.86$, $p = .092$, $\eta_p^2 = .006$), however patterns were consistent with expectations. Participants of Color scored

[10] Detailed analyses using the originally hypothesized comparison between URM and ORG participants are presented in the SOM. Patterns are mainly consistent with that reported in text, though less often significant.

significantly higher on the quiz in the MC compared to the CB condition ($F(1, 299) = 4.09$, $p = .044$, $\eta_p^2 = .013$), whereas White participants' quiz scores did not differ between the CB and MC conditions ($F(1, 141) = 0.44$, $p = .507$, $\eta_p^2 = .003$). See Figure 3.

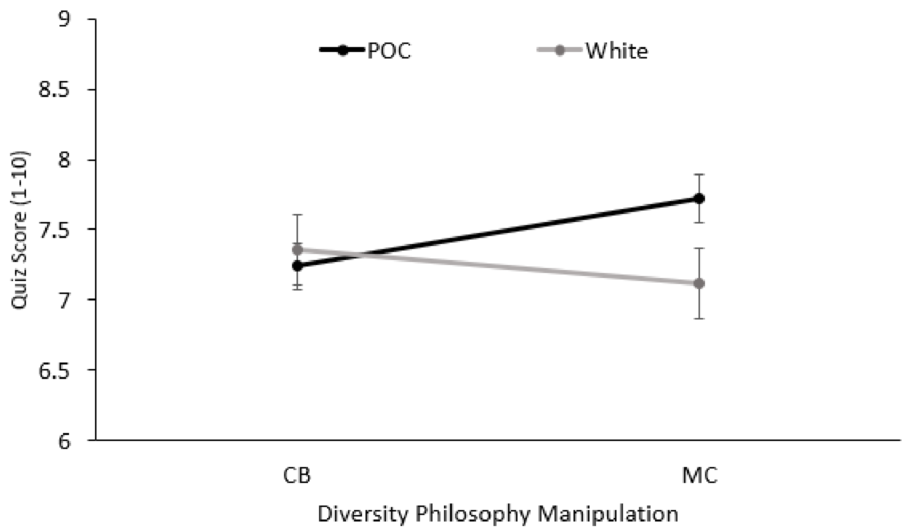

Figure 3. Non-significant interaction of participant race and diversity philosophy condition on quiz score (*p* = .09). Standard error bars are presented.

Belonging. We conducted the same ANOVA as above, this time with belonging as the dependent variable. We again found a main effect of lesson ($F(2, 440) = 4.01$, $p = .019$, $\eta_p^2 = .018$). Mirroring the pattern for quiz score, participants reported greatest belonging in the Math lesson ($M = 5.20$, $se = .08$), then Chemistry ($M = 4.98$, $se = .08$; LSD $M_D = .29$, $se = .10$, $p = .005$, 95% CI [.09, .50]) and then Physics lessons ($M = 4.87$, $se = .09$; LSD $M_D = .09$, $se = .11$, $p = .399$, 95% CI [-.12, .31]). This pattern did not differ as a function of participant race ($F(2, 440) = 1.66$, $p = .19$, $\eta_p^2 = .008$).

Additionally, the interaction of race and condition was significant ($F(1, 440) = 5.31$, $p = .022$, $\eta_p^2 = .012$). POC participants reported nonsignificantly higher belonging in the MC condition (as compared to the CB condition ($F(1, 299) = 1.63$, $p = .202$, $\eta_p^2 = .005$), whereas White participants reported nonsignificantly higher belonging in the CB condition as compared to the MC condition ($F(1, 141) = 3.52$, $p = .063$, $\eta_p^2 = .024$).

See Figure 4.

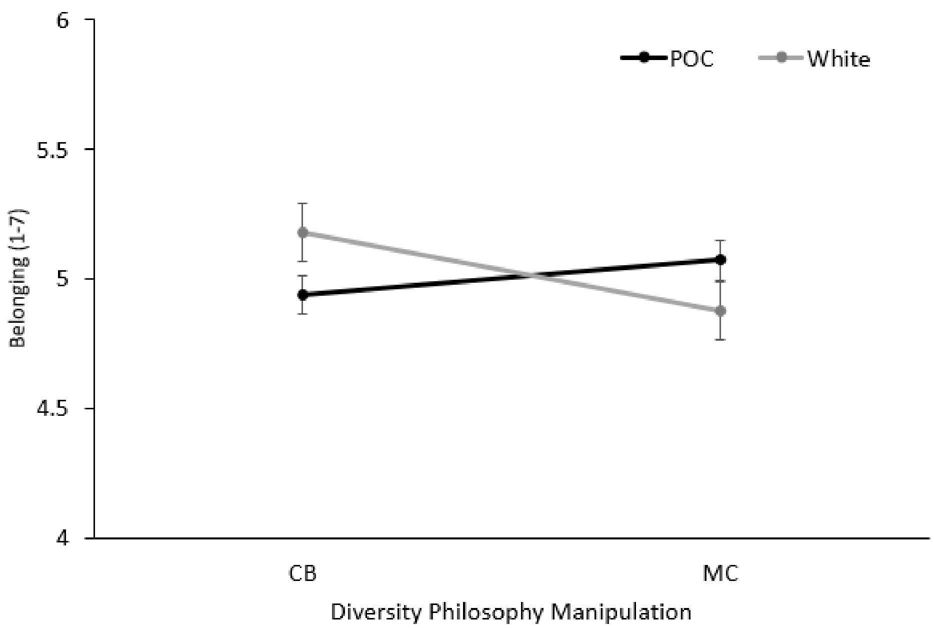

Figure 4. Significant interaction of diversity philosophy condition and participant race on feelings of belonging. Standard error bars are presented.

Perceived Instructor Bias. We again conducted the same ANOVA as above, this time with perceived instructor bias as the dependent variable. We found a significant participant race by condition interaction for perceived instructor bias ($F(1, 440) = 7.30$, $p = .007$, $\eta_p^2 = .016$). POC perceived more instructor bias in the CB condition as compared to the MC condition (F(1,

299) = 7.31, p = .007, η_p^2 = .024). White participants did not detect significantly different levels of instructor bias based on condition (F(1, 141) = 2.25, p = .136, η_p^2 = .016). See Figure 5.

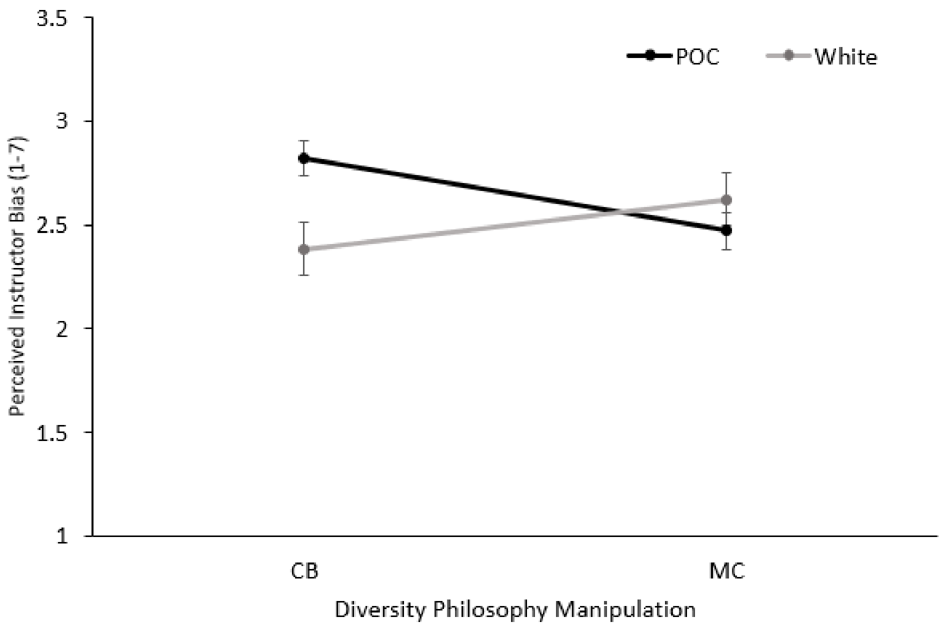

Figure 5. Significant interaction of participant race and diversity philosophy condition on perceived instructor bias. Standard error bars are presented.

Exploratory Analyses[11]

Control Condition. Much of the experimental literature on diversity philosophies directly compares MC to CB without including a control, and researchers have debated what constitutes an appropriate control (Apfelbaum, Phillips, & Richeson, 2014). In the present study,

[11] Our pre-registered analyses included exploratory testing to determine whether participant gender moderated the effects of diversity philosophy condition or race. We conducted three 2 (gender: woman, man) x 2 (race: POC, White) x 2 (condition: MC, CB) x 3 (lesson: chemistry, physics, math) ANOVAs for quiz score, belonging, and perceived instructor bias. Gender only moderated the effect of race on perceived instructor bias. Men of Color perceived higher instructor bias as compared to White men, while there were no differences between Women of Color and White women. Detailed analyses including any main effects of gender or interactions between gender and STEM lesson are reported in the SOM.

we found it important to test whether a control condition that included no mention of diversity, group differences, or similarities would result in different participant responses compared to the MC and CB conditions. We therefore conducted three 2 (race: POC, White) x 3 (condition: MC, CB, Control) x 3 (lesson: chemistry, physics, math) ANOVAs with quiz score, belonging, and instructor bias as the dependent variables.

As in the analyses presented earlier, we observed a significant interaction of condition and participant race for perceived instructor bias ($F(2, 670) = 10.46$, $p < .001$, $\eta_p^2 = .030$), but not for belonging ($F(2, 670) = 2.61$, $p = .074$, $\eta_p^2 = .008$) or quiz score ($F(2, 670) = 1.96$, $p = .141$, $\eta_p^2 = .006$). Cell-level comparisons are presented in Table 2. The Control condition significantly differed from the experimental conditions for only one variable; POC rated the instructor as significantly more biased in the Control condition compared to MC, whereas White participants rated the instructor as significantly less biased in the Control condition compared to MC.

Table 2

Post hoc Comparisons of the Control condition to the CB and MC Conditions

				CB	Control	MC			
	95% *CI*	*p*	*LSD* M_D	*M*	*M*	*M*	*LSD* M_D	*p*	95% *CI*
Quiz Score									
POC	[-.48, .45]	.956	-.01 (.24)	7.24 (.17)	7.27 (.17)	7.72 (.18)	-.48 (.24)	.050	[-.95, .00]
White	[-.53, .77]	.715	.12 (.33)	7.36 (.24)	7.41 (.23)	7.12 (.25)	.46 (.33)	.166	[-.19, 1.11]
Belonging									
POC	[-.15, .26]	.577	.06 (.10)	4.94 (.07)	5.00 (.08)	5.07 (.08)	-.09 (.11)	.380	[-.30, .12]
White	[-.40, .21]	.550	-.09 (.16)	5.18 (.12)	5.05 (.11)	4.87 (.12)	.23 (.16)	.138	[-.08, .54]
Instructor Bias									
POC	[-.16, .31]	.536	.07 (.12)	2.82 (.08)	2.89 (.09)	2.47 (.09)	.42 (.12)	.001	[.18, .66]
White	[-.48, .08]	.165	-.20 (.14)	2.38 (.11)	2.14 (.10)	2.62 (.11)	-.45 (.14)	.002	[-.73, -.17]

Note. Estimated marginal means are presented with standard error in parentheses. Control condition is centered (grey), with post hoc Least Significant Difference (LSD) comparisons on each side, testing whether the MC and CB conditions differ from Control.

Discussion

Overall, results showed that an instructor's use of an MC diversity philosophy can signal inclusion for marginalized racial and ethnic groups; POC showed higher quiz performance and less perceived instructor bias in the MC compared to the CB condition. Feelings of belonging mirrored this pattern but did not reach significance. Moreover, for POC, performance and perceptions of bias in the Control condition seemed to more closely align with the CB condition than the MC condition; this suggests that MC, in particular, may signal inclusion relative to Control rather than CB specifically signaling exclusion. Thus, contrary to White instructors' beliefs that talking about race may make them seem biased, we found that POC evaluated the instructor as least biased when he utilized MC language in the course. POC students interpreted no mention of diversity (Control condition) as equivalently biased as the use of CB language; POC therefore may interpret an instructor who does not acknowledge diversity at all as having a CB ideology.

In contrast, White participants generally showed the opposite pattern of results compared to POC, though these differences rarely reached significance. Notably, White participants perceived the instructor to be most biased when using MC language and least biased when not mentioning diversity at all (Control condition). This is consistent with prior work showing that White individuals are hesitant to talk about race at all and consider mentioning race to be racist (Apfelbaum et al., 2008; Goff et al., 2013).

The present research has important implications for equity and inclusion in higher education. First, by having participants learn novel STEM content we demonstrate that situational cues of exclusion do not just disrupt in-the-moment recall of prior learned information (Wilton et al., 2015) or ability to focus cognitive resources on the task at hand (as in stereotype threat; Schmader & Johns, 2003). Indeed, we show that diversity messaging can impact the learning process from the beginning, either strengthening learning through an MC cue of inclusion, or impairing learning through a CB cue of exclusion. This is particularly important given that undergraduates of color tend to make the decision to drop out of STEM majors based on these early learning experiences in college (Riegle-Crumb et al., 2019; Syed, 2010). Second, we show that White instructors' refusal to discuss race in the classroom due to a fear of appearing racist is 1) not warranted (at least for Students of Color), and 2) detrimental to Students of Color. Our findings are especially troubling given that the majority of the professoriate is White, White teachers are especially likely to endorse a CB approach to diversity (Johnson, 2002; Lewis, 2001; Sleeter, 2001), and STEM faculty in particular may be less likely than faculty in other disciplines to discuss diversity in their classes if questions of identity do not "naturally" occur within the curriculum (Laird, 2011). Thus, POC entering college and taking their first introductory-level STEM courses likely receive either no mention of diversity or a CB approach from their faculty, potentially resulting in greater perceived bias and lower performance.

The present study used a large-scale, nationally representative sample of first-year undergraduates. We created realistic, yet experimentally controlled, course materials and online course interface in order to as closely as possible simulate students' real experiences in their early college STEM courses. We found small effect sizes yet consistent patterns across our self-report measures. Note that our manipulation was subtle; the CB and MC conditions differed only in substituting the word "similarities" for "differences" 5 times across the content spoken by the instructor or written in the syllabus. We argue that the observed effects showcase the level of vigilance shown by students in attending to cues about race. Even subtle variations send strong signals regarding the instructor's views. In a semester-long, face-to-face course, effects of recurring diversity language (or lack thereof) may be compounded.

In the present study we show consistently positive effects of MC for Students of Color. However, diversity philosophy researchers have shown that MC is not the best policy in all possible situations (Apfelbaum, Stephens, & Reagans, 2016). If students are severely underrepresented in a class, MC language may lead them to feel tokenized or only valued for their identity, resulting in lower feelings of belonging within the course or perceptions that their instructor is more biased. While much of the research has been on White instructors and their efforts to talk about diversity in the classroom, it is important to consider other faculty identities as well. In the present study, the only information given about the professor's identity was his name and voice. We posited that students likely guessed he was White; indeed, post hoc testing with an independent sample ($N = 454$) revealed that participants rated the instructor as more likely to be White as compared to Black, Asian or Latino. Perceptions of the instructor's race did not vary based on the diversity philosophy condition or the participants' race (see the SOM for details). Would the effects observed in the present study be the same regardless of instructor race and gender, or might faculty identity moderate the impact of diversity language? Future research should address this question.

Conclusion

Using a realistic online course experience and a large, diverse sample of first-year college students, we found that POC experienced more positive outcomes in a chemistry, math, or physics course when the instructor utilized MC as compared to CB language. Ours is the first study to show that instructors' use of diversity philosophies can affect the learning of novel STEM content, rather than simply disrupt their performance for previously learned material. We conclude that educators must be sensitive to the effects of their language and the way they communicate their thoughts about diversity rather than shying away from discussion or falsely equating varying student experiences and backgrounds. If we want to remove barriers to marginalized students' success in STEM, we must continue to study signals of exclusion and inclusion delivered by STEM authorities in the classroom.

References

Apfelbaum, E. P., Sommers, R. S., & Norton, M.I., (2008). Seeing race and seeming racist? Evaluating strategic colorblindness in social interaction. *Journal of Personality and Social Psychology, 95*, 918-932. doi:10.1037/a0011990

Apfelbaum, E. P., Phillips, K. W., & Richeson, J. A. (2014). Rethinking the baseline in diversity research: Should we be explaining the effects of homogeneity? *Psychological Science, 9*, 235-244. doi: 10.1177/1745691614527466

Apfelbaum, E. P., Stephens, N. & Reagans, R. (2016). Beyond one-size-fits-all: Tailoring diversity approaches to the representation of social groups. Journal of Personality and Social Psychology, 4, 111, 547-566. doi: 10.1037/pspi0000071

Apfelbaum, E. P., Grunberg, R., Halevy, N., & Kang, S. (2017). From ignorance to intolerance: Perceived intentionality of racial discrimination shapes preferences for colorblindness versus multiculturalism. *Journal of Experimental Social Psychology*, *69*, 86–101. https://doi.org/10.1016/j.jesp.2016.08.002

Aragón, O. R., Dovidio, J. F., & Graham, M. J. (2016). Colorblind and multicultural ideologies are associated with faculty adoption of inclusive teaching practices. *Journal of Diversity in Higher Education, 10*, 201-215. doi: 10.1037/dhe0000026

Authors. (2019). Differential effects of multiculturalism vs. colorblindness: Content or confound? *Manuscript submitted for publication.*

Celeste, L., Baysu, G., Phalet, K., Meeussen, L., & Kende, J. (2019). Can school diversity policies reduce belonging and achievement gaps between minority and majority youth? Multiculturalism, colorblindness, and assimilationism assessed. *Personality and Social Psychology Bulletin, 45*, 1603-1618. doi: 10.1177/0146167219838577

Cho, J., Morris, M. W., Slepian, M. L., & Tadmor, C. T. (2017). Choosing fusion: The effects of diversity ideologies on preference for culturally mixed experiences. *Journal of Experimental Social Psychology*, *69*, 163–171. https://doi.org/10.1016/j.jesp.2016.06.013

Cohen, G. J., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313*, 1307-1310. doi: 10.1126/science.1128317

Dasgupta, N., & Stout, J. G. (2014). Girls and women in science, technology, engineering, and mathematics: STEMing the tide and broadening participation in STEM careers. *Policy Insights from the Behavioral and Brain Sciences, 1*, 21-29. doi: 10.1177/2372732214549471

Dierker, L., Alexander, J., Cooper, J. L., Selya, A., Rose, J., & Dasgupta, N. (2016). Engaging diverse students in statistical inquiry: A comparison of learning experiences and outcomes of under-represented and non-underrepresented students enrolled in a multidisciplinary project-based statistics course. *International Journal for the Scholarship of Teaching and Learning*, *10*. doi: 10.20429/ijsotl.2016.100102

Dover, T. L., Major, B., & Kaiser, C. R., (2016). Members of high-status groups are threatened by pro-diversity organizational messages. *Journal of Experimental Social Psychology, 62*, 58-67. doi:10.1016/jesp.2015.10.006

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191. doi: 10.3758/BF03193146

Ferrare. J. J., & Lee, Y. G. (2014). Should we still be talking about leaving? A comparative examination of social inequality in undergraduates' major switching patterns (*WCER*

Working Paper No. 2014-5). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research website: http://www.wcer.wisc.edu/publications/workingPapers/papers.php

Goff, P. A., Jackson. M. C., Nichols, A. H., & Di Leone, B. A. (2013). Anything but race: Avoiding racial discourse to avoid hurting you or me. *Psychology, 4*, 335-339. doi: 10.4236/psych.2013.43A048

Haak, D. C. HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in Introductory Biology, *Science, 332*, 1213-1216. doi: 10.1126/science.1204820

Hachfeld, A., Hahn, A., Schroeder, S., Anders, Y., & Kunter, M. (2015). Should teachers be colorblind? How multicultural and egalitarian beliefs differentially relate to teachers' professional competence for teaching in diverse classrooms. *Teaching and Teacher Education, 48*, 44-55. doi: 10.1016/j.tate.2015.02.001

Hess, C., Gault, B., & Yi, Y. (2013). Accelerating change for faculty women of color in STEM: Policy, action, and collaboration. *Institute for Women's Policy Research.* Retrieved from http://www.iwpr.org/publications/pubs/accelerating-change-for-women-faculty-of-color-in-stem-policy-action-and-collaboration

Johnson, L. (2002). My eyes have been opened: White teachers and racial awareness. *Journal of Teacher Education, 52*, 153-167. doi: 10.1177/0022487102053002007.

Laird, T. (2011). Measuring the diversity inclusivity of college courses. *Research in Higher Education, 52*, 6, 572-588. doi: 10.1007/s11162-010-9210-3

Leslie, S., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science, 347*, 262-265. doi: 10.1126/science.1261375

Lewis, A. E. (2001). There is no 'race' in the schoolyard: color-blind ideology in an (almost) all-White school. *American Educational Research Journal, 38*, 781-811. doi: 10.3102/00028312038004781.

Lin, M. H., Kwan, V. S., Cheung, A., & Fiske, S. (2005). Stereotype content model explains prejudice for an envied outgroup: Scale of anti-Asian American stereotypes. *Personality and Social Psychology Bulletin, 31*, 34-47. doi: 10.1177/0146167204271320

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. doi: 10.1037/apl0000022

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences, 109*, 16474-16479. doi: 10.1073/pnas.1211286109

NSF National Center for Science and Engineering Statistics. (2019). *Women, minorities, and persons with disabilities in science and engineering.* Retrieved from https://ncses.nsf.gov/pubs/nsf19304/digest/about-this-report

Ogunwole, S. U., Drewery, M. P. Jr., & Rios-Vargas, M. (2012). The population with a bachelor's degree or higher by race and Hispanic origin: 2005-2010. *America Community Survey Briefs, U.S. Census Bureau*. Retrieved from: https://www2.census.gov/library/publications/2012/acs/acsbr10-19.pdf

Perry, S. P., Murphy, M. C., & Dovidio, J. F. (2015). Modern prejudice: Subtle, but unconscious? The role of bias awareness in Whites' perceptions of personal and others' biases. *Journal of Experimental Social Psychology, 61*, 64-78. doi: 10.1016/j.jesp.2015.06.007

Plaut, V. C. (2010). Diversity science: Why and how difference makes a difference. *Psychological Inquiry, 21,* 77–99. doi: 10.1080/10478401003676501

Plaut, V. C., Buffardi, L. E., Garnett, F. G., & Sanchez Burks, J. (2011). What about me? Perceptions of exclusion and whites' reactions to multiculturalism. *Journal of Personality and Social Psychology, 101*, 337-353. doi: 10.1037/a0022832

Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Ditlmann, R., & Crosby, J. R. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions. *Journal of Personality and Social Psychology*, *94*(4), 615–630. https://doi.org/10.1037/0022-3514.94.4.615

Ramsey, L. R., Betz, D. E., & Sekaquaptewa, D. (2013). The effects of an academic environment intervention on science identification among women in STEM. *Social Psychology of Education*, *16*, 377-397. doi: 10.1007/s11218-013-9218-6.

Riegle-Crumb, C., King, B., & Irizarry, Y. (2019). Does STEM stand out? Examining racial/ethnic gaps in persistence across postsecondary fields. *Educational Researcher, 48*, 133-144. doi: 10.3102/0013189X19831006

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*, 44-452. doi: 10.1037/0022-3514.85.3.440

Sleeter, C. (2001). Preparing teachers for culturally diverse schools: research and the overwhelming presence of whiteness. *Journal of Teacher Education, 52*, 94-106. doi: 10.1177/0022487101052002002.

Syed, M. (2010). Developing an integrated self: Academic and ethnic identities among ethnically diverse college students. *Developmental Psychology, 46*, 1590-1604. doi: 10.1037/a0020738

Walton, G. M., & Cohen. G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology, 92*, 82-96. doi: 10.1037/0022-3514.92.1.82

Wilson, J. L., Meyer, K. A., & McNeal, L. (2012). Mission and diversity statements: What they do and do not say. *Innovative Higher Education, 37*, 125-139. doi: 10.1007/s10755-011-9194-8

Wilton, L. S., Good, J. J., Moss-Racusin, C. A., & Sanchez, D. T. (2015). Communicating more than diversity: The effect of institutional diversity statements on expectations and performance as a function of race and gender. *Cultural Diversity and Ethnic Minority Psychology, 21*, 315-325. doi: 10.1037/a0037883

Whitley, B. E., & Webster, G. D. (2019). The relationships of intergroup ideologies to ethnic prejudice: A meta-analysis. *Personality and Social Psychology Review, 23*, 207-237. doi: 10.1177/1088868318761423